UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS MATEMÁTICAS
DEPARTAMENTO DE ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

CONSTRUCCIÓN DE ÍNDICES CONTEXTUALES DE DESARROLLO INFANTIL TEMPRANO BASADOS EN FAMILIAS CONSISTENTES DE OPERADORES DE AGREGACIÓN.

TESIS DOCTORAL DE:

KARINA ROJAS PATUELLI

BAJO LA DIRECCIÓN DE:

JAVIER MONTERO DE JUAN
DANIEL GÓMEZ GONZÁLEZ

Madrid, 2013

Universidad Complutense de Madrid

FACULTAD DE CIENCIAS MATEMÁTICAS

Departamento de Estadística e Investigación Operativa

Construcción de índices contextuales de desarrollo infantil temprano basados en familias consistentes de operadores de agregación.

Doctorando: Karina Rojas Patuelli.

Directores:
Javier Montero de Juan.
Daniel Gómez González.

Madrid, Marzo 2013

En primer lugar, quiero expresarles un agradecimiento muy especial a los profesores Javier Montero y Daniel Gómez por aceptar el compromiso de trabajar con ellos durante este periodo, confiando en mí y valorando mi proyecto de investigación. Ambos, junto con el profesor Tinguaro Rodríguez, me han entregado muchas y valiosas sugerencias y apoyo, que sin duda han hecho de mi trabajo una rica experiencia tanto en nuevos conocimientos y aprendizajes, como en momentos de amistad, comprensión y compañía.

Gracias a Estela Ortiz, que me inició en la temática del desarrollo infantil temprano al formar parte del equipo de estudios de la *Junta Nacional de Jardines Infantiles del Gobierno de Chile*, patrocinando mis estudios doctorales durante su dirección. Así mismo, al *Programa de Formación de Capital Humano Avanzado* de CONICYT-BECAS CHILE, por otorgarme la Beca que permitió realizar mi proyecto de investigación.

Agradezco también a mis compañeros y amigos Martín López, Paola Viviani, Rosa Alonso, Javier Martín, Gregorio Tirado y Tinguaro Rodriguez, uno de mis principales apoyos en los momentos de dificultad. Infinitas gracias por la agradable, saludable, cálida y alegre compañía que me han brindado a lo largo de estos cuatro años, ciertamente habría sido muy duro el trabajo sin vuestra activa presencia.

A mi madre y a mi padre por el gran esfuerzo que les ha significado permitirme llegar hasta acá. Sé que no ha sido fácil, y se los agradezco profunda y sinceramente. A mis amigas y amigos de Chile, que desde la distancia han seguido año a año mi historia, y me han apoyado con consejos, palabras de aliento, confianza y cariño. Y a mi hija Julieta, que me da la energía para seguir adelante sólo con una mirada y una sonrisa. Gracias July por todos los momentos de felicidad que me has dado, asi como por todo lo que he aprendido gracias a tu inocente, simple y grandiosa sabiduría.

Y termino agradeciendo a Pablo, que junto a sus hijas y a sus padres, han hecho de este último tiempo uno de mis mejores años. Gracias por la seguridad y el amor que me has hecho sentir, por el apoyo y el ímpetu, ha sido fundamental para lograr la etapa final de mi proyecto. Gracias familia!!.

Prólogo

Este trabajo se centra en la generación de un método de construcción de índices en el marco de un problema de agregación y de clasificación borrosa no supervisada, cuando la información que se agrega tiene un grado de incertidumbre y está estructurada jerárquicamente.

Este tema fue motivado por la necesidad de generar índices de desarrollo infantil temprano en el contexto de evaluación de políticas públicas dirigidas a la primera infancia. En este escenario, dado el probado impacto que tienen las intervenciones educativas dirigidas a esta población, muchos estudios se han focalizado en identificar los factores que inciden en el logro del desarrollo infantil, siendo la calidad del entorno del hogar de los niños, un factor diferenciador ante el logro de su desarrollo, y evidenciando por tanto, la necesidad de considerar el contexto del niño al momento de medir su desarrollo. Sin embargo, las frecuentes medidas nítidas, índices basados en técnicas lineales y la comparación de resultados de cada niño con una población de referencia estándar para evaluar el logro de su desarrollo, no representan adecuadamente esta realidad.

Frente a esto, se desarrolla un método de generación de índices contextuales que puede ser utilizado en distintas áreas, particularmente, lo aplicamos para evaluar el desarrollo infantil temprano de niños chilenos que se encuentran en condición de vulnerabilidad. Para la generación del método, se estudian y desarrollan tres temas, a saber; ***representación del***

conocimiento bajo incertidumbre, ***robustez de un proceso de agregación con información jerárquica priorizada***, y ***clasificación borrosa no supervisada***. Esto, a fin de conseguir una representación más adecuada del conocimiento bajo una realidad psicosocial, generar índices robustos que consideren la naturaleza del conjunto de datos de entrada, e identificar una población de referencia que comparta características del entorno social de la población evaluada a la hora de interpretar sus resultados.

