ACADEMIA ASTURIANA DE CIENCIA E INGENIERÍA

ANÁLISIS ESTADÍSTICO DE DATOS *FUZZY*: RIGOR CIENTÍFICO Y UTILIDAD PRÁCTICA

DISCURSO PRESENTADO EN EL ACTO DE SU INCORPORACIÓN COMO ACADÉMICA DE NÚMERO POR LA

PROF. MARÍA ÁNGELES GIL ÁLVAREZ

Y CONTESTACIÓN DE LA

ILMA. PROF. CONSUELO MARTÍNEZ LÓPEZ

Académica de número de la
Academia Asturiana de Ciencia e Ingeniería

EL DÍA 4 DE JULIO DE 2025

AACI
ACADEMIA ASTURIANA DE CIENCIA E INGENIERIA

c/ San Francisco. Edificio Histórico - Universidad de Oviedo
OVIEDO

Ilustrísimo Sr. presidente de la Academia Asturiana de Ciencia e Ingeniería (AACI)
Ilustrísimos Sres. presidentes de las Academias Asturianas de Medicina y de Jurisprudencia y Sr. director del Real Instituto de Estudios Asturianos
Ilustrísimos Sres. académicos y colegas
Autoridades, patrocinadores de AACI, compañeros, amigos, familia, señoras y señores

Agradecimientos iniciales

Aunque debo expresar mi gratitud a muchas personas por estar hoy aquí y por tener el honor de ingresar como miembro de número en esta Academia Asturiana de Ciencia e Ingeniería, aún con poco más de tres años de vida, sé que todos entenderéis sobradamente que comience manifestando el más profundo agradecimiento hacia mi hermano Pedro Gil.

Por Pedro vine a trabajar a Asturias según acabé la licenciatura de Matemáticas en la Universidad de Valladolid. Por Pedro supe de la importancia que tenía la investigación si quería seguirse una carrera profesional dentro de la universidad. Y, por Pedro, conocí los conjuntos *fuzzy* y la posibilidad potencial de incorporarlos a la Probabilidad y la Estadística, el tema de este discurso.

Es cierto que, desde que se marchó, aprovecho cualquier oportunidad para integrar a Pedro en charlas, conferencias o discursos que imparto, pero en este caso es especialmente justo y obligado hacerlo. Él no llegaría a ser testigo ni de la gestación ni del nacimiento de la AACI, aun cuando es una iniciativa que habría aplaudido con entusiasmo, máxime dado que el principal impulsor de la misma fue uno de sus primeros alumnos en la Universidad de Oviedo, en cursos de doctorado.

Mis sinceras gracias también a ese alumno de Pedro, el presidente de esta Academia, Mario Díaz. Cuando hace algo más de tres años, en los tiempos del distanciamiento social tras el confinamiento, me llamó para preguntarme si podíamos hablar, adiviné fácilmente (téngase en cuenta que pertenecí a su equipo cuando ejerció como vicerrector de investigación de la Universidad de Oviedo) que él había dedicado buena parte de esos meses atípicos que estábamos viviendo a madurar algún proyecto que estimaba fascinante.

Me quedaban pocos años para jubilarme y tenía bastante claros los planes para mantener la cabeza activa y no dejarme arrastrar por ninguna nostalgia una vez llegada la jubilación. Si, por casualidad, lo que Mario quería contarme implicaba involucrarme en lo que se le hubiera ocurrido, no pensaba dejarme embaucar por atractivo que fuera el proyecto. Pero me enredé yo sola. Aun siendo una empresa ambiciosa y exigente, era más ilusionante y desafiante de lo que habría podido imaginar: una Academia Asturiana de Ciencia e Ingeniería. Al inmenso honor que supondría ser miembro de una institución de tal envergadura, se uniría la oportunidad de ser testigo de primera línea y partícipe de su crecimiento, aunque sin olvidar que conllevaría una importante responsabilidad añadida de gestión y trabajo. Me habló de compañeros que se incorporarían a la iniciativa… Imposible resistirse.

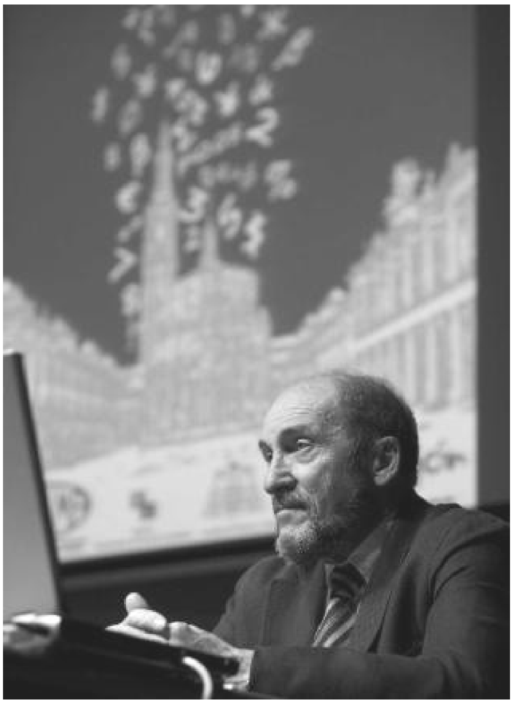
Quedan muchos otros a quienes debería incluir en los agradecimientos de este discurso, pero o bien irán revelándose a lo largo de su lectura (colaboradores, compañeros, asesores, inspiradores, etc.) o aparecerán al final de este.

ÍNDICE

1. Introducción histórica de la investigación recogida en el discurso

Cuando hace algo más de 48 años concluí los estudios de la licenciatura en Matemáticas por la Universidad de Valladolid, **Pedro Gil Álvarez** acababa de trasladarse a Asturias tras obtener la agregaduría en Investigación Operativa por la Universidad de Oviedo. Iba a dirigir el Departamento de Matemáticas de la Facultad de Ciencias y buscaba matemáticos, que por entonces escaseaban un tanto en la región. Y yo buscaba empleo. Seguramente con una buena dosis de inconsciencia y de candidez, le dije que podía presentarme a una de las plazas disponibles. Sopesé que la docencia en primer curso de Químicas, que era la que se me asignaría, no parecía revestir grandes problemas y, para mayor fortuna, los otros grupos del mismo curso correrían a cargo de Teófilo Brezmes, con quien había coincidido en algunas asignaturas de la carrera en Valladolid y de quien me constaba que era una gran persona, si bien luego descubrí que era mucho más que eso. Aquel 1976-1977 fue un curso muy especial para los dos, más por la transición personal que suponía el paso directo de alumnos a profesores, que por la transición política por la que atravesaba el país.

Aunque ya existía un Departamento de Matemáticas en la Facultad de Ciencias, Pedro tuvo que impulsarlo hasta darle un carácter netamente matemático. No existía entonces una licenciatura en Matemáticas en la Universidad de Oviedo, lo que constituía un hándicap en muchos aspectos. Había muy pocos libros de análisis, de álgebra, de matemáticas generales, de probabilidad o de estadística y, menos aún, de temas más especializados.



Pedro Gil Álvarez

En aquellos primeros momentos desconocíamos hasta dónde alcanzaban nuestras responsabilidades como profesores universitarios, más allá de la labor docente que, a pesar de la carencia de material bibliográfico, nos gustaba mucho. Pero Pedro nos inculcó desde el primer día que había mucho más que hacer y, de forma indolora, nos fue introduciendo en el mundo de la investigación. Si el número de libros especializados era muy reducido, el de revistas científicas aún lo era más. Realizar una tesis matemática en esas condiciones y con tan limitados mimbres, resultó muy audaz, pero nuestra confianza en la fortaleza del director era ciega y fuimos tirando hacia delante, doctorándonos primero y asumiendo después gradualmente labores de cosupervisión/supervisión.

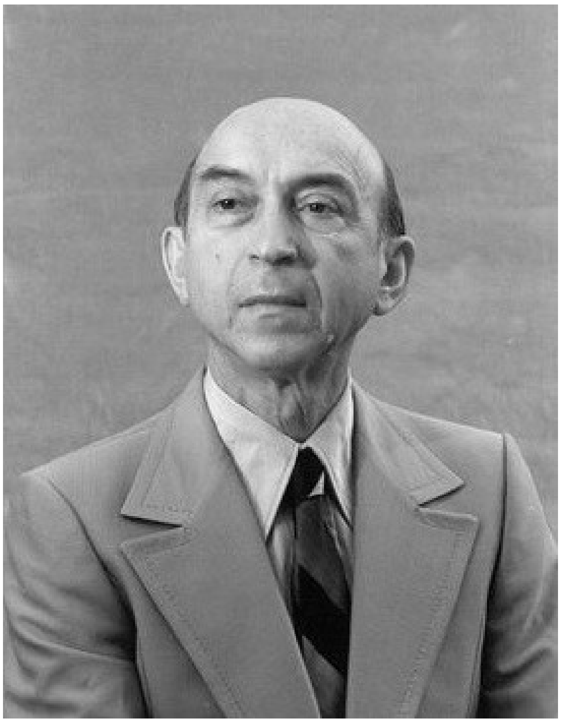
Las primeras investigaciones concernieron a la teoría de la información en relación con la estadística: medidas de entropía, incertidumbre o diversidad, medidas de desigualdad, medidas de información mutua, medidas de información generalizada, entre otras, junto con su estimación en distintos tipos de muestreo, su aplicación a la comparación de experimentos aleatorios, etc. De ellas derivaron cinco tesis doctorales dirigidas por Pedro Gil y otras tres codirigidas conmigo.

Desde su llegada a Oviedo, Pedro nos había sugerido que echáramos alguna mirada a trabajos relativos a los conjuntos *fuzzy*. En 1976, el quinto ejercicio de una agregaduría, paso inmediatamente anterior al del acceso a cátedra, consistía en el desarrollo por escrito de un tema escogido por sorteo entre cierto número de ellos propuestos por el tribunal al principio de la oposición.

Su mentor y director de tesis doctoral, el profesor **Sixto Ríos García**, formaba parte de ese tribunal y había incluido entre estos temas el artículo seminal de la teoría de conjuntos *fuzzy*, *"Fuzzy sets"*, del que era autor Lotfi A. Zadeh, publicado en la revista *Information and Control* en 1965 y uno de los trabajos científicos más citados de todos los tiempos y campos. Finalmente, no fue el tema que le correspondió a Pedro en el sorteo del ejercicio de su oposición, pero dejó una huella importante en él.

Sixto Ríos García

Lotfi A. Zadeh, azerbaiyano-estadounidense, introdujo el concepto de conjunto *fuzzy* para acuñar una clase de elementos que estuviera caracterizada por cierta propiedad definida de forma imprecisa (o *ill-defined*, como el propio Zadeh calificaba en contraposición al matemático *well-defined*), en el sentido de que su cumplimiento no necesariamente se satisficiera de forma inequívoca, sino que admitiera 'gradación'. Esta gradación asumiría valores entre 0 y 1, representando 0 y 1 el incumplimiento y el cumplimiento incontestables, respectivamente.

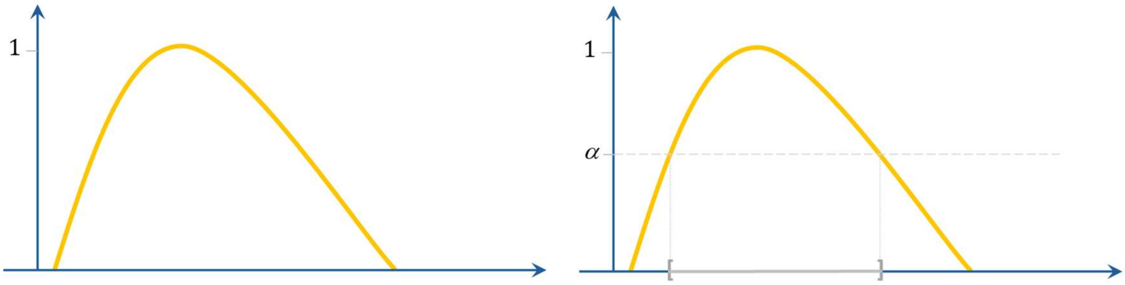


Lotfi A. Zadeh

Se había formado Zadeh como ingeniero eléctrico y electrónico doctorado en el MIT y era un reconocido experto en teoría de sistemas, pero a su vez tenía una raigambre manifiesta en las matemáticas (en especial, en la estadística). De este modo, cuando estableció la teoría de conjuntos *fuzzy* no descuidó en ningún momento los aspectos matemáticos, formalizando un **conjunto *fuzzy*** como una función sobre un referencial y con valores en el intervalo [0,1], interpretándose el valor asignado a cada elemento del referencial como el 'grado de compatibilidad' de ese elemento con la propiedad que caracterizara al conjunto.

Habitualmente, aunque no de forma exclusiva, el referencial es el espacio de los números reales o un espacio euclídeo de dimensión cualquiera, suponiéndose a menudo que cada **α-nivel** de un conjunto *fuzzy* de ese referencial (colección de valores del espacio con compatibilidad al menos α con la propiedad que lo caracteriza), con $\alpha \in [0,1]$, es un conjunto compacto y convexo del espacio correspondiente.

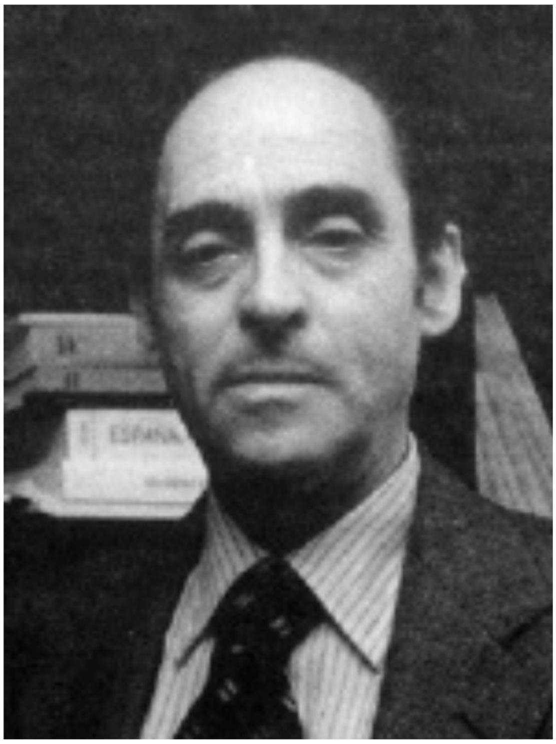
El caso unidimensional da lugar a los llamados **números *fuzzy*** y es el más común en la práctica. Desde los primeros trabajos se apuntaba ya su motivación y aplicación inmediatas a situaciones que involucraran valoraciones humanas.



Ejemplo de número *fuzzy* (izquierda) y de su intervalo α-nivel (derecha)

Los conjuntos *fuzzy* surgieron como extensión y como contrapartida de la rigidez de la lógica y de la toma de decisiones binarias o multivaluadas. Han tendido puentes entre las matemáticas y la forma de evaluar, valorar, percibir, juzgar y razonar del pensamiento humano, que muy a menudo es imprecisa, bien sea por falta de instrumentos de medida o de instrumentos suficientemente precisos o por concernir a magnitudes que son intrínsecamente ambiguas.

Volviendo al hecho del tema que se incluyó en el sorteo del quinto ejercicio de la oposición a agrageado de Pedro, es muy posible que en la propuesta de Sixto Ríos incidieran tanto el interés creciente que en esos años iba adquiriendo la teoría de conjuntos *fuzzy* y sus aplicaciones, como la atracción ostensible que empezaba a mostrar por esta teoría su antiguo compañero y buen amigo, el profesor **Francisco Azorín Poch**, catedrático de Estadística de la Universidad Autónoma de Madrid (antes de la de Santiago de Compostela) y que entre 1977 y 1982 ejerció como presidente del Instituto Nacional de Estadística.



Francisco Azorín Poch

Tal atracción se evidenció en el discurso de su toma de posesión a finales de 1981 como académico numerario de la Real Academia de Ciencias Exactas, Físicas y Naturales (ahora de España, y a la sazón de Madrid), RAC, *"Conjuntos Borrosos, Estadística y Probabilidad"*, al que dio contestación el entonces ya veterano académico Ríos. Que para una ocasión tan señalada como la de dicho ingreso, alguien que era un experto reconocido internacionalmente en cuestiones como, por ejemplo, las de los diversos tipos y métodos de muestreo y sus aplicaciones, eligiera como tema uno más bien incipiente y, en cualquier caso, embrionario en lo relativo a la estadística y la probabilidad, es buena prueba de que lo encontraba muy prometedor.

Pedro supo del contenido del discurso, así como del "*Glosario de Conjuntos Borrosos en relación con la Estadística"*, también elaborado por Azorín Poch y publicado por el INE en 1980. Y en 1982 emprendió de forma manifiesta una suerte de campaña de concienciación acerca del potencial de la teoría de conjuntos *fuzzy* en conexión con las de probabilidad y estadística.

Esa campaña dio los frutos perseguidos en nuestra pequeña comunidad departamental y, a partir de entonces, se empezaron a abrir algunas líneas al respecto.

1.1 – Pincelada histórica sobre la investigación en el análisis de datos *fuzzy*: primera aproximación

Una de ellas encarnó un primer acercamiento a la estadística con datos *fuzzy*. Hasta aquel momento, la literatura sobre la posible generación aleatoria de los datos *fuzzy* era muy reducida. Prácticamente se limitaba a unos pocos estudios sobre tres modelos distintos. Dos de estos, los llamados "sistemas de información *fuzzy*" (véanse, por ejemplo, Tanaka, Okuda y Asai, 1976, Asai, Tanaka y Okuda, 1977, Okuda, Tanaka y Asai, 1978) y las "variables aleatorias *fuzzy*" (Kwakernaak, 1978, 1979), partían de considerar experimentos aleatorios que involucraban la medición de variables numéricas asociadas y para los que la observación o percepción del valor de esa variable en cada medición se suponía que no era exacta, aunque sí podía expresarse mediante una cantidad imprecisa modelizada a través de conjuntos *fuzzy*.

En el otro modelo, el de los "conjuntos *fuzzy* aleatorios" (Féron, 1976ab, 1979ab), el experimento aleatorio implicaba directamente la medición de una variable intrínsecamente imprecisa para la que dichos valores podían describirse mediante conjuntos *fuzzy*.

Tanto los conjuntos *fuzzy* aleatorios como las variables aleatorias *fuzzy* tuvieron un recorrido un tanto corto. Posiblemente, porque su formalización carecía de la especificación o clarificación de algunos de sus componentes esenciales. No obstante, en años posteriores esa formalización fue depurada y perfeccionada por Puri y Ralescu (1985, 1986) en cuanto a los primeros y por Kruse y Meyer (1987) en cuanto a las segundas. Como se explicará más tarde, en la primera de estas se fundamenta la segunda aproximación al análisis de datos *fuzzy* que ocupará la mayor parte de las secciones y los capítulos siguientes del discurso.

Los **sistemas de información *fuzzy***, introducidos en la segunda parte de los setenta del siglo XX por el ingeniero eléctrico japonés **Hideo Tanaka**, junto a su mentor Kiyoji Asai y su discípulo Tetsuji Okuda, 'agrupaban' los valores de la variable numérica subyacente mediante 'particiones *fuzzy*' en el sentido definido por el matemático argentino-estadounidense **Enrique H. Ruspini** (1969, 1970), empleando la 'probabilidad asociada a un conjunto *fuzzy*' según la definición dada por Lotfi A. Zadeh (1968). Se establecía de este modo un espacio de probabilidad, si bien a costa de exigir que las percepciones *fuzzy* de los valores de la variable numérica constituyeran una partición 'à la Ruspini' y de asumir que la probabilidad de Zadeh representara una especie de probabilidad inducida mediante las percepciones *fuzzy*.


Enrique Ruspini **Hideo Tanaka**

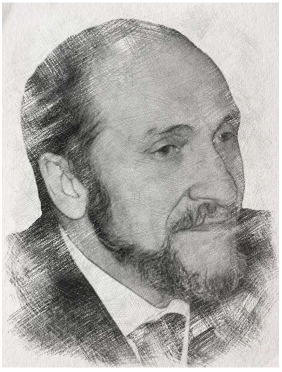
Sobre la base de ese modelo se desarrollaron diversos estudios. Entre ellos:

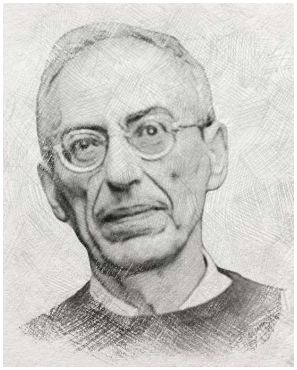
- Resolución de problemas de decisión e información estadística con datos *fuzzy*:
  - criterio de comparación de sistemas de información *fuzzy* basado en el valor esperado de la información muestral;
  - criterios de comparación de sistemas de información *fuzzy* basados en medidas de la información muestral tipo Shannon y tipo Fisher;
  - criterio de comparación de sistemas de información *fuzzy* basado en los conjuntos de pérdida accesibles;
  - relaciones entre los distintos criterios de comparación de sistemas de información *fuzzy*;
  - análisis del efecto asociado a la imprecisión *fuzzy* y la pérdida de información según los diferentes criterios; especial examen de la conexión con la suficiencia estadística.
- Resolución de problemas de estimación paramétrica en la distribución de la variable numérica subyacente a partir de datos *fuzzy*:
  - estimación puntual bayesiana, como problema especial de decisión estadística con datos *fuzzy*;
  - extensión del método de máxima verosimilitud a partir de datos *fuzzy*, basada en la minimización de cierta medida de inexactitud tipo Shannon;
  - introducción a la estimación paramétrica intervalar con datos *fuzzy*.

- Resolución de problemas de contrastes de hipótesis estadísticas acerca de la distribución de la variable numérica subyacente a partir de datos *fuzzy*:
  - Contraste de hipótesis bayesiano, como problema especial de decisión estadística con datos *fuzzy*;
  - Extensión del contraste de Neyman y Pearson con datos *fuzzy*;
  - Extensión del contraste de la razón de verosimilitudes y combinación con las estimaciones de mínima inexactitud con datos *fuzzy*;

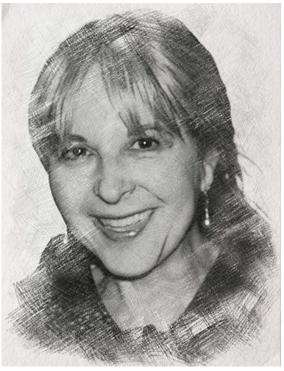
La investigación en esta línea dio lugar a tres tesis doctorales dirigidas por Pedro Gil (dos codirigidas conmigo) y a un buen número de artículos en revistas y libros internacionales que han servido como referencia a varios trabajos de otros grupos.

Además de Pedro y yo misma, nos acompañaron en esta parte del camino cinco compañeros del Departamento de Matemáticas de la Facultad de Ciencias, que a finales de los ochenta por la aplicación de la L.R.U. en la Universidad de Oviedo pasó en su mayoría a integrarse en el área de conocimiento de Estadística e Investigación Operativa del Departamento de Matemáticas de la Universidad de Oviedo. *Norberto Corral Blanco*, *Mª Teresa López García* y *Mª Rosa Casals Varela* lo hicieron con la realización de sus tesis y los trabajos derivados, mientras que *Teófilo Brezmes Brezmes* y *José Manuel Álvarez Garrido* colaboraron en varios de los trabajos sobre el tema.

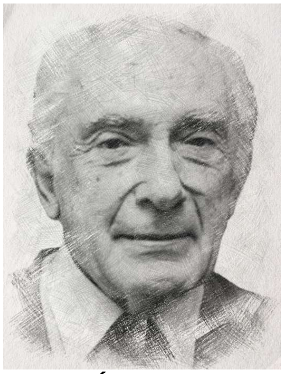

Pedro Gil Álvarez


Norberto Corral Blanco


Mª Teresa López García


Mª Rosa Casals Varela


Teófilo Brezmes Brezmes


José M. Álvarez Garrido

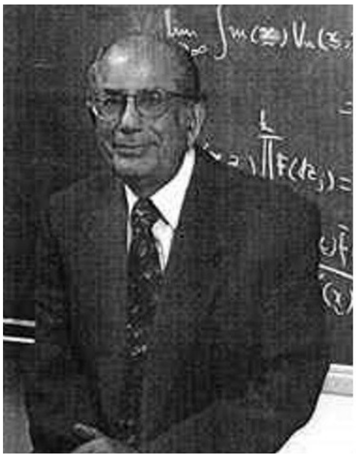
Compañeros del Departamento de Matemáticas que participaron, y con los que colaboré, en el primer acercamiento al análisis de datos *fuzzy*

1.2 – Pincelada histórica sobre la investigación en el análisis de datos *fuzzy*: segunda aproximación

La investigación anterior resultó en cierta medida pionera en su tiempo y, a día de hoy, aún existen estudios que se basan en ella, pero entre finales de los ochenta y principios de los noventa optamos por abandonarla. A pesar de que se dispusiera de un espacio de probabilidad, este se había obtenido recurriendo a dos definiciones/restricciones que no se veían suficientemente naturales, justificables e interpretables en un contexto probabilístico. Más bien, parecían imposiciones artificiales forzadas para llegar a disponer de tal espacio.

Además, mi estancia durante cerca de dos años, entre 1988 y 1990, en la Universidad de California en Berkeley, supuso un revulsivo al tener al alcance una biblioteca universitaria de su talla: con bases de datos enormemente avanzadas para la época, con prácticamente todos los libros y revistas que pudieran necesitarse y soñarse, y con las primeras conexiones de red *bitnet*, de cuya existencia carecía de conocimiento previo. Aunque en nuestra biblioteca en Oviedo se hubiera incrementado ligeramente el número de libros y de revistas científicas, el cambio fue drástico y enriquecedor.

En este periodo, fue un hallazgo crucial el de los trabajos conjuntos del matemático indo-estadounidense **Madan L. Puri** con su doctorando, el matemático rumano-estadounidense **Dan A. Ralescu**, publicados en el segundo lustro de los ochenta del siglo XX.


Madan L. Puri


Dan A. Ralescu

En ellos establecieron un modelo que, como señalaremos en el siguiente capítulo, completaba en muy buena parte el introducido por Robert Féron. Se trataba de un modelo para el mecanismo aleatorio que generaba de forma directa valores de conjunto *fuzzy*. En otras palabras, para las magnitudes aleatorias con valores intrínsecamente imprecisos, cuando estos se modelizaban a través de conjuntos *fuzzy*.

Puri y Ralescu acuñaron dicho mecanismo aleatorio como variable aleatoria *fuzzy*, el mismo término que había adoptado Kwakernaak para un concepto asociado a un escenario diferente. Si bien se mantuvo así en todos los primeros trabajos de índole probabilística, paulatinamente pareció más coherente optar por el de conjunto *fuzzy* aleatorio que le asignó Féron.

Los **conjuntos *fuzzy* aleatorios** asociados a un espacio de probabilidad que caracterizaba un experimento aleatorio se introdujeron inicialmente como aplicaciones del espacio de resultados experimentales en un conjunto de valores de conjunto *fuzzy* tal que nivel a nivel constituyeran conjuntos aleatorios asociados a dicho espacio (concepto que se había establecido a lo largo de los setenta).

