

El análisis de intervalos. Aplicaciones en la ingeniería

Palabras claves: datos de intervalo, análisis de datos simbólicos (ADS), aritmética de intervalos, incertidumbre, análisis de errores.

Resumen: Aunque el *análisis de intervalos* (AI) tiene sus orígenes hace más de 2.200 años en el mundo griego de Arquímedes, es solo a partir de 1960 cuando se convierte en un foco activo de investigación y se produce un desarrollo impresionante en el área de los métodos y aplicaciones con los datos de intervalo.

Este artículo realiza en primer lugar un recorrido por las diferentes aproximaciones al AI, incluyendo la más reciente del análisis de datos simbólicos (ADS), y muestra algunas aplicaciones en Ingeniería. Posteriormente presenta la aritmética con intervalos y aborda la cuestión clave de definir distancias entre intervalos. A continuación desarrolla los fundamentos de la Estadística Descriptiva con datos de intervalo. Por último, se concluye el trabajo con un énfasis especial en las aplicaciones actuales del AI, así como en las posibilidades que ofrece para algunos temas del futuro de la Ingeniería.

Key words: interval-valued data, symbolic data analysis (SDA), interval arithmetic, uncertainty, error analysis.

Abstract: *It is generally assumed that Interval Analysis (IA) had its origins 2200 years ago in the classical world of Archimedes. However, it is only since 1960 that it has become a focus of active research and since then there has been an impressive development in the field of methods and applications of interval data.*

This article provides an overview of the different approaches to IA, including the most recent developments in symbolic data analysis (SDA), and describes some applications in Engineering. Subsequently, some key mathematical issues on intervals, such as arithmetic operations and distance, are discussed. In addition, a development of the fundamentals of Descriptive Statistics with interval data will be presented. Finally, this work will consider current applications of IA and exciting possibilities for the future of some Engineering subjects.



Carlos Maté Jiménez

Profesor Propio de la ETSI (ICAI), de la Universidad Pontificia Comillas de Madrid, adscrito al Departamento de Organización Industrial y al Instituto de Investigación Tecnológica. Doctor en Ciencias Matemáticas y Diplomado en Ciencias Económicas y Empresariales por la Universidad Complutense. Ha escrito varios libros sobre estadística y publicado diversos artículos sobre aplicaciones de los métodos estadísticos en prestigiosas revistas internacionales y nacionales, tanto del ámbito industrial y de organización como del económico y financiero.

Introducción

El *Análisis de Intervalos* (AI), también llamado Análisis Intervalar, tiene sus orígenes en el mundo clásico, ya que se debe a Arquímedes la propuesta de un método para proporcionar una sucesión de intervalos que contenga al número π con suficiente precisión.

Bajo la hipótesis de que las observaciones y las estimaciones en el mundo real son incompletas para representar con exactitud los datos reales, Ramon Moore introdujo el AI, en el año 1959, como una herramienta para el control automático de los errores. De hecho, si se requiere usar una aproximación numérica a un problema matemático y se exige precisión, los datos se deben representar por intervalos, ya que cualquier resultado calculado incorporará un error en la entrada, errores de redondeo durante el cálculo y errores de truncamiento.

Veamos el ejemplo de la medición de la **presión arterial** en una persona. Al tomar la presión arterial se conoce el resultado de la presión que ejerce la sangre contra las paredes de las arterias. El resultado de la lectura de la presión arterial se da en 2 cifras. La mayor de ellas es la sistólica. La menor de ellas es la llamada diastólica y es el segundo número en la lectura. Un ejemplo de la lectura de la presión arterial sería 130/70 (130 sobre 70), en la cual 130 es el número sistólico y 70 es el número diastólico. Estos números deberían llevar asociado un error del instrumento de medida. Por ejemplo, si las medidas se realizan con un 5% de error, realmente la medición sistólica sería 130 ± 6.5 , mientras que la medición diastólica sería 70 ± 3.5 .

Desde 1960, el AI ha sido un foco activo de investigación y hemos asistido a un desarrollo impresionante en el área de los métodos y aplicaciones con los datos de intervalo (ver, por ejemplo, <http://www.cs.utep.edu/interval-comp/main.html>, la web de la comunidad de cálculos con intervalos). Una revisión completa desde un punto de vista matemático y computacional se puede ver en Alefeld y Mayer (2000).

Moore (1966) estableció las bases para el AI. Posteriormente Moore

(1979) desarrolló un conjunto importante de técnicas que proporcionaban un completo y riguroso análisis del error en los resultados computacionales. Recientemente, Moore et ál. (2009) han presentado una introducción actualizada al AI, incluyendo ejemplos con la herramienta de *software* INTLAB, una toolbox desarrollada en MATLAB. Según Maté (2011), esta obra resulta clave para cualquiera que quiera introducirse en este campo, aunque algunos temas importantes del AI se han tratado de forma breve.

Desde el punto de vista estadístico y de minería de datos, el AI se enmarca dentro del llamado *Análisis de Datos Simbólicos* (ADS). En el análisis clásico una variable en cada individuo o unidad solo puede tomar un valor, ya sea categórico o numérico. En el ADS una variable en cada individuo o unidad puede ser algo no tan simple como un intervalo de números reales, en lo que se denomina variable valorada por intervalos. También puede tomar como valor un conjunto de categorías con sus pesos, una lista de categorías, un histograma o una distribución de probabilidad. Por tanto, se trata de un campo del conocimiento relacionado con el análisis multivariante, el reconocimiento de patrones y la inteligencia artificial, cuyos objetivos son introducir nuevos métodos de análisis de la información y extender las técnicas clásicas de análisis de datos a los llamados datos simbólicos —véase Bock y Diday (2000), Billard y Diday (2003, 2006) y Noirhomme-Fraiture y Brito (2011)—.