Desde nuestro punto de vista, y rescatando el objetivo principal que se persigue a la hora de generar índices; "*describir características, comportamientos o fenómenos de la realidad, a través de la evolución de una variable o el establecimiento de una relación entre variables, las que comparada con períodos anteriores, productos similares o una meta o compromiso, permita evaluar el desempeño y su evolución en el tiempo*" [63], se considera esencial acercarse de mejor manera a la realidad que se pretende describir, mediante la representación de la información de una manera más precisa, una agregación robusta de la información, acorde a las características del conjunto de datos de entrada, y una clasificación de elementos con límites difusos que se adecúe a conjuntos de datos con algún nivel de concentración. Es en este sentido, en el que formulamos el problema desde la *lógica borrosa*, ya que entrega mayor flexibilidad tanto en el manejo de los conceptos abstractos mediante variables lingüísticas asociadas a conjuntos borrosos, como en la formulación de relaciones entre variables mediante procesos de agregación, y la identificación de patrones adecuados mediante índices de calidad de sistemas de clasificación borrosa no supervisada.

En el primer capítulo de este trabajo, se contextualiza el estudio, se define el problema y se plantean los objetivos, para luego mostrar en el segundo capítulo, un breve estado del arte acerca de enfoques y métodos de medición del desarrollo infantil, lógica borrosa, proceso de agregación y sistemas de clasificación borrosa no supervisada.

Si bien el enfoque de este trabajo ya es un aporte en el desarrollo de índices bajo un escenario psicosocial, en tanto se le otorga importancia al contexto con el objetivo de relativizar las medidas de evaluación, claramente, el principal aporte se presenta en el tercer capítulo. En este capítulo, se evidencia la necesidad de establecer restricciones en la definición de ***familia de operadores de agregación (FAO)***, a fin se asegurar algún nivel de consistencia del proceso de agregación en el que participa, y de esta forma, garantizar resultados agregados robustos. Frente a esto, se definen los conceptos de ***consistencia*** y ***estabilidad*** en el contexto de un problema de agregación, y se presentan definiciones de distintos niveles de ***estabilidad estricta*** para una familia de operadores de agregación. Tales definiciones, consideran la naturaleza del conjunto de datos de entrada en términos de su estructura, focalizándonos en datos desestructurados, en datos con estructura lineal y jerárquica priorizada. El análisis de estas propiedades nos permite conocer a priori el nivel de estabilidad que tendrán los resultados, luego de un proceso de agregación bajo determinada estructura del conjunto de datos. A modo de ejemplo, en este capítulo se presenta un tabla resumen con las familias de operadores de agregación comúnmente utilizadas y su nivel de estabilidad, además de ejercicios de simulación que muestran el comportamiento de algunas familias bajo distintos tamaños muestrales, a fin de identificar el tamaño mínimo necesario para asegurar cierto nivel de estabilidad.

Otro aporte de este trabajo se muestra en el cuatro capítulo, en el cual se utilizan familias consistentes de operadores de agregación para definir los criterios de ***relevancia***, ***redundancia*** y ***cobertura*** de las clases de una clasificación borrosa de un conjunto de datos. En base a la combinación de estos tres criterios, se genera un método de evaluación de calidad de un sistema de clasificación borrosa no supervisada. Una de las principales ventajas de este método, es que **no exige la definición previa de la cantidad de clases que tendrá la clasificación**, en tanto evalúa la calidad de un conjunto de clases borrosas determinadas por los posibles patrones de clase de una clasificación, permitiendo **seleccionar las mejores clases en función de criterios de calidad** . Este conjunto de posibles patrones se definen en una rejilla, de tal manera que cubran los valores que toman los variables de entrada, considerando además, todas las permutaciones que existen entre dichas variables. Este proceso de búsqueda de los mejores patrones, y por consiguiente, de una buena clasificación borrosa del conjunto de datos, va eliminando clases en la medida que no satisfacen las restricciones establecidas sobre los criterios de relevancia, redundancia y cobertura.

Finalmente, en el *Capítulo* 5 se establecen pautas para generar ***índices contextuales*** en base a la agregación de la información con incertidumbre bajo procesos robustos estudiado en el *Capítulo* 3, y la identificación de una clasificación borrosa de calidad estudiada en el *Capítulo* 4. En este capítulo, se define el ***contexto*** de un elemento del conjunto de datos de entrada, sobre el cual es posible construir distintos tipos de índices que consideren la información de dicho contexto en la evaluación de logro de cada elemento.

El ámbito de aplicación de este trabajo se ejemplifica al final de los Capítulos 3, 4 y 5, en los cuales se aplica el método de construcción de índices a la realidad chilena. Primero se desarrollan índices de entorno del hogar del niño, luego se desarrolla un algoritmo para determinar patrones que dan lugar a un sistema de clasificación borrosa no supervisada, cuyo análisis permite la definición de índices contextuales. Esta clasificación se puede utilizar para acotar el contexto de los niños, en tanto a través de ella, se identifica la población infantil que comparte características respecto de su entorno inmediato, población que servirá de estándar relativo para una evaluación de desarrollo contextual. Todo esto, bajo el objetivo de generar índices contextuales de desarrollo infantil temprano de dicha población.

Finalmente, en el último capítulo se presentan las conclusiones, aportes y futuras líneas de investigación. Las principales conclusiones de este trabajo, así como también las futuras líneas de investigación, se pueden resumir en los siguientes tres aspectos relacionados con cada uno de los desafíos planteados al inicio de la investigación:

- ***Conseguir una representación más adecuada del conocimiento bajo una realidad psicosocial.***

La naturaleza imprecisa de los conceptos que se estudian en las ciencias sociales y del comportamiento, dificulta el tránsito entre el desarrollo teórico de los conceptos y la representación numérica de los mismos. En este trabajo, el problema de la representación es facilitado al formularlo desde la ***lógica borrosa***, en tanto permite representar los conceptos a través de ***variables lingüísticas*** asociados a ***conjuntos borrosos***, en la cual el valor de verdad de sus estados rescata la vaguedad de dichos conceptos.