Se veía como modelo oportuno para las valoraciones humanas esencialmente imprecisas, sin requerir que subyaciera una valoración numérica. Como ejemplo de motivación al que podía adaptarse de forma bastante innata, se encontraba el problema de decisión cuando la asignación de utilidades o pérdidas resultaba imprecisa. En este sentido, James O. Berger, experto en teoría de la decisión estadística bayesiana y académico extranjero de la RAC, refiriéndose a las pérdidas y a la distribución *a priori*, señalaba (1985) que *«En muchos problemas, estas cantidades serán muy imprecisas o incluso no únicas.»*

Esta fue la razón por la que el primer estudio estadístico en el que incorporamos ese modelo fue el de la comparación de experimentos aleatorios asociados a un mismo problema de decisión unietápico. Se consideraba como novedad que, para cada acción, las pérdidas se comportaban como un conjunto *fuzzy* aleatorio asociado al espacio de estados del problema y se extendía el criterio basado en el valor esperado de la información muestral. Este estudio se desarrolló durante mi estancia en Berkeley, si bien su publicación tuvo lugar una vez de vuelta a España.

En esa vuelta, se retomaron por un breve tiempo unas pocas investigaciones sobre la pérdida de información atribuible a la imprecisión inherente a los sistemas de información *fuzzy* y, sobre todo, algunas relativas a la estimación y al contraste de hipótesis sobre medidas de diversidad, información mutua y desigualdad.

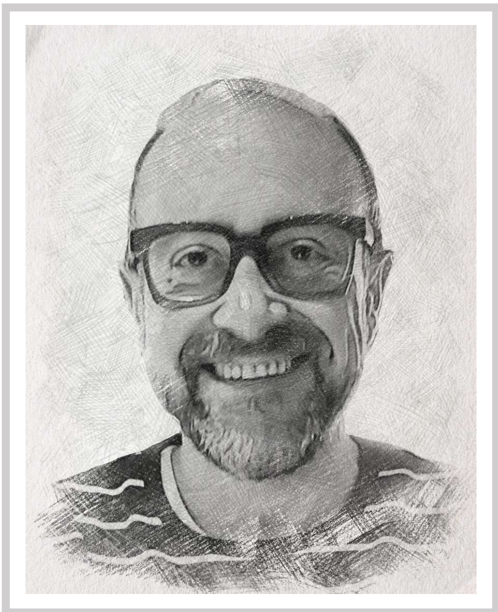
Entre 1991 y 1994 hay que destacar una serie de coyunturas que contribuyeron a decantar la balanza de la dedicación hacia el tema que ocupa esta sección y el contenido mayoritario de este documento:

- En el curso 1991-1992 obtuve la cátedra de Estadística e Investigación Operativa en la Universidad de Oviedo. Siguiendo la política que había aprendido de su maestro, el profesor Sixto Ríos García, Pedro me animó a emprender líneas de trabajo nuevas y a asumir independencia, sin eludir posibles colaboraciones esporádicas.
- En ese mismo curso se incorporó al área de Estadística e Investigación Operativa del Departamento de Matemáticas de la Universidad de Oviedo Miguel López Díaz, recién licenciado en Matemáticas por la Universidad de Zaragoza, y se me designó directora de su tesis doctoral.
- Por aquellos años, en encuentros en algunos congresos, los profesores Lotfi Zadeh (que supervisó mi estancia en Berkeley) y Enric Trillas (matemático que había sido presidente del CSIC, dirigía en aquellos momentos el Instituto Nacional de Técnica Aeroespacial y, junto con Azorín Poch, era uno de los principales impulsores, divulgadores y expertos de la teoría de conjuntos *fuzzy* en España) nos alentaron a continuar con la investigación en el análisis de datos *fuzzy*.

Fue, sin duda, la tesis doctoral de **Miguel López Díaz** el principal detonante para continuar con los conjuntos *fuzzy* aleatorios. El germen de la misma consistía en la formalización del problema de decisión estadística con utilidades o pérdidas con valores *fuzzy*, el establecimiento de un conjunto de axiomas de racionalidad que garantizaran la existencia de una función de utilidad *fuzzy* concordante con las preferencias del decisor, la extensión del análisis bayesiano y las equivalencias de sus formas normal y extensiva en ese contexto.

Personalmente, el análisis de decisión y el enfoque de la estadística bayesianos me habían llamado poderosamente la atención desde la licenciatura. Seguramente influyera en ello que el primer libro de decisión que estudiábamos en los setenta del siglo XX, en la especialidad de Estadística en la Universidad de Valladolid, fue el escrito por Meyer A. Girshick y David H. Blackwell (1954), este último también supervisor de mi estancia en Berkeley y académico de la National Academy of Sciences. A ello se unió, sin duda, mi admiración profunda por los profesores expertos en decisión Javier Girón y Sixto Ríos García, académicos históricos de la RAC, el primero amigo y compañero de doctorado de Pedro Gil y el segundo, como ya se ha dicho, su mentor y maestro. Con los años, Javier sería nuestro asesor principal para la docencia de teoría de la decisión en la licenciatura de Matemáticas en Oviedo, que compartimos Miguel y yo.

En el desarrollo temprano de la tesis de Miguel, se revelaron varios de los rasgos más característicos del Miguel matemático: la necesidad imperativa y loable de un rigor absoluto, su gusto por los problemas que concernieran a contextos generales y su inclinación notoria por los aspectos probabilísticos y analíticos. Como consecuencia, la tesis fue adquiriendo unas proporciones muy por encima de las previstas que involucraron, entre otras, el examen de diversas cuestiones sobre la medibilidad, la integración iterada y el intercambio del orden de integración para los conjuntos *fuzzy* aleatorios y fue el punto de partida de la línea de investigación sobre los fundamentos probabilísticos de los mismos dentro del departamento a la que se referirá el Capítulo 2.



Miguel López Díaz desempeñó un papel clave en el establecimiento de muchos de los fundamentos probabilísticos para el análisis de datos *fuzzy*

Siempre he juzgado conveniente la codirección de trabajos, por cuanto he entendido que tal codirección permitía aprovechar una sinergia de los conocimientos, las experiencias y las pericias de los codirectores. De hecho, la única tesis doctoral que supuestamente no codirigí fue la de Miguel. Añado ‘supuestamente’, porque siempre he declarado que, en realidad, también se trataba de una tesis ‘codirigida’: Miguel había actuado como ‘codirector’.

Miguel mostró, desde bien joven, una gran madurez científica que el tiempo fue realzando. Como ejemplo ilustrativo de esa afirmación, un año después de defender su propia tesis ya había codirigido otra. En total, codirigimos cinco tesis en las que imprimió su sello de profesionalidad inconfundible. En lo relativo a los resultados probabilísticos en ellas, su repercusión fue determinante para configurar un marco formal impecable y sin fisuras, del que se ha nutrido el análisis de datos *fuzzy* posterior.

Investigadores que han colaborado en los estudios probabilísticos desarrollados por el Departamento de Estadística e I.O. y D.M. de la Universidad de Oviedo sobre los conjuntos *fuzzy* aleatorios. Entre 1997 y 2004, las tesis doctorales de los cinco primeros fueron codirigidas por Miguel López Díaz y la del último dirigida por él

Los problemas relativos a los **fundamentos probabilísticos de los conjuntos *fuzzy* aleatorios** que fueron resolviéndose a lo largo de tesis y trabajos en el departamento y que se resumirán en el Capítulo 2 de este discurso son:

- El establecimiento de métricas oportunas entre datos *fuzzy*.
- El establecimiento de un encaje isométrico entre el espacio de datos *fuzzy* y un espacio de funciones ('identificando' cada dato *fuzzy* con cierta función de un espacio de Hilbert, la métrica sobre el primer espacio con una métrica del segundo e induciendo la aritmética entre datos *fuzzy* a partir de la aritmética usual entre funciones).
- El establecimiento de la equivalencia entre la medibilidad nivel a nivel de los conjuntos *fuzzy* aleatorios introducida por Puri y Ralescu y la medibilidad tipo Borel sugerida por Maurice R. Fréchet (1948) para los elementos aleatorios particularizada al caso de valores de conjunto *fuzzy*. Como consecuencia de la equivalencia, se derivan de forma inmediata la distribución de los conjuntos *fuzzy* aleatorios, su independencia estocástica, etc.

Aplicando la política de emancipación aprendida de mis ancestros científicos Sixto Ríos García y Pedro Gil, desde mediados de la primera década del siglo XXI Miguel se dedicó de forma casi exclusiva al estudio de los órdenes estocásticos. Era evidente su predilección por las matemáticas requeridas para dicho estudio, pero su tremendo sentido de la lealtad le 'conminaba' a permanecer trabajando en el tema de los conjuntos *fuzzy* aleatorios. Gracias a Pedro, pudimos convencerle de que los órdenes estocásticos supondrían una línea nueva muy interesante para el Departamento, de modo que sus contribuciones y colaboraciones en el tema anterior ya solo fueron puntuales (algunos artículos recopilatorios o la edición de algún número especial de revista).

A finales del siglo XX, se iniciaron varios estudios propios del análisis de datos *fuzzy*. Mientras que para los de naturaleza probabilística se había contado como punto de partida con modelos, conceptos y resultados establecidos por Puri y Ralescu, para los del análisis de datos no se disponía de antecedentes.

Podría considerarse que los trabajos pioneros al respecto partieron de nuestro Departamento de Estadística e I.O. y D.M. en la Universidad de Oviedo y, de forma simultánea, del grupo del Profesor Wolfgang Näther en la Universidad de Freiberg en Alemania. Desafortunadamente, aunque se estableció contacto con Näther y uno de sus discípulos, Ralf Körner, la jubilación del primero y el paso del otro al mundo de las finanzas, apenas permitieron esbozar una colaboración que sin duda habría sido muy provechosa. Aunque ellos se habían ocupado mayoritariamente de estudios de regresión lineal, coincidíamos en varios intereses investigadores en mente y en el afán por la formalización matemáticamente rigurosa.

Aparte de los problemas de decisión estadística con pérdidas/utilidades *fuzzy*, ya referidos, otros **problemas del análisis de datos *fuzzy*** estudiados han sido:

- Desarrollo de métodos inferenciales sobre medidas resumen de la distribución de los conjuntos *fuzzy* aleatorios: la media tipo Aumann, las medidas de desigualdad y la varianza tipo Fréchet.
- Definición de medidas resumen robustas de localización y escala para conjuntos *fuzzy* aleatorios y análisis de su robustez frente a cambios de datos o presencia de valores 'atípicos'.
- Análisis de sensibilidad de las medidas resumen de la distribución de los conjuntos *fuzzy* aleatorios respecto de la 'forma' de los datos *fuzzy*.
- Análisis estadístico comparativo entre las medidas resumen de la distribución de los conjuntos *fuzzy* aleatorios con las basadas en otras escalas de valoración para datos imprecisos.

Investigadores que han participado/colaborado en los estudios estadísticos con datos *fuzzy* desarrollados en el Departamento de Estadística e I.O. y D.M. de la Universidad de Oviedo. Quienes aparecen en las dos filas superiores realizaron sus tesis doctorales en el tema. Van Aelst, Filzmoser, Bertoluzza, Coppi y Trutschnig han codirigido algunas de esas tesis

1.3 – Un indicador de la repercusión hasta el momento de la investigación en el discurso

La repercusión y las contribuciones principales de la investigación sobre el análisis de datos *fuzzy* que se han descrito muy sucintamente y se detallarán algo más en los capítulos siguientes, son las relativas a la segunda aproximación, tanto en lo concerniente a los resultados probabilísticos como a los métodos inferenciales desarrollados.

El impacto científico queda puesto de manifiesto en el hecho de la aparición de términos/palabras clave vinculados/as a dicha aproximación entre los nuevos códigos de las dos últimas versiones de la *Mathematics Subject Classification*. Esta consiste en una clasificación alfanumérica derivada de la colaboración de expertos de las dos bases de datos más relevantes en Matemáticas: *Mathematical Reviews* y *Zentralblatt MATH*.

En la actualización realizada sobre la MSC2000 para elaborar la MSC2010 (https://mathscinet.ams.org/mathscinet/msc/msc2010.html), se incorporaron códigos nuevos referidos a *fuzzy* y *fuzziness* en Probability y Statistics. Concretamente:

MSC2010	Descripción
60-XX	**Probability theory and stochastic processes**
60A86	*Fuzzy probability*
62-XX	**Statistics**
62A86	*Fuzzy analysis in statistics*
62B86	*Fuzziness, sufficiency, and information*
62C86	*Decision theory and fuzziness*
62E86	*Fuzziness in connection with the topics on distributions in this section*
62F86	*Parametric inference and fuzziness*
62G86	*Nonparametric inference and fuzziness*
62H86	*Multivariate analysis and fuzziness*
62J86	*Fuzziness, and linear inference and regression*
62K86	*Fuzziness and design of experiments*
62L86	*Fuzziness and sequential methods*
62M86	*Inference from stochastic processes and fuzziness*
62N86	*Fuzziness, and survival analysis and censored data*

En la MSC2020 (https://mathscinet.ams.org/mathscinet/msc/msc2020.html) se ha mantenido esa clasificación, que parece responder al crecimiento claro de la investigación sobre el tema y en la que las bases de datos muestran la incidencia de los trabajos a los que va a referirse este discurso.

1.4 – Acerca del contenido del discurso

La orientación y la elección del contenido del discurso han obedecido a razones diversas. Entre ellas, cabe reseñar haber querido hacer el reconocimiento inexcusable a una tarea desarrollada mayoritariamente por nuestro Grupo de Investigación SMIRE⚡CoDiRE (https://bellman.ciencias.uniovi.es/smire+codire/) y por su predecesor SMIRE y que ha sido pionera en su enfoque y tratamiento.

Por ello, los antecedentes propiamente dichos se reducen esencialmente a los fundamentos probabilísticos de los conjuntos *fuzzy* aleatorios en las dos primeras secciones del Capítulo 2, puesto que las otras dos secciones de ese mismo capítulo y el Capítulo 3 resumen las contribuciones del grupo y su impacto a través de su potencial de investigación y transferencia.

Además, se ha optado por aludir a los principales conceptos y métodos, tanto los de apoyo como los relativos a las contribuciones al análisis de datos *fuzzy*, sin apenas entrar en el detalle de los pormenores matemáticos detrás de los mismos.

Esto último no se ha debido solamente a procurar imprimir un carácter bastante divulgativo al discurso, sino que también subraya algo inherente a la estadística: **el desarrollo** de técnicas para el análisis estadístico de datos (cualquiera que sea su tipología) requiere una formación matemática (concretamente, probabilística) muy sólida por parte de los investigadores que las establecen, si bien **la aplicación** de esas técnicas no precisa de una formación de ese nivel por parte de los usuarios. Para el análisis de datos *fuzzy* y la aplicación de métodos estadísticos, con el objeto de automatizar la organización y presentación de los datos, así como la ejecución de los cálculos asociados a los métodos, se trabaja en la puesta a punto de 'paquetes estadísticos' de programas diseñados ex profeso.

Sin embargo, el *software* estadístico no es una panacea. Es una ayuda prácticamente indispensable hoy en día, máxime cuando la tipología de los datos es compleja, pero los programas difícilmente pueden controlar por sí solos todas las vicisitudes que atañen a los datos y a los métodos (especialmente, las posibles condiciones que deban cumplirse para la aplicación oportuna de tales métodos a esos datos). En este sentido, los usuarios pueden permitirse desconocer los fundamentos teóricos de cada procedimiento de análisis de datos, aunque ello no les exime en ningún momento de tener que verificar las condiciones que deben satisfacer los datos según el procedimiento aplicado para que las conclusiones estadísticas sean fiables, si bien en el caso al que se refiere el Capítulo 3, esas condiciones son a menudo muy poco limitantes.

2. Los elementos aleatorios con valores *fuzzy*. Aspectos probabilísticos

La analítica de datos, sea cual sea su tipología y complejidad, se basa en técnicas estadísticas, fundamentadas a su vez en modelos y en proposiciones y teoremas de carácter probabilístico.

Gracias a este fundamento, pueden establecerse de forma singular y universal medidas del grado de certeza o del grado de error de las conclusiones del análisis estadístico de los datos, así como desarrollarse enfoques y métodos para obtener tales conclusiones.

Aunque en sus inicios los estudios probabilísticos se reducían a colecciones de cálculos sistemáticos relacionados mayoritariamente con juegos de azar, a partir del siglo XVIII empezaron a examinarse aspectos formales, a investigarse resultados teóricos y a ampliar los problemas a los que aplicarlos.

Hacia los años treinta y cuarenta del siglo XX, la probabilidad devino en una rama inequívoca de las matemáticas, al verificarse que las ideas principales de la teoría de la probabilidad podían formularse en términos de una especialidad del análisis matemático, la teoría de la medida, que había tenido un desarrollo notable en los primeros años de ese siglo.

El ingrediente inicial para la aplicación de la teoría de la probabilidad al análisis de datos lo determinan los modelos matemáticos asociados al proceso de la generación aleatoria de tales datos.

El primero de estos modelos se basa en el paradigma del *experimento aleatorio*, entendiendo por tal un proceso cuyo resultado no puede predecirse con certeza antes de que se ejecute. Ese modelo corresponde al *espacio de probabilidad* que constituye una de las aplicaciones más sobresalientes de los espacios de medida.

El concepto matemático de medida es una unificación formal de las medidas de índole geométrica, física e incluso más abstractas. Si bien la noción de medida se remonta a la antigua Grecia, no fue hasta finales del siglo XIX y principios del XX que su estudio pasó a constituir una rama de las matemáticas: la teoría de la medida. Las bases de la teoría moderna de la medida fueron sentándose, entre otros, por Giuseppe Peano, Camille Jordan, **Émile Borel**, **Henri L. Lebesgue**, Johann Radon, Constantin Carathéodory y Maurice R. Fréchet. El trabajo de Bombal (1991) y el libro de Bressoud (2008) ofrecen revisiones excelentes y bien detalladas de la evolución de esta teoría, así como de su conexión con la integración de funciones.

Paul Lévy escribió el primer libro sobre probabilidad basado en ideas de teoría de la medida ("*Calcul des Probabilités*", 1925) y, más tarde, lamentó haber dejado pasar la oportunidad de introducir la formalización siguiente.

El ***espacio de probabilidad*** asociado a un experimento aleatorio se caracteriza por tres componentes:

- el conjunto de los resultados potenciales del experimento (*espacio muestral*),

- una clase no vacía de subconjuntos de resultados, que son aquellos que se pueden e interesa medir en cada situación (a los que se hace referencia como *conjuntos medibles* o *sucesos*) y que se supondrá cerrada para las operaciones unión numerable y complementación de conjuntos (es decir, que es una σ-álgebra), y
- la *probabilidad* de ocurrencia de los sucesos, que se supone que es una medida 'normalizada', es decir, que a cada conjunto de la clase anterior le asigna un valor no negativo, que la probabilidad de que ocurra al menos uno de una sucesión de sucesos que no pueden simultanearse dos a dos es la suma de las probabilidades de esos sucesos, y (como rasgo distintivo de la probabilidad frente a las medidas en general) la de que ocurra un resultado dentro de los potenciales es igual a 1.

Émile Borel, Henri Lebesgue, Sergei Bernstein, Paul Lévy y Andréi Kolmogórov

Esta caracterización es la publicada por **Andréi N. Kolmogórov** (1933) y sirvió, y sigue sirviendo, como fundamento de la teoría de la probabilidad. Maurice R. Fréchet (1938) destacó la habilidad de Kolmogórov al organizar y sintetizar de una forma muy adecuada una teoría que Émile Borel había creado muchos años antes, combinando la aditividad numerable con la probabilidad clásica, y a la que muchos otros fueron añadiendo aportaciones formales (entre ellos, **Sergei Bernstein** llevó a cabo en 1917 un primer acercamiento a la axiomatización de la probabilidad). Fréchet señalaba que en el momento (1909) en que Borel incorporó la aditividad numerable en el cálculo de probabilidades se habían reunido todos los elementos necesarios para formular explícitamente todo el cuerpo de axiomas de la teoría de la probabilidad, si bien Kolmogórov había tenido el acierto de formalizarlo y corroborar que con tales condiciones axiomáticas no se necesitaba nada más para construir la teoría.

La necesidad de una 'axiomatización' de la probabilidad había sido planteada por David Hilbert a principios del siglo XX. Correspondía al sexto de la lista de 23 problemas matemáticos elaborada por Hilbert para el Segundo Congreso Internacional de Matemáticos, celebrado en 1900 en París. Ese sexto problema consideraba la axiomatización de la probabilidad como parte de la axiomatización de las ciencias físicas.

En realidad, más que una axiomatización propiamente dicha, Kolmogórov introdujo una definición. Postuló que cualquier medida que cumpliera ciertas condiciones (a las que tradicionalmente suele referirse como 'axiomas' de no negatividad, normalización y σ-aditividad) debía denominarse *medida de probabilidad*. Y el concepto de medida se había presentado previamente como una función que satisfacía las dos primeras de esas condiciones.

Si bien esa definición impone restricciones o propiedades mínimas exigibles para la probabilidad, ni configura su interpretación filosófica ni ofrece pautas para su cálculo práctico, como ya Hilbert defendía que debía ser. Pero las interpretaciones filosóficas principales (clásica, frecuentista y subjetiva) son consistentes con los 'axiomas' de Kolmogórov.

El segundo modelo matemático es el ligado al mecanismo aleatorio generador de los datos y concierne a las '*magnitudes' aleatorias*, entendiéndose estas como procesos que asignan a cada resultado experimental (es decir, a cada elemento del espacio muestral) un 'valor' según la escala de medición adoptada. Su formalización en conexión con el espacio de probabilidad que modeliza el experimento subyacente, consiste en caracterizar la magnitud aleatoria mediante una función definida sobre el espacio muestral de dicho experimento y con llegada en el conjunto de los valores potenciales con la escala de medición considerada.

Para garantizar que puedan calcularse las probabilidades de que una magnitud aleatoria tome valores en subconjuntos de interés del espacio de llegada (probabilidades que determinan la denominada ***distribución*** de esa magnitud) se impone una condición sobre la función correspondiente, también inspirada en la teoría de la medida: su *medibilidad*. Una *función medible*, noción introducida por **Henri Lebesgue**, es aquella que preserva la estructura entre dos espacios de medida, en el sentido de que la preimagen (también llamada imagen inversa o anti-imagen) de cualquier conjunto medible en el espacio de llegada, que es la colección de los elementos del espacio muestral de partida cuya imagen por la función pertenece a dicho conjunto, sea a su vez un conjunto medible en el de partida. Cuando se particulariza a espacios de probabilidad, la medibilidad a la que se recurre es la llamada ***medibilidad Borel***, en la que se supone que la σ-álgebra sobre el espacio de llegada es la *σ-álgebra de Borel* asociada al espacio muestral correspondiente.

Cuando la magnitud aleatoria toma valores numéricos, se habla de ***variable aleatoria*** y la σ-álgebra de Borel consiste en la menor σ-álgebra que contiene a todos los conjuntos abiertos reales, entendiendo por conjunto abierto real aquel para el que cualquiera que sea el elemento en el mismo existe un entorno abierto (colección de valores numéricos que distan de dicho elemento menos de un número positivo arbitrario) que está totalmente contenido en él. Y la probabilidad de que la variable tome valores en un conjunto cualquiera de esa σ-álgebra de Borel (o *probabilidad inducida* de dicho conjunto) vendrá dada por la probabilidad en el espacio de probabilidad original de la preimagen de dicho conjunto por la variable aleatoria.

Entre las medidas resumen de la distribución de una variable aleatoria, merece mencionarse que la integración de esta respecto de la medida de probabilidad original conduce a la definición de la *media* o *valor esperado de su distribución*. A su vez, coincide con el valor (si existe) que minimiza la media cuadrática (entendida como la raíz cuadrada de la media de los cuadrados) de la distancia usual (la euclídea), cuyo valor mínimo constituye la *desviación estándar de la distribución* de la variable. La media resulta, por tanto, una medida resumen de centralización o localización de la distribución, y cuando en la minimización se reemplaza la media cuadrática por la media, se obtiene la llamada *mediana de la distribución*, que es menos sensible (es decir, más robusta) que la media a los posibles cambios y a la presencia de valores extremos o atípicos.

Los estudios probabilísticos relativos a las variables aleatorias y sus medidas resumen experimentales/poblacionales (parámetros) conciernen, entre otros, a:

- la caracterización funcional de su distribución (a través de las funciones de probabilidad/masa o de densidad, de la función de distribución o de la función característica);
- los modelos de distribuciones, o clases paramétricas, que modelizan de forma exacta, límite o idealizada las de muchas variables de la vida real y facilitan numerosos desarrollos y cálculos prácticos;
- y un buen número de resultados límite, como las leyes de los grandes números, los teoremas del límite central y, en general, los referidos a grandes muestras o las aproximaciones *bootstrap*, con las implicaciones inferenciales correspondientes.

2.1 – Los elementos aleatorios de naturaleza cualquiera según Fréchet

La noción de variable aleatoria se fue extendiendo a lo largo de los años. La más inmediata de esas extensiones fue la asociada a las magnitudes que involucraban a su vez la observación/medición simultánea de un número finito de magnitudes modelizadas mediante variables aleatorias. El modelo matemático dio lugar a los denominados *vectores aleatorios* (o variables aleatorias multidimensionales), en los que la σ-álgebra del espacio de llegada sería la de Borel en el espacio euclídeo de dimensión igual al número de observaciones/mediciones simultáneas y entendiendo por conjunto abierto aquel para el que cualquiera que sea el elemento en el mismo existe una bola abierta (colección de valores vectoriales que distan de dicho elemento menos de un número positivo prefijado) que está totalmente contenida en él. Este modelo supuso un fundamento crucial para, por un lado, el análisis de datos multivariantes y, por otro, la formalización del modelo base de la inferencia estadística: la muestra aleatoria.

En su artículo de 1948 *"Les éléments aléatoires de nature quelconque dans un espace distancié"*, en la revista *Annales de l'Institut Henri Poincaré*, **Maurice R. Fréchet** señalaba que la teoría de la probabilidad se había ocupado sucesivamente de los números aleatorios, los puntos aleatorios en el plano o en el espacio, etc. (i.e., los vectores aleatorios) y que en los últimos años se habían comenzado a estudiar elementos aleatorios más complejos como las funciones aleatorias. Apuntaba que, si bien parecía que en cada caso se tenía que construir una teoría específica para el nuevo tipo de elementos aleatorios, en realidad bastaba con reproducir de forma muy simple un esquema común cuyo rasgo diferencial radicaba en la introducción de una distancia entre dos valores potenciales cualesquiera de tales elementos aleatorios.

De hecho, con el propósito de construir un marco abstracto general para el estudio de los límites, la continuidad y la convergencia, en la investigación conducente a su tesis doctoral, *"Sur quelques points du calcul fonctionnel"*, Fréchet (1906) había introducido los espacios métricos (término acuñado pocos años más tarde por Felix Hausdorff en su obra sobre los fundamentos de la teoría de conjuntos, *"Grundzüge der Mengenlehre"*).

A principios del siglo XX, el enfoque habitual de las matemáticas era más bien poco abstracto y axiomático. Como consecuencia, dependiendo de los espacios con los que se trabajara, podían aparecer diferentes nociones de convergencia, si bien existían algunas similitudes entre tales nociones, pero no había una comprensión general del término. La introducción por Fréchet de la idea de ***espacio métrico***, basada en la definición axiomática de ***métrica*** (o ***distancia***) dio pie al establecimiento de conceptos y a la demostración de resultados asociados generales que serían aplicables a todos los ejemplos específicos conocidos previamente.