Volviendo al ejemplo de la presión arterial, si consideramos el dato $[70, 130]$, que es tal cual aparece en la realidad, es decir, no un único número, estamos considerando un dato de intervalo o un dato de una variable valorada por intervalos. Los datos de intervalo son un caso particular de los llamados datos simbólicos. Para ver la evolución de la presión arterial de una persona en un periodo de tiempo realizaríamos n mediciones que darían lugar a n intervalos. Para poder analizar esta información necesitamos la aritmética de intervalos y evaluar la distancia entre intervalos.

Este artículo presenta un recorrido por las diferentes aproximaciones al AI, centrándose en las aplicaciones en ingeniería. Posteriormente se abordan las cuestiones de la notación, aritmética, distancia entre intervalos y se proporcionan las medidas de estadística descriptiva para datos de intervalo. Por último, se concluye el trabajo con un énfasis especial en las aplicaciones actuales del AI, así como en las posibilidades que ofrece para algunos temas del futuro de la Ingeniería.

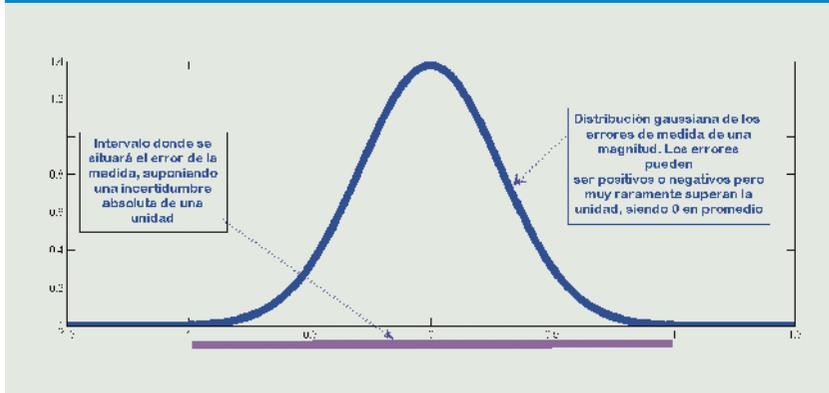
Intervalos en los errores e incertidumbre asociados a las medidas de ingeniería

Como es bien sabido, los aparatos de medición en ingeniería, como los sensores, nunca son absolutamente precisos. El valor \tilde{x} medido por un sensor es en general diferente del valor real (desconocido) x de la correspondiente cantidad física. Cuando se procesan datos de sensores es importante tener en cuenta de forma adecuada los errores resultantes de las medidas $\Delta x = \tilde{x} - x$.

El enfoque tradicional de la ciencia y la ingeniería para manejar los errores considera que conocemos la distribución de probabilidad del error de una medida, suponiéndose en muchos casos que el error sigue una distribución gaussiana, como se muestra en la Figura 1. Por tanto, se pueden aplicar técnicas estadísticas estándares para manejar dicha incertidumbre.

En muchas situaciones reales no conocemos la distribución anterior, o bien no puede asegurarse que sea normal, y lo único que se conoce es una cota superior Δ sobre el error de medida: $|\Delta x| \leq \Delta$. Esta cota se llama **incertidumbre absoluta** asociada a la medida. En este caso, una vez que conocemos el resultado de la medida \tilde{x} , la única información que tenemos sobre el valor real de x es que pertenece a un intervalo cuyo límite inferior (superior), notado por $x_L(x_U)$, viene dado por $x_L = \tilde{x} - \Delta$ ($x_U = \tilde{x} + \Delta$). Es habitual entonces designar el valor real de x por el intervalo $[x] = [x_L, x_U]$. La Figura 1 nos muestra como alternativa a la distribución gaussiana el plantea-

Figura 1. Distribución gaussiana de los errores frente a dato de intervalo



miento del AI según el cual el error se encuentra dentro del intervalo $[-1, 1]$, ya que la incertidumbre absoluta es de una unidad.

El dato de intervalo en una medida se relaciona directamente con el concepto de precisión. La precisión es la tolerancia de medida o de transmisión del instrumento y define los límites de los errores cometidos cuando el instrumento se emplea en condiciones normales de servicio. Hay varias formas para expresar la precisión de una medida que proporciona un intervalo. Veámoslo al medir una temperatura de $200\text{ }^{\circ}\text{C}$ cuyo campo de medida es de $300\text{ }^{\circ}\text{C}$.

a) Directamente, en unidades de la variable medida. Ejemplo: precisión de $\pm 1\text{ }^{\circ}\text{C}$. Las medidas de temperatura mostrarían una distribución de los errores como la de la Figura 1.

b) Tanto por ciento de la lectura efectuada. Ejemplo: Precisión de $\pm 1\%$ de $200\text{ }^{\circ}\text{C}$, es decir $\pm 2\text{ }^{\circ}\text{C}$.

c) Tanto por ciento del valor máximo del campo de medida. Ejemplo: precisión de $\pm 0,5\%$ de $300\text{ }^{\circ}\text{C} = \pm 1,5\text{ }^{\circ}\text{C}$.

La precisión varía en cada punto del campo de medida, si bien el fabricante la especifica en todo el margen del instrumento indicando a veces su valor en algunas zonas de la escala. Por ejemplo, un manómetro puede tener una precisión de $\pm 1\%$ en toda la escala y de $\pm 0,5\%$ en la zona central.

La necesidad de operar medidas, calcular distancias entre medidas y calcular medidas estadísticas para varias medidas nos introduce en la

aritmética de intervalos, el cálculo de distancias entre intervalos y el cálculo de medidas estadísticas para datos de intervalo. Todo ello se analiza en la segunda parte del artículo.

Por tanto, podemos decir que el AI tiene sus raíces en los problemas relacionados con los errores en las medidas. Desde un punto de vista formal y analítico estos problemas son estudiados por las disciplinas del Análisis Numérico y la Matemática Computacional.