El ***entorno del hogar de un niño*** es un concepto impreciso, que está compuesto por distintos aspectos, todos imprecisos. Tales aspectos constituyen dimensiones del concepto o factores que inciden en la consecución de un entorno adecuado para el desarrollo del niño, tales como: *la condición económica y familiar en la cual vive el niño*, *el entorno material y las rutinas que tiene el niño*, asi como también, *las pautas de estimulación del desarrollo y de definición de límites de conducta que manifiesta su familia*, y *la calidad de las interacciones comunicativas y afectivas dentro y fuera de su hogar*. Cada una de estas dimensiones o factores, a su vez, están constituidos por sub dimensiones, que presentan una estructura priorizada respecto del aporte en el factor del cual forman parte. Por ejemplo, la condición económica y familiar del hogar del niño, está compuesta por: *organización equilibrada de roles parentales*, *clima familiar no violento* y *seguridad económica*, sub dimensiones que inciden con distinta fuerza en el logro de un entorno facilitador del desarrollo, siendo más importante la organización familiar que la seguridad económica a la hora de estudiar el desarrollo infantil. Y nuevamente, estas sub dimensiones se constituyen por un conjunto de variables lingüísticas que también presentan una estructura de prioridades, generándose de esta forma, una ***estructura jerárquica priorizada*** de variables lingüísticas borrosas.

Con esto podemos ver, que el intento por medir algunos conceptos imprecisos se traduce en una descomposición del mismo, en la cual cada uno de sus componentes también presentan un grado de imprecisión, constituyéndose una estructura anidada de conceptos que llevan consigo un establecimiento de prioridades. En este trabajo, se utilizó una técnica cualitativa llamada "***juicio experto***", que corresponde a una serie de entrevistas semi estructuradas dirigidas a expertos, la cual nos permitió establecer una definición teórica del entorno del hogar del niño en función de la información disponible, y de las distintas líneas teóricas que estudian el desarrollo en primera infancia y factores asociados. El producto final de esta serie de entrevistas es una definición teórica del entorno del hogar del niño, plasmada en una estructura jerárquica priorizada de los componentes que la constituyen y su correspondiente orden de prioridades, la que se utiliza para la generación de variables lingüísticas asociadas a conjuntos borrosos.

Ahora bien, la representación numérica de los conceptos abstractos medidos mediante técnicas cualitativas, aún tiene muchos ámbitos de mejoramiento. En este trabajo, se utiliza información cualitativa que ha sido rescatada desde los discursos de las personas, sin embargo, tales discursos han sido "*cuantificados*" mediante variables lingüísticas cuyos estados o categorías son nítidas y presentan el mismo nivel de importancia ante al concepto medido. Frente a esto, a fin de intentar "*rescatar*" la riqueza de los discursos, se realizó un trabajo conjunto con expertos en desarrollo infantil, el cual permitió asociar las variables lingüísticas a conjuntos borrosos que representan los distintos estados de dichas variables, estados que además tienen diferentes grados de importancia frente a su correspondiente variable lingüística. De esta forma, cada niño se asocia a un grado de pertenencia a cada uno de los estados o categorías de las variables lingüísticas, aportando una mayor flexibilidad a la representación del concepto.

Al respecto, una línea de trabajo futuro es contribuir al mejoramiento de la representación numérica de este tipo de conceptos, sobre la base de los discursos originales extraídos mediante técnicas cualitativas, y a través de esta, poder definir los estados de las variables lingüísticas y sus correspondientes grados de pertenencia a conjuntos borrosos asociados desde la fuente original.

- ***Generar índices robustos que consideren la naturaleza del conjunto de datos de entrada.***

Manteniendo el marco de la lógica borrosa, la generación de índices puede ser visto como un proceso de agregación de variables lingüísticas asociadas a conjuntos borrosos, en tanto reduce la dimensionalidad de un conjunto de datos mediante operadores de agregación que relacionan dichas variables. No obstante, los resultados que se obtienen en un índice deben presentar cierto nivel de robustez que no afecte su validez, en tanto constituyen una herramienta para la toma de decisiones.

Esto nos llevó a estudiar los conceptos de ***consistencia*** y ***estabilidad*** en el contexto de un proceso de agregación, concluyendo que mediante algunas propiedades de estabilidad definidas para familias de operadores de agregación, es posible relacionar un conjunto de variables lingüísticas borrosas, a fin de obtener un valor agregado que asegure estabilidad ante posibles cambios de cardinalidad del conjunto de datos de entrada. Uno de los ámbitos de aplicación más directos es el usual problema de agregación cuando existe pérdida de la información en un elemento del conjunto de datos de entrada, ya que este problema queda resuelto si el proceso de agregación se ha definido en base a ***familias consistentes de operadores de agregación***.