Maurice R. Fréchet

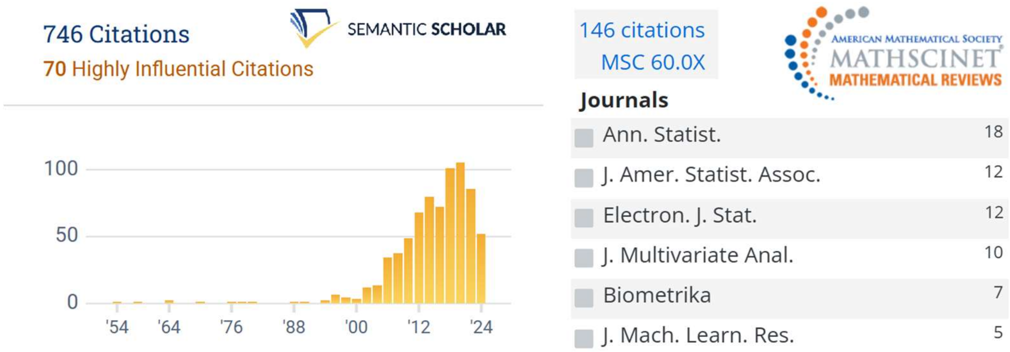
La mera consideración de un espacio de elementos de cualquier naturaleza, dotado de una métrica oportuna (más concretamente de una de las denominadas tipo L^2), permitía extender los rasgos esenciales de las magnitudes aleatorias introducidas hasta ese momento, definiéndose los ***elementos aleatorios generales*** sin dificultades formales, de manera que:

- la **medibilidad Borel**, en la que la σ-álgebra sobre el espacio de llegada, que incluiría a los valores potenciales del elemento aleatorio, sería la menor σ-álgebra que contuviera a todos los conjuntos abiertos, entendiendo por tales aquellos para los que, cualquiera que fuera el punto (elemento) considerado en dicho espacio, el conjunto de puntos del mismo que con la métrica adoptada distaran del punto inicial menos de un número positivo arbitrario estuviera contenido totalmente en él;
- al igual que para las variables aleatorias, la **distribución del elemento aleatorio** se induciría de forma inmediata a partir de dicha medibilidad Borel, sin necesidad de recurrir a una definición nueva establecida de forma expresa; y, como consecuencia, se establecerían las tres siguientes medidas:
  - el **error cuadrático medio asociado al elemento aleatorio** respecto de un punto genérico del espacio de llegada y para la métrica L^2, vendría dado por la media de la variable aleatoria consistente en el cuadrado de la distancia de los valores del elemento aleatorio al punto genérico;
  - la **media del elemento aleatorio** correspondería al punto, si existiera, en el espacio de llegada en el que se alcanzara el mínimo absoluto de la función real error cuadrático medio asociado al elemento aleatorio;

– la **varianza del elemento aleatorio** sería el mínimo absoluto del error cuadrático medio anterior.

Si bien muchas de las contribuciones más sobresalientes para el progreso de las matemáticas se han dirigido a resolver grandes preguntas, conjeturas, etc., otras han ido ligadas a la propuesta de áreas de investigación nuevas. Fréchet reconoció que su interés investigador se encuadraba dentro de esta última categoría. Como ejemplo, con la realización de su tesis doctoral inició en 1906 el estudio sobre funcionales en un espacio métrico, lo que sirvió de punto de partida de un tema tan fructífero como el análisis funcional. Y otro ejemplo muy destacable es el trabajo aludido de 1948, que sirvió y sigue sirviendo como fundamento obligado para formalizar las magnitudes aleatorias que generan datos de cualquier tipología.

De hecho, puede señalarse que tanto en *Mathematical Reviews* como en *Zentralblatt MATH*, o en la revisión más general *Semantic Scholar*, los dos ejemplos mencionados son los dos trabajos de Maurice Fréchet que han recibido más citas. Más específicamente, en lo concerniente a *"Les éléments aléatoires de nature quelconque dans un espace distancié"*, la herramienta de búsqueda de literatura científica *Semantic Scholar* evidencia (véase la imagen izquierda de la siguiente figura) que la mayoría de las citas recogidas lo han sido en el siglo XXI, en buena parte desde 2012. Y de las citas en revistas matemáticas recogidas en la base *MathScinet*, buena parte de ellas provienen de las más prestigiosas en Estadística y Probabilidad (véase la imagen derecha de la figura siguiente).



Evolución temporal de las citas recibidas hasta mayo de 2025 por el trabajo en el que Maurice Fréchet introdujo los elementos aleatorios según *Semantic Scholar* (izquierda) y revistas matemáticas con más citas a dicho trabajo según *MathScinet* (derecha)

2.2 – Los elementos aleatorios con valores *fuzzy*

Hacia finales de los cuarenta del siglo pasado, cerca de veinte años antes de que Zadeh introdujera la teoría de conjuntos *fuzzy*, Fréchet señalaba en su trabajo de 1948 ya referido que:

«... la naturaleza, la ciencia y la tecnología ofrecen muchos ejemplos de elementos aleatorios que no son ni números, ni series, ni vectores, ni funciones…»

«... También podemos encontrar elementos que aún no se han descrito matemáticamente.»

«... podemos considerar también elementos asociados a ciudades elegidas al azar y que no pueden describirse mediante las nociones matemáticas habituales: número, función, curva, etc. Por ejemplo, la moralidad de su población, su tendencia política, la impresión de belleza que transmite, etc.
Se trata de una categoría nueva de elementos aleatorios.»

Parecía que, en buena parte, Fréchet estuviera anticipando con esos últimos ejemplos los elementos aleatorios con valores *fuzzy* que irían formalizándose en la segunda parte del siglo XX. De hecho, los primeros estudios sobre estos elementos los llevó a cabo a principios y mediados de los setenta Robert L. Féron, quien fuera discípulo de Fréchet y que en 1954 defendió en la Universidad de París su tesis doctoral *"Information, Régression, Corrélation"*, realizada bajo la supervisión de Maurice Fréchet y George Darmois.

Robert L. Féron introdujo la noción de *conjunto fuzzy aleatorio* (1976ab, 1979) como una aplicación del espacio medible asociado al experimento aleatorio en otro espacio medible, a través de dos definiciones. Una de ellas seguía el enfoque de los elementos aleatorios según Fréchet, exigiendo que el espacio de llegada estuviera establecido sobre un espacio métrico y que se satisficiera la medibilidad de la aplicación entre los espacios involucrados. No obstante, ni se especificaba la métrica, ni se especificaba que la medibilidad debía ser la correspondiente a la σ-álgebra de Borel para la topología inducida por la métrica sobre el espacio de llegada. La otra definición extendía nivel a nivel el concepto de conjunto aleatorio, según la noción dada por Robert Fortet (también discípulo de Fréchet, varios años antes, y supervisor, años más tarde, de la tesis doctoral de George Matheron, uno de los pioneros en el estudio de esos elementos aleatorios) y Mehri Kambouzia (1975). Tampoco se explicaban las relaciones entre ambas aproximaciones, ni se proponían nociones sobre media, varianza, etc.

Hacia el inicio de los ochenta, **Madan L. Puri** y **Dan A. Ralescu** (el primero doctorando de Erich L. Lehmann por la Universidad de California en Berkeley en 1962 y el segundo doctorando del primero por la Universidad de Indiana en Bloomington en 1980) iniciaron su colaboración en el tema de los conjuntos *fuzzy* aleatorios (que, como ya se comentó, acuñaron con el nombre de *variables aleatorias fuzzy*). Sus tesis doctorales habían versado sobre estadística matemática (contrastes de hipótesis y estimación), aunque Ralescu había colaborado en Bucarest durante algunos años con Constantin V. Negoita en diversos estudios matemáticos sobre la lógica *fuzzy* y el análisis de sistemas. Por su parte, Puri se había centrado en distintos problemas de probabilidad y de estadística, y sus primeras investigaciones sobre funciones con valores *fuzzy* datan de algo después de que Ralescu se uniera a su grupo de trabajo.

Robert Féron, Lotfi Zadeh, Erich Klement, Dan Ralescu y Madan Puri

A lo largo de su estancia en la Universidad de Indiana en Bloomington (entre 1977 y 1980), Dan Ralescu inició una colaboración muy fructífera con Madan Puri. Tras defender en septiembre de 1980 su tesis doctoral *"Admissibility of Estimators in the One Parameter Exponential Family and in Multivariate Location Problems"* bajo la dirección de Puri, Ralescu se unió a la Universidad de Cincinnati en Ohio, sin abandonar su investigación sobre funciones con valores de conjunto *fuzzy*, su diferenciación e integración y algunos encajes isométricos del espacio de esos valores, bien fuera solo o en compañía, principalmente de Puri. En los primeros trabajos conjuntos de Puri y Ralescu sobre el tema aparecían como referencias ineludibles tanto el trabajo de Robert Féron de 1976 como un *preprint* del que su publicación tras aceptación se demoró algunos años y, a día de hoy, es con seguridad su artículo más citado (en torno a las 1500 citas en las bases de datos bibliométricas habituales –en *Google Scholar* figuran más de 2650 en mayo de 2025–).

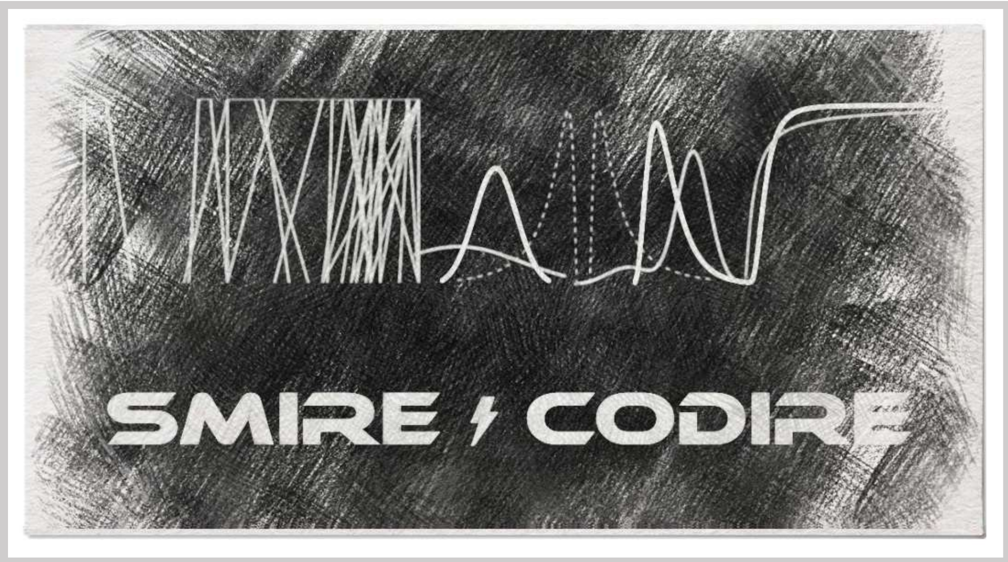
En el trabajo *"Fuzzy random variables"* de Madan L. Puri y Dan A. Ralescu, publicado en 1986, sus autores formalizaron tales variables como los **conjuntos *fuzzy* aleatorios** (término que encontramos más adecuado y que emplearemos en lo que sigue) introducidos por Robert Féron según la segunda definición mencionada, es decir, como funciones con valores de conjunto *fuzzy* tales que sus funciones nivel fueran conjuntos aleatorios, entendiendo por ello que fueran medibles Borel con respecto de la conocida distancia de Hausdorff sobre el referencial. Además de esta concreción, que no había explicitado Féron, Puri y Ralescu establecieron el concepto de valor medio (o **media tipo Aumann**) de un conjunto *fuzzy* aleatorio como aquel conjunto *fuzzy* cuyo α-nivel coincidiera con la integral de Aumann del conjunto aleatorio α-nivel.

No resultaba evidente que los conjuntos *fuzzy* aleatorios según tal definición fueran elementos aleatorios en el sentido de Fréchet. Aunque no se hiciera referencia expresa al trabajo de Fréchet (1948) sobre los elementos aleatorios, los trabajos de Puri y Ralescu (1985) y el conjunto de ellos con Erich P. Klement (1986) parecían querer dar respuesta a la cuestión de si la medibilidad Borel nivel a nivel exigida en la definición anterior era equivalente a la medibilidad Borel global respecto de alguna métrica sobre el espacio de valores *fuzzy* considerado. Esos artículos, que se escribieron con posterioridad al de Puri y Ralescu de 1986 a pesar de publicarse con antelación, o al mismo tiempo, definieron un conjunto *fuzzy* aleatorio como una función con valores de conjunto *fuzzy* que fuera medible Borel (i.e., elemento aleatorio) respecto de la métrica del supremo, o supremo respecto de $\alpha \in [0,1]$ de las distancias Hausdorff entre los α-niveles de los valores *fuzzy* correspondientes.

No obstante, no examinaron en estos trabajos si las dos definiciones eran o no equivalentes, probándose años más tarde que no lo eran. Es más, la métrica del supremo no confería de separabilidad al espacio de valores *fuzzy*, lo que implicaba algunos inconvenientes, si bien permitía encajar el espacio de los conjuntos *fuzzy* en $\mathbb{R}^p$ en el espacio de Banach de funciones continuas en $[0,1] \times \mathbb{S}^{p-1}$, donde $\mathbb{S}^{p-1}$ es la esfera unidad en $\mathbb{R}^p$ (es decir, de los elementos en $\mathbb{R}^p$ con norma euclídea no superior a 1). Este encaje se llevaba a cabo a través de la **función soporte de un conjunto *fuzzy*** (Puri y Ralescu, 1985), inspirada en la función soporte de conjuntos cerrados, convexos y no vacíos introducida en 1903 por Minkowski.

2.3 – Contribuciones principales sobre aspectos probabilísticos de los conjuntos *fuzzy* aleatorios del Grupo de Investigación SMIREϟCoDiRE de la Universidad de Oviedo

Aunque, como se ha señalado, el primer acercamiento de nuestro grupo en la Universidad de Oviedo a los conjuntos *fuzzy* aleatorios según Puri y Ralescu tuvo lugar al considerarlos como base para estudios de problemas de decisión estadística con pérdidas/utilidades *fuzzy*, la formalización y el desarrollo metodológico de tales problemas desencadenaron varios de índole probabilística. El grupo de investigación responsable de estos estudios, y de los posteriores de carácter estadístico, ha adoptado con los años el acrónimo SMIREϟCoDiRE.



El Grupo de Investigación de la Universidad de Oviedo SMIREϟCoDiRE, que he coordinado hasta mi jubilación (https://bellman.ciencias.uniovi.es/smire+codire/), aglutina buena parte de los trabajos sobre el análisis estadístico de datos *fuzzy* y sobre varios de sus fundamentos probabilísticos

Entre estos merecen atención especial los relativos a dilucidar si los conjuntos *fuzzy* aleatorios según la definición nivel a nivel de Puri y Ralescu eran elementos aleatorios en el sentido de Fréchet. El interés primordial de esta discusión residía en el hecho de que, de ser así, nociones como, entre otras, la de la **distribución de un conjunto *fuzzy* aleatorio** o la de la **independencia de conjuntos *fuzzy* aleatorios** asociados a un mismo experimento se inducirían de forma inmediata, sin necesidad de establecerlas de forma más bien artificial.

Estudios mediante métricas complejas condujeron a la conclusión de que la medibilidad Borel nivel a nivel equivalía a la medibilidad Borel global con métricas como, entre ellas, la mid/spr definida por SMIREϟCoDiRE, con una interpretación muy intuitiva y gran sencillez de cálculo, principalmente en el caso unidimensional. Además, la métrica dotaría de separabilidad al espacio de valores *fuzzy*, de manera que la σ-álgebra se generaría a partir de un conjunto numerable de bolas abiertas del espacio de valores.

Como implicación notable del hecho de que los conjuntos *fuzzy* aleatorios fueran elementos aleatorios en el sentido de Fréchet (1948) con la métrica tipo L^2 mid/spr, en SMIRE⚡CoDiRE se comprobó la coherencia de la media tipo Aumann de un conjunto *fuzzy* aleatorio definida por Puri y Ralescu (1986) con el enfoque de Fréchet, ya que se trataba del argumento de la minimización de la función error cuadrático medio basada en dicha distancia, definiéndose la varianza del conjunto *fuzzy* aleatorio como el valor numérico del error cuadrático medio mínimo.

Una vez disponible el modelo para los mecanismos que generan aleatoriamente datos con valor de conjunto *fuzzy*, modelo que cuenta con la aquiescencia plena del marco probabilístico, el grupo SMIRE⚡CoDiRE emprendió el desarrollo de una metodología para el análisis estadístico de tales datos. Como en el caso de datos numéricos, el objetivo final era extraer conclusiones acerca de la distribución de uno o varios conjuntos *fuzzy* aleatorios observados en una o varias poblaciones, sobre la base de datos obtenidos por muestreo de esas poblaciones.

Inicialmente, la idea directriz era conservar en lo posible los conceptos y los métodos del análisis estadístico inferencial con datos numéricos, de manera que la filosofía que sustenta a la estadística no se viera afectada por el tipo de datos a tratar. Sin embargo, hay una serie de rasgos del caso *fuzzy* que lo distinguen sustancialmente del numérico, de modo que la metodología estadística con datos numéricos no podría extenderse, ni directa ni fácilmente, al tratamiento de datos con valores de conjunto *fuzzy*. Entre estos rasgos distintivos, cabe señalar:

- ⊗ no puede establecerse un 'operador diferencia' en el espacio de conjuntos *fuzzy*, de forma que siempre esté bien definido (es decir, que esté definido para cualquier par de valores del espacio) y que dote de estructura lineal a dicho espacio;
- ⊗ no existe un orden completo en ese espacio (es decir, que permita comparar cualquier par de valores del espacio) que sea universalmente aceptable;
- ⊗ no se han definido aún modelos de distribución suficientemente generales para los conjuntos *fuzzy* aleatorios, ni tan siquiera en el caso unidimensional; si bien Puri y Ralescu (1985) establecieron la noción de conjunto *fuzzy* aleatorio normal, se trataría de un modelo poco realista y con muy escasa aplicación;
- ⊗ no se dispone de teoremas límite, tipo los teoremas del límite central, que sean aplicables directamente con fines inferenciales.

El primero de estos inconvenientes, y también el cuarto como consecuencia de algunos resultados de aproximación adicionales, se ha soslayado en muchos de los problemas estadísticos recurriendo a una adaptación del encaje isométrico del espacio de valores de conjunto *fuzzy* en un espacio de Banach de funciones continuas, que habían establecido Puri y Ralescu (1985) a través de la *función soporte* y de la métrica del supremo. De este modo, el grupo SMIRE⚡CoDiRE ha corroborado que ese encaje seguía siendo válido cuando la métrica adoptada era la mid/spr, con la propiedad añadida de que al manejarse una distancia tipo L^2 el encaje lleva a un espacio de Hilbert separable (y, por lo tanto, lineal y equipado con un producto interior) de funciones de cuadrado integrable. Más concretamente, la imagen por la función soporte del espacio de valores de conjunto *fuzzy* resulta ser un cono convexo (es decir, cerrado bajo combinaciones lineales con coeficientes positivos) de dicho espacio de Hilbert.

Como se verá en el capítulo siguiente, este resultado es crítico en muchos de los métodos del análisis de datos *fuzzy*, pero sus implicaciones incumben también a los fundamentos de los conjuntos *fuzzy* y de los conjuntos *fuzzy* aleatorios.

Entre las consecuencias más relevantes derivadas del encaje isométrico establecido por SMIRE⚡CoDiRE, pueden destacarse:

- cada valor de conjunto *fuzzy* del espacio de partida puede identificarse biyectivamente con una función de un espacio de Hilbert (en particular, con una función del cono convexo en que se encaja por medio de la función soporte), coincidiendo la norma de la diferencia entre dos funciones del cono con la distancia mid/spr entre sus antecedentes;
- la **aritmética** (suma y producto por un número real) entre valores de conjunto *fuzzy* que induciría la aritmética usual entre sus funciones soporte correspondientes coincide con la aritmética que anticipó Zadeh (1975ab) sobre la base de su *principio de extensión*;
- una función con valores de conjunto *fuzzy* es un conjunto *fuzzy* aleatorio si y solo si su composición con la función soporte es un elemento aleatorio con valores en un cono convexo de un espacio de Hilbert de funciones;

lo que, unido a las ideas de Fréchet (1948), ha permitido a SMIRE⚡CoDiRE:

- verificar que la integral de Bochner de dicho elemento aleatorio coincide con el valor de la función soporte en la **media tipo Aumann** del conjunto *fuzzy* aleatorio;
- definir la **varianza tipo Fréchet** del conjunto *fuzzy* aleatorio como la distancia mid/spr cuadrática media mínima en el espacio de valores *fuzzy* considerado; ese valor coincide, a su vez, con la varianza del elemento aleatorio obtenido tras composición con la función soporte.

Al mismo tiempo, al ser la métrica mid/spr tipo L^2 y llegar mediante la función soporte a un espacio de Hilbert, la existencia del producto interior en el mismo ha posibilitado definir la **covarianza** de dos conjuntos *fuzzy* aleatorios como la correspondiente a los elementos aleatorios imagen.

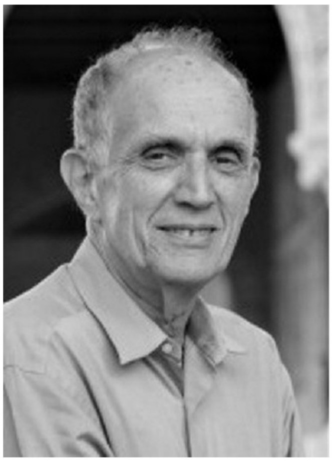
Otros resultados de interés para el análisis de datos *fuzzy* han sido los relativos a teoremas límite. Entre estos, para refrendar la idoneidad de la media tipo Aumann, en SMIRE⚡CoDiRE se extendió la *Ley Fuerte de los Grandes Números* para conjuntos *fuzzy* aleatorios con la métrica del supremo, la cual implicaba el cumplimiento de dicha ley con la métrica mid/spr y prácticamente cualquier otra métrica que pudiera considerarse en el espacio de valores *fuzzy*. De ello se deriva que la media tipo Aumann es casi seguro a través de la métrica considerada el límite de la sucesión de las medias muestrales que se obtienen con la aritmética que recurre al principio de extensión de Zadeh y será útil con fines de garantizar la estimación consistente de la media poblacional mediante la muestral.

2.4 – Aproximación crucial del teorema del límite central para conjuntos *fuzzy* aleatorios

Por otro lado, algunos autores (véanse, por ejemplo, Proske y Puri, 2002, y Krätschmer, 2002) desarrollaron *teoremas del límite central* para conjuntos *fuzzy* aleatorios.

De acuerdo con ellos, se obtendría como límite en distribución de la imagen por la función soporte de la media muestral a través de la métrica del supremo un proceso gaussiano, o a través de métricas tipo L^1/L^2 un elemento aleatorio gaussiano, con valores en el espacio de Banach/Hilbert de funciones en el que se encaja el espacio de valores *fuzzy* por medio de la función soporte. Sin embargo, no puede garantizarse que el proceso o elemento aleatorio gaussiano obtenido tome valores en el cono convexo del encaje isométrico. De hecho, difícilmente el límite correspondería a la composición de la función soporte con un conjunto *fuzzy* aleatorio.

Esto redundaría en la inviabilidad de aplicar directamente el teorema del límite central con propósitos inferenciales para los conjuntos *fuzzy* aleatorios. Tal contrariedad va a poder sortearse fácilmente si se tiene en cuenta un resultado establecido por **Evarist Giné** y **Joel Zinn** en 1990, que ofrece una **aproximación *bootstrap* del teorema del límite central** para procesos empíricos generales y, en particular, para elementos aleatorios con valores de función en el espacio de Hilbert en que se encaja isométricamente el espacio de valores de conjunto *fuzzy*.



Bradley Efron

Evarist Giné

Joel Zinn

La ventaja de la particularización de este resultado es que, mediante el *bootstrap* (introducido años antes por **Bradley Efron**, 1979, como una técnica que no requería conocer la distribución poblacional, ni tratar con muestras grandes, sino que bastaría con realizar remuestreo a partir de la muestra de datos disponible) la aproximación del elemento aleatorio límite sería una media muestral a partir de la imagen por la función soporte de un conjunto *fuzzy* aleatorio. Se trataría así de una media muestral de un elemento aleatorio con valores en el cono convexo del encaje y, por lo tanto, involucraría operaciones cerradas dentro de dicho cono, por lo que su valor no se saldría del mismo. En otras palabras, el teorema del límite central para conjuntos *fuzzy* aleatorios no aseguraría en sí mismo que el límite en distribución fuera un conjunto *fuzzy* aleatorio, mientras que sí lo haría su aproximación *bootstrap*.

3. El análisis estadístico de datos *fuzzy* y primeras aplicaciones

Como se ha subrayado anteriormente, el análisis estadístico de datos *fuzzy* no solo parte del tratamiento de datos con una complejidad diferencial, sino con peculiaridades muy específicas. Aunque tales datos puedan identificarse con funciones de un espacio de Hilbert, cualquier método que los analice no puede discurrir fuera del cono convexo del espacio en el que se encajan, y hay que poder capear la falta de linealidad, la de un orden total aceptable y la de modelos de distribuciones. Por contar con los avales probabilísticos que permiten conservar las nociones fundamentales de la estadística con datos numéricos (por ejemplo, el muestreo aleatorio simple, la estimación insesgada o la consistente, el p-valor de un contraste de hipótesis, etc.), se considerarán los conjuntos *fuzzy* aleatorios como los mecanismos generadores de los datos *fuzzy*.

Al igual que en el caso de datos numéricos/vectoriales, el objetivo de los procedimientos de análisis de datos *fuzzy* generados por uno o varios conjuntos *fuzzy* aleatorios sobre una población (o varias) es, por un lado, describir y resumir la distribución de estos elementos aleatorios en una muestra de individuos seleccionados de la población (en muestras elegidas cada una de una de las poblaciones) y, por otro lado, aprovechar esa información muestral para inferir conclusiones sobre la distribución de los conjuntos *fuzzy* aleatorios en la población (en las poblaciones).

Este capítulo recoge de forma sucinta la investigación y las aplicaciones que ha desarrollado y continúa desarrollando SMIRE⚡CoDiRE.

3.1 – Metodología desarrollada para el análisis estadístico de datos *fuzzy* basada en los conjuntos *fuzzy* aleatorios

Una vez disponibles el modelo para los mecanismos que generan aleatoriamente los datos *fuzzy* y los resultados probabilísticos apuntados, desde hace algo más de tres décadas SMIRE⚡CoDiRE ha llevado y continúa llevando a cabo estudios para analizar tales datos, en su mayoría de tipo inferencial.

Los enfoques seguidos para la construcción de los métodos de análisis de datos *fuzzy* son:

- la particularización de procedimientos existentes, o que vayan estableciéndose expresamente, para el Análisis de Datos Funcionales; este enfoque será válido siempre que el proceso involucrado en el procedimiento considerado se mueva dentro del cono convexo en que se encaje isométricamente el espacio de valores de conjunto *fuzzy*;
- el desarrollo de procedimientos *ad hoc*, habitualmente realizado cuando la particularización anterior falla; esos procedimientos recurren a menudo a técnicas de aproximación basadas en la teoría de grandes muestras para variables aleatorias uni- o multidimensionales.