Los profesores y el personal de los laboratorios de cualquier escuela de ingeniería conocen bien la importancia del análisis de los errores en los diferentes tipos de medidas mecánicas, eléctricas, etc., que se realizan en el mundo real. Sin embargo, solo recientemente se ha propuesto de forma muy completa explorar la aplicabilidad del AI en los laboratorios. Rothwell y Cloud (2012) presentan una técnica para el análisis automático de errores utilizando la matemática de los intervalos. Concluyen, al desarrollar dos casos típicos de experimentos de laboratorio como medir la permitividad de placas base y medir la impedancia RF de cargas coaxiales, que la aproximación del AI proporciona estimaciones de los errores comparables a los clásicos métodos de propagación del error, pero con mucho menos esfuerzo. Es decir, que utilizar el AI en laboratorios de medidas es mucho más sencillo que implantar las fórmulas de propagación del error, especialmente cuando las expresiones son

complicadas o entran en juego varias variables.

Intervalos en el modelado de la incertidumbre en sistemas de ingeniería

Por lo general, cualquier sistema de ingeniería debe incorporar para cada variable incluida en el mismo la incertidumbre asociada a dicha variable. En este sentido, se han publicado recientemente algunos trabajos que modelan dicha incertidumbre a través de intervalos.

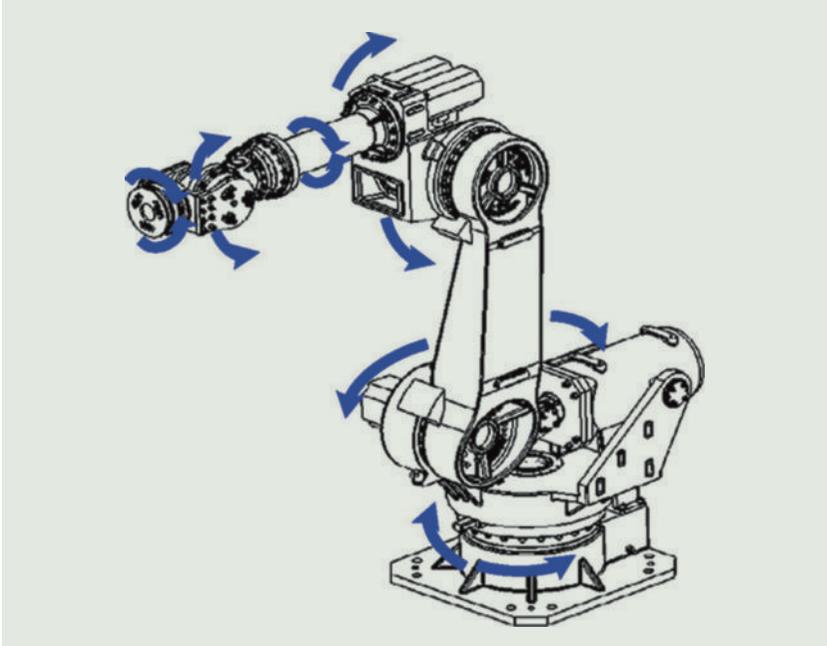
Lei (2012) indica que las aplicaciones del AI a la programación de la producción apenas han sido investigadas. Sobre esta base, propone un algoritmo genético basado en el AI que resulta ser eficiente y novedoso en el empleo de intervalos en el problema de programación de tareas. Argumenta a favor del AI que tiene un bajo coste para modelar la incertidumbre.

Mousavi et al. (2011) analizan la importante cuestión de la evaluación del riesgo en grandes proyectos de ingeniería. Como parte de la metodología desarrollada, para cada riesgo importante incluyen el cálculo de la puntuación del intervalo de riesgo –interval risk score (IRS)– como producto de la probabilidad de que ocurra dicho riesgo, expresada en intervalo, por el criterio de impacto, también expresado en intervalo. Es decir, el sistema propuesto por Mousavi et al. (2011) modela la incertidumbre mediante intervalos empleando el producto de intervalos (véase en la página 25 la definición correspondiente).

Qiu et al. (2009) proponen modelar los parámetros de incertidumbre de sistemas de vibración no lineales mediante intervalos, a efectos de analizar la respuesta dinámica de dichos sistemas. Nuevamente nos encontramos ante un trabajo donde se sustituye la distribución de probabilidad sobre parámetros por el rango de los mismos, obteniéndose soluciones factibles y de menor complejidad que las aproximaciones clásicas.

Un campo donde se ve cada vez más necesario el uso del AI es en robótica, ya que las operaciones de un robot son siempre por definición con

Figura 2. Robot de soldadura con operaciones sujetas a mediciones en intervalos debido al error asociado a cada mecanismo



incertidumbre. Por ejemplo, el robot de soldadura de la Figura 2 –tomando de Xiao et ál. (2012)– realiza un conjunto de operaciones y en cada una de ellas el valor de la variable considerada viene acompañado de una incertidumbre. Por ejemplo, un giro no es un ángulo exacto sino que será un intervalo de ángulos.

Merlet (2010) presenta el uso del AI en la resolución de los problemas cinemáticos de los robots paralelos y en el análisis del espacio de trabajo y la planificación del movimiento. Además, indica la viabilidad y eficiencia del AI en el diseño óptimo de robots y en el análisis del desempeño de los mismos frente a aproximaciones clásicas.

Intervalos en la borrosidad de la realidad

La lógica borrosa (*fuzzy logic*) fue propuesta por el profesor Zadeh en 1964 para dar respuesta a los problemas asociados con la lógica de los sistemas que hasta ese momento era la lógica clásica. Según Zadeh (1965), al programar el funcionamiento de un aparato de aire acondicionado según la lógica clásica es necesario dar una regla para cada temperatura. Sin embargo, él

planteó una lógica que permitiera programar el funcionamiento de dicho aparato con sentencias del tipo “trabaja un poco más rápido cuando percibas un ambiente más caluroso”. Esta aproximación borrosa (¿qué se entiende exactamente con “un poco más rápido”? o con cierta vaguedad (¿qué se entiende exactamente por “un ambiente más caluroso”?) a la realidad dio lugar a los conjuntos y sistemas borrosos.