En este trabajo se definen tres niveles de ***estabilidad estricta*** para familias de operadores de agregación, en términos absolutos, en el límite y en probabilidad, tanto para agregar datos sin una estructura inherente, como para agregar datos linealmente ordenados y con estructura jerárquica priorizada. Estas propiedades se basan en la idea que el resultado obtenido en base a un operador definido para n ítems, no difiera del resultado obtenido en base a un operador definido para $n-1$ ítems, cuando el elemento adicionado corresponde a la agregación de los elementos ya contenidos en el conjunto de datos. De esta forma, dada una estructura determinada del conjunto de datos de entrada, es posible definir las familias consistentes de operadores que se utilizarán en el proceso de agregación en base al nivel de estabilidad que se requiere para asegurar resultados robustos.

Los tres niveles de estabilidad estricta de familias de operadores de agregación que son utilizadas para agregar datos con una estructura lineal, se han definido para datos cuyo orden lineal va desde la izquierda hacia la derecha como ***R-estrictamente estable***, y por tanto, el último elemento que se agrega es la agregación de los anteriores, y para aquellos que van desde la derecha hacia la izquierda como ***L-estrictamente estable***, y por tanto, el primer elemento es la agregación de los siguientes. Además, para el caso de conjuntos de datos que cambian de sentido a partir de cierto momento, se define el cumplimiento de esta propiedad desde la derecha y desde la izquierda como ***LR-estrictamente estable***.

Ciertamente, una línea de trabajo futuro es extender estas propiedades para el caso en que se agregue un elemento en la posición i-ésima desde la derecha o j-ésimo desde la izquierda, y de esta forma definir las propiedades ***iR-estrictamente estable*** y ***jL-estrictamente estable*** para cada uno de los niveles de estabilidad (absoluta, asintótica y en probabilidad). Esta extensión resuelve el problema de información faltante de alguna variable de un elemento del conjunto de datos de entrada, **independientemente de la posición que tenga**, en tanto permite generar un operador de la misma familia pero de una dimensión menor, manteniendo la estabilidad de los resultados de dicho elemento, sin importar si es el primero o el último.

- ***Identificar una población de referencia que comparta características del entorno social del niño a la hora de interpretar los resultados de su desarrollo.***

En construcción de índices, además de establecer un proceso que disminuya la dimensionalidad del conjunto de datos mediante operadores que agreguen la información, se requiere la definición de parámetros que otorguen una valoración a tales resultados para su posterior interpretación. Por ejemplo, en el contexto del desarrollo infantil, los índices de logro de un niño suelen ser comparados con baremos obtenidos en base a una población estándar a fin de valorar su nivel de desarrollo.

A nuestro juicio, es metodológicamente correcto tomar en cuenta las condicionantes que intervienen en la medida de logro a la hora de generar índices, es por esto que nos referimos a ***índices contextuales***. Por ejemplo, respecto de la temática de estudio, el proceso de desarrollo de un niño no es el mismo para todos los sujetos, en tanto existen variables socio demográficas y económicas que inciden en el logro de su desarrollo, asi como también, variables relativas a la dinámica familiar en términos de conductas, actitudes y organización en torno al fomento del desarrollo infantil, requiriéndose por tanto, baremos relativos, que consideran las características del entorno del hogar de cada niño ante el logro de su desarrollo.

La identificación de entornos familiares bajo el enfoque de este trabajo, corresponde a un problema de clasificación borrosa no supervisada. Los métodos comúnmente utilizados en clasificación no supervisada, requieren la definición a priori de la cantidad de clases que tendrá la clasificación borrosa del conjunto de datos, cuestión que se ha resuelto en gran medida con la definición de índices que evalúan la calidad de las particiones para distinto número de clases, a fin de seleccionar la partición mejor evaluada. No obstante, estas medidas se suelen basar sólo en un criterio, tendiendo a valorar aquellas particiones borrosas que no presentan solapamiento.

En este trabajo, se establece un método de selección de clases borrosas en base a la búsqueda de los mejores patrones borrosos según algunos criterios de calidad. Dicha búsqueda se realiza sobre un conjunto de posibles patrones definidos sobre una rejilla, a fin de abarcar la amplitud de valores que toman las variables del conjunto de datos de entrada. A partir de dichos patrones, se generan las clases borrosas que serán evaluadas en función de los índices de ***relevancia***, ***redundancia*** y ***cobertura***.

Por lo tanto, además de no requerir una definición a priori de la cantidad de clases de la clasificación, este método garantiza que las clases seleccionadas son las mejor evaluadas respecto de un conjunto de clases que cubren la amplitud de los valores que toman las variables de entrada.

Mediante la aplicación de este método de búsqueda de patrones, se identificaron cinco entornos del hogar de niños chilenos en condición de vulnerabilidad:

- *Precarios en su organización pero con pautas de crianza favorables para el desarrollo*,
- *Favorables para el desarrollo*,
- *Con carencias en las relaciones familiares y condiciones económicas, además de pocos recursos para la promoción y desarrollo de los niños*,
- *Con precaria socialización fuera del hogar, pero con relevancia en el manejo disciplinar adecuado para el desarrollo del niño* y,
- *Con dificultades para poner límites y con un potencial clima familiar violento.*

Por lo tanto, el entorno del hogar de cada niño se constituye a través del grado de pertenencia a cada uno de estos cinco entornos. Esta información es utilizada para establecer su población normativa, y a través de esto, se genera una medida que indica el logro de su desarrollo considerando su propio contexto, el que es definido mediante las característica de su entorno familiar.

Ahora bien, además de otros ámbitos de aplicación del método de construcción de índices contextuales presentado, nos planteamos tres posibles líneas futuras de investigación.