Como ya anticipamos, entre los métodos y estudios establecidos hasta el momento cabe destacar:

- **Inferencias sobre medidas resumen de la distribución de los conjuntos *fuzzy* aleatorios:**
  - estimación 'puntual' insesgada y fuertemente consistente de la media *fuzzy* tipo Aumann poblacional mediante la media muestral correspondiente; estimación por intervalo;
  - estimación puntual insesgada y fuertemente consistente de la varianza numérica tipo Fréchet poblacional mediante la varianza muestral correspondiente; estimación por intervalo;
  - contrastes de hipótesis (bilaterales) de igualdad de medias *fuzzy* tipo Aumann uni-, bimuestrales y ANOVA, los dos últimos para muestras independientes o ligadas;
  - contrastes de hipótesis (uni- y bilaterales) sobre varianzas tipo Fréchet uni- y k-muestrales para muestras independientes.

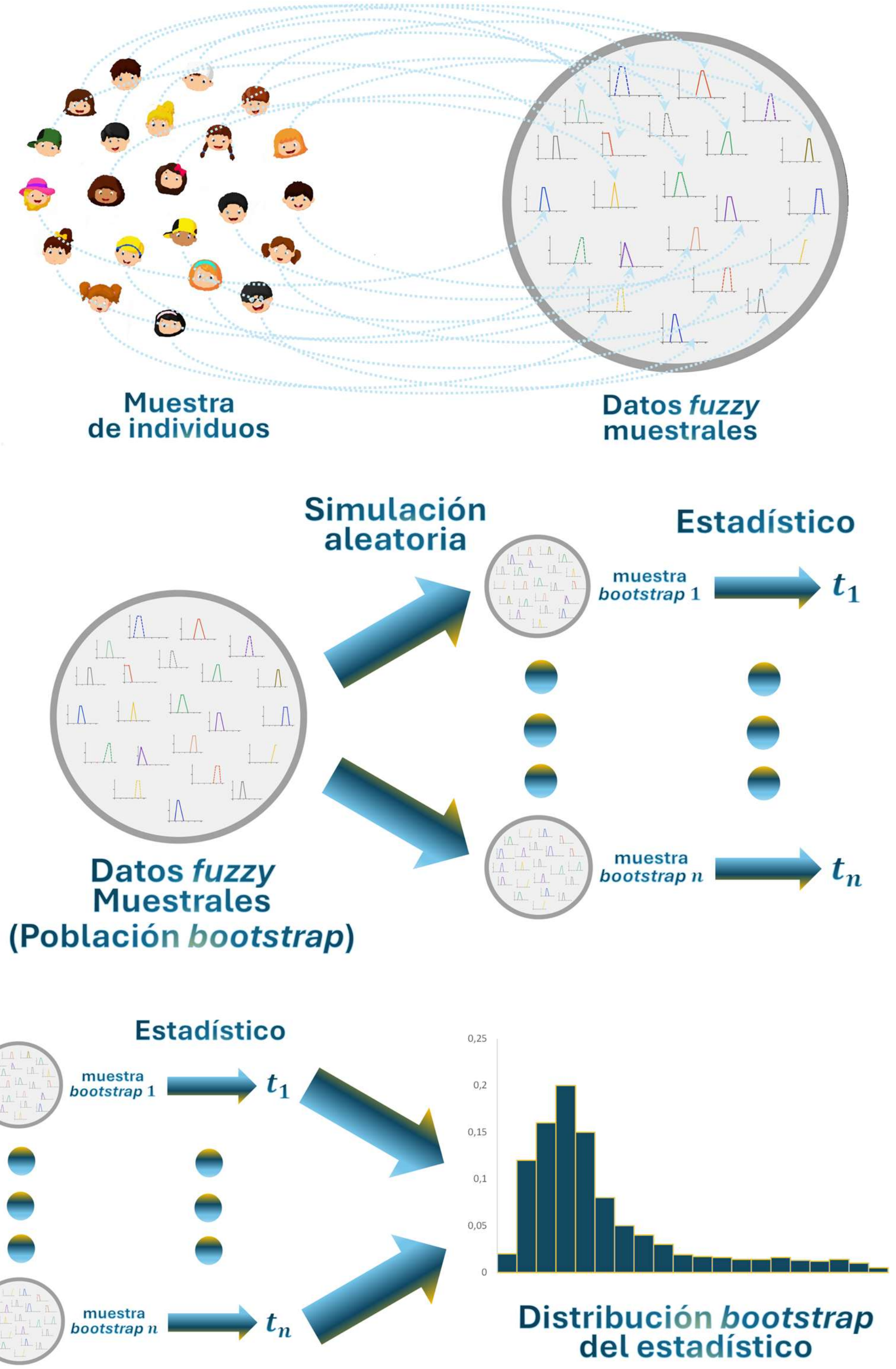
  Se han realizado también algunos estudios de estimación 'puntual' insesgada y fuertemente consistente de la desigualdad hiperbólica (tanto *fuzzy* como numérica) mediante la medida muestral correspondiente, de estimación por intervalo y de contrastes de hipótesis uni- y bilaterales.

- **Definición y propiedades de medidas resumen robustas de localización y escala para conjuntos *fuzzy* aleatorios y análisis de su robustez frente a contaminación o presencia de valores 'atípicos':**
  - definición y propiedades de las medianas *fuzzy* como medidas de tendencia central o localización para conjuntos *fuzzy* aleatorios**,** basadas en la minimización de medias de distintas métricas; análisis de sus propiedades y de su robustez estadística;
  - definición y propiedades de las medias *fuzzy* recortadas como medidas de localización para conjuntos *fuzzy* aleatorios**,** basadas en la eliminación de proporciones prefijadas (recortes) de valores 'extremos' sobre la base de distintas métricas; análisis de sus propiedades y de su robustez estadística;
  - definición y propiedades de los M-estimadores *fuzzy* de localización para conjuntos *fuzzy* aleatorios**,** basadas en la ponderación diferenciada de los valores en función de su atipicidad, sobre la base de distintas métricas y funciones de pérdida; análisis de sus propiedades y de su robustez estadística;
  - estudios comparativos entre las medidas anteriores;
  - definición y propiedades de medidas robustas de dispersión/escala para conjuntos *fuzzy* aleatorios**,** referidas a medidas de localización robustas o independientes de ellas; análisis de sus propiedades y de su robustez estadística; estudios comparativos.

Algunos de estos estudios y de las cuestiones planteadas a continuación se han apoyado total o parcialmente en ***simulaciones*** de uno de los dos tipos siguientes:

- o bien simulaciones aleatorias a partir de los datos *fuzzy* muestrales disponibles,
- o bien simulaciones de datos de número *fuzzy* con forma prefijada dependiente de ciertos puntos reales (habitualmente los extremos del 0- y 1-nivel) obtenidos a partir de simulaciones de modelos tradicionales o modelos obtenidos por ajuste a ejemplos del mundo real.

Un ejemplo de las del primer tipo son las que intervienen en el esquema siguiente para la determinación de la distribución *bootstrap* de estadísticos con valores reales basados en muestras de datos *fuzzy*, como la que se considera en el contraste unimuestral sobre la media tipo Aumann de un número *fuzzy* aleatorio (conjunto *fuzzy* aleatorio con valores de número *fuzzy*).



Esquema de la determinación *bootstrap* de la distribución de un estadístico (con valores reales) basado en una muestra de *datos fuzzy*

Aunque con menor profundidad, se realizaron también algunos estudios introductorios relativos a regresión y correlación lineales, análisis discriminante y otros problemas sobre la distribución de conjuntos *fuzzy* aleatorios.

Los algoritmos que llevan a cabo los cálculos asociados a las técnicas anteriores se han implementado en el ***software* estadístico** mediante paquetes de código abierto como, entre otros, SAFD (Statistical Analysis of Fuzzy Data), FuzzyStatTra (Statistical Methods for Trapezoidal Fuzzy Numbers) e IntervalQuestionStat (Tools to Deal with Interval-Valued Responses in Questionnaires).

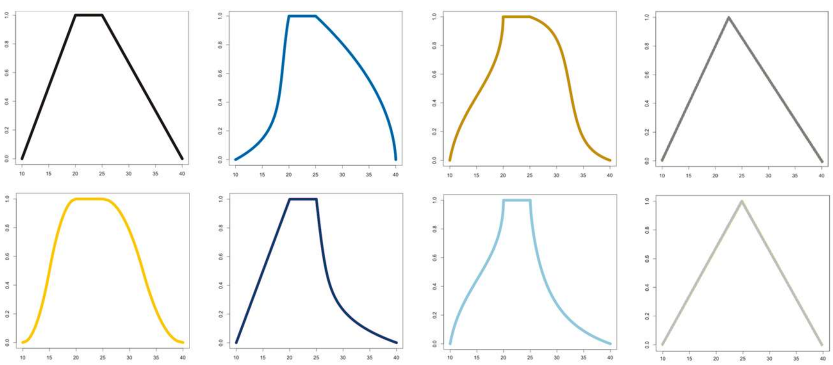
En relación con la metodología desarrollada, se han planteado varias cuestiones. Entre ellas merece atención especial, dado su interés práctico, la consistente en analizar si la 'forma' de los números *fuzzy* elegida para expresar las valoraciones imprecisas influye o no sobre las conclusiones estadísticas que se extraen al aplicar dicha metodología. A diferencia de los resultados que conducen al establecimiento de los métodos anteriores, que son de carácter general, los relativos a tal análisis se han basado en estudios de simulación ya que, aunque no es posible garantizar universalidad, sí puede evidenciarse predominancia. Los procedimientos de simulación empleados están inspirados en ejemplos del mundo real tratados por el grupo, de modo que mimeticen una amplia diversidad de valoraciones humanas.

- **Análisis de sensibilidad del efecto de la 'forma' de los datos *fuzzy* en las conclusiones estadísticas:**

  Una de las implicaciones principales derivadas de este análisis es la de comprobar empíricamente que el empleo de 'formas trapezoidales' para expresar una valoración imprecisa, en lugar de otras curvas que representen valoraciones con interpretaciones muy próximas (por ejemplo, que compartan su intervalo 0-nivel y su intervalo 1-nivel o una aproximación unipuntual del mismo, pero no compartan los 'brazos'), apenas influye en las conclusiones estadísticas obtenidas.

  Siendo las formas trapezoidales muy versátiles y sencillas de dibujar, interpretar y explicitar, supone una ventaja práctica notable la escasa significación de tal influencia.

  El análisis de sensibilidad se llevó a cabo tanto desde una perspectiva descriptiva como inferencial. Desde el punto de vista descriptivo, se examinó el comportamiento de las distintas medidas de localización y de escala simulando muestras de cuaternas ordenadas, generadas de acuerdo con los procedimientos antedichos, y considerando las ocho 'formas' en la figura siguiente.



Ejemplo operativo de ocho números *fuzzy* considerados en el análisis de sensibilidad, pertenecientes a clases cerradas para la aritmética usual y con significado e interpretación muy cercanos

Para cada una de las medidas se observó que la variabilidad de su valor para las distintas formas era apenas perceptible, si acaso ligeramente mayor respecto de las formas triangulares.

El enfoque inferencial corroboró la conclusión anterior, mediante los contrastes bimuestrales de las hipótesis de igualdad de medias tipo Aumann, por un lado, y las varianzas tipo Fréchet, por otro, a partir de muestras simuladas y calculando, para cada una de las muestras y cada uno de los test, el p-valor correspondiente al caso trapezoidal frente a cualquiera de los otros siete casos. Los p-valores obtenidos fueron bastante altos (los menores, los asociados a las formas triangulares pero, en cualquier caso, por encima de 0.5).

3.2 – Aplicación de la metodología con datos *fuzzy* a la inferencia estadística acerca de la distribución de variables aleatorias unidimensionales

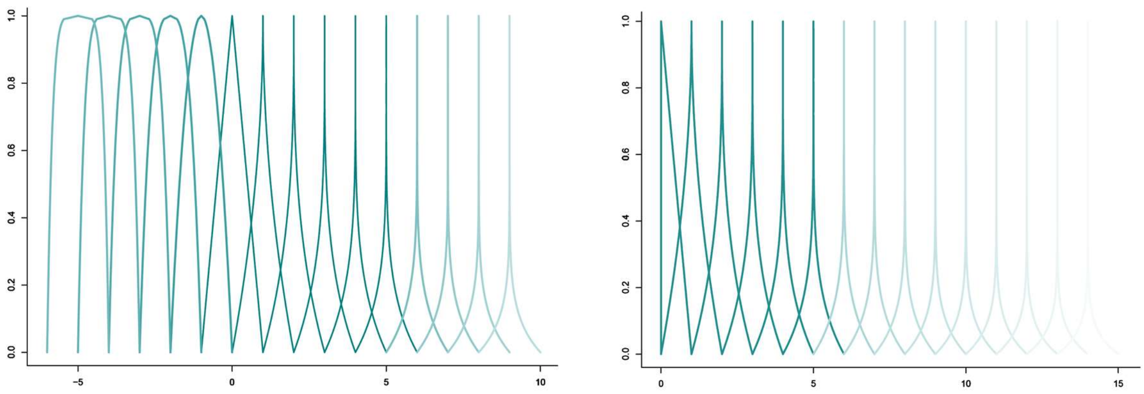
Entre las aplicaciones de la metodología con datos *fuzzy* hay una de naturaleza teórica, cuyas implicaciones estadísticas están pendientes aún de un estudio más exhaustivo.

Se indicó en la introducción del Capítulo 2 que la distribución de una variable aleatoria unidimensional (elemento aleatorio con valores en el espacio de los números reales con la distancia euclídea) se podía caracterizar funcionalmente mediante la función de distribución o la función característica (o la función de probabilidad/masa o la de densidad, en los casos de variables discretas y continuas, respectivamente).

SMIRE⚡CoDiRE ha establecido una **caracterización funcional alternativa de la distribución de una variable aleatoria unidimensional** que recurre a ciertos números *fuzzy* aleatorios (representaciones *fuzzy* de los valores de la variable) y a su media tipo Aumann. Esta caracterización conlleva el valor añadido que supone estar fundamentada en un valor medio que, entre otras, le otorga propiedades de consistencia.

Más concretamente, dada una variable aleatoria unidimensional, existen familias de funciones que transforman cada número real en un número *fuzzy*, de manera que al componer una cualquiera de esas transformaciones con la variable aleatoria se obtiene un número *fuzzy* aleatorio cuya media tipo Aumann es exclusiva para la distribución de dicha variable. En otras palabras, al componer dos variables aleatorias con una de tales transformaciones, las medias tipo Aumann de los números *fuzzy* aleatorios resultantes coinciden si y solo si lo hacen sus distribuciones.

La figura siguiente muestra dos ejemplos de esas transformaciones. En ambos ejemplos se supone que las variables aleatorias sobre las que se aplican son discretas con valores enteros (en el de la izquierda) y con valores enteros no negativos (en el de la derecha), si bien podrían extenderse a variables continuas sin ningún problema.



Ejemplos de dos ‘transformaciones caracterizadoras’ para variables aleatorias discretas: para variables que toman valores enteros (el de la izquierda) y para variables que toman valores enteros no negativos (el de la derecha)

En la parte derecha de la figura que se incluye a continuación se representa la media tipo Aumann de la composición de la transformación caracterizadora, a la derecha de la figura anterior, con una variable con distribución de Poisson de parámetro 1 (cuyo diagrama de barras aparece en la parte izquierda).

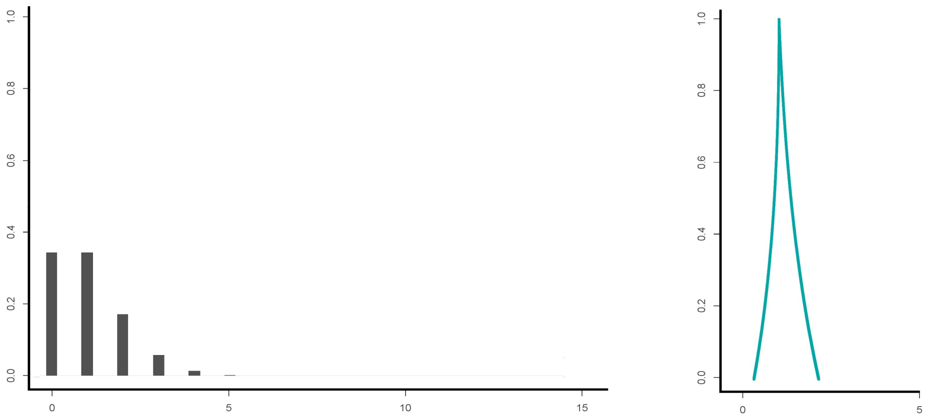


Diagrama de barras de la distribución de Poisson de parámetro 1 (izquierda) y media tipo Aumann del número *fuzzy* aleatorio obtenido por su composición con la transformación caracterizadora de la derecha en la figura anterior

Esta caracterización funcional alternativa junto con los contrastes uni-, bi- y k-muestrales sobre la igualdad de medias tipo Aumann, que forman parte de la metodología resumida en la sección precedente, pueden emplearse con fines inferenciales relativos a distribuciones de variables aleatorias unidimensionales. Así, por ejemplo, su aplicación conjunta permite desarrollar:

– test de hipótesis de bondad de ajuste a una distribución específica;
– test ANOVA de distribuciones (contraste de la igualdad de las distribuciones de varias variables), tanto a partir de muestras independientes como ligadas;
– test de simetría de la distribución de una variable respecto de un punto prefijado.

3.3 – Métodos en busca de datos: aplicación de la metodología con datos *fuzzy* al diseño y análisis de respuestas en cuestionarios y motivación de nuevos estudios teórico-prácticos

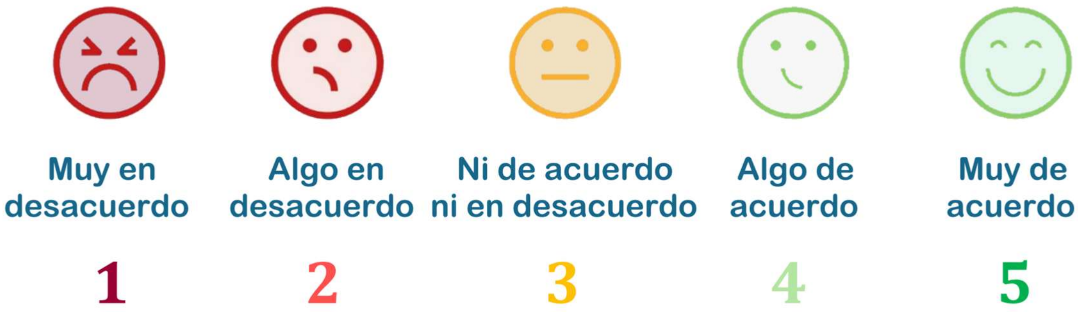
En los primeros trabajos que se apuntaron en la sección 3.1, los ejemplos ilustrativos eran un tanto artificiales y, en su mayoría, los datos estaban referidos a una **escala de tipo lingüístico *fuzzy*** (es decir, a una codificación de los valores en una de las conocidas escalas tipo Likert mediante números *fuzzy*). Sin embargo, en los últimos estudios el análisis de datos se ha centrado en ejemplos del mundo real que involucran cuestionarios en los que varias o todas sus cuestiones demandan un tipo de respuesta que, de forma prácticamente consustancial, puede expresarse recurriendo a una escala *fuzzy* libre.

Un cuestionario consiste en varias series de preguntas e indicaciones con el propósito de obtener información de las personas consultadas. Cada una de estas series de preguntas concierne a menudo a rasgos latentes, actitudes u opiniones, cuyas respuestas idóneas vendrían dadas por grados de acuerdo con ciertas afirmaciones, grados de satisfacción con ciertas decisiones o formas de actuar, etc. No se trata de cuestiones susceptibles de respuesta dicotómica o de respuesta numérica, ni tampoco de respuesta totalmente abierta, sino de respuesta imprecisa que podría 'medirse' en escalas que admitieran gradación en sus valoraciones.

En este respecto cobran sentido especialmente revelador las palabras escritas por el filósofo de la ciencia austro-británico Karl R. Popper en su autobiografía (1974, 2005): *«... tanto la precisión como la certidumbre son ideales falsos. Son imposibles de alcanzar, y por lo tanto peligrosamente engañosas si se aceptan como guías sin ningún cuestionamiento.» «... no se debería nunca intentar ser más preciso de lo que requiere la situación del problema.»*

Entre las escalas que admiten gradación en sus valoraciones, las más populares y utilizadas hasta el momento son las denominadas escalas tipo Likert. Una **escala tipo Likert** (introducida en 1932 por Rensis Likert) es una escala de medición psicométrica que se utiliza para preguntar a una persona sobre su nivel de conformidad con una afirmación, de satisfacción, de frecuencia, etc.. Cuando se responde a una cuestión y se utiliza una escala tipo Likert, el respondiente elige entre un número prefijado (usualmente entre 3 y 10) de opciones (habitualmente 'etiquetas lingüísticas' o 'símbolos gráficos que las traducen') la más acorde con su valoración. Se emplea con mucha asiduidad en las investigaciones de ciencias sociales que implican cuestionarios y, a día de hoy, sigue siendo la escala a la que más se recurre para ordenar por gradación las respuestas en la investigación por encuestas.

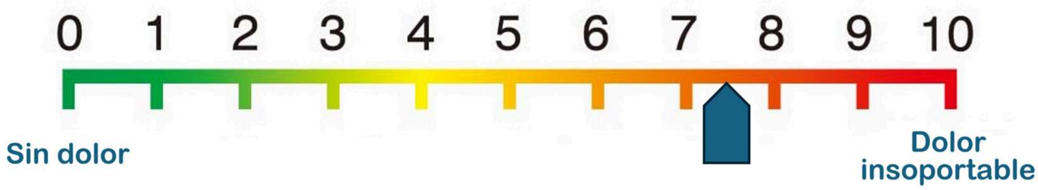
Cuando quieren analizarse desde una perspectiva estadística las respuestas a un ítem o a un conjunto de ítems muy relacionados y que están dirigidos a valorar cierto rasgo latente o 'constructo', los elementos de la escala suelen codificarse numéricamente, con frecuencia a través de números consecutivamente equidistantes (casi siempre enteros) para facilitar la aplicación de las técnicas estadísticas clásicas. Este proceder asume implícitamente que la fuerza e intensidad del grado de acuerdo/satisfacción/frecuencia/… es lineal, desde un acuerdo total a un desacuerdo absoluto, y que las distancias entre cada dos opciones consecutivas son iguales cualesquiera que sean tales opciones.



Ejemplo de escala tipo Likert de 5 puntos.
En la parte superior la modalidad simbólica, en la intermedia la modalidad lingüística y en la inferior su codificación numérica más usual

Estas suposiciones no resultan suficientemente avaladas por la percepción real y la captación de las diferencias individuales naturales se ve muy limitada al reducirse la elección a un número pequeño de opciones posibles, lo que conlleva en la mayoría de los casos una pérdida de información estadísticamente relevante.

Esta limitación puede soslayarse a través de la llamada **escala tipo visual analógica**, en la que el respondiente especifica su valoración/calificación indicando una posición a lo largo de una línea continua entre dos puntos extremos, que a menudo aparecen etiquetados. La escala visual analógica se atribuye como utilizada por primera vez a Hayes y Patterson (1921), si bien algunos trabajos matizan algo más su origen (véase, por ejemplo, Shiina, 2021). Se utiliza bastante en investigación epidemiológica y clínica para medir la intensidad o frecuencia de varios síntomas como el dolor, que desde la percepción del paciente se mueve dentro de un espectro continuo.

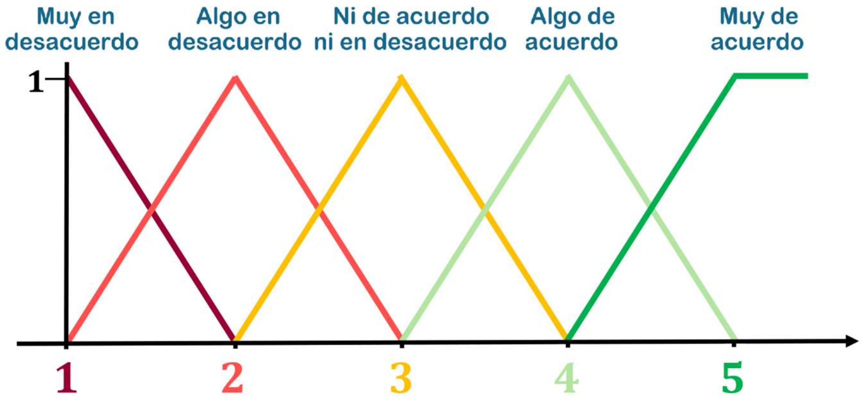


Ejemplo de valoración del dolor sobre la base de una escala visual analógica

La continuidad de la escala permite reflejar fielmente diferencias individuales en los respondientes, así como aplicar una gama muy amplia de métodos estadísticos para analizar las respuestas. No obstante, la precisión que implica la consideración de las escalas visuales analógicas no parece inherente a la respuesta a preguntas relativas a rasgos latentes, actitudes, opiniones, etc.

En la búsqueda de escalas que admitan gradación en sus valoraciones y que recojan la imprecisión intrínseca a las respuestas, se ha ido considerando en la literatura la posibilidad de incorporar los números *fuzzy* a esas escalas.

Una de las vías más habituales de esa incorporación ha sido la de las llamadas **variables lingüísticas *fuzzy***, y sus escalas asociadas, que vienen a ser una codificación mediante números *fuzzy* de las etiquetas de una escala tipo Likert.



Ejemplo de una escala lingüística *fuzzy* asociada a una variable lingüística que codifica mediante números *fuzzy* una escala tipo Likert de 5 puntos

Introducidas por Lotfi A. Zadeh (1975), las codificaciones subyacentes suelen llevarse a cabo por expertos y existen varias que se ajustan razonablemente para una misma variable lingüística. Son muchos los estudios realizados sobre las mismas y sus aplicaciones en diversos problemas del mundo real. Algunos programas de plataformas de programación y cálculo numérico, como puede ser MATLAB, incluyen herramientas para su representación gráfica.

Aunque podría ser una alternativa a considerar frente a las dos anteriores y se trate de uno de los instrumentos de la lógica *fuzzy* con más implicaciones prácticas, en el contexto de los cuestionarios los aspectos críticos apuntados para las escalas tipo Likert seguirían siendo aplicables a las lingüísticas *fuzzy*. Especialmente reseñable sería el hecho de que, al igual que ocurría con las de tipo Likert, una escala lingüística *fuzzy* apenas permitiría reflejar las diferencias individuales e identificaría valoraciones a las que, de manera casi instintiva, el respondiente se vería en la necesidad de añadir matices.

Las tres escalas descritas muestran algún inconveniente esencial cuando se procura analizar estadística y psicométricamente las respuestas a cuestionarios. Por un lado, mediante un lenguaje natural, inevitablemente finito, o su codificación (numérica o de número *fuzzy*) es imposible captar apropiadamente las diferencias individuales entre los respondientes. Por otro lado, la escala visual analógica no capta la imprecisión natural ligada a buena parte de las respuestas a preguntas que admitan gradación.