La Figura 3 muestra la diferencia entre la lógica clásica y la lógica borrosa. En la lógica clásica el conjunto A corresponde a todos los puntos de un intervalo que tienen la misma densidad de probabilidad de aparecer. Es decir, cualquier subintervalo de amplitud una milésima tiene la misma probabilidad de ser el resultado de nuestra medición. En la lógica borrosa el conjunto A corresponde a todos los puntos de un intervalo que tienen diferente posibilidad de aparecer. Es decir, la posibilidad que tiene cada valor de medición dentro del intervalo A aumenta a medida que nos acercamos al punto medio de dicho intervalo. Razonamientos análogos pueden hacerse para el complementario o contrario, \bar{A} , la intersección de A con su complementario o la unión de A con su complementario.

La conexión entre la teoría de los conjuntos borrosos y el AI tiene una larga historia y se propuso por primera vez en los años setenta del siglo XX. Para más detalles de esta conexión y líneas actuales de investigación véase Maté (2011).

Intervalos en el análisis temporal de la realidad

Un campo donde se han realizado avances muy importantes en los últimos años es el del análisis de una información temporal en forma de

Figura 3. Lógica clásica versus lógica borrosa

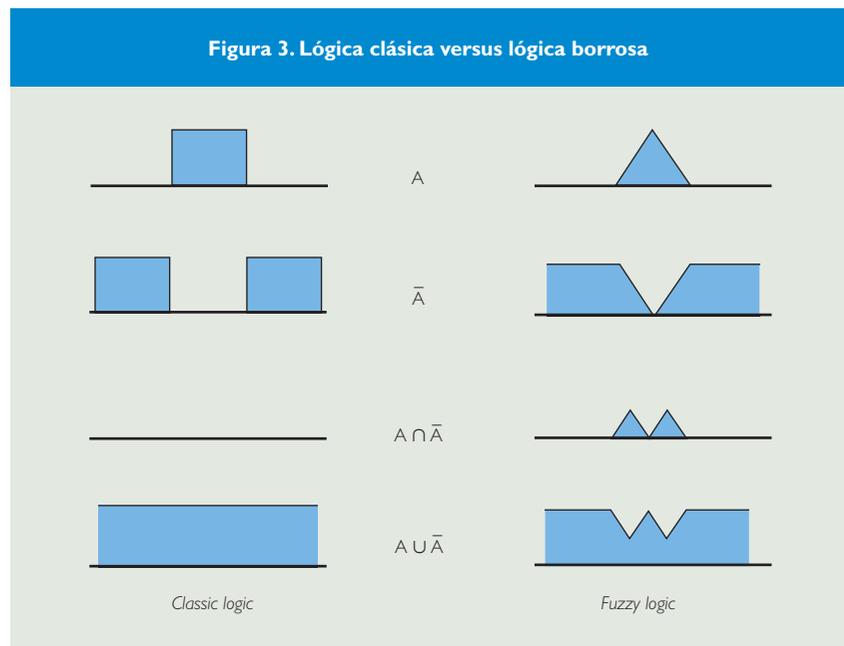


Tabla I. Datos de la Figura 4

Mes	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SEP	OCT	NOV	DIC
Dda. mínima	20238	19920	19137	17679	17582	18000	194	18198	18909	18141	19342	19523
Dda. máxima	26091	24219	23513	23491	21373	23445	23609	22008	22842	21768	24956	26302

intervalo. Ello da lugar a las llamadas series temporales de intervalos (STI) –interval time series (ITS) en inglés–. Como consecuencia de la aparición de este concepto surge la necesidad de analizar estas series y de proponer métodos de predicción para las mismas –véase Muñoz-San-Roque et ál. (2007), Arroyo y Maté (2009), Arroyo et ál. (2011a, 2011b)–.

Por ejemplo, García-Ascanio y Maté (2010) analizan la evolución de la demanda mensual de energía eléctrica en Mwh en cada hora del día, durante el año 2000. La Figura 3 del citado artículo, que reproducimos aquí como figura 4, muestra la evolución del intervalo de demanda mínima y demanda máxima en la hora 1 del día a lo largo de los diferentes meses del año 2000. Dichos intervalos quedan recogidos en la Tabla I.

Dichos autores concluyen que los métodos de predicción de STI constituyen una herramienta a considerar para los sistemas de planificación de la energía. Su potencial uso conducirá a una reducción del riesgo en las

decisiones de operación de dichos sistemas.

Algunas consideraciones sobre la notación para datos de intervalo

Las cuestiones de la notación y la terminología son siempre críticas en la comprensión y difusión de la ciencia.

En el caso del AI y del ADS podemos encontrar diferentes formas de notar los intervalos y el conjunto de intervalos, según el campo en el que nos movamos, aunque todas ellas coinciden en que un intervalo se considera cerrado, incluyendo como parte del mismo sus límites inferior y superior.

Si consultamos documentos procedentes del campo de la matemática de los intervalos y algunos del ADS como Noirhomme-Fraiture y Brito (2011), es habitual que los intervalos se noten con letras mayúsculas, aunque no hay unanimidad. Por ejemplo, Neumaier (1990) y Han et ál. (2008) utilizan letras minúsculas.

Si consultamos documentos procedentes de algunos grupos que in-

vestigan el análisis de las STI, o del ámbito de la ingeniería, los intervalos se representan con minúsculas entre corchetes.