La primera orientada a definir una medida de similitud entre vectores en las cuales sus coordenadas representan grados de pertenencia, y que por tanto, tal medida debe considerar que cada coordenada está asociada a un concepto diferente, implicando relevar la posición del dato en el vector de coordenadas de pertenencias. La segunda, orientada a comparar los resultados obtenidos en base al método de clasificación borrosa no supervisada propuesto en este trabajo con distintos métodos comúnmente utilizados, y bajo distintos ámbitos de aplicación. Y finalmente, desarrollar índices para evaluar los resultados de una clasificación borrosa no supervisada en base a un marco de referencia borroso.

Los desarrollos de este trabajo nos han permitido realizar presentaciones en congresos nacionales e internacionales (*FuzzyMad 2010-2011-2012, EUROFUSE 2011, IPMU 2012* y *ESTYLF 2012*), así como también, publicar dos artículos en la revista *Fuzzy Sets and Systems* (uno de ellos aún se encuentra en segunda revisión), y ser parte de una selección de trabajos publicados en dos libros. Tales logros se sintetizan a continuación:

- Rojas, K., Gómez, D., Rodríguez, J.T., Montero, J., Valdivia, A., Paiva, F. Development of child's home environment indexes based on consistent families of aggregation operators with prioritized hierarchical information. Fuzzy Sets and Systems: Special Issue EUROFUSE 2011. (in second revision) (2013).

- Rojas, K., Gómez, D., Rodríguez, J.T., Montero, J. Strictly stable families of aggregation operators. Fuzzy Sets and Systems.

http://dx.doi.org/10.1016/j.fss.2012.12.010 (2013).

- Rojas, K., Gómez, D., Rodríguez, J.T., Montero, J. Stability in aggregation operators. In Greco, S. et al. (Eds.) Advances in Computational Intelligence. Springer. 299, 317-325 (2012).

- Rojas, K., Gómez, D., Rodríguez, J.T., Montero, J. Some properties of consistency in the families of aggregation operators. In B. de Baets (Ed.) et al., Eurofuse 2011, AISC. Springer. 107, 169-176 (2011).

Índice general

Capítulo 1

Introducción.

1.1. Contextualización y definición del problema.

Las intervenciones dirigidas a la primera infancia son de gran interés en política educativa, ya que en este período se puede producir mayor impacto en desarrollo humano. La calidad del entorno social de los niños en edad temprana es la influencia más importante a considerar en el logro del desarrollo infantil, afectando a lo largo de toda la vida escolar. Es por esta razón, que el desarrollo de índices de entorno del hogar del niño aparece de forma natural en investigaciones relativas a política educacional y evaluación de programas sociales dirigidos a la primera infancia, siendo por tanto, un insumo necesario para la generación de índices de desarrollo infantil temprano. Sin embargo, las medidas nítidas, los índices basados en técnicas lineales habituales y la comparación con una población de referencia estándar para evaluar el logro en desarrollo infantil, no aseguran una representación adecuada de la realidad, en tanto ésta tiene una naturaleza borrosa y un comportamiento no lineal que depende de las características del entorno.

Frente a este escenario se evidencian tres desafíos. En primer lugar, ***conseguir una representación más adecuada del conocimiento bajo una realidad psicosocial***, en segundo lugar, ***generar índices robustos que consideren la naturaleza del conjunto de datos de entrada***, y por último, ***identificar una población de referencia que comparta características del entorno social del niño a la hora de interpretar los resultados de su desarrollo***.

Atendiendo al primer desafío, el problema surge en que ***la mayoría, si no todos los conceptos de las ciencias sociales son imprecisos***, en el sentido que a menudo es muy difícil asignar objetos a categorías definidas exactamente [88]. Según Lazarsfeld [56], en algunos casos solo se observan los síntomas, detrás de la cual se asume una realidad permanente, en otros casos, el fin investigativo es tan grande, que sólo podemos analizar algunos aspectos, como algún patrón de la cultura. El problema parece requerir un tipo de formulación más flexible. Así mismo, Luce [61] comparte que el lenguaje de conjuntos no es adecuado para formular problemas psicológicos, mencionando que los límites de muchos de ”mis conjuntos” con los que trabajo, son más borrosos que los de los conjuntos matemáticos. Las categorías son inciertas, no son conjuntos bien definidos, y su borrosidad no es bien representada por la noción de probabilidad.

Frente a este problema de representación, Zadeh [95], interesado en proveer un lenguaje formal para las ciencias sociales y del comportamiento, propuso un nuevo modelo matemático que ayuda a formalizar la vaguedad y graduación de los conceptos lingüísticos a través de *conjuntos borrosos*. Este enfoque permite modelizar la realidad de manera más flexible, adecuándose a las variables o factores psicosociales. Por lo tanto, frente a la imprecisión de la información relativa a las características del entorno del hogar del niño, ***se requiere un modelado basado en la lógica borrosa, en tanto la flexibilidad del enfoque rescata de mejor manera la riqueza de los discursos***.