Se hace imprescindible en ese marco disponer de una escala rica y expresiva en la que *«... algo puede tener significado, aunque no pueda etiquetarse con un nombre...»* (según la frase oportuna que se introdujo –en la revisión histórica de las 19:02 horas del 5 de septiembre de 2008– en Wikipedia, al explicar el término *fuzzy concept* https://en.wikipedia.org/wiki/Special:Contributions/212.182.183.8). Esto concuerda con el propósito de Zadeh de matizar el lenguaje natural tratando las mediciones asociadas como objetos de computación.

En este mismo sentido, Zadeh (2008) afirma que *«Paradójicamente, una de las contribuciones principales de la lógica fuzzy [...] es su elevado poder de* ***'precisiar'*** *lo que es impreciso»* (una especie de remedo de la cita atribuida a Galileo Galilei, con cierta controversia sobre si es literal o apócrifa, *«Cuenta lo que se puede contar, mide lo que es medible y haz medible lo que no lo sea»*). El término ***'precisiar',*** no admitido por la RAE, se ha empleado como traducción del ***'precisiation'*** usado por Zadeh y que a su vez no forma parte del vocabulario en inglés, si bien se entiende su significado. Parece como si Zadeh emulara aquella estrofa del cantautor y artista Jacques Brel: *«Je t'inventarai des mots insensés, que tu comprendras».*

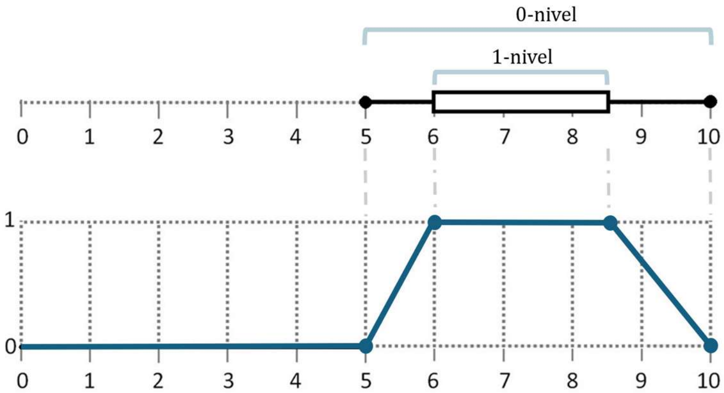
En este sentido, el científico computacional Tim Hesketh, el psicólogo del trabajo Robert Pryor y la psicómetra Beryl Hesketh, los tres australianos, introdujeron (1988ab) las llamadas ***fuzzy rating scales*** (**escalas de valoración *fuzzy*** libre), que aprovechan tres de las principales capacidades de los números *fuzzy*, a saber:

– la de formalizar matemáticamente valoraciones imprecisas,
– la de **'precisiarlas'** de forma continua, permitiendo infinitos matices,
– y la de desarrollar cálculos matemáticos con ellas.

Tim Hesketh

Robert Pryor

Beryl Hesketh



Ejemplo de un dato basado en una escala de valoración *fuzzy*
cuyo 0-nivel está incluido en el intervalo [0,10] y esquema/interpretación de su trazado

A diferencia de los ítems de cuestionarios que involucran escalas tipo Likert o codificaciones numéricas/*fuzzy* de las mismas, para los que las respuestas a las cuestiones están limitadas a elegir una dentro de una lista prefijada y reducida de etiquetas o valores (codificaciones), los ítems basados en escalas de valoración *fuzzy* tienen un formato libre y continuo al igual que los que emplean escalas visuales analógicas. Esta libertad implica una ganancia y una fidelidad de información que son cruciales para las conclusiones del análisis estadístico de las respuestas.

Y, a diferencia de los ítems basados en escalas visuales analógicas, los ítems que recurren a escalas de valoración *fuzzy* permiten captar también de forma continua y plenamente expresiva la imprecisión en las respuestas.

En conexión con el análisis estadístico de datos provenientes del uso de escalas de valoración *fuzzy*, Beryl Hesketh y sus colaboradores desarrollaron algunos estudios de índole descriptiva, que en realidad correspondían a estudios separados de las variables aleatorias reales asociadas a los extremos de los 0- y 1-niveles. No obstante, eran conscientes de que esos estudios no trataban cada dato *fuzzy* como un todo, y Hesketh y otros (2011) señalaban que *«Estamos a la espera de paquetes de programas que permitan a los investigadores utilizar conceptos fuzzy y aplicar análisis estadísticos y otros para contrastar hipótesis garantizando que se recoja apropiadamente el significado.»*

La metodología resumida en la sección 3.1 ofrece una solución con soporte matemático sólido al objetivo planteado por Hesketh y otros. Y no solo desde una perspectiva descriptiva, sino también con varios procedimientos inferenciales.

Una cuestión esencial en este punto es que, si bien la consideración de la escala de valoración *fuzzy* parece muy conveniente y bien justificada en el marco del que se ocupa esta sección, las conclusiones estadísticas a las que se llega cuando se recurre a la misma pueden diferir de manera importante o apenas lo hacen de aquellas que se obtendrían con otras escalas menos complejas. Se ha respondido a esta cuestión a través de la consideración de diversos ejemplos del mundo real en los que varios ítems se contestaban de acuerdo con dos escalas de valoración/calificación y mediante algunos análisis estadísticos de las diferencias en las respuestas según la escala de valoración, así como estableciendo cierto 'orden' entre las cuatro escalas mencionadas a través de un indicador psicométrico relevante sobre la base de simulaciones.

3.3.1. *Algunos estudios del mundo real*

Para examinar si las conclusiones estadísticas podían ser en ocasiones distintas en función de la escala de valoración considerada, y a su vez indagar sobre las posibles dificultades asociadas a la cumplimentación de cuestionarios con ítems basados en la escala de valoración *fuzzy*, se llevaron a cabo el diseño y la ejecución de diversos cuestionarios en los que simultáneamente se requería la respuesta de acuerdo con una escala tipo Likert y con una escala de valoración *fuzzy*.

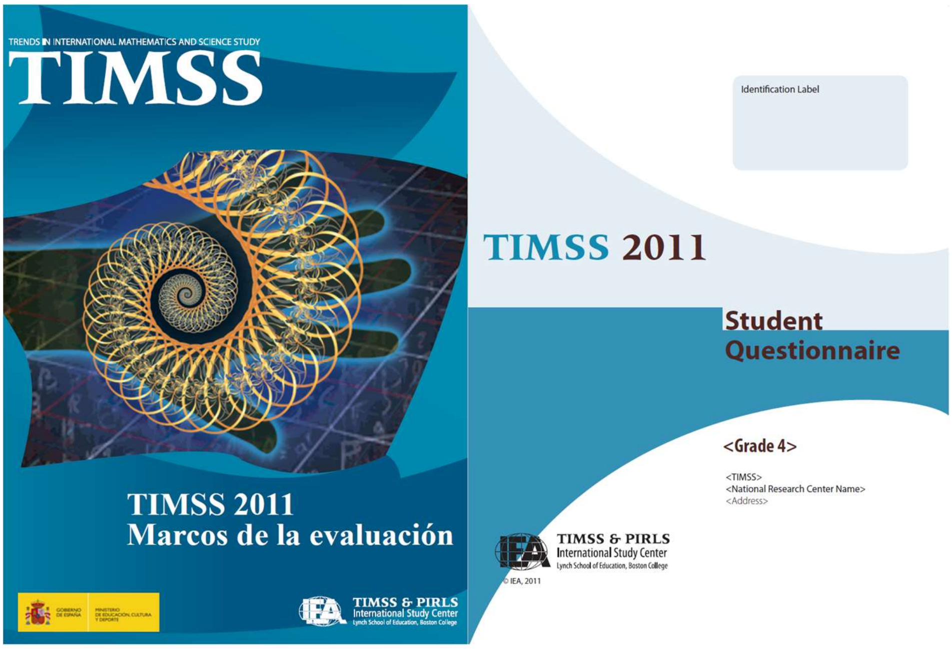
En primer lugar, los respondientes fueron estudiantes del *Máster on Soft Computing and Intelligent Data Analysis* de la Universidad de Oviedo, impartido y coordinado entre 2009 y 2012 en colaboración con el European Centre for Soft Computing (Mieres, 2006-2016). La formación recibida por esos estudiantes hacía muy sencilla su comprensión a la hora de cubrir los diferentes ítems, de manera que apenas se necesitaba de explicación y de entrenamiento. En estudios posteriores intervinieron personas con una formación más variada y el tiempo de explicación requerido fue algo mayor, variando principalmente tal tiempo según su afinidad con las matemáticas y su receptividad.

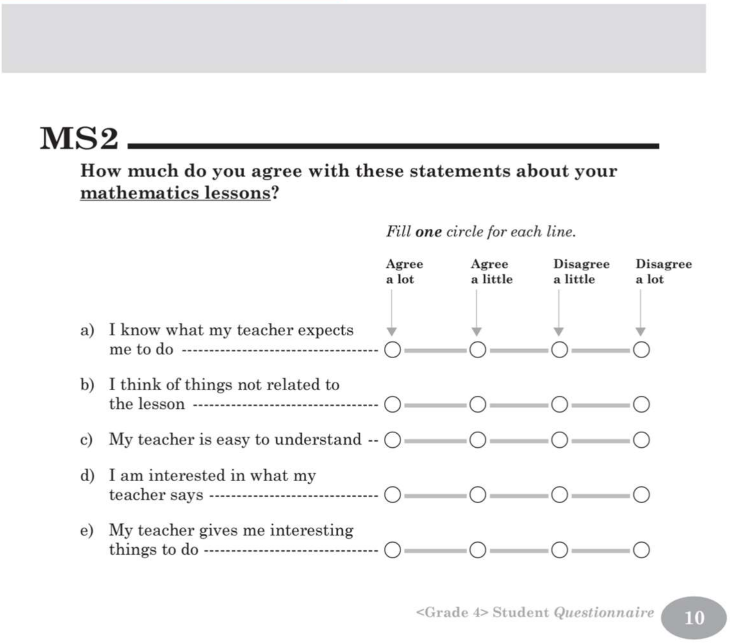
Un ejemplo que ha dado lugar a un buen número de trabajos, tanto por su análisis estadístico como por las implicaciones/motivaciones formales derivadas del mismo, es el de la adaptación de algunos de los ítems de los conocidos estudios internacionales de progreso en comprensión lectora (PIRLS), y en matemáticas y ciencias (TIMSS) de la IEA (International Association for the Evaluation of Educational Achievement).

TIMSS es una evaluación de alumnos de 4º y 8º grado (en España, 4º de Educación Primaria y 2º de ESO) que recopila datos de tendencias, opiniones y nivel de aprendizaje cada cuatro años desde 1995 y en él participa un elevado número de centros educativos de cerca de 70 países. PIRLS es una evaluación de alumnos de 4º de Primaria que se lleva a cabo cada cinco años desde 2001. Los países participantes tienen en cuenta ambas evaluaciones para valorar la eficacia de sus sistemas educativos, de modo que sus conclusiones tienen una trascendencia notable.

Cada evaluación consta de cuatro cuestionarios (del entorno familiar, del centro, del profesorado y del alumnado) cuyos ítems incluyen distintas escalas (mayoritariamente dicotómicas, numéricas y tipo Likert).

En 2011 coincidieron los ciclos de las dos evaluaciones y una parte del informe estadístico de los datos recogidos en España fue elaborado por compañeros del Departamento de Estadística e Investigación Operativa y Didáctica de la Matemática de la Universidad de Oviedo, bajo la dirección del miembro de SMIRE⚡CoDiRE Norberto Corral (véase Corral Blanco *et al.* 2013). Como ejemplo de las cuestiones formuladas en el Cuestionario del Estudiante de 4º de Primaria 2011 cinco de las de matemáticas (que se conservan en el de 2023) aparecen a continuación.





MS2

How much do you agree with these statements about your mathematics lessons?

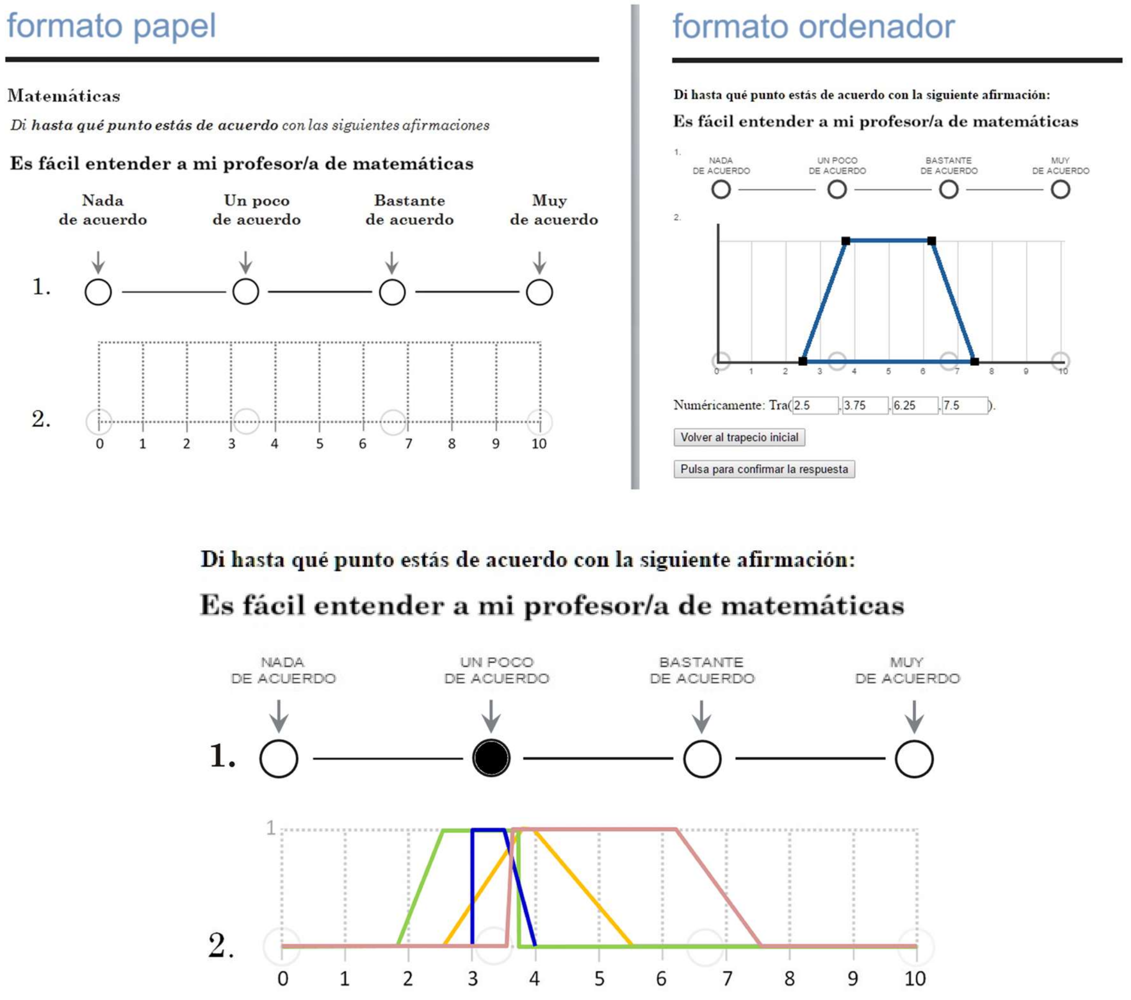
*Fill **one** circle for each line.*

	Agree a lot	Agree a little	Disagree a little	Disagree a lot
a) I know what my teacher expects me to do	○	○	○	○
b) I think of things not related to the lesson	○	○	○	○
c) My teacher is easy to understand	○	○	○	○
d) I am interested in what my teacher says	○	○	○	○
e) My teacher gives me interesting things to do	○	○	○	○

<Grade 4> Student *Questionnaire* 10

Ejemplo de cinco de los ítems relativos a las matemáticas
del Cuestionario TIMSS 2011 del Estudiante

La pérdida de información que para un análisis estadístico conlleva reducir a cuatro (los puntos de la escala tipo Likert empleada) el número de respuestas posibles, repercute en muchos casos en conclusiones bastante dispares a las que se extraerían con una escala más rica y expresiva como la de valoración *fuzzy*. Para comprobarlo, se adaptaron algunas de las preguntas extraídas del cuestionario del estudiante, de modo que admitieran doble respuesta: una elegida según la escala tipo Likert de cuatro puntos del cuestionario original y otra *fuzzy* trapezoidal (dado que, según se ha comentado en el análisis de sensibilidad de la sección 3.1, la forma adoptada por los datos/respuestas *fuzzy* apenas afecta a las conclusiones finales).



Ejemplo de ítem de doble escala del Cuestionario TIMSS 2011 del Estudiante.
Arriba, a la izquierda, la versión para cumplimentar a mano,
y a la derecha la versión para cumplimentar en ordenador con vértices ■ móviles.
Debajo, ejemplo de las respuestas dobles de 4 alumnos cuyas
respuestas Likert coincidieron ('UN POCO DE ACUERDO') pero las *fuzzy* libres difirieron

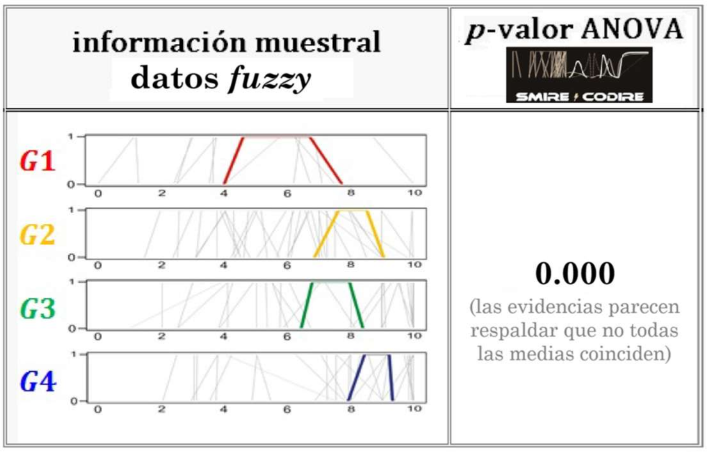
El cuestionario así diseñado se aplicó en sus dos formatos, manual y en ordenador, en un grupo de unos 70 alumnos de 4º de Primaria del Colegio San Ignacio de Oviedo. Gracias al apoyo de la dirección del centro y al de los profesores de ese curso, la experiencia resultó muy interesante desde una doble perspectiva.

- Por un lado, permitió corroborar estadísticamente la intuición previa: en muchos casos las diferencias, tanto descriptivas como inferenciales, podían ser bastante importantes. Como ejemplo ilustrativo:

información muestral datos tipo Likert | ***p*-valor Kruskal-Wallis**

	1	2	3	4
G1	0	2	3	2
G2	0	1	6	11
G3	2	0	5	9
G4	0	1	4	14

0.167
(no hay evidencias suficientes de que las medias no coincidan)



Diferencia de conclusiones en test ANOVA unifactoriales de igualdad de medias para la cuestión de la figura anterior con respuesta tipo Likert (original) y de valoración *fuzzy* con el factor relativo a la nota obtenida en la última evaluación de matemáticas, agrupada en los niveles ***G1*** := de 1 a 6, ***G2*** := de 6 a 8, ***G3*** := de 8 a 9 y ***G4*** := de 9 a 10

- Por otro lado, también sirvió para verificar empíricamente que la formación de los respondientes no era un factor decisivo, ya que todos los alumnos, con edad en torno a los 9 años, fueron muy receptivos y captaron sin problema las instrucciones indicadas, a pesar de desconocer nociones como la de función y simplemente saber desde hacía poco lo que representaba un trapecio.

A partir de esta última experiencia se realizaron otras más, entre ellas un breve estudio sobre la posible influencia de la especialidad médica en la percepción sobre los pacientes mentales, llevado a cabo en el Hospital Son Llatzer de Palma de Mallorca (Lubiano *et al.*, 2018).

Además, se inició una colaboración estrecha con el equipo del catedrático en Psicología Social de la Universidad de Oviedo Antonio León García Izquierdo traducida en la cooperación conjunta en Ayudas GRUPIN del Principado de Asturias desde 2018, constituyendo parte importante del grupo SMABSS (https://bellman.ciencias.uniovi.es/smabss/) y en una primera transferencia de la tecnología descrita en este capítulo. Más concretamente, en relación con esa transferencia se ha colaborado, en varios de los estudios y en contratos de investigación suscritos entre el centro de I+D de ARCELORMITTAL GLOBAL R&D y la CÁTEDRA ASTURIAS PREVENCIÓN dirigida por el profesor García Izquierdo.

3.3.2. *Estudios de análisis de diferencias y comparativos entre escalas de valoración basados en simulaciones*

Para ratificar la influencia potencial de la escala de valoración de forma más general se ha recurrido a comparaciones de las medias y varianzas con escala de valoración *fuzzy* frente a una de las dos codificaciones de escalas tipo Likert, mediante test de hipótesis bimuestral para muestras ligadas en el caso de las medias y mediante el análisis de las distancias dos a dos en el caso de las varianzas. Si bien no es posible obtener conclusiones plenamente generales, sí ha podido llegarse a conclusiones bastante mayoritarias sobre la base de simulaciones de datos *fuzzy* según el segundo procedimiento indicado en la sección 3.1, es decir, simulaciones de números *fuzzy* trapezoidales donde los extremos de los intervalos 0- y 1-nivel se generan a partir de simulaciones de modelos de distribuciones de variables aleatorias reales obtenidos por ajuste a los ejemplos del mundo real como los de la subsección 3.3.1.

Se han simulado así las 'respuestas' según la escala de valoración *fuzzy* y cada una de ellas se ha vinculado con una respuesta tipo Likert de h puntos (habitualmente $h = 4$ o 5) mediante un criterio de 'Likertización' consistente en la minimización de la distancia de la respuesta *fuzzy* a los códigos numéricos establecidos de acuerdo con el intervalo de referencia de la *fuzzy*. Después, cada dato 'Likertizado' se ha codificado mediante una escala lingüística *fuzzy*. De esta manera, con las respuestas disponibles en las tres escalas, se ha calculado el valor de los estadísticos o de las distancias para comparar la media de la escala de valoración *fuzzy* con la codificada tipo Likert y las varianzas de ambas, respectivamente. Las conclusiones (véanse, por ejemplo, Lubiano *et al.*, 2017b, Arellano *et al.*, 2019) han avalado de forma predominante que la escala de valoración influye sensiblemente: p-valores pequeños en el test de igualdad de medias y porcentajes elevados de distancias pequeñas/moderadas en la comparación de varianzas.

La extensión de ese índice al caso de datos *fuzzy* ha permitido visualizar tendencias mayoritarias inequívocas, tanto a partir de situaciones de la vida real (arriba en la figura de la página siguiente) como en ejemplos basados en simulaciones (abajo en la figura de la página siguiente).

No obstante, si bien nuestra impresión es que esas diferencias son debidas a que la escala de valoración *fuzzy* es más informativa estadísticamente, estos resultados comparativos no ofrecen garantía de su superioridad efectiva. Para confirmar esa intuición se ha acudido a una herramienta de comparación popular relacionada con los cuestionarios: el índice α de Cronbach, indicador de la consistencia interna/fiabilidad de los ítems de los constructos, entendiendo por estos últimos conceptos que directamente no son ni observables ni medibles, pero sí pueden aproximarse mediante ítems o preguntas en los que subyacen. Un valor alto del índice α (que equivale a un valor próximo a 1) implica que los ítems del constructo presentan una excelente consistencia interna o interrelación.

Escala de valoración	ejemplo de marco de valoración	ejemplo de respuesta	capta imprecisión intrínseca	capta diferencias individuales	tendencia mayoritaria de α
Likert codificada lingüística *fuzzy*	1, 0; 1 2 3 4	1, 0; 1 2 3 4	parcialmente	NO ✘	
Likert codificada numérica	Nada de acuerdo, Un poco de acuerdo, Bastante de acuerdo, Muy de acuerdo; 1 2 3 4	Nada de acuerdo, Un poco de acuerdo, Bastante de acuerdo, Muy de acuerdo; 1 2 3 4	NO ✘	NO ✘	$\wedge$
Visual analógica	Nada de acuerdo, Muy de acuerdo; 0 1 2 3 4 5 6 7 8 9 10	Nada de acuerdo, Bastante de acuerdo, Muy de acuerdo; 0 1 2 3 4 5 6 7 8 9 10	NO ✘	SÍ ✔	$\wedge$
De valoración *fuzzy* (libre)	Nada de acuerdo, Un poco de acuerdo, Bastante de acuerdo, Muy de acuerdo; 0 1 2 3 4 5 6 7 8 9 10	Nada de acuerdo, Muy de acuerdo; 0 1 2 3 4 5 6 7 8 9 10	SÍ ✔	SÍ ✔	$\wedge$

porcentajes	$n = 30$			$n = 50$			$n = 100$			$n = 300$		
Comparación por α de Cronbach ↓	$k = 10$	$k = 20$	$k = 30$	$k = 10$	$k = 20$	$k = 30$	$k = 10$	$k = 20$	$k = 30$	$k = 10$	$k = 20$	$k = 30$
$\alpha_{FRS} > \alpha_{VAS}$	97.5	99.6	99.7	99.3	99.9	100	99.9	100	100	100	100	100
$\alpha_{VAS} > \alpha_{NEL}$	85.5	95.4	98.3	89.8	98.0	99.5	96.0	99.8	99.9	99.1	100	100
$\alpha_{NEL} > \alpha_{FLS}$	79.8	89.5	93.4	86.2	94.3	97.8	91.5	98.0	99.6	97.9	99.9	100

En el estudio de las tendencias en cuanto al empleo de las distintas escalas de valoración (FRS = escala de valoración *fuzzy* libre, NEL = Likert codificada numérica, FLS = Likert codificada lingüística *fuzzy* y VAS = escala visual analógica) predominan muy mayoritariamente valores de consistencia interna más alta para la FRS, incluso con respecto de la VAS, para distintos números n de respondientes, y distintos números k de ítems

La consecuencia primordial de los estudios en esta sección es que, siempre que las implicaciones de índole social, sanitaria, tecnológica, etc. del análisis estadístico de las respuestas a los cuestionarios sean capitales, el uso de las escalas de valoración *fuzzy* es aconsejable por cuanto va a proporcionar conclusiones más fieles y fiables. La complejidad añadida no debería retraer a los usuarios de su empleo, máxime cuando el desarrollo de *software* para obtener las mismas está y continuará creciendo.

3.4 – Datos en busca de métodos: aplicación de la metodología robusta con datos *fuzzy* al tratamiento de valoraciones por expertos en el análisis de riesgo de rotura de embalses

En las aplicaciones del mundo real de la subsección 3.3.1, quienes gestionaron el análisis de las respuestas de los ítems de los cuestionarios basados en escala de valoración *fuzzy*, diseñaron estos de manera que el formato y las instrucciones para su cumplimentación vinieran dados de antemano para los respondientes. Es decir, se suscitó el tipo de dato que iba a analizarse.

Recientemente, la situación en la que están trabajando varios miembros de SMIRE⚡CoDiRE ha surgido en dirección contraria. La información recibida ha venido dada por ternas ordenadas, cuya obtención se explica a continuación, y el significado expreso de tales ternas ha motivado de manera inmediata su adaptación a datos *fuzzy* triangulares.