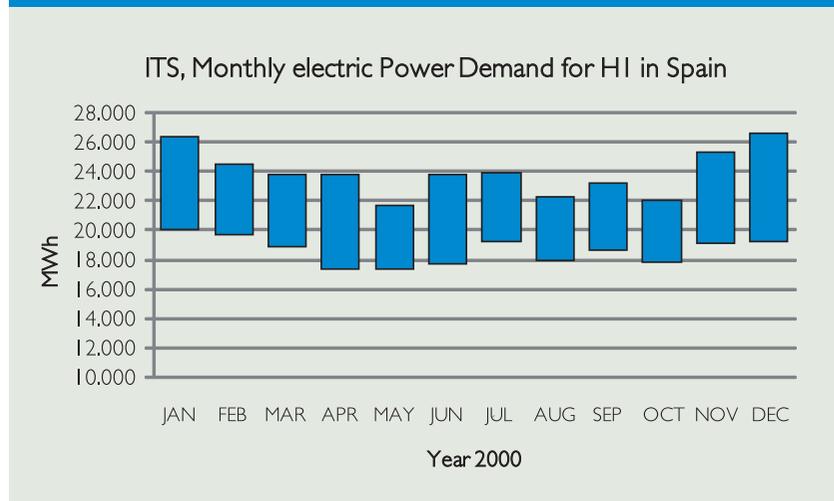
La notación con letras mayúsculas tiene en mi opinión varios inconvenientes. Entre otros tendríamos la confusión con la notación de conjuntos y la confusión con la notación que en estadística se asigna a las variables aleatorias. La notación con letras minúsculas también tiene varios inconvenientes como la confusión con la notación de los propios números reales y la confusión con la notación lógica de los valores reales que hay dentro de un intervalo.

Otra cuestión a considerar es la notación de los extremos o límites de un intervalo. A veces, es habitual hacerlo con una barra inferior para el límite inferior y una barra superior para el límite superior. Por ejemplo, un intervalo que analizara el comportamiento de una magnitud en enero sería notado $E = [\underline{E}, \bar{E}]$ o bien $[e] = [e, \bar{e}]$. Esta notación es peligrosa porque la barra superior encima de una letra representa en estadística el concepto de media, de manera que \bar{E} podría hacer referencia al límite superior del intervalo E, o bien a la media muestral de la variable aleatoria E.

Para evitar el mayor número de controversias, notaremos los intervalos entre corchetes con letras minúsculas. Un elemento cualquiera del intervalo será notado por la letra sin corchetes. Los límites superior e inferior serán notados por la letra sin corchete con el subíndice L para inferior (de Lower) y el subíndice U para superior (de Upper).

Una representación o parametrización del intervalo $[x]$ es la que considera su límite superior x_U e inferior x_L . Es decir, $[x] = [x_L, x_U] = \{x \in \mathbb{R} / -\infty < x_L \leq x \leq x_U < \infty\}$

Figura 4. Serie temporal de intervalo (ITS). Tomada de García-Ascanio y Maté (2010)



y a esta caracterización nos referiremos como (L, U).

Otra representación del intervalo $[x]$ considera su centro x_C y su radio x_R . Es decir, $[x] = \langle x_C, x_R \rangle$ donde $x_C = (x_L + x_U)/2$ y $x_R = (x_U - x_L)/2$, siendo $x_R > 0$. A esta segunda parametrización nos referiremos como $\langle C, R \rangle$. La anchura es definida por $w([x]) = x_U - x_L$.

Las dos representaciones anteriores para intervalos reales son las elegidas habitualmente para el tratamiento de los datos de intervalo. Obviamente, se podría utilizar cualquier otro par de puntos que permita calcular los extremos del intervalo, como el límite superior y el radio, o el centro y el límite inferior.

La parametrización (L, U) no es fácil de emplear cuando se persigue un análisis estadístico, debido a la restricción de orden que conlleva. Por tanto, será recomendable emplear la caracterización $\langle C, R \rangle$, ya que solo exige una restricción de no negatividad sobre el segundo valor.

El conjunto de todos los intervalos de la recta real ha sido representado de varias formas. Una de ellas es $\mathbf{I}\mathbb{R}$, otra $\mathbf{I}(\mathbb{R})$.

La aritmética de intervalos

Sea \circ la forma de notar una de las operaciones básicas "suma", "diferencia", "producto" y "división", para números reales, i.e., $\circ \in \{+, -, \cdot, /\}$. Entonces definimos las correspondientes operaciones para intervalos $[x] = [x_L, x_U]$, $[y] = [y_L, y_U]$ por $[x] \circ [y] = \{x \circ y \mid x \in [x], y \in [y]\}$ donde suponemos que $0 \notin [y]$ en el caso de la división.

Se puede demostrar que el conjunto $\mathbf{I}(\mathbb{R})$ de los intervalos compactos reales es cerrado con respecto a estas operaciones. Además, se verifica que es posible expresar las citadas operaciones utilizando los extremos de los intervalos, según se indica en la Tabla 2.

Estas operaciones se aplican cuando uno de los intervalos es un único número y son críticas en algunas de las aplicaciones descritas en los epígrafes anteriores. Por ejemplo, se utilizarán en el análisis de la propagación de errores ya que muchas medidas son producto o cociente de otras medidas.

Tabla 2. Operaciones entre 2 intervalos

Operaciones entre dos intervalos $[x] = [x_L, x_U], [y] = [y_L, y_U]$	Intervalo resultante
Suma	$[x] + [y] = [x_L + y_L, x_U + y_U]$
Diferencia	$[x] - [y] = [x_L - y_U, x_U - y_L]$
Producto	$[x] \cdot [y] = [\min S, \max S]$ donde S es el conjunto $\{x_L y_L, x_U y_L, x_L y_U, x_U y_U\}$
Inverso	$1/[x] = [1/x_U, 1/x_L]$
Cociente	$[y]/[x] = [y] \times (1/[x])$

El gran inconveniente de la aritmética de intervalos es que la diferencia entre un intervalo consigo mismo depende de la anchura del intervalo. Es decir, aplicando la definición anterior de diferencia resulta que

$$[x] - [x] = [-1, 1] \cdot w([x])$$

En términos matemáticos, el espacio $(\mathbf{I}(\mathbb{R}), +, -)$ no es lineal, sino semilineal, debido a la ausencia de un elemento simétrico con respecto a la suma.

Para un compendio detallado de muchas operaciones aritméticas con intervalos véase Neumaier (1990).

Distancias entre intervalos

Uno de los problemas más importantes en cualquier espacio matemático es el de la distancia entre dos de sus elementos. Introducida una métrica, queda establecida la correspondiente topología y es posible abordar la gran mayoría de los problemas prácticos que surjan en dicho espacio. En estadística y minería de datos algunos problemas prácticos muy importantes son la formación de conglomerados, la reducción de las dimensiones del problema y el análisis de la regresión. Todos ellos se pueden plantear si tenemos previamente establecida una métrica con buenas propiedades.