Manteniendo este enfoque y atendiendo al segundo desafío, bajo un contexto borroso, ***el desarrollo de índices puede ser analizado como un proceso de agregación***, en tanto mediante un operador de agregación se puede reducir la dimensionalidad de un conjunto de datos que contiene información con incertidumbre, generando un valor agregado unidimensional. El objetivo que se persigue ante un problema de agregación es facilitar el análisis y la toma de decisiones, de aquí la importancia de la robustez del proceso en términos de los resultados obtenidos. Sin embargo, el cumplimiento de este objetivo se ve afectado por el *problema de dimensionalidad*, en tanto a lo largo del proceso de agregación se pueden producir diversos cambios de cardinalidad en los datos de entrada, ante lo cual, cada vez que se produce un cambio, es necesario actualizar el valor agregado a partir del nuevo conjunto de datos. Esto ocurre con frecuencia cuando por ejemplo no existe información en una o más variables de un elemento del conjunto de datos, frente a esto, es necesario agregar la información bajo un operador de menor dimensión. En este caso, el problema de agregación ya no involucra sólo un operador para un n determinado, sino que a una secuencia de operadores $A_2, A_3, ..., A_n$, que constituyen una *familia de operadores de agregación (FAO)*, también conocida como *función de agregación extendida (EAF)* por otros autores [45, 16]. Sin embargo, la definición de FAO no establece restricciones que garanticen que el nuevo operador de agregación utilizado en dicha actualización, mantenga la consistencia lógica respecto de la agregación previa. Por lo tanto, a fin de asegurar un proceso de agregación robusto ante cambios de dimensionalidad, ***es necesario analizar la consistencia de las FAOs de manera que se tenga en cuenta las características de los datos de entrada, y se asegure un nivel de estabilidad en los resultados asociados a un proceso de agregación***.

Por último, y dado que el desarrollo óptimo de un niño requiere un ambiente que garantice la satisfacción de todas las necesidades físicas básicas y las disposiciones para el cuidado de la salud y la seguridad (Caldwell y Bradley, 1984), ***es metodológicamente correcto considerar el entorno del hogar del niño al momento de estudiar y comparar su nivel de logro mediante un índice contextual de desarrollo infantil***, en tanto constituye un factor diferenciador. Así mismo, psicólogos y educadores han señalado la inadecuación de los test tradicionales para revelar la capacidad de aprender en niños con dificultades de aprendizaje o en aquellos en los que factores no intelectuales son la causa de sus fallos [87]. Desde este punto se avanza en la evaluación del *potencial de aprendizaje* en la línea de Vygotzki, donde su objetivo no es medir tan sólo la ejecución de los sujetos, sino su posibilidad de aprendizaje. Su fin no es el pronóstico académico, entendido del modo tradicional, sino la estimación de la posibilidad de aprovechamiento de diferentes programas de entrenamiento cognitivo.

Siguiendo esta línea, en el *Capítulo* 5 se establece el contexto del niño mediante la definición de un conjunto compuesto por niños que comparten algunas características de su entorno familiar. Este conjunto constituirá la población normativa del niño, que a través de un operador de agregación (como por ejemplo el máximo o el percentil 95), permite estimar el ***potencial de desarrollo*** del niño, y con esto, obtener **una medida de desarrollo que representa la proporción del potencial de desarrollo logrado por el niño al momento de la medición**. De esta manera, se obtiene una medida de desarrollo contextualizada, donde el grupo de referencia o normativo comparte características del entorno del hogar del niño evaluado.

Este problema puede ser visto como una clasificación borrosa no supervisada, en tanto se requiere identificar patrones representativos de los distintos entornos familiares, y disponer de algún grado de similitud entre los niños respecto de las características de su hogar. Sin embargo, las técnicas usuales de conglomerados no entregan buenos resultados cuando el conjunto de datos de entrada tiene un alto nivel de concentración, como suele ocurrir en investigación social y del comportamiento, en tanto identifican patrones con poca disimilitud. Además, requieren la definición de la cantidad de clases del sistema de clasificación, información que *a priori* es desconocida. Por lo tanto, a fin de generar un sistema de clasificación no supervisada adecuado bajo un conjunto de datos concentrados, ***es necesario generar un procedimiento de búsqueda de patrones en base a distintos***

criterios de calidad de un sistema de clasificación no supervisado en ambiente borroso.

En síntesis, el problema de este trabajo se centra en la utilización de información con incertidumbre para el desarrollo de un sistema de clasificación borrosa no supervisada de entornos familiares de niños en base a familias consistentes de operadores de agregación, en el cual, las clases de dicho sistema representan la población de referencia o normativa para la generación de índices contextuales de desarrollo infantil temprano. Para llevar a cabo esta tarea, cabe destacar que se ha hecho un avance significativo en los problemas de agregación incorporando las ideas de estabilidad y consistencia en familias de operadores de agregación para la posterior elaboración de índices contextuales.

1.2. Objetivos.

Los tres desafíos planteados anteriormente se recogen transversalmente en los siguientes dos objetivos, el primero desde una perspectiva teórica y el segundo desde una perspectiva aplicada:

Objetivo 1: Desarrollar ún método de construcción de índices.

- Formular el problema de construcción de índices como un problema de agregación en ambiente borroso.
- Representar de manera más precisa la información dada por el conjunto de datos de entrada, intentando rescatar las valoraciones reales de las variables lingüísticas frente al concepto medido.
- Establecer restricciones que garanticen estabilidad en los resultados de un proceso de

agregación, en concordancia con las características del conjunto de datos de entrada.

- Desarrollar un método de búsqueda de patrones en base a la evaluación de calidad de una clasificación borrosa no supervisada.

Objetivo 2: Generar un modelo borroso del desarrollo infantil temprano aplicado a la realidad chilena.