Los especialistas en materia de seguridad de presas se enfrentan a menudo a la tarea de valorar sus expectativas o percepciones sobre el riesgo de fallo (o de rotura) potencial de una presa. La información disponible sobre cada presa difícilmente cuenta con datos históricos de fallos que permitan evaluar con criterios probabilísticos plausibles sus características de seguridad. Para superar este inconveniente, e inspirándose en algunos estudios previos sobre fiabilidad de distintos sistemas, el Tomo I de la Guía Técnica de Explotación de Presas y Embalses, dedicado al Análisis de Riesgos Aplicado a la Seguridad de Presas y Embalses y elaborado en 2012 por el Comité Nacional Español de Grandes Presas, propuso un enfoque alternativo consistente en involucrar a expertos para emitir una suerte de posibilidad de fallo con valores en el intervalo [0,1], donde 0 indicaría la imposibilidad de que ocurriera el fallo de la presa y 1 la posibilidad potencial máxima de tal ocurrencia.

Como los juicios o las valoraciones de esos expertos siempre involucran una imprecisión intrínseca a la naturaleza de lo que se juzga o valora, esta imprecisión debe tenerse en cuenta a la hora de la emisión. En este sentido, el enfoque alternativo propone como práctica de evaluación habitual la consistente en que cada experto establezca tres valores que caractericen su percepción, a saber:

– los dos valores (mínimo y máximo) entre los que considera que se encuentra su valoración (el mínimo y el máximo plausible/posible/compatible con ella)
– y un tercer valor entre los dos anteriores, que represente el más plausible/posible/compatible con su percepción.

Si bien en ocasiones se recurre al empleo de tablas de descriptores verbales (como las conocidas de Reagan, Mosteller y Youtz, 1989), en las que expresiones probabilísticas verbales se 'traducen' en términos de ternas de valores específicos en [0,1], otra opción que se ajusta muy apropiadamente y cuya interpretación capta la esencia de la mayoría de los juicios y las valoraciones humanas son los conjuntos (en este caso números) *fuzzy*. En la guía citada anteriormente se señala el interés potencial de la lógica *fuzzy* en la agregación de resultados, pero también lo tiene en la modelización de las valoraciones emitidas por los expertos.

De este modo, mediante la escala de números *fuzzy* con intervalo 1-nivel contenido en el intervalo [0,1] se tendría un ejemplo muy útil de una escala de valoración *fuzzy*. La libertad de esa escala le confiere una riqueza y una expresividad que le permiten reflejar distintos matices, a saber: las diferencias individuales de cada experto ante cada escenario y la imprecisión de la valoración a través de sendas escalas numéricas continuas. Por ejemplo, dos expertos podrían percibir como MUY VEROSÍMIL/PROBABLE/POSIBLE la ruptura de una presa (o un mismo experto podría calificar dos presas con ese mismo descriptor verbal), pero asignarles dos ternas diferentes.

Además, como ya se ha comentado en la sección anterior, la escala de valoración *fuzzy* resulta psicométrica y sensiblemente superior a las escalas codificadas numéricas o *fuzzy* de las de tipo Likert (como las que se obtendrían a partir de la tabla de Reagan, Mosteller y Youtz, 1989, o de otras análogas).

Arriba, ejemplo de valoración de un experto respecto del riesgo de rotura de cierta presa. Debajo, se incluye una aplicación diseñada en R para la recopilación potencial de las respuestas expresadas como ternas de probabilidades/posibilidades de rotura en términos de porcentajes

En la práctica, al evaluar el riesgo de fallo de una misma presa en un instante dado, expertos distintos proporcionan habitualmente valoraciones diferentes. Sin embargo, el objetivo último de estos estudios es el de llegar a cierta valoración de 'consenso' que resuma mediante algún criterio de agregación una especie de posición central de esas valoraciones.

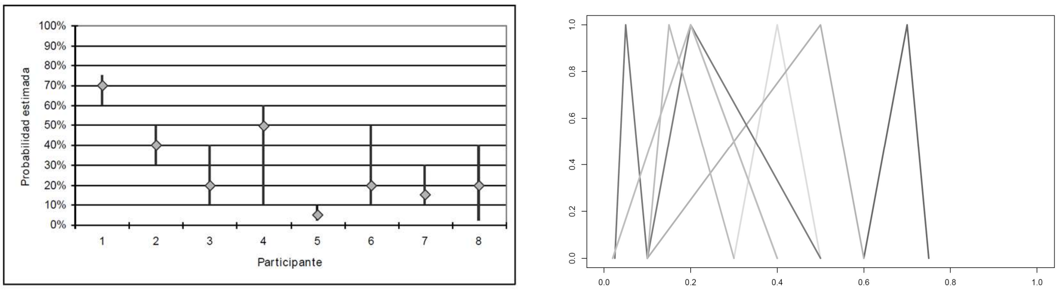
A la hora de resumir la información de las valoraciones emitidas y recogidas para una misma presa a través de cierto valor central, una forma inmediata de proceder sería la correspondiente a considerar la media de sus cotas inferiores, la de sus valores más posibles (o mejores estimaciones) y la de las cotas superiores. La adopción de ese procedimiento podría justificarse formalmente, ya que coincidiría con la media tipo Aumann correspondiente, si bien tendría en cuenta con la misma relevancia las percepciones de todos los expertos, incluso aquellas que pudieran resultar atípicas o muy discrepantes.

En la guía técnica a la que se ha hecho referencia, se apunta a que si algún experto ha emitido una valoración que se salga del 'rango' del grupo de expertos para una presa, convendría que argumentara sus razones, pero admitiéndose que pueden existir discrepancias claras que no dejan de ser sino un reflejo de la diversidad de capacidades y percepciones de los expertos.

Con el objeto de evitar que la influencia en la valoración central de esas percepciones discrepantes sea mayor de lo deseable, sería conveniente recurrir al empleo de indicadores de centralización 'robustos'. La medición robusta de la tendencia central, o localización, para datos numéricos es un tema sobre el que se han desarrollado muchos estudios en las décadas más recientes y que constituye aún un desafío notable en estadística. Su extensión al problema de la estimación robusta de la tendencia central para valoraciones imprecisas, como las que se analizarán en el estudio que se ha iniciado, no es en absoluto ni directa ni trivial.

La tentación de considerar medidas robustas de tendencia central separadamente para cada una de las tres componentes que caracterizan cada valoración no solo no iría acompañada de una justificación formal, sino que no tendría en cuenta las vinculaciones entre las tres componentes que determinan cada una de las valoraciones. En otras palabras, no contemplaría cada valoración de un experto como un todo (componentes ligadas), sino que las manejaría de forma independiente, lo que podría llevar a perder información relevante. Estos inconvenientes pueden soslayarse con la modelización de las valoraciones imprecisas por medio de números *fuzzy* triangulares, que recogerían plenamente la información disponible y, además, se adaptarían de forma muy natural a la interpretación de los números *fuzzy*.

La gráfica siguiente muestra a la izquierda un ejemplo de comparativa en escala lineal de las valoraciones emitidas por juicio de experto en el Apéndice B de la guía técnica y a la derecha su 'traducción' mediante números *fuzzy* triangulares.



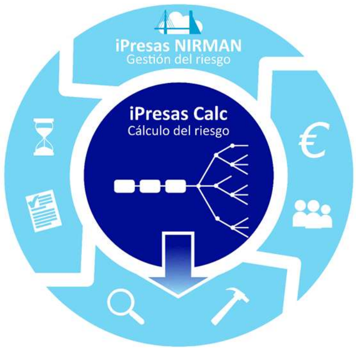
Ejemplo de comparativa en escala lineal de valoraciones de 8 expertos (izquierda)
y representación de sus correspondientes valoraciones *fuzzy* triangulares (derecha)

Como se ha subrayado anteriormente, el hecho de que la forma del número *fuzzy* que sirve como modelo de la valoración sea triangular apenas afecta a la valoración de la estimación de la tendencia central o localización.

Como la media tipo Aumann es una medida resumen de consenso pero es excesivamente sensible frente a posibles errores o cambios en las evaluaciones de los expertos o a posibles evaluaciones discrepantes y, al tiempo, no estaría bien visto prescindir de ninguna valoración, se está optando por aplicar la metodología de los M-estimadores de localización, en lugar de 'recortar las valoraciones situadas en las colas', y hallar su media. Al ser, además, el número de expertos en la evaluación del riesgo de rotura de las presas muy reducido para cada presa, la opción de considerar las M-estimaciones resulta también más aconsejable que la de las medias recortadas. Habitualmente, en las primeras se tienen en cuenta (con mayor o menor peso) las opiniones de todos los expertos, lo cual no ocurriría en las medias recortadas (salvo que no se 'recortara' ninguna observación).

Esta transferencia de tecnología se ha materializado por el momento en el contrato *Desarrollo de una Metodología para Combinar Percepciones de Expertos en Seguridad de Presas y Valoración de Sesgos¸* que han suscrito varios miembros de SMIRE⚡CoDiRE con la empresa de base tecnológica y *spin-off* de la Universitat Politècnica de València iPresas, cuyo objetivo es el apoyo a la gestión integral de riesgos de inundación y la seguridad de presas, así como de otras infraestructuras.

iPresas
Risk Analysis



4. Epílogo

Los números *fuzzy* aleatorios y su extensión (los conjuntos *fuzzy* aleatorios) constituyen un concepto formalizado sólidamente dentro del marco probabilístico.

Su empleo para modelizar la generación aleatoria de datos imprecisos con valores *fuzzy* permite mantener la gran mayoría de las nociones e ideas del razonamiento estadístico para datos numéricos (o vectoriales).

Cuando la repercusión de las conclusiones del análisis estadístico de datos imprecisos que admiten gradación tiene trascendencia para la vida diaria, resulta oportuno y recomendable recurrir a las escalas de valoración *fuzzy* y a la metodología estadística basada en los números *fuzzy* aleatorios.

El discurso ha resumido los antecedentes probabilísticos de esa metodología, algunos de ellos del grupo de investigación SMIRE⚡CoDiRE, así como las principales contribuciones del grupo a la misma desarrolladas hasta el momento.

Como sucedía en el caso del análisis estadístico de datos numéricos o vectoriales, la matemática (la teoría de la probabilidad) tras la metodología estadística para datos *fuzzy* está fundamentada en conceptos y resultados en absoluto rudimentarios y, a menudo, muy complejos de demostrar (espacios de Hilbert, medibilidades Borel, encajes isométricos, etc.). Tales fundamentos no se han pormenorizado en el discurso, obviándolos intencionadamente para llegar mejor a una audiencia y a unos usuarios que no tendrían por qué conocerlos ni necesitarían seguirlos puntualmente.

Afortunadamente, esa complejidad matemática apenas afecta a la aplicabilidad de los métodos, menos aún cuanto más extenso sea el progreso y desarrollo del *software* adecuado.

Dedicatoria

Suele decirse que *«siempre se van los mejores»* y, en nuestro departamento, así ha sido. Tuve la gran suerte de poder investigar con los tres, de poder admirarles y de aprender algo de ellos, aunque no todo lo que me habría gustado.

Este discurso, ¡va por ellos!

Con mis añorados
Teófilo Brezmes (arriba, en 1977),
Pedro Gil (abajo a la izquierda, en 1995)
y *Miguel López* (abajo a la derecha, en 2001)

Más agradecimientos

Comencé este discurso con el agradecimiento a mi hermano Pedro y al presidente de la AACI, Mario. Anticipé que habría muchas más personas a las que agradecer y, con seguridad, no podré mencionar a tantas como debería hacerlo.

Gracias a aquellos con quienes he tenido la fortuna de colaborar en alguna ocasión, o en muchas, entre ellos los miembros de los grupos SMIRE, SMIRE⚡CoDiRE y SMABSS. Mi sentimiento de grupo está muy arraigado y la única pena es que son ya bastantes las personas que nos han ido dejando, aunque en su mayoría lo han hecho por razones de jubilación. Menos mal que aún quedan algunos jóvenes con suficiente bravura para seguir dándole alegrías al tema de investigación y a las demás líneas del equipo. Gracias, de forma muy especial, a Asun Lubiano, Beatriz Sinova, Carlos Carleos, Sonia Pérez y José García, pero también Manolo, Norberto, Teresa, Antonia, Rosa, Ana Colubi, Gil, Ana Belén, Sara, Ángela, Marta, … y a todos los de los primeros años de trabajo, entre ellos, Rigo, Covadonga, Nacho Martínez, Luis José y Nacho Cascos, así como a Gloria y Conchita por estar siempre tan cerca.

Gracias al Departamento de Estadística e Investigación Operativa y Didáctica de la Matemática, creado por Pedro Gil, y a sus equipos de dirección, con Pedro, Manolo, Susana Montes y Luis José al frente de los mismos.

Gracias a la Facultad de Ciencias, a sus decanos y equipos y a sus alumnos, a los que siempre ha sido un privilegio dar clase, especialmente gratificante estos últimos años no solo por el nivel académico, sino aún más por la categoría humana y su empatía entendiendo que el PDI tiene su corazoncito. También a las Facultades de Biología y Química donde tantos años impartí clase y de cuyos estudiantes, e igual modo, aprendí mucho y guardo un recuerdo muy grato.

En resumen, ¡gracias a la Universidad de Oviedo por todo ello!

Así mismo, quiero manifestar mi gratitud a la Real Academia de Ciencias Exactas, Físicas y Naturales (RAC), por incluirme entre sus académicos numerarios. Aún no sé muy bien por qué, pero en su momento dije que mi objetivo primordial era aprender de los maestros, y puedo declarar sin ambages que así está siendo y que estoy disfrutándolo a conciencia.

Y, como consecuencia de estar en la RAC, el Real Instituto de Estudios Asturianos (RIDEA) me ha incluido como miembro de honor, otra oportunidad que me reportará enseñanzas muy singulares sobre una Asturias a la que debo tanto.

En esta Academia Asturiana de Ciencia e Ingeniería he entrado menos asustada que en la RAC o en el RIDEA. Hay tanto por hacer y tantos amigos, que no ha habido ni tiempo ni espacio para el susto. Hoy tengo especial suerte y es un honor contar como respondiente con mi compañera de Facultad Consuelo Martínez, Chelo.

Gracias a los asistentes y a los amigos que han podido acercarse a este acto (compañeros de la carrera en Valladolid, de la Estadística e Investigación Operativa en España, amigos corales, amigos de Geras de Gordón y amigos pequeños y grandes del Colegio La Toba de Avilés, entre otros). Saben que valoro mucho el acompañamiento también en los momentos alegres como es este. Y a tantas personas que me han ayudado, aunque no estén aquí hoy.

Por último, pero no menos importante, gracias a mi familia. A mis padres, hermanos, sobrinos, tíos, primos y acoplados, a quienes tanto quiero y que nunca me abandonan. A mis dos hijas del alma, por ser y estar. Y mi agradecimiento menos impreciso: ¡Gracias, José Manuel, por cómo me quieres y cuidas, por cuánto me inspiras cada día y por hacerme tan fácil quererte!

Concluyendo

Para acabar, creo que una forma oportuna de finalizar este discurso es leyendo un poema que guarda bastante relación con su contenido. En su visita a Asturias en septiembre de 2024 a recoger el Premio EL COMERCIO de Ciencia y en noviembre del mismo año para presentar su último libro de divulgación hasta entonces, *"La levedad de las libélulas"*, mi admirado y querido (de hecho, nuestro admirado y querido) Carlos López Otín nos hizo un bonito regalo al advertirnos sobre un poema que yo debería haber conocido, aunque no era así. Está escrito por la Premio Nobel de Literatura 1996, la polaca Wisława Szymborska, aparece incluido en su libro "*Instante"* (*Chwila*) e incluye algunas valoraciones *fuzzy*.

Contribución a la Estadística

De cada cien personas,
– las que todo lo saben mejor:
cincuenta y dos,
– las inseguras de cada paso:
casi todo el resto,
– las prontas a ayudar, siempre que no dure mucho:
hasta cuarenta y nueve,
– las buenas siempre, porque no pueden de otra forma:
cuatro, o quizá cinco,
– las dispuestas a admirar sin envidia:
dieciocho,
– las que viven continuamente angustiadas por algo o por alguien:
setenta y siete,
– las capaces de ser felices:
como mucho, veintitantas,
– las inofensivas de una en una, pero salvajes en grupo:
más de la mitad seguro,
– las crueles cuando las circunstancias obligan:
eso mejor no saberlo ni siquiera aproximadamente,
– las sabias a posteriori:
no muchas más que las sabias a priori,
– las que de la vida no quieren nada más que cosas:
cuarenta, aunque quisiera equivocarme,
– las encorvadas, doloridas y sin linterna en lo oscuro:
ochenta y tres, tarde o temprano,
– las dignas de compasión:
noventa y nueve,
– las mortales:
cien de cien. Cifra que por ahora no sufre ningún cambio.

A esto último podríamos añadir: y ello, a pesar del progreso de la investigación sobre el envejecimiento que el propio Carlos ha llevado a cabo con muchos de sus discípulos.

¡Gracias a todos! ¡De corazón!

5. Referencias

Referencias y bibliografía de consulta de trabajos de los grupos SMIRE, SMIRE+CoDiRE y SMABSS

Alonso, M.C.; Brezmes, T.; Lubiano, M.A.; Bertoluzza, C. (2001). A generalized real-valued measure of the inequality associated with a fuzzy random variable. *International Journal of Approximate Reasoning*, Vol. 26, nº 1, pp. 47-66.

Arellano, I.; Sinova, B.; De la Rosa de Sáa S.; Lubiano M.A.; Gil M.Á. (2019). Descriptive comparison of the rating scales through different scale estimates: simulation-based analysis. En: Destercke, S.; Denoeux, T.; Gil, M.Á.; Grzegorzewski, P.; Hryniewicz, O. (Eds.) *Uncertainty Modelling in Data Science. Advances in Intelligent Systems and Computing*, Vol. 832, Springer, pp. 9-16.

Bertoluzza, C.; Corral, N.; Salas, A. (1995). On a new class of distances between fuzzy numbers. *Mathware and Soft Computing*, Vol. II, nº 2, pp. 71-84.

Blanco-Fernández, A.; Casals, R.M.; Colubi, A.; Coppi, R.; Corral, N.; De la Rosa de Sáa, S.; D'Urso, P.P.; Ferraro, M.B.; García-Bárzana, M.; Gil, M.Á.; Giordani, P.; González-Rodríguez, G.; López, M.T.; Lubiano, M.A.; Montenegro, M.; Nakama, T.; Ramos-Guajardo, A.B.; Sinova, B.; Trutschnig, W. (2013). Arithmetic and distance-based approach to the statistical analysis of imprecisely valued data. En: Borgelt, C.; Gil, M.Á.; Sousa, J.M.C.; Verleysen, M. (Eds.) *Towards Advanced Data Analysis by Combining Soft Computing and Statistics. Studies in Fuzziness and Soft Computing*, Vol. 285, Springer, pp. 1-18.

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. (2013). Random fuzzy sets: a mathematical tool to develop statistical fuzzy data analysis. *Iranian Journal on Fuzzy Systems*, Vol. 10, nº 2, pp. 1-28.

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. (2014a). A distance-based statistical analysis of fuzzy number-valued data. *International Journal of Approximate Reasoning*, Vol. 55, pp. 1487-1501.

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. (2014b). Rejoinder on "A distance-based statistical analysis of fuzzy number-valued data". *International Journal of Approximate Reasoning*, Vol. 55, pp. 1601-1605.

Blanco-Fernández, A.; Colubi, A.; Corral, N.; González-Rodríguez, G. (2008). On a linear independence test for interval-valued random sets. En: Dubois, D.; Lubiano, M.A.; Prade, H.; Gil, M.Á.; Grzegorzewski, P.; Hryniewicz, O. (Eds.) *Soft Methods for Handling Variability and Imprecision. Advances in Soft Computing*, Vol. 48, Springer, pp. 111-117.

Blanco-Fernández, A.; Colubi, A.; García-Bárzana, M.; Montenegro, M. (2013). A linear regression model for interval-valued data based on set arithmetic. En: 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. Advances in Intelligent Systems and Computing*, Vol. 190, Springer, pp. 105-113.

Blanco-Fernández, A., Corral, N., González-Rodríguez, G. (2011). Estimation of a flexible simple linear model for interval data based on set arithmetic. *Computational Statistics and Data Analysis*, Vol. 55, nº 9, pp. 2568-2578.

Blanco-Fernández, A.; Corral, N.; González-Rodríguez, G.; Lubiano, M.A. (2008). Some properties of the d_K-variance for interval-valued random sets. En: Dubois, D.; Lubiano, M.A.; Prade, H.; Gil, M.Á.; Grzegorzewski, P.; Hryniewicz, O. (Eds.) *Soft Methods for Handling Variability and Imprecision. Advances in Soft Computing*, Vol. 48, Springer, pp. 331-337.

Blanco-Fernández, A.; Corral, N.; González-Rodríguez, G.; Palacio, A. (2010). On some confidence regions to estimate a linear regression model for interval data. En: 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. Advances in Intelligent and Soft Computing*, Vol. 77, Springer, pp. 33-40.

Casals, M.R.; Corral, N.; Gil, M.Á.; López, M.T.; Lubiano, M.A.; Montenegro, M.; Naval, G.; Salas, A. (2013). Bertoluzza et al.'s metric as a basis for analyzing fuzzy data. *Metron*, Vol. 71, nº 3, pp. 307-322.

Casals, M.R.; Gil, M.Á. (1989). A note on the operativeness of Neyman-Pearson tests with fuzzy information. *Fuzzy Sets and Systems*, Vol. 30, nº 2, pp. 215-220.

Casals, M.R.; Gil, M.Á.; Gil, P. (1986a). The fuzzy decision problem: an approach to the problem of testing statistical hypotheses with fuzzy information*. European Journal of Operational Research*, Vol. 27, nº 3, pp. 371-382.

Casals, M.R.; Gil, M.Á.; Gil, P. (1986b). On the use of Zadeh's probabilistic definition for testing statistical hypotheses from fuzzy information. *Fuzzy Sets and Systems*, Vol. 20, nº 2, 175-190.

Casals, M.R.; Gil, P. (1990). Contraste bayesiano de hipótesis difusas compuestas a partir de información difusa. *Revista de la Real Academia de Ciencias de Madrid*, Vol. LXXXIV, pp. 441-451.

Casals, M.R.; Salas, A. (1988). Sequential Bayesian test from fuzzy experimental information. En: Bouchon-Meunier, B., Saitta, L., Yager, R.R. (Eds.) *Uncertainty and Intelligent Systems*. IPMU 1988. *Lecture Notes in Computer Science*, Vol. 313, Springer, pp. 314-321.

Cascos, I.; López-Díaz, M.; Gil, M.Á. (2002). On the variation of the f-inequality of a random variable. En: Grzegorzewski, P.; Hryniewicz, O.; Gil, M.Á. (Eds.) *Soft Methods in Probability, Statistics and Data Analysis. Advances in Soft Computing*, Vol. 16, Physica-Verlag, pp. 98-104.

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

Cascos, I.; Pandolfo, G.; Sinova, B. (2023). The zonoid region parameter depth. *Statistical Papers*, Vol. 64, pp. 2183-2205.

Castaño-Pérez, A.M.; Lubiano, M.A.; García-Izquierdo, A.L. (2020). Gendered beliefs in STEM undergraduates: A comparative analysis of fuzzy rating versus Likert scales. *Sustainability*, Vol. 12, nº 15, 6227.

Colubi, A.; Coppi, R.; D'Urso, P.; Gil, M.Á. (2007). Statistics with fuzzy random variables. *Metron*, Vol. LXV, nº 3, pp. 277-303.

Colubi, A; Domínguez-Menchero, J.S.; López-Díaz, M.; Körner, R. (2001). A method to derive strong laws of large numbers for random upper semicontinuous functions. *Statistics & Probability Letters*, Vol. 53, nº 3, pp. 269-275.

Colubi, A; Domínguez-Menchero, J.S.; López-Díaz, M.; Ralescu, D.A. (2001). On the formalization of fuzzy random variables. *Information Sciences*, Vol. 133, nº 1-2, pp. 3-6.

Colubi, A.; Domínguez-Menchero, J.S.; López-Díaz, M.; Ralescu, D.A. (2002). A D[0,1]-representation of random upper semicontinuous functions. *Proceedings of the American Mathematical Society*, Vol. 130, nº 11, pp. 3237-3242.

Colubi, A.; Fernández-García, C.; Gil, M.Á. (2002). Simulation of random fuzzy variables: an empirical approach to statistical/probabilistic studies with fuzzy experimental data. *IEEE Transactions on Fuzzy Systems*, Vol. 10, nº 3, pp. 384-390.

Colubi, A.; Gil, M.Á.; González-Rodríguez, G., López, M.T. (2008). Empirical comparisons of goodness-of-fit tests for binomial distributions based on fuzzy representations. En: Dubois, D.; Lubiano, M.A.; Prade, H.; Gil, M.Á.; Grzegorzewski, P.; Hryniewicz, O. (Eds.) *Soft Methods for Handling Variability and Imprecision. Advances in Soft Computing*, Vol. 48, Springer, pp. 190-197.

Colubi, A.; González Rodríguez, G.; Gil, M.Á.; Trutschnig, W. (2011). Nonparametric criteria for supervised classification of fuzzy data. *International Journal of Approximate Reasoning*, Vol. 52, nº 9, pp. 1272-1282.

Colubi, A.; González-Rodríguez, G.; Lubiano, M.A.; Montenegro, M. (2006). Exploratory analysis of random variables based on fuzzification. En: Lawry, J.; Miranda, E.; Bugarin, A.; Li, S.; Gil, M.Á.; Grzegorzewski, P.; Hryniewicz, O. (Eds.) *Soft Methods for Integrated Uncertainty Modelling. Advances in Soft Computing*, Vol. 37, Springer, pp. 95-102.

Colubi, A.; López-Díaz, M.; Domínguez-Menchero, J.S.; Gil, M.Á. (1999). A generalized Strong Law of Large Numbers. *Probability Theory and Related Fields*, Vol. 114, pp. 401-417.

Coppi, R.; Gil, M.Á.; Kiers, H.A.L., Guest Editors (2006). The fuzzy approach to statistical analysis. *Computational Statistics and Data Analysis*, Vol. 51, nº 1, pp. 1-14.

Coppi, R.; Gil, M.A; Kiers, H.A.L. (2007). Book review on The Fuzzy Approach to Statistical Analysis (A Special Issue of the journal of 'Computational Statistics and Data Analysis', Vol. 51, nº 1, ISSN 0167-9473, 2006, pp. 1-452). *Fuzzy Sets and Systems*, Vol. 158, nº 19, pp. 2203-2208.

Corral, N.; Gil, M.Á. (1984). The minimum inaccuracy fuzzy estimation: an extension of the maximum likelihood principle. *Stochastica*, Vol. VIII, pp. 63-81.

Corral, N.; Gil, M.Á. (1988). A note on interval estimation with fuzzy data. *Fuzzy Sets and Systems*, Vol. 28, nº 2, pp. 209-215.

Corral, N.; Gil, M.Á.; López, M.T., Salas, A.; Bertoluzza, C. (1998). Statistical models. En: Ruspini, E.H.; Bonissone, P.; Pedrzyck, W. (Eds.). *Handbook of Fuzzy Computation*, Section C: 2.3, Institute of Physics, pp. 1-18.

Corral, N.; Gil, M.Á.; López-García, H. (1996). The fuzzy hyperbolic inequality index of fuzzy random variables in finite populations. *Mathware and Soft Computing*, Vol. III, nº 3, pp. 329-339.