La cuantificación de la distancia entre intervalos ha sido un problema que ha tenido distintos enfoques en los últimos años. En este sentido, consideremos dos intervalos $[x] = \langle x_C, x_R \rangle$ e $[y] = \langle y_C, y_R \rangle$.

Una de las primeras distancias se debe a Hausdorff y también es conocida

como la distancia Hausdorff-Euclidiana y como la distancia de Manhattan, calculándose como la suma de la distancia euclídea entre centros con la distancia euclídea entre radios y se notará $d_H([x], [y])$. Es decir,

$$d_H([x], [y]) = |x_C - y_C| + |x_R - y_R| \quad (1)$$

Esta distancia no es muy adecuada, como vemos a continuación en el siguiente ejemplo. Sean los intervalos $[x] = [0, 5]$ e $[y] = [6, 7]$ y por otro lado $[x] = [0, 5]$ y $[z] = [6, 10]$. Entonces $d_H([x], [y]) = 6 = d_H([x], [z])$. Obviamente, parece que los primeros están más próximos entre sí que los segundos y esta distancia no informaría sobre ello.

Se demuestra que

$$d_H([x], [y]) = |x_C - y_C| + |x_R - y_R| = \max\{|x_L - y_L|, |x_U - y_U|\} \quad (2)$$

La llamada distancia basada en el núcleo, utilizada en diferentes trabajos prácticos -Arroyo (2008), Arroyo et ál. (2011),...-, viene definida como

$$d_{\text{núcleo}}([x], [y]) = \frac{1}{\sqrt{2}} \sqrt{(x_L - y_L)^2 + (x_U - y_U)^2} = \sqrt{(x_C - y_C)^2 + (x_R - y_R)^2} \quad (3)$$

Con el objetivo de ponderar la distancia euclídea al cuadrado entre los radios de los intervalos, se ha propuesto extenderla incorporando un parámetro θ . Es decir,

$$d_\theta([x], [y]) = \sqrt{(x_C - y_C)^2 + \theta(x_R - y_R)^2}, \quad \theta > 0. \quad (4)$$

Según Sinova et ál. (2012), se trata de una métrica del tipo L_2 sobre $\mathbf{I}(\mathbb{R})$ de manera que $(\mathbf{I}(\mathbb{R}), d_\theta)$ es un espacio métrico separable. Además, esta métrica posee buenas propiedades

operacionales y generaliza otras distancias bien conocidas entre intervalos (para $\theta = 1$ aparece la distancia del núcleo, mientras que para $\theta = 0$ se anulan los radios). El parámetro θ , por tanto, nos permite elegir la importancia relativa que se dará según la distancia euclídea a la distancia entre los radios. En concreto, $d_{1/3}$ se emplea a menudo en las aplicaciones, ya que corresponde a la llamada métrica de Bertoluzza.

Se verifica que cualquiera de las distancias anteriores satisface las conocidas propiedades de una métrica.

En términos matemáticos, el espacio $I(\mathbb{R})$ es un espacio métrico.

Medidas de estadística descriptiva con datos de intervalo

Consideremos una variable valorada por intervalos, $[X]$, de la que se han tomado n datos.

$$\{[x_i], i = 1, 2, \dots, n\} = \{[x_{Li}, x_{Ui}], i = 1, \dots, n\} = \{[x_{Ci}, x_{Ri}], i = 1, \dots, n\}$$

Por ejemplo, si consideramos los datos de la tabla 1 para ilustrar las medidas que vamos a definir, entonces n es 12. Se han seguido fundamentalmente dos líneas de desarrollo: estadísticos univariantes y estadísticos bivariantes.

A. Medidas estadísticas clásicas univariantes.

Desarrolladas por Billard y Diday (2003, 2006) deducen que la media muestral, inicialmente notada por \bar{X} , lo que puede llevar a confusión con la variable aleatoria media muestral, por lo que nosotros la notaremos por \bar{x}_I , indicando que es el valor medio que se asigna a una muestra de n intervalos para diferenciarlo de \bar{x} , notación habitual del promedio de n datos clásicos, vendrá dada por

$$\bar{x}_I = \frac{1}{2n} \sum_{i=1}^n (x_{Li} + x_{Ui}) = \frac{1}{n} \sum_{i=1}^n x_{Ci} \quad (5)$$

Es decir, la media de un conjunto de intervalos es la media de sus centros. Obsérvese que si los intervalos fuesen números reales únicos, aparece la conocida media aritmética de n números. Para los datos de la Tabla 1 resulta que

$$\bar{x}_I = 21.241 \text{ Mwh}$$

La varianza muestral, que aunque inicialmente se notó por S^2 , lo que puede llevar a confusión con la variable aleatoria varianza muestral, nosotros notaremos por s_I^2 , indicando que es el valor de varianza que se asigna a una muestra de n intervalos para diferenciarlo de s^2 , notación habitual de la varianza muestral de n datos clásicos, vendrá dada por

$$s_I^2 = \frac{1}{3n} \sum_{i=1}^n (x_{Ui}^2 + x_{Ui}x_{Li} + x_{Li}^2) - \frac{1}{4n^2} \left[\sum_{i=1}^n (x_{Ui} + x_{Li}) \right]^2 \quad (6)$$