- Construir un modelo teórico del entorno del hogar del niño, en función de las condicionantes que inciden en el logro de su potencial de desarrollo.
- Generar índices de factores asociados a un entorno familiar que fomenta el desarrollo del niño.
- Identificar patrones representativos de los distintos entornos del hogar del niño.
- Generar índices contextuales de desarrollo infantil temprano.

1.3. Extended abstract with conclusions.

1.3.1. Problem description.

The interventions which are targeted to early childhood are of great interest for educational policy, as this early period permits to generate a greater impact on human development. The quality of social environment in early childhood is the most important influence that has to be considered for the child's development, showing an effect throughout the whole school life. This is why the development of indexes evaluating the surroundings of the child's home appears naturally in the studies related with educational policy and evaluation of social programmes targeted to early childhood, becoming a necessary input for the generation of early childhood indexes. However, clear measures and indexes based on the usual linear techniques and on the comparison with a standard reference population for the evaluation of success in child development do not ensure a proper representation of reality, because reality is blurry in nature and behaves in a non-linear way, which depends on the characteristics of its surroundings.

Facing this scenario, three challenges become clear: First, ***achieving a more appropriate representation of knowledge in a psycho-social reality***, second, ***generating sturdy indexes which take into account the nature of the whole set of data inputs***, and finally, ***identifying a reference population with similar characteristics as the child's social environment when interpreting the results regarding the child's development***.

1.3.2. Objetives.

The three challenges mentioned before are set out transversally in the following two objectives; in the first one, from a theoretical perspective and in the second one from a more applied point of view:

Objective 1: Developing a method for the construction of indexes.

- Formulating the problem of construction of indexes as a problem related with aggregation in a fuzzy environment.
- Representing more precisely the information given in the input data set, trying to save

the real valuations of the linguistic variables regarding the measured concept.

- Setting restrictions in order to ensure stability in the results of an aggregation process, in accordance with the characteristics of the input data set.
- Developing a method for the search of patterns based on the quality evaluation of an unsupervised fuzzy classification.

Objective 2: Generating a fuzzy model of early child development applied to the Chilean reality.

- Constructing a theoretical model of the environment of the child's home, depending on the conditions that may have an effect on the child reaching its development potential.
- Generating indexes of factors associated with family environments which promote the child's development.
- Identifying patterns which are representative for the different environments of the child's home.
- Generating contextual indexes of early child development.

1.3.3. Conclusions.

The main conclusions of this study may be summarized within the following three aspects, related with the three challenges that were set at the beginning of the study:

- ***Achieving a more appropriate representation of knowledge in a psycho-social reality.***

The imprecise nature of the concepts that are studied in social and behavioural sciences makes it more difficult to pass from the theoretical development of the concepts to their numerical representation. In this study, the problem of representation becomes easier as it is set within the ***fuzzy logic***, as it permits to represent the concepts with ***linguistic variables*** associated with ***fuzzy sets***, which in the value of truth of its states are rescuing the vagueness of these concepts.

The ***environment of the child's home*** is an imprecise concept, which is composed of different aspects, all of them imprecise in nature. These aspects are dimensions of the concept or factors which have an effect on reaching an appropriate surrounding for the child's development, such as: *economic and family status children live with, material surrounding and the routines the child has, and also patterns of development stimulation and of setting limits on behaviour shown by its family*, and *the quality of the communicative and affective interactions within and outside of its home.* In turn, each one of these dimensions or factors are made up by sub-dimensions, which show a prioritized structure with regard to the contribution they make to the factor they are a part of. For example, the economic and family status of the child's home is composed of: *balanced organization of parental roles, non-violent family atmosphere* and *economic security*, sub-dimensions which have a different weight regarding the effect they have on reaching an enabling environment for development, appearing the family status as more important than the economic security when studying child development. And, again, these sub-dimensions are made up by a set of linguistic variables which also show a structure of priorities. In this way, a ***prioritized hierarchy structure*** of fuzzy linguistic variables is created.

So, we can see that the attempt to measure some imprecise concepts turns into a decomposition of these concepts, showing each of its components a certain degree of imprecision, too, which becomes a nested structure of concepts, meaning the set up of priorities. In this study, a qualitative technique called ***expert judgement*** has been used, which consists of a series of semi-structured interviews with experts, which permitted us to make up a theoretical definition of the environment of the child's home, based on the available information and on the different existing theoretical perspectives in the studies of early childhood development and associated factors. The final product of these interviews is a theoretical definition of the surrounding of the child's home, reflected in an hierarchical prioritized structure of its components and the related order of priorities, which is used for generating linguistic variables associated with fuzzy groups.

That said, the numerical representation of the abstract concepts measured with qualitative techniques still has many areas that may be improved. In this study, the qualitative information used is the one that is taken from the speeches of people. However, these speeches have been ”*quantified*” through linguistic variables and the states or categories of these variables are clear and show the same level of importance with regard to the measured concept. To address this and in order to try to ”*rescue*” the richness of the speeches, the work was done together with experts in child development, which permitted the association of the linguistic variables with fuzzy groups, representing the different states of these variables, states which also have different levels of importance with respect to its corresponding linguistic variable. In this way, each child is associated with a degree of belonging to each one of the states or categories of the linguistic variables, contributing with greater flexibility to the representation of the concept. On this matter, future work should contribute to improve the numerical representation of this kind of concepts, based on the original speeches obtained with qualitative techniques and so be able to define the states of the linguistic variables and their associated degrees of belonging to fuzzy groups associated from the original source.