Corral, N.; Gil, M.Á.; Gil. P. (2011). Interval and fuzzy-valued approaches to the statistical management of imprecise data. En: Pardo, L.; Balakrishnan, N.; Gil, M.Á. (Eds.) *Modern Mathematical Tools and Techniques in Capturing Complexity. Understanding Complex Systems*, Vol. 2011, Springer, pp. 453-468.

Corral Blanco, N.; Zurbano Fernández, E.; Blanco Fernández, Á.; García Honrado, I.; Ramos Guajardo, A.B. (2013). Capítulo 1. Estructura del entorno educativo familiar: Su influencia sobre el rendimiento y el rendimiento diferencial. En: *PIRLS - TIMSS 2011. Estudio Internacional de progreso en comprensión lectora, matemáticas y ciencias. IEA. Volumen II. Informe Español. Análisis Secundario* Ministerio de Educación, Cultura y Deporte.

De la Rosa de Sáa, S.; Carleos, C.; López, M.T.; Montenegro, M. (2018). 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. Studies in Systems, Decision and Control*, Vol. 142, Springer, pp. 709-719.

De la Rosa de Sáa, S.; Filzmoser, P.; Gil, M.Á.; Lubiano, M.A. (2015). On the robustness of absolute deviations with fuzzy data. En: Grzegorzewski, P.; Gagolewski, M.; Hryniewicz, O.; Gil, M.Á. (Eds.) *Strengthening Links Between Data Analysis and Soft Computing. Advances in Intelligent Systems and Computing*, Vol. 315, Springer, pp. 133-141.

De la Rosa de Sáa, S.; Gil, M.Á.; González-Rodríguez, G.; López, M.T.; Lubiano, M.A. (2015). Fuzzy rating scale-based questionnaires and their statistical analysis. *IEEE Transactions on Fuzzy Systems*, Vol. 23, nº 1, pp. 111-126.

De la Rosa de Sáa, S.; Gil, M.A.; López, M.T.; Lubiano, M.A. (2013). Fuzzy rating vs. fuzzy conversion scales: an empirical comparison through the MSE. En: 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. Advances in Intelligent Systems and Computing*, Vol. 190, Springer, pp. 135-143.

De la Rosa de Sáa, S.; Lubiano, M.A.; Sinova, B.; Filzmoser, P.; Gil, M.Á. (2021). Location-free robust scale estimates for fuzzy data. *IEEE Transactions on Fuzzy Systems*, Vol. 29, nº 6, pp. 1682-1694.

De la Rosa de Sáa, S.; Lubiano, M.A.; Sinova, B.; Filzmoser, P. (2017). Robust scale estimators for fuzzy data. *Advances in Data Analysis and Clasification*, Vol. 11, nº 4, pp. 731-758.

De la Rosa de Sáa, S.; Van Aelst, S. (2013). Comparing the representativeness of the 1-norm median for Likert and free-response fuzzy scales. En: Borgelt, C.; Gil, M.A.; Sousa, J.M.C.; Verleysen, M. (Eds.) *Towards Advanced Data Analysis by Combining Soft Computing and Statistics. Studies in Fuzziness and Soft Computing*, Vol. 285, Springer, pp. 87-98.

D'Urso, P.; Gil, M.Á. (2013). Fuzzy Statistical Analysis: methods and Applications. *Metron*, Vol. 71, nº 3, pp. 197-199.

D'Urso, P.; Gil, M.Á. (2017). Fuzzy data analysis and classification. *Advances in Data Analysis and Clasification*, Vol. 11, nº 4, pp. 645–657.

García, D.; Lubiano, M.A.; Alonso, M.C. (2001). Estimating the expected value of fuzzy random variables in the stratified random sampling from finite populations. *Information Sciences*, Vol. 138, nº 1-4, pp. 165-184.

García-García, J.; Lubiano, M.A. (2022, 2025). *IntervalQuestionStat: Tools to Deal with Interval-Valued Responses in Questionnaires*. Version 0.2.0. https://cran.r-project.org/web/packages/IntervalQuestionStat, https://cran.r-project.org/web/packages/IntervalQuestionStat/IntervalQuestionStat.pdf.

García-García, J.; Gil, M.Á.; Lubiano, M.A. (2024a). On some properties of Cronbach's α coefficient for interval-valued data in questionnaires. *Advances in Data Analysis and Classification*, doi:10.1007/s11634-024-00601-w.

García-García, J.; Gil, M.Á.; Lubiano, M.A. (2024b). Sensitivity analysis on the choice of the metric on Cronbach's α coefficient for interval-valued data in questionnaires. En: Ansari, J.; Fuchs, S.; Trutschnig, W.; Lubiano, M.A.; Gil, M.Á.; Grzegorzewski, P.; Hryniewicz, O. (Eds.) *Combining, Modelling and Analyzing Imprecision, Randomness and Dependence. Advances in Intelligent Systems and Computing*, Vol. 1458, Springer, pp. 149-157.

García-Izquierdo, A.L.; Ramos-Villagrasa, P.J.; Lubiano, M.A. (2020). Developing biodata for public manager selection purposes: A comparison between fuzzy logic and traditional methods. *Journal of Work and Organizational Psychology*, Vol. 36, nº 3, pp. 231-242.

Gebhardt, J.; Gil, M.Á.; Kruse, R. (1998). Fuzzy set-theoretic methods in Statistics. En: Slowinski, R. (Ed.) *Fuzzy Sets in Decision Analysis, Operations Research and Statistics*. Kluwer Academic Publishers, pp. 311-347.

Gil, M.Á. (1987). Fuzziness and loss of information in statistical problems. *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 17, pp. 1016-1025.

Gil, M.Á. (1988a). Variation in the mean squared error of estimates due to fuzziness in the information system. *Fuzzy Systems and Mathematics*, Vol. 2, pp. 46-56.

Gil, M.Á. (1988b). On the loss of information due to fuzziness in experimental observations. *Annals of the Institute of Statistical Mathematics*, Vol. 44, nº 3, pp. 451-462.

Gil, M.Á. (1988c). Bayesian decision making with previous probabilistic uncertainty and actual fuzzy imprecision. *Kybernetes*, Vol. 17, nº 3, pp. 52-66.

Gil, M.Á. (1992a). A note on the connection between fuzzy numbers and random intervals. *Statistics & Probability Letters*, Vol. 13, nº 4, pp. 311-319.

Gil, M.Á. (1992b). Sufficiency and fuzziness in random experiments. *Annals of the Institute of Statistical Mathematics*, Vol. 44, nº 3, pp. 451-462.

Gil, M.Á. (1993a). Analyzing the meaning of fuzziness in random experiments. En: *Fuzzy Logic: State of the Art*. Kluwer Academic Publishers, pp. 429-439.

Gil, M.Á. (1993b). Statistical management of fuzzy elements in random experiments. Part 1: A discussion on treating fuzziness as a kind of randomness. *Information Sciences*, Vol. 69, nº 3, pp. 229-242.

Gil, M.Á. (1993c). Análisis y tratamiento estadístico de elementos difusos en experimentos aleatorios (artículo invitado con discusiones). *Estadística Española*, Vol. 35, nº 134, pp. 477-525.

Gil, M.Á. (2001). Fuzzy random variables. *Information Sciences*, Vol. 133, nº 1-2, pp. 1-2.

Gil, M.Á. (2003). 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, Hermes Science Publications, pp. 205-243.

Gil, M.Á. (2013). On a meeting point between Fuzzy Sets and Statistics. En: Seising, R., Trillas, E., Moraga, C., Termini, S. (Eds.) *On Fuzziness. Studies in Fuzziness and Soft Computing*, Vol. 298, Springer, pp. 199-204.

Gil, M.Á. (2014). Comments on "Statistical reasoning with set-valued information: Ontic *vs.* epistemic views". *International Journal of Approximate Reasoning*, Vol. 55, pp. 1580-1582.

Gil, M.Á. (2018). Fuzzy random variables à la Kruse & Meyer and à la Puri & Ralescu: Key differences and coincidences. En: Mostaghim, S.; Nürnberger, A.; Borgelt, C. (Eds.) *Frontiers in Computational Intelligence. Studies in Computational Intelligence*, Vol. 739, Springer, pp. 21-29.

Gil, M.Á. (2023). *Las Matemáticas del Análisis de Datos Imprecisos*. Discurso leído el 25 de enero de 2023 en el acto de su recepción como académica de número en la Real Academia de Ciencias Exactas, Físicas y Naturales. Contestación: F. Javier Girón González-Torre.

Gil, M.Á.; Brezmes, T. (1984). La quietud esperada: un criterio para comparar sistemas de información difusos. *Revista de la Real Academia de Ciencias de Zaragoza*, Vol. 39, pp. 13-23.

Gil, M.Á.; Brezmes, T. (1985). Fuzzified Blackwell's method to compare experiments. *R.A.I.R.O.-Recherche Opérationnelle*, Vol. 19, pp. 105-111.

Gil, M.Á.; Casals, M.R. (1988). An operative extension of the likelihood ratio test from fuzzy data. *Statistical Papers*, Vol. 29, nº 3, pp. 191-203.

Gil, M.Á.; Colubi, A.; Terán, P. (2014). Random fuzzy sets: why, when, how. *Boletín de Estadística e Investigación Operativa*, Vol. 30, nº 1, pp. 5-29.

Gil, M.Á.; Corral, N.; Casals, M.R. (1989). The likelihood ratio test for goodness of fit with fuzzy experimental observations. *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 19, nº 4, pp. 771-779.

Gil, M.Á.; Corral, N.; Gil, P. (1985). The fuzzy decision problem: an approach to the point estimation problem with fuzzy information. *European Journal of Operational Research*, Vol. 26, nº 1, pp. 26-34.

Gil, M.Á.; Corral, N.; Gil, P. (1988). The minimum inaccuracy estimates in χ^2 tests for goodness of fit with fuzzy observations. *Journal of Statistical Planning and Inference*, Vol. 19, nº 1, pp. 95-115.

Gil, M.Á.; Gil, P. (1988a). Sobre el tamaño de muestra para experimentos aleatorios con imprecisión difusa. *Trabajos de Estadística*, Vol. 3, pp. 33-53.

Gil, M.Á.; Gil, P. (1988b). Los sistemas de información difusos y la definición probabilística de Zadeh. *Revista de la Real Academia de Ciencias de Madrid*, Vol. LXXXII, pp. 405-424.

Gil, M.Á.; Gil, P. (1992). Fuzziness in the experimental outcomes: comparing experiments and removing the loss of information. *Journal of Statistical Planning and Inference*, Vol. 31, nº 1, pp. 93-111.

Gil, M.Á.; Gil, P. (2015). Randomness and fuzziness: combined better than unified. En: Magdalena, L.; Verdegay, J.L.; Esteva, F. (Eds.) *Enric Trillas: A Passion for Fuzzy Sets. Studies in Fuzziness and Soft Computing*, Vol. 322, Springer, pp. 213-222.

Gil, M.Á.; Gil, P.; Ralescu, D.A. (1999). Fuzzy random variables: modeling linguistic statistical data. En: Zadeh, L.A.; Kacprzyk, J. (Eds.) *Computing with Words in Information/Intelligent Systems 2*, Physica-Verlag, pp. 137-157.

Gil, M.Á.; González-Rodríguez, G. (2012). 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. Studies in Fuzziness and Soft Computing*, Vol. 271, Springer, pp. 407-420.

Gil, M.Á.; González-Rodríguez, G.; Colubi, A.; Montenegro, M. (2007). Testing linear independence in linear models with interval-valued data. *Computational Statistics and Data Analysis*, Vol. 51, nº 6, pp. 3002-3015.

Gil, M.Á., González-Rodríguez, G.; Kruse, R., Guest Editors (2013). Editorial of the Special Issue "Statistics with Imperfect Data". *Information Sciences*, Vol. 245, pp. 1-3.

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

Gil, M.Á.; Hryniewicz, O. (2023). Statistics with imprecise data. En: Lin, T.Y.; Liau, C.J.; Kacprzyk, J. (Eds.) *Granular, Fuzzy, and Soft Computing.* Part of the *Encyclopedia of Complexity and Systems Science Series book series*, Springer, pp. 895-909.

Gil, M.Á.; Jain, P. (1992). Comparison of experiments in decision problems with fuzzy utilities. *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 22, nº 4, p. 662-670.

Gil, M.Á.; López, M.T. (1993). Statistical management of fuzzy elements in random experiments. Part 2: The Fisher Information associated with a Fuzzy Information System. *Information Sciences*, Vol. 69, nº 3, pp. 243-257.

Gil, M.Á.; López, M.T.; Garrido, J.M.A. (1987). An extensive-form analysis for comparing fuzzy information systems by means of the worth and quietness of information. *Fuzzy Sets and Systems*, Vol. 23, nº 2, pp. 239-255.

Gil, M.Á.; López, M.T.; Gil, P. (1984). Comparison between fuzzy information systems. *Kybernetes*, Vol. 13, pp. 245-251.

Gil, M.Á.; López, M.T.; Gil, P. (1985a). Quantity of information; Comparison between information systems: 1. Non-fuzzy states. *Fuzzy Sets and Systems*, Vol. 15, nº 1, pp. 65-78.

Gil, M.Á.; López, M.T.; Gil, P. (1985b). Quantity of information; Comparison between information systems: 2. Fuzzy states. *Fuzzy Sets and Systems*, Vol. 15, nº 2, pp. 129-145.

Gil, M.Á.; López, M.T.; Lubiano, M.A.; Montenegro, M. (2001). Regression and correlation analyses of a linear relation between random intervals. *Test*, Vol. 10, nº 1, pp. 183-201.

Gil, M.Á.; López-Díaz, M. (1996). Fundamentals and Bayesian analyses of decision problems with fuzzy-valued utilities. *International Journal of Approximate Reasoning*, Vol. 15, nº 3, pp. 203-224.

Gil, M.Á.; López-Díaz, M.; López-García, H. (1998). The fuzzy hyperbolic inequality index associated with fuzzy random variables. *European Journal of Operational Research*, Vol. 110, nº 2, pp. 377-391.

Gil, M.Á.; López-Díaz, M.; Ralescu, D.A., Guest Editors (2006a). Editorial of the Special Issue "Fuzzy Sets and Probability/Statistics Theories". *Fuzzy Sets and Systems*, Vol. 157, nº 19, p. 2545.

Gil, M.Á.; López-Díaz, M.; Ralescu, D.A. (2006b). Overview on the development of fuzzy random variables. *Fuzzy Sets and Systems*, Vol. 157, nº 19, pp. 2546-2557.

Gil, M.Á.; López-Díaz, M.; Rodríguez-Muñiz, L.J. (1998). An improvement of a comparison of experiments in statistical decision problems with fuzzy utilities. *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 28, nº 6, pp. 856-864.

Gil, M.Á.; Lubiano, M. A.; De la Rosa de Sáa, S.; Sinova, B. (2015). Analyzing data from a fuzzy rating scale-based questionnaire. A case study. *Psicothema*, Vol. 27, nº 2, pp. 182-191.

Gil, M.Á.; Lubiano, M.A.; Montenegro, M.; López, M.T. (2002). Least squares fitting of an affine function and strength of association for interval-valued data. *Metrika*, Vol. 56, nº 2, pp. 97-111.

Gil, M.Á.; Montenegro, M.; González-Rodríguez, G.; Colubi, A.; Casals, R. (2006). Bootstrap approach to the multi-sample test of means with imprecise data. *Computational Statistics and Data Analysis*, Vol. 51, nº 1, pp. 148-162.

Gil, P.; Gil, M.Á. (1998). Análisis Bayesiano de problemas de decisión con valoración difusa de las consecuencias. *Revista de la Real Academia de Ciencias de Madrid*, Vol. XCII, pp. 355-361.

Gil, P.; Gil, M.Á.; Menéndez, M.L., Pardo, L. (1990). Connections between some criteria to compare fuzzy information systems. *Fuzzy Sets and Systems*, Vol. 37, nº 2, pp. 183-192.

González-Rodríguez, G.; Blanco, A.; Colubi, A.; Lubiano, M.A. (2009). Estimation of a simple linear regression model for fuzzy random variables. *Fuzzy Sets and Systems*, Vol. 160, nº 3, pp. 357-370.

González-Rodríguez, G.; Blanco, A.; Corral, N.; Colubi, A. (2007). Least squares estimation of linear regression models for convex compact random sets. *Advances in Data Analysis and Classification*, Vol. 1, nº 1, pp. 67-81.

González-Rodríguez, G.; Colubi, A.; D'Urso, P.; Montenegro, M. (2009). Multi-sample test-based clustering for fuzzy random variables. *International Journal of Approximate Reasoning*, Vol. 50, nº 5, pp. 721?731.

González-Rodríguez, G.; Colubi, A.; Gil, M.Á. (2006). A fuzzy representation of random variables: an operational tool in exploratory analysis and hypothesis testing. *Computational Statistics and Data Analysis*, Vol. 51, nº 1, pp. 163 – 176.

González-Rodríguez, G.; Colubi, A.; Gil, M.Á. (2012). Fuzzy data treated as functional data. A one-way ANOVA test approach. *Computational Statistics and Data Analysis*, Vol. 56, nº 4, pp. 943-955.

González-Rodríguez, G.; Colubi, A.; Gil, M.Á.; Coppi, R. (2006). A method to simulate fuzzy random variables. En: Lawry, J.; Miranda, E.; Bugarin, A.; Li, S.; Gil, M.Á.; Grzegorzewski, P.; Hryniewicz, O. (Eds.) *Soft Methods for Integrated Uncertainty Modelling. Advances in Soft Computing*, Vol. 37, Springer, pp. 103-110.

González-Rodríguez, G.; Colubi, A.; Gil, M.Á.; Lubiano, M.A. (2012). A new way of quantifying the symmetry of a random variable: estimation and hypothesis testing. *Journal of Statistical Planning and Inference*, Vol. 142, nº 12, pp. 3061-3072.

González-Rodríguez, G.; Montenegro, M.; Colubi, A.; Gil, M.Á. (2006). Bootstrap techniques and fuzzy random variables: synergy in hypothesis testing with fuzzy data. *Fuzzy Sets and Systems*, Vol. 157, nº 19, pp. 2608-2613.

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

López, M.T.; Gil, P. (1986). Una medida de incertidumbre probabilística para sucesos difusos. *Trabajos de Estadística*, Vol. 1, pp. 60-69.

López-Díaz, M.; Gil, M.Á. (1997). Constructive definitions of fuzzy random variables. *Statistics & Probability Letters*, Vol. 36, nº 2, pp. 135-143.

López-Díaz, M.; Gil, M.Á. (1998a). Reversing the order of integration in iterated expectations of fuzzy random variables, and statistical applications. *Journal of Statistical Planning and Inference*, Vol. 74, nº 1, pp. 11-29.

López-Díaz, M.; Gil, M.Á. (1998b). Approximating integrably bounded fuzzy random variables in terms of the 'generalized' Hausdorff metric. *Information Sciences*, Vol. 104, nº 2, pp. 279-291.

López-Díaz, M.; Gil, M.Á. (1998c). The lambda-average value of the expected value of a fuzzy random variable. *Fuzzy Sets and Systems*, Vol. 99, nº 3, pp. 347-352.

López-Díaz, M.; Gil, M.Á. (1999). An extension of Fubini's Theorem for fuzzy random variables. *Information Sciences*, Vol. 115, nº 1-4, pp. 29-41.

López-Díaz, M.; Ralescu, D.A. (2006). Tools for fuzzy random variables: Embeddings and measurabilities. *Computational Statistics and Data Analysis*, Vol. 51, nº 1, pp. 109-114.

López-Díaz, M.; Rodríguez-Muñiz, L.J. (2004). Addition nodes in influence diagrams with fuzzy-valued utilities and variables. En: López-Díaz, M.: Gil, M.Á.; Grzegorzewski, P.; Hryniewicz, O.; Lawry, J. (Eds.) *Soft Methodology and Random Information Systems, Advances in Soft Computing*, Vol. 26, Springer, pp. 621-627.

López-Díaz, M.; Rodríguez-Muñiz, L.J. (2007). Influence diagrams with super value nodes involving imprecise information. *European Journal of Operational Research*, Vol. 179, nº 1, pp. 203-219.

López-García, H.; Gil, M.Á.; Corral, N.; López, M.T. (1998). Estimating the fuzzy inequality associated with a fuzzy random variable in random samplings from finite populations. *Kybernetika*, Vol. 34, nº 2, pp. 149-161.

López-García, H.; López-Díaz, M.; Gil, M.Á. (2000). Interval-valued quantification of the inequality associated with a random set. *Statistics & Probability Letters*, Vol. 46, nº 2, pp. 149-159.

Lubiano M.A.; Carleos, C.; Montenegro, M.; Gil M.Á. (2019). 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. Advances in Intelligent Systems and Computing*, Springer, Vol. 832, Springer, pp. 157-165.

Lubiano, M.A.; Colubi, A.; González-Rodríguez, G. (2008). Empirical results concerning a fuzzy-based measure of symmetry of real random variables. En: Dubois, D.; Lubiano, M.A.; Prade, H.; Gil, M.Á.; Grzegorzewski, P.; Hryniewicz, O. (Eds.) *Soft Methods for Handling Variability and Imprecision. Advances in Soft Computing*, Vol. 48, Springer, pp. 147-154.

Lubiano, M.A.; De la Rosa de Sáa (2017, 2025). *FuzzyStatTra: Statistical Methods for Trapezoidal Fuzzy Numbers*. Version 1.0. https://cran.r-project.org/web/packages/FuzzyStatTra, https://cran.r-project.org/web/packages/FuzzyStatTra/FuzzyStatTra.pdf.

Lubiano, M.A.; De la Rosa de Sáa, S., Montenegro, M., Sinova, B., Gil, M.Á. (2016). Descriptive analysis of responses to items in questionnaires. Why not a fuzzy rating scale? *Information Sciences*, Vol. 360, pp. 131-148.

Lubiano, M. A.; De la Rosa de Sáa, S.; Sinova, B.; Gil, M.Á. (2015). Empirical sensitivity analysis on the influence of the shape of fuzzy data on the estimation of some statistical measures. En: Grzegorzewski, P.; Gagolewski, M.; Hryniewicz, O.; Gil, M.Á. (Eds.) *Strengthening Links Between Data Analysis and Soft Computing. Advances in Intelligent Systems and Computing*, Vol. 315, Springer, pp. 123-131.

Lubiano, M.A.; García-García, J.; García-Izquierdo, A.L.; Castaño, A.M. (2023). The extended version of Cohen's d index for interval-valued data. En: García-Escudero, L.A.; Gordaliza, A.; Mayo, A.; Lubiano, M.A.; Gil, M.Á.; Grzegorzewski, P.; Hryniewicz, O. (Eds.) *Building Bridges between Soft and Statistical Methodologies for Data Science. Advances in Intelligent Systems and Computing*, Vol. 1433, Springer, pp. 263-270.

Lubiano, M.A.; García-Izquierdo, A.L.; Gil, M.Á. (2021). Fuzzy rating scales: does internal consistency of a measurement scale benefit from coping with imprecision and individual differences in psychological rating? *Information Sciences*, Vol. 550, pp. 91-108.

Lubiano, M.A.; Gil, M.Á. (1999). Estimating the expected value of fuzzy random variables in random samplings from finite populations. *Statistical Papers*, Vol. 40, nº 3, pp. 277-295.

Lubiano, M.A.; Gil, M.Á. (2002). *f*-Inequality indices for fuzzy random variables. En: Bertoluzza, C.; Gil, M.Á.; Ralescu, D.A. (Eds.) *Statistical Modeling, Analysis and Management of Fuzzy Data*, Physica-Verlag, pp. 43-63.

Lubiano, M.A.; Gil, M.Á.; López-Díaz, M. (1999). On the Rao-Blackwell theorem for fuzzy random variables. *Kybernetika*, Vol. 35, nº 2, pp. 167-175.

Lubiano, M.A.; Gil, M.Á.; López-Díaz, M.; López, M.T. (2000). The lambda-mean squared dispersion associated with a fuzzy random variable. *Fuzzy Sets and Systems*, Vol. 111, nº 3, pp. 307-317.

Lubiano, M.A.; Gil, M.Á.; Sinova, B.; Casals, M.R., López, M.T. (2017a). Measuring the dissimilarity between the distributions of two random fuzzy numbers. En: Ferraro, M.B.; Giordani, P.; Vantaggi, B.; Gagolewski, M.; Gil, M.Á.; Grzegorzewski, P.; Hryniewicz, O. (Eds.) *Soft Methods for Data Science. Advances in Intelligent Systems and Computing*, Vol. 456, Springer, pp. 319-327.

Lubiano, M.A.; González-Gil, P.; Sánchez-Pastor, H.; Pradas, C.; Arnillas, H. (2018). 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. Studies in Systems, Decision and Control*, Vol. 142, Springer, pp. 653-662.

Lubiano, M.A.; Montenegro, M.; Pérez-Fernández, S.; Gil, M.Á. (2023). Analyzing the influence of the rating scale for items in a questionnaire on Cronbach coefficient alpha. En: Balakrishnan, N.; Gil, M.Á.; Martín, N.; Morales, D.; Pardo, M.d.C. (Eds.) *Trends in Mathematical, Information and Data Sciences. Studies in Systems, Decision and Control*, Vol. 445, Springer, pp. 377-388.

Lubiano, M.A., Montenegro, M., Sinova, B., De la Rosa de Sáa, S., Gil, M.Á. (2016). Hypothesis testing for means in connection with fuzzy rating scale-based data: algorithms and applications. *European Journal of Operational Research*, Vol. 251, pp. 918-929.

Lubiano, M.A.; Salas, A.; Carleos, C.; De la Rosa de Sáa, S.; Gil, M.Á. (2017b). Hypothesis testing-based comparative analysis between rating scales for intrinsically imprecise data. *International Journal of Approximate Reasoning*, Vol. 88, pp. 128-147.

Lubiano, M.A.; Salas, A.; De la Rosa de Sáa, S.; Montenegro, M.; Gil, M.Á. (2017c). An empirical analysis of the coherence between fuzzy rating scale- and Likert scale-based responses to questionnaires. En: Ferraro, M.B.; Giordani, P.; Vantaggi, B.; Gagolewski, M.; Gil, M.Á.; Grzegorzewski, P.; Hryniewicz, O. (Eds.) *Soft Methods for Data Science. Advances in Intelligent Systems and Computing*, Vol. 456, Springer, pp. 329-337.

Lubiano, M.A.; Salas, A.; Gil, M.Á. (2017d). A hypothesis testing-based discussion on the sensitivity of means of fuzzy data with respect to data shape. *Fuzzy Sets and Systems*, Vol. 328, nº 1, pp. 54-69.

Lubiano, M.A.; Trutschnig, W. (2010). ANOVA for fuzzy random variables using the R-package SAFD. En: 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. Advances in Intelligent and Soft Computing*, Vol. 77, Springer, pp. 449-456.

Montenegro, M.; Casals, M.R.; Colubi, A.; Gil, M.Á. (2008). Testing 'two-sided' hypothesis about the mean of an interval-valued random set. En: Dubois, D.; Lubiano, M.A.; Prade, H.; Gil, M.Á.; Grzegorzewski, P.; Hryniewicz, O. (Eds.) *Soft Methods for Handling Variability and Imprecision. Advances in Soft Computing*, Vol. 48, Springer, pp. 133-139.

Montenegro, M.; Casals, M.R.; Lubiano, M.A.; Gil, M.Á. (2001). Two-sample hypothesis tests of means of a fuzzy random variable. *Information Sciences*, Vol. 133, nº 1-2, pp. 89-100.