Obsérvese que si los intervalos fuesen n números reales únicos (*crisp data*, en inglés), aparece la conocida expresión de la varianza como diferencia entre la media de los cuadrados y el cuadrado de la media. Para los datos de la Tabla 1 resulta que

$$s_I^2 = 454.417.746,7 - 451.173.000,7 = 3.244.745,972 \text{ (Mwh)}^2$$

La raíz cuadrada será

$$s_I = \sqrt{3.244.745,972} = 1.801,32 \text{ Mwh}$$

B. Medidas estadísticas en forma de intervalo.

Han et ál. (2008) proponen el intervalo medio cuya notación y expresión de cálculo se dan a continuación

$$\begin{aligned} [\bar{x}] &= [\bar{x}_L, \bar{x}_U] = \frac{1}{n} \sum_{i=1}^n [x_i] = \frac{1}{n} \sum_{i=1}^n [x_{Li}, x_{Ui}] = \\ &= \left[\frac{\sum_{i=1}^n x_{Li}}{n}, \frac{\sum_{i=1}^n x_{Ui}}{n} \right] \quad (7) \end{aligned}$$

$$\langle \bar{x} \rangle = \langle \bar{x}_C, \bar{x}_R \rangle = \left\langle \frac{\sum_{i=1}^n x_{Ci}}{n}, \frac{\sum_{i=1}^n x_{Ri}}{n} \right\rangle \quad (8)$$

Es decir, según este planteamiento la media de un conjunto de intervalos es otro intervalo cuyos atributos resultan promediándose los atributos de los intervalos considerados. Para los datos de la Tabla 1 resulta que

$$\begin{aligned} [\bar{x}] &= [18.846,91667; 23.634,75]; \langle \bar{x}_C, \bar{x}_R \rangle \\ &= \langle 21.240,83333; 2.393,916667 \rangle \end{aligned}$$

C. Estadística descriptiva bidimensional con datos de intervalo.

Sean dos variables valoradas por intervalos, $[X]$ e $[Y]$, de la que se han tomado n datos

$$\begin{aligned} \{([x_i], [y_i]), i = 1, 2, \dots, n\} &= \\ = \{([x_{Li}, x_{Ui}], [y_{Li}, y_{Ui}]), i = 1, \dots, n\} &= \\ = \{([x_{Ci}, x_{Ri}], [y_{Ci}, y_{Ri}]), i = 1, \dots, n\}. \end{aligned} \quad (9)$$

Para calcular la covarianza, después de diferentes aproximaciones por varios autores, Billard (2008) concluyó que, si calculamos el promedio de cada variable de intervalo, es decir, la media de los intervalos $[X]$ e $[Y]$ que serán

$$\begin{aligned} \bar{x}_I &= \frac{1}{2n} \sum_{i=1}^n (x_{Li} + x_{Ui}), \bar{y}_I = \\ &= \frac{1}{2n} \sum_{i=1}^n (y_{Li} + y_{Ui}) \end{aligned}$$

la covarianza vendrá dada por

$$Cov([X], [Y]) = A + B$$

$$\begin{aligned} A &= \frac{1}{6n} \sum_{i=1}^n [2(x_{Li} - \bar{x}_I)(y_{Li} - \bar{y}_I) + (x_{Li} - \bar{x}_I)(y_{Ui} - \bar{y}_I)] \\ B &= \frac{1}{6n} \sum_{i=1}^n [(x_{Ui} - \bar{x}_I)(y_{Li} - \bar{y}_I) + 2(x_{Ui} - \bar{x}_I)(y_{Ui} - \bar{y}_I)] \end{aligned} \quad (10)$$

Si los intervalos son escalares (i.e.,

$$\begin{aligned} x_{Li} &= x_{Ui} = x_i; y_{Li} = y_{Ui} = \\ &= y_i; \bar{x}_I = \bar{x}; \bar{y}_I = \bar{y} \end{aligned} \quad),$$

aparece la conocida fórmula para la covarianza con datos clásicos. Es decir,

$$Cov(X, Y) = \frac{1}{n} \sum_{i=1}^n [(x_i - \bar{x})(y_i - \bar{y})]$$

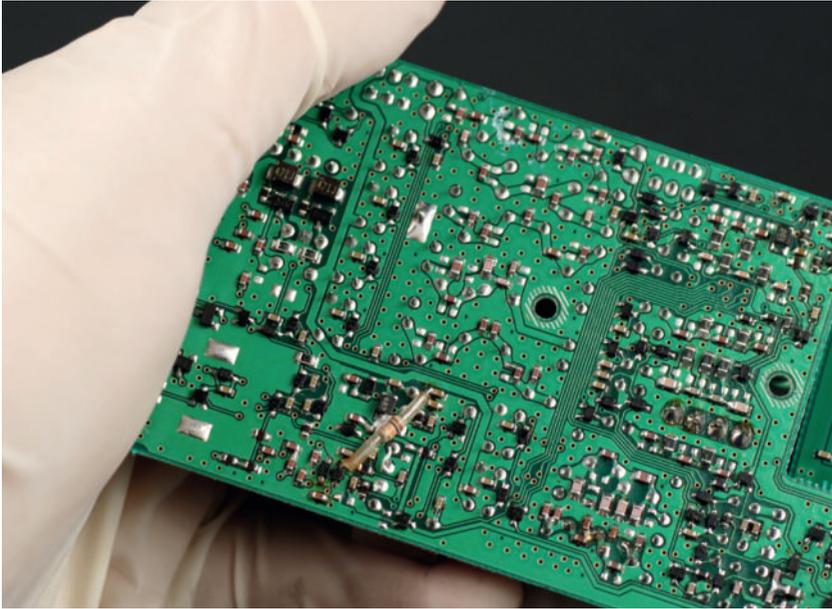
Definida la covarianza es posible establecer un coeficiente de correlación para datos de intervalo como extensión de la conocida fórmula para datos clásicos. Llamemos $s_{[X]}^2$ a la expresión (6) para los datos de intervalo de $[X]$ y $s_{[Y]}^2$ a la expresión (6) para los datos de intervalo de $[Y]$. Entonces

$$r_I = \frac{Cov([X], [Y])}{\sqrt{s_{[X]}^2 s_{[Y]}^2}} \quad (11)$$

De esta manera, todas las consideraciones propias del análisis e interpretación del coeficiente de correlación lineal de Pearson son aplicables a este coeficiente para datos de intervalo.