- ***Generating sturdy indexes which take into account the nature of the whole set of data inputs.***

Maintaining the frame of the fuzzy logic, the generation of indexes may be seen as a process of aggregation of linguistic variables, associates with fuzzy groups, as it reduces the dimensionality of a set of data with aggregation operators, which connect these variables. However, the results obtained by an index must have a certain level of sturdiness without affecting its validity, as they turn into a tool for making decisions.

This has led us to research the concepts of ***consistency*** and ***stability*** in the context of an aggregation process, concluding that through some stability properties defined for families of aggregation operators, it is possible to relate a group of fuzzy linguistic variables, with the aim of obtaining an aggregated value ensuring stability facing possible changes of cardinality of the group of input data. One of the more direct fields of application is the common aggregation problem when there is a loss of information in one of the elements of the data entry group, as this problem is resolved if the aggregation process is defined based on ***consistent families of aggregation operators***.

In this study, three levels of ***strict stability*** are defined for families of aggregation operators, in absolute terms, on the border and in probability; this is for aggregating data which lack an inherent structure, as well as for aggregating data which are ordered lineally and count on a hierarchical prioritized structure. These properties are based on the idea that the result obtained by a certain operator for n items should not be different from the result based on an operator that has been defined for $n-1$ items, when the added element falls within the scope of the aggregation of the already contained elements in the data set. In this way, when we have a given structure of the data entry set, it is possible to define the consistent families of operators which will be used in the process of aggregation, based on the level of stability required for ensuring sturdy results.

The three levels of strict sturdiness for the families of aggregation operators used to aggregate data with a lineal structure have been defined for data with a lineal order moving from the left to the right as ***R-strictly stable***, and so the last aggregated element is the aggregate of the previous ones. And for those moving from the right to the left as ***L-strictly stable*** , and so the first element is the aggregation of the following ones.

Indeed, future work should be about extending these properties for the case in which one element is added on the i-position from the right or the j-position from the left, and so defining the properties ***iR-strictly stable*** and ***jL-strictly stable*** for each of the stability levels (absolute, asymptotic and probabilistic). This extension resolves the problem of missing information of some variable of en element of the data entry set, ***regardless of its position***, as it permits the generation of an operator of the same family, but of a minor dimension, maintaining the stability of the results of this element, no matter if it's the first or the last one.

- ***Identifying a reference population with similar characteristics as the child's social environment when interpreting the results regarding the child's development.***

When constructing indexes, in addition to the establishment of a process which decreases the dimensionality of the data set with operators aggregating the information, for its interpretation we also need the definition of parameters giving a valuation to the results. For example, in the context of child development, the indexes of success of a child are usually compared with scales obtained based on a standard population.

In our view, it is methodologically correct if we take into account the conditioning factors in the measurement of success when generating indexes; this is why we talk about ***contextual indexes***. For examples, with regard to the researched subject, the development process of a child is not the same for all individuals, as there are socio-demographic and economical variables which have an effect on achieving development, as there also are variables related with family dynamics in terms of behaviour, attitudes and organization around fostering the child's development. So, we need relative scales, which consider the characteristics of the surroundings of each child's home in the achievements of its development.

The identification of family surroundings in the view of this research is a problem of unsupervised fuzzy classification. The methods that are usually used in unsupervised classification require the a-priori definition of the number of classes the fuzzy partition of hte data set will have, an issue that has been partially solved with the definition of indexes for the evaluation of the quality of the partitions for different numbers of classes, with the aim of selecting the partition which receives the best evaluation. However, these measures usually are based on just one criterion, tending to valuate higher the partitions without overlapping.

In this study, a method of selection of fuzzy types of a classification is set up, based on the search for the best fuzzy patterns. This search is based on a set of possible patterns which are defined on a grating, aiming to cover the wide extend of values taken by the variables of the entry data set. Based on these patterns, the fuzzy types are generated and they will be evaluated with regard to its indexes of ***relevance, redundancy*** and ***coverage***.

Therefore, in addition to the lack of need to define a-priori the number of classification types, this method makes sure that the selected types are the ones which were best evaluated with regard to a set of types, which cover the spread of values the entry variables adopt.

Using this method in the search of patterns, we have identified five surroundings of homes of Chilean children in vulnerable situations:

- *Precarious regarding their organization but with childrearing guidelines which are favourable for development,*
- *Favourable for development,*
- *With deficiencies in family relations and economic state, and besides with few resources dedicated to promotion and development of children,*
- *With precarious socialization outside of home, but with discipline management which is appropriate for the child's development,* and
- *With difficulties in setting limits and a family atmosphere which is potentially violent.*

Therefore, the surrounding of the home of each child is formed throughout the degree of belonging to each of these five surroundings. This information is used to establish its normative population and by doing this, we generate a measure which indicates the achievement of its development considering its own context, which is defined by the characteristics of its family environment.

That said, in addition to other areas in which the presented method of construction of contextual indexes may be applied, we are thinking about three possible paths for future research.

The first one is oriented towards defining a measure of similitude between vectors, where the coordinates represent degrees of belonging; so, this measure has to take into account that each coordinate is associated with a different concept, which means relieving the position of the data on the vector of belonging coordinates. The second one aims to compare the results obtained with the method of fuzzy unsupervised classification suggested in this study with other different methods which are usually used, and doing this in different application areas. And, finally, developing indexes for