Montenegro, M.; Colubi, A.; Casals, R.; Gil, M.Á. (2002). Test of one-sided hypotheses on the expected value of a fuzzy random variable. En: Grzegorzewski, P.; Hryniewicz, O.; Gil, M.Á. (Eds.) *Soft Methods in Probability, Statistics and Data Analysis. Advances in Soft Computing*, Vol. 16, Physica-Verlag, pp. 228-235.

Montenegro, M.; Colubi, A.; Casals, M.R.; Gil, M.Á. (2004). Asymptotic and Bootstrap techniques for testing the expected value of a fuzzy random variable. *Metrika*, Vol. 59, nº 1, pp. 31-49.

Montenegro, M.; Gil, M.Á.; López, M.T.; Lubiano, M.A (2002). Quantifying the linear correlation between two interval-valued random sets. En: Bouchon-Meunier, B.; Gutiérrez-Ríos, J.; Magdalena, L.; Yager, R.R. (Eds.) *Technologies for Constructing Intelligent Systems 2 - Tools*, Springer, pp. 99-108.

Montenegro, M.; González-Rodríguez, G.; Colubi, A.; Gil, M.Á. (2004). Bootstrap approach to test the linear independence between interval-valued random sets. . En: López-Díaz, M.: Gil, M.Á.; Grzegorzewski, P.; Hryniewicz, O.; Lawry, J. (Eds.) *Soft Methodology and Random Information Systems, Advances in Soft Computing*, Vol. 26, Springer, pp. 431-438.

Montenegro, M.; González-Rodríguez, G.; Gil, M.Á.; Colubi, A.; Casals, R. (2004). Introduction to ANOVA with fuzzy random variables. . En: López-Díaz, M.: Gil, M.Á.; Grzegorzewski, P.; Hryniewicz, O.; Lawry, J. (Eds.) *Soft Methodology and Random Information Systems, Advances in Soft Computing*, Vol. 26, Springer, pp. 487-494.

Nakama, T.; Colubi, A.; Lubiano, M.A. (2010). Two-way analysis of variance for interval-valued data. En: 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. Advances in Intelligent and Soft Computing*, Vol. 77, Springer, pp. 475-482.

Ramos-Guajardo, A.B.; Colubi, A.; González-Rodríguez, G.; Gil, M.Á. (2010). One sample tests for a generalized Fréchet variance of a fuzzy random variable. *Metrika*, Vol. 71, nº 2, pp. 185-202.

Ramos-Guajardo, A.B.; González-Rodríguez, G.; Colubi, A.; Gil, M.Á. (2008). Asymptotic tests for the variance of a fuzzy random variable using the D_K-metric. En: Dubois, D.; Lubiano, M.A.; Prade, H.; Gil, M.Á.; Grzegorzewski, P.; Hryniewicz, O. (Eds.) *Soft Methods for Handling Variability and Imprecision. Advances in Soft Computing*, Vol. 48, Springer, pp. 140-146.

Ramos-Guajardo, A.B.; González-Rodríguez, G.; Montenegro, M.; López, M.T. (2010). Power analysis of the homoscedasticity test for random fuzzy sets. En: 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. Advances in Intelligent and Soft Computing*, Vol. 77, Springer, pp. 537-544.

Ramos-Guajardo, A.B.; Lubiano, M.A. (2012). K-sample tests for equality of variances of random fuzzy sets. *Computational Statistics and Data Analysis*, Vol. 56, nº 4, pp. 956-966.

Ramos-Guajardo, A.B.; Lubiano, M.A.; González-Rodríguez, G. (2013). Bootstrap comparison of statistics for testing the homoscedasticity of random fuzzy sets. En: 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. Advances in Intelligent Systems and Computing*, Vol. 190, Springer, pp. 125-133.

Rodríguez-Muñiz, L.J.; López-Díaz, M. (2003). Hukukara derivative of the fuzzy expected value. *Fuzzy Sets and Systems*, Vol. 183, nº 3, pp. 593-600.

Rodríguez-Muñiz, L.J.; López-Díaz, M. (2007). On the exchange of iterated expectations of random upper semicontinuous functions. *Statistics & Probability Letters*, Vol. 77, pp. 1628-1635.

Rodríguez-Muñiz, L.J.; López-Díaz, M. (2008). A new framework for the Bayesian analysis of single-stage decision problems with imprecise utilities. *Fuzzy Sets and Systems*, Vol. 159, nº 24, pp. 3271-3280.

Rodríguez-Muñiz, L.J.; López-Díaz, M. (2010). Different models with fuzzy random variables in single-stage decision problems. En: Hüllermeier, E.; Kruse, R.; Hoffmann, F. (Eds.) *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Applications. Communications in Computer and Information Science*, Vol. 81, Springer, pp. 298-305.

Rodríguez-Muñiz, L.J.; López-Díaz, M.; Gil, M.Á. (2003). Differentiating random upper semicontinuous functions under the integral sign. *Test*, Vol. 12, nº 1, pp. 241-258.

Rodríguez-Muñiz, L.J.; López-Díaz, M.; Gil, M.Á. (2002). Reversing the order of integration in iterated expectations of compact convex random sets. En: Grzegorzewski, P.; Hryniewicz, O.; Gil, M.Á. (Eds.) *Soft Methods in Probability, Statistics and Data Analysis. Advances in Soft Computing*, Vol. 16, Physica-Verlag, pp. 140-145.

Rodríguez-Muñiz, L.J.; López-Díaz, M.; Gil, M.Á. (2005). Solving influence diagrams with fuzzy chance and value nodes. *European Journal of Operational Research*, Vol. 167, nº 2, pp. 444-460.

Rodríguez-Muñiz, L.J.; López-Díaz, M.; Gil, M.Á.; Ralescu, D.A. (2003). The s-differentiability of a fuzzy-valued mapping. *Information Sciences*, Vol. 151, pp. 283-299.

Ruiz, M.M.; Gil, M.Á. (1992). La afinidad no probabilística: una medida de borrosidad de un subconjunto difuso. *Revista de la Real Academia de Ciencias de Madrid*, Vol. LXXXVI, pp. 251-262.

Salas, A.; Corral, N.; Bertoluzza, C. (2002). Linear regresión in a fuzzy context. The least square method. En: Bertoluzza, C.; Gil, M.Á.; Ralescu, D.A. (Eds.). *Statistical Modeling, Analysis and Management of Fuzzy Data*, Physica-Verlag, pp. 255-281.

Sinova, B. (2016). M-estimators of location for interval-valued random elements. *Chemometrics and Intelligent Laboratory Systems*, Vol. 156, pp. 115-127.

Sinova, B. (2018). Scale equivariant alternative for fuzzy M-Estimators of location. En: Gil, E.; Gil, E.; Gil, J.; Gil, M.Á. (Eds.) *The Mathematics of the Uncertain: A Tribute to Pedro Gil. Studies in Systems, Decision and Control*, Vol. 142, Springer, pp. 733-743.

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

Sinova, B. (2023a). Empirical analysis of the Maxbias of the fuzzy MDD-based Hampel M-estimator of location. En: Balakrishnan, N.; Gil, M.Á.; Martín, N.; Morales, D.; Pardo, M.d.C. (Eds.) *Trends in Mathematical, Information and Data Sciences. Studies in Systems, Decision and Control*, Vol. 445, Springer, pp. 447-456.

Sinova, B. (2023b). The d_θ depth-based interval trimmed mean. En: García-Escudero, L.A.; Gordaliza, A.; Mayo, A.; Lubiano, M.A.; Gil, M.Á.; Grzegorzewski, P.; Hryniewicz, O. (Eds.) *Building Bridges between Soft and Statistical Methodologies for Data Science. Advances in Intelligent Systems and Computing*, Vol. 1433, Springer, pp. 350-357.

Sinova, B.; Casals, M.R.; Colubi, A.; Gil, M.Á. (2010). The median of a random interval. En: 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. Advances in Intelligent and Soft Computing*, Vol. 77, Springer, pp. 575-583.

Sinova, S.; Casals, M.R.; Gil, M.Á. (2014). Central tendency for symmetric random fuzzy numbers. *Information Sciences,* Vol. 278, pp. 599-613.

Sinova, B.; Casals, M.R.; Gil, M.Á.; Lubiano, M.A. (2015). The fuzzy characterizing function of the distribution of a random fuzzy number. *Applied Mathematical Modelling*, Vol. 39, nº 14, pp. 4044-4056.

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

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

Sinova, B.; De la Rosa de Sáa, S.; Gil, M.Á. (2013). A generalized L^1-type metric between fuzzy numbers for an approach to central tendency of fuzzy data. *Information Sciences*, Vol. 242, pp. 22-34.

Sinova, B.; De la Rosa de Sáa, S.; Lubiano, M.A.; Gil, M.Á. (2015). An overview on the statistical central tendency for fuzzy datasets. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 23 (Suppl. 1), pp. 105-132.

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

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

Sinova, B., Gil, M.Á., Van Aelst, S. (2016). M-estimates of location for the robust central tendency of fuzzy data. *IEEE Transactions on Fuzzy Systems*, Vol. 24, nº 4, pp. 945-956.

Sinova, B.; González-Rodríguez, G.; Van Aelst, S. (2013). An alternative approach to the median of a random interval using an L^2 metric. En: 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. Advances in Intelligent Systems and Computing*, Vol. 190, Springer, pp. 273-281.

Sinova, B.; González-Rodríguez, G.; Van Aelst, S. (2018). M-estimators of location for functional data. *Bernoulli*, Vol. 24, nº 3, pp. 2328-2357.

Sinova, B.; Palacio Vega, S.; Gil, M.Á. (2024). Sensitivity analysis on the choice of the metric on cronbach's α coefficient for interval-valued data in questionnaires. En: Ansari, J.; Fuchs, S.; Trutschnig, W.; Lubiano, M.A.; Gil, M.Á.; Grzegorzewski, P.; Hryniewicz, O. (Eds.) *Combining, Modelling and Analyzing Imprecision, Randomness and Dependence. Advances in Intelligent Systems and Computing*, Vol. 1458, Springer, pp. 466-473.

Sinova, B.; Pérez-Fernández, S.; Montenegro, M. (2015). The wabl/ldev/rdev median of a random fuzzy number and statistical properties. En: Grzegorzewski, P.; Gagolewski, M.; Hryniewicz, O.; Gil, M.Á. (Eds.) *Strengthening Links Between Data Analysis and Soft Computing. Advances in Intelligent Systems and Computing*, Vol. 315, Springer, pp. 143-150.

Sinova, B.; Van Aelst, S. (2013). Comparing the medians of a random interval defined by means of two different L^1 metrics. En: Borgelt, C.; Gil, M.Á.; Sousa, J.M.C.; Verleysen, M. (Eds.) *Towards Advanced Data Analysis by Combining Soft Computing and Statistics. Studies in Fuzziness and Soft Computing*, Vol. 285, Springer, pp. 75-86.

Sinova, B.; Van Aelst, S. (2015). On the consistency of a spatial-type interval-valued median for random intervals. *Statistics & Probability Letters*, Vol. 100, pp. 130-136.

Sinova, B.; Van Aelst, S. (2017). Tukey's biweight loss function for fuzzy set-valued M-estimators of location. En: Ferraro, M.B.; Giordani, P.; Vantaggi, B.; Gagolewski, M.; Gil, M.Á.; Grzegorzewski, P.; Hryniewicz, O. (Eds.) *Soft Methods for Data Science. Advances in Intelligent Systems and Computing*, Vol. 456, Springer, pp. 447-454.

Sinova, B.; Van Aelst, S. (2018a). Advantages of M-estimators of location for fuzzy numbers based on Tukey's biweight loss function. *International Journal of Approximate Reasoning*, Vol. 93, pp. 219-237.

Sinova, B.; Van Aelst, S. (2018b). A spatial-type interval-valued median for random intervals. *Statistics*, Vol. 52, nº 3, pp. 479-502.

Sinova, B.; Van Aelst, S. (2019a). Empirical analysis of the maximum asymptotic bias of location estimators for fuzzy number-valued data. *International Journal of Approximate Reasoning*, Vol. 133, pp. 1-13.

Sinova, B.; Van Aelst, S. (2019b). Empirical comparison of the performance of location estimates of fuzzy number-valued data. En: Destercke, S.; Denoeux, T.; Gil, M.Á.; Grzegorzewski, P.; Hryniewicz, O. (Eds.) *Uncertainty Modelling in Data Science. Advances in Intelligent Systems and Computing*, Vol. 832, Springer, pp. 191-199.

Sinova, B.; Van Aelst, S. (2024). Fuzzy S-estimators. En: Ansari, J.; Fuchs, S.; Trutschnig, W.; Lubiano, M.A.; Gil, M.Á.; Grzegorzewski, P.; Hryniewicz, O. (Eds.) *Combining, Modelling and Analyzing Imprecision, Randomness and Dependence. Advances in Intelligent Systems and Computing*, Vol. 1458, Springer, pp. 474–482.

Sinova, B.; Van Aelst, S.; Terán, P. (2021). M-estimators and trimmed means: from Hilbert-valued to fuzzy set-valued data. *Advances in Data Analysis and Classification*, Vol. 15, pp. 267-288.

Trutschnig, W.; González-Rodríguez, G.; Colubi, A.; Gil, M.Á. (2009). A new family of metrics for compact, convex (fuzzy) sets based on a generalized concept of mid and spread. *Information Sciences, Vol.* 179, nº 23, pp. 3964-3972.

Trutschnig, W.; Lubiano, M.A. (2019, 2025). *SAFD: Statistical Analysis of Fuzzy Data*. Version 2.1. https://cran.r-project.org/web/packages/SAFD, https://cran.r-project.org/web/packages/SAFD/SAFD.pdf.

Trutschnig, W.; Lubiano, M.A.; Lastra, J. (2013). SAFD-An R package for statistical analysis of fuzzy data. En: Borgelt, C.; Gil, M.A.; Sousa, J.M.C.; Verleysen, M. (Eds.) *Towards Advanced Data Analysis by Combining Soft Computing and Statistics. Studies in Fuzziness and Soft Computing*, Vol. 285, Springer, pp. 107-118.

Otras referencias y bibliografía de consulta

Asai, K.; Tanaka, H.; Okuda, T. (1977). On discrimination of fuzzy states in probability space, *Kybernetes*, Vol. 6, nº 3, pp. 185-192.

Azorín Poch, F. (1981). *Conjuntos Borrosos, Estadística y Probabilidad*. Discurso leído el 2 de diciembre de 1981 en el acto de su recepción como académico de número en la Real Academia de Ciencias Exactas, Físicas y Naturales de Madrid. Contestación: Sixto Ríos García.

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer.

Bernstein, S. (1917). Опыт аксиоматического обоснования теории вероятностей (On the axiomatic foundation of the theory of probability). *Communications of the Kharkiv Mathematical Society*, Vol. 15, pp. 209-274.

Bombal, F. (1991). *La Teoría de la Medida: 1875-1925.* En: Seminario de Historia de la Matemática. Universidad Complutense de Madrid. (http://blogs.mat.ucm.es/bombal/wp-content/uploads/sites/40/2018/11/HIS-La.Teoria.de_.la_.Medida._1875-1925.pdf).

Borel, E. (1909). Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo*, Vol. 28, pp. 247-271.

Bressoud, D.M. (2008). *A Radical Approach to Lebesgue's Theory of Integration*. Cambridge University Press, Cambridge.

Comité Nacional Español de Grandes Presas (2012). *Guías Técnicas de Seguridad de Presas. Guía Técnica de Explotación de Presas y Embalses. Tomo 1*. Exce Consulting Group.

Efron, B. (1979). Bootstrap methods: Another look at the Jackknife. *Annals of Statististics*, Vol. 7, nº 1, pp. 1-26.

Efron, B. (1981). Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika*, Vol. 68, nº 3, pp. 589-599.

Efron, B. (2000). The sootstrap and modern statistics. *Journal of the American Statistical Association*, Vol. 95, nº 452, pp. 1293-1296.

Efron, B. (2013). A 250-year argument: Belief, behavior, and the bootstrap. *Bulletin of the American Mathematical Society*, Vol. 50, nº 1, pp. 129-146.

Féron, R. (1954). *Information, Régression, Corrélation*. Thèse de Doctorat, Université de Paris, Faculté des Sciences (Directeurs: Maurice Fréchet et Georges Darmois).

Féron, R. (1976a). Ensembles aléatoires flous. *Comptes Rendus de l'Académie des Sciences, Série A*, Vol. 282, pp. 903-906.

Féron, R. (1976b). Ensembles flous, ensembles aléatoires flous, et économie aléatoire floue. *Publications Econométriques*, Vol. IX, nº 1, pp. 25-64.

Féron, R. (1979a). Ensembles aléatoires flous dont la fonction d'appartennance prendses valeurs dans un treillis distributif fermé. *Publications Econométriques*, Vol. XII, nº 1, pp. 81-118; Bibliographical addendum, Vol. XII, nº 2, pp. 63-67.

Féron, R. (1979b). Sur les notions de distance et d'écart dans une structure floue etleurs applications aux ensembles aléatoires flous. Cas où le référentiel n'est pas métrique. *Comptes Rendus de l'Académie des Sciences, Série A*, Vol. 289, pp. 35-38.

Féron, R; Féron, M. (1988). Fuzzy specifications and random fuzzy events considered as basic tools for statistical prediction. *Fuzzy Sets and Systems*, Vol. 28, nº 3, pp. 285-293.

Fortet, R.; Kambouzia, M. (1975). Ensemble saléatoires, répartitions ponctuelles aléatoires, problèmes de recouvrement. *Annales de l'Institut Henri Poincaré*, *Section B*, Tome 11, nº 4, pp. 299-316.

Fréchet, M. (1906). Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Matematico di Palermo*, Vol. 22, pp. 1-72.

Fréchet, M. (1938). Exposé et discussion de quelques recherches récentes sur les fondements du calcul des probabilités. *Colloque Consacré à la Théorie des Probabilités (Genève), Actualités Scientifiques et Industrielles*, Vol. 735, pp. 23-55.

Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut Henri Poincaré*, Vol. 10, nº 4, pp. 215-310.

Fréchet, M. (1950). Conferencias sobre los elementos aleatorios de naturaleza cualquiera. *Trabajos de Estadística*, Vol. 1, nº 2, pp. 157-181.

Fréchet, M. (1957). Une nouvelle théorie: celle des éléments aléatoires abstraits. *Revue Philosophique de la France et de l'Étranger*, Vol. 147, pp. 145-158.

Freitas Pinto, P. de J.; dos Santos Grecco, C.H.; Nunes Cosenza, C.A. (2020). Fuzzy Model for the Priorization Analysis of Variable Quality Performance: An Approach in Shipbuilding. *Fuzzy Information and Engineering*, Vol. 12, nº 2, pp. 181-203.

Giné, E.; Hahn, M.G.; Zinn, J. (1983). 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, Springer, pp. 112-135.

Giné, E.; Zinn, J. (1984). Some limit theorems for empirical processes. *Annals of Probability*, Vol. 12, nº 4, pp. 929-989.

Giné, E.; Zinn, J. (1990). Bootstrapping general empirical measures. *Annals of Probability*, Vol. 18, nº 2, pp. 851-869.

Girshick, M.A.; Blackwell, D. (1954). *Theory of Games and Statistical Decisions*. John Wiley and Sons.

Hausdorff, F. (1914). *Grundzüge der Mengenlehre*. Leipzig Viet.

Hayes M.H.S.; Patterson D.G. (1921). Experimental development of the graphic rating method. *Psychological Bulletin*, Vol. XVIII, pp. 98-99.

Hesketh, B.; Griffin, B.; Loh, V. (2011). A future-oriented retirement transition adjustment framework. *Journal of Vocational Behavior*, Vol. 79, nº 2, pp. 303-314.

Hesketh, B.; McLachlan, K.; Gardner, D. (1992). Work adjustment theory: An empirical test using a fuzzy rating scale. *Journal of Vocational Behavior*, Vol. 40, nº 3, pp. 318-337.

Hesketh, B..; Pryor, R.; Gleitzman, M.; Hesketh, T. (1988a). Practical applications and psychometric evaluation of a computerised fuzzy graphic rating scale, *Advances in Psychology*, Vol. 56, C, pp. 425-454.

Hesketh, T.; Hesketh, B. (1994). Computerized fuzzy ratings: The concept of a fuzzy class. *Behavior Research Methods, Instruments, & Computers*, Vol. 26, nº 3, pp. 272-277.

Hesketh, T.; Pryor, R.; Hesketh, B. (1988b). An application of a computerized fuzzy graphic rating scale to the psychological measurement of individual differences. *International Journal of Man-Machine Studies*, Vol. 29, nº 1, pp. 21-35.

Hilbert, D. (1900). Mathematische Probleme. Vortrag, gehalten auf dem internationalen Mathematike-Congress zu Paris 1900. *Göttingen Nachrichten*, pp. 253-297 [Translation: Mathematical Problems. *Bulletin of the American Mathematical Society*, Vol. 8, pp. 437-479 (1902)].

Kendall, D.G. (1974). Foundations of a theory of random sets. En: Harding, E.F.; Kendall, D.G. (Eds.) *Stochastic Geometry*. John Wiley and Sons.

Klement, E.P.; Puri, M.L.; Ralescu, D.A. (1986). Limit theorems for fuzzy random variables. *Proceedings of the Royal Society of London A*, Vol. 407, pp. 171-182.

Kolmogorov, A.N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer, Berlin.

Körner, R. (1997). On the variance of fuzzy random variables. *Fuzzy Sets and Systems*, Vol. 92, nº 1, pp. 83-93.

Körner, R.; Näther, W. (1998). Linear regression with random fuzzy variables: Extended classical estimates, best linear estimates, least squares estimates. *Information Sciences*, Vol. 109, nº 1-4, pp. 95-118.

Krätschmer, V. (2002). Limit theorems for fuzzy random variables. *Fuzzy Sets and Systems*, Vol. 126, nº 2, pp. 253-263.

Kruse, R.; Meyer, D. (1987). *Statistics with Vague Data.* D. Reidel Publishing Co., Theory and Decision Library, B: Mathematical and Statistical Methods.

Kwakernaak, H. (1978). Fuzzy random variables-I. definitions and theorems. *Information Sciences*, Vol. 15, nº 1, pp. 1-29.

Kwakernaak, H. (1979). Fuzzy random variables-II. Algorithms and examples for the discrete case, *Information Sciences*, Vol. 17, nº 3, pp. 253-278.

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

Likert, R. (1932). The method of constructing an attitude scale. *Archives of Psychology*, Vol. 140, pp. 44-53.

Matheron, G. (1965). *Les Variables Régionalisées et leur estimation: Une Application de la Théorie des Fonctions Aléatoires aux Sciences de la Nature*. Thèse de Doctorat, Université de Paris, Faculté des Sciences (Directeur: Robert Fortet). Publicada por Masson, Paris.

Matheron, G. (1975). *Random Sets and Integral Geometry*. John Wiley and Sons.

Minkowski, H. (1903). Volumen und Oberfläche. *Mathematische Annalen*, Vol. 57, nº 4, pp. 447-495.

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

Näther, W. (1997). Linear statistical inference for random fuzzy data. *Statistics*, Vol. 29, nº 3, pp. 221-240.

Näther, W. (2000). On random fuzzy variables of second order and their application to linear statistical inference with fuzzy data. *Metrika*, Vol. 51, nº 3, pp. 201-221.

Näther, W.; Albrecht, M. (1990). Linear regression with random fuzzy observations. *Statistics*, Vol. 21, nº 4, pp. 521-531.

Negoita, C.V.; Ralescu, D.A. (1975). *Applications of Fuzzy Sets to Systems Analysis*. Interdisciplinary Systems Research Series. Vol. 11, Birkhaeuser, Basel, Stuttgart and Halsted Press.

Okuda, T.; Tanaka, H.; Asai, K. (1978). A formulation of fuzzy decision problems with fuzzy information using probability measures of fuzzy events. *Information & Control*, Vol. 38, nº 2, pp. 135–147.

Popper, K. (1974, 2005). *Unended Quest. An Intellectual Autobiography*. Library of Living Phylosophers, Inc.-edición 1974, Taylor and Francis e-Library-edición 2005.

Proske, F.N.; Puri, M.L. (2002). Central limit theorem for Banach space valued fuzzy random variables. *Proceedings of the American Mathematical Society*, Vol. 130, nº 5, pp. 1493-1501.

Puri M.L.; Ralescu, D.A. (1985). The concept of normality for fuzzy random variables. *Annals of Probability*, Vol. 13, nº 4, pp. 1373-1379.

Puri M.L.; Ralescu, D.A. (1986). Fuzzy random variables. *Journal of Mathematical Analysis and Applications*, Vol. 114, nº 2, pp. 409-422.

Puri M.L.; Ralescu, D.A. (1991). Convergence theorem for fuzzy martingales, *Journal of Mathematical Analysis and Applications*, Vol. 160, nº 1, pp. 107-122.

Ralescu, D.A. (1980). *Admissibility of Estimators in the One Parameter Exponential Family and in Multivariate Location Problems*. PhD Thesis, Indiana University, Department of Mathematics.

Reagan, R.T.; Mosteller, F.; Youtz, C. (1989). Quantitative meanings of verbal probability expressions. *Journal of Applied Psychology*, Vol. 74, nº 3, pp. 433-442.

Ruspini, E.H. (1969). A new approach to clustering. *Information & Control*, Vol. 15, nº 1, pp. 22-32.

Ruspini, E.H. (1970). Numerical methods for fuzzy clustering. *Information Sciences*, Vol. 2, nº 3, pp. 319-350.

Shafer, G.; Vovk, V. (2006). The sources of Kolmogorov's Grundbegriffe. *Statistical Science*, Vol. 21, nº 1, pp. 70-98.

Shiina, (2021). A Commentary on "The Historical Roots of Visual Analog Scale in Psychology as Revealed by Reference Publication Year Spectroscopy" (Frontiers in Human Neurosience, 2019). *Frontiers in Human Neuroscience*, Vol. 15, 711691.

Szymborska, W. (2002). *Chwila*. Kraków: Wydawn. Znak [Traducción: *Instante*, Ediciones Igitur (2004)].

Tanaka, H.; Okuda, T.; Asai, K. (1976). A formulation of fuzzy decision problems and its application to an investment problem, *Kybernetes*, Vol. 5, nº 1, pp. 25-30

Wang, M.; Kang, J.; Liu, W., Li, M.; Su, J.; Fang, Z.; Li, X.; Shang, L.; Zhang, F.; Guo, C. (2014). Design and study of mine silo drainage method based on fuzzy control and avoiding peak filling valley strategy. *Scientific Reports*, Vol. 14, 9300.

Zadeh, L.A. (1965). Fuzzy sets. *Information & Control*, Vol. 8, nº 3, pp. 338-353.

Zadeh, L.A. (1968). Probability measures of Fuzzy events. *Journal of Mathematical Analysis and Applications*, Vol. 23, nº 2, pp. 421-427.

Zadeh, L.A. (1975). The concept of a linguistic variable and its application to approximate reasoning. Part 1. *Information Sciences*, Vol. 8, pp. 199-249; Part 2. *Information Sciences*, Vol. 8, pp. 301-353; Part 3. *Information Sciences*, Vol. 9, pp. 43-80.

Zadeh, L.A. (2008). Is there a need for fuzzy logic? *Information Sciences*, Vol. 178, pp. 2751-2779.