Conclusiones

Este artículo ha mostrado aproximaciones del AI a la ingeniería que incluyen los errores y la incertidumbre



en las medidas, la incertidumbre en los sistemas de ingeniería, la borrosidad en dichos sistemas y el análisis temporal de la realidad. En varias de ellas los

autores concluyen que el AI proporciona soluciones factibles y de menor complejidad que las aproximaciones clásicas basadas en distribuciones.

Por otro lado, se han abordado los problemas de la notación, las operaciones aritméticas y la distancia entre intervalos. Además, se han incluido las medidas de estadística descriptiva unidimensional y bidimensional para datos de intervalo, las cuales generalizan las conocidas para datos clásicos.

De la misma forma que se han constatado importantes avances en robótica mediante el uso del AI, es de esperar que algunas áreas críticas en la evolución de la ingeniería del siglo XXI, como la nanotecnología, las redes inteligentes o la telemática, se beneficien de un enfoque basado en el AI.

Mención aparte merecen las ciudades inteligentes (*smart cities*) donde la aproximación del AI permitirá a los ciudadanos manejar la información en rangos, lo que reducirá la incertidumbre en la toma de decisiones y mejorará la eficiencia de los sistemas al controlar mejor el riesgo. ■

Bibliografía

- Alefeld, G., y Mayer, G. (2000). Interval analysis: theory and applications. *Journal of Computational and Applied Mathematics*, 121, 421-464.
- Arroyo, J. (2008). Métodos de Predicción para Series Temporales de Intervalos e Histogramas. Ph.D. Thesis, Universidad Pontificia Comillas, Madrid, Spain.
- Arroyo, J., Maté, C. (2009). Forecasting histogram time series with k-nearest neighbours methods. *International Journal of Forecasting*, 25 (1), 192-207.
- Arroyo, J., González-Rivera, G., y Maté, C. (2011). Forecasting with interval and histogram data. Some financial applications. *Handbook of empirical economics and finance*, A. Ullah, D. Giles (Eds.), Chapman & Hall/CRC, 247-279.
- Arroyo, J., Espinola, R., y Maté, C. (2011). Different Approaches to Forecast Interval Time Series: A Comparison in Finance. *Computational Economics*, 37 (2), 169-191.
- Billard, L. (2008). Sample Covariance Functions for Complex Quantitative Data. Processing, World Conferences International Association of Statistical Computing 2008, Yokohama, Japan.
- Billard, L., Diday, E. (2003). From the statistics of data to the statistics of knowledge: Symbolic data analysis. *Journal of the American Statistical Association*, 98 (462), 470-487.
- Billard, L., Diday, E. (2006). *Symbolic Data Analysis: Conceptual Statistics and Data Mining*. Chichester, Wiley.
- Bock, H.-H., Diday, E. (2000). *Analysis of Symbolic Data, Exploratory Methods for Extracting Statistical In-*
- formation from Complex Data*. Eds., Berlin-Heidelberg, Springer-Verlag
- García-Ascanio, C., y Maté, C. (2010). Electric power demand forecasting using interval time series: A comparison between VAR and iMLP. *Energy Policy* 38, 715-725.
- Han, A., Hong, Y., Lai, K. K., Wang, S. (2008). Interval time series analysis with an application to the sterling-dollar exchange rate. *Journal of Systems Science and Complexity*, 21 (4), 550-565.
- Lei, D. (2012). Interval job shop scheduling problems. *International Journal of Advanced Manufacturing Technology*, 60 (1-4), 291-301.
- Maté, C. G. (2011). Book Review: Ramon E. Moore, Ralph B. Kearfott, Michael J. Cloud (Eds.), Introduction to interval analysis, SIAM, USA, 2009, 223pp., ISBN: 978-0-898716-69-6 (hardcover), \$72. *Fuzzy Sets and Systems*, Vol. 177, Issue 1, 95-97.
- Merlet, J.-P. (2010). Interval analysis and robotics. Springer Tracts in Advanced Robotics, 66 (STAR), 147-156.
- Moore, R. E. (1966). *Interval Analysis*. Prentice-Hall, New York.
- Moore, R. E. (1979). *Methods and Applications of Interval Analysis*. Philadelphia: Society for Industrial and Applied Mathematics, New York.
- Moore, R. E., Kearfott, R. B., Cloud, M. J. (2009). *Introduction to Interval Analysis*. SIAM, USA.
- Mousavi, S. M., Tavakkoli-Moghaddam, R., Hashemi, H., Mojtahedi, S. M. H. (2011). A novel approach based on non-parametric resampling with interval analysis for large engineering project risks. *Safety Science*, 49 (10), 1340-1348.
- Muñoz San Roque, A.; Maté, C.; Sarabia, A.; Arroyo, J. (2007). iMLP: applying multi-layer perceptron to interval-valued data. *Neural Processing Letters*, 157-169.
- Neumaier, A. (1990). *Interval Methods for Systems of Equations*. Cambridge University Press, Cambridge.
- Noirhomme-Fraiture, M., Brito, P. (2011). Far Beyond the Classical Data Models: Symbolic Data Analysis. *Statistical Analysis and Data Mining*, 4 (2), 157-170.
- Qiu, Z., Ma, L., Wang, X. (2009). Non-probabilistic interval analysis method for dynamic response analysis of nonlinear systems with uncertainty. *Journal of Sound and Vibration*, 319 (1-2), 531-540.
- Rothwell, E. J.; y Cloud, M. J.; (2012). Automatic Error Analysis Using Intervals. *IEEE Transactions on Education*, vol. 55, no. 1, 9-15.
- Sinova, B., Colubi, A., Gil, M. A., González-Rodríguez, G. (2012). Interval arithmetic-based simple linear regression between interval data: Discussion and sensitivity analysis on the choice of the metric. *Information Sciences*, 199, 109-124.
- Xiao, Y.; Du, Z.; Dong, W. (2012). Smooth and near time-optimal trajectory planning of industrial robots for online applications. *Industrial Robot: An International Journal*, Vol. 39, Issue: 2, 169 - 177.
- Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control* 8, 338-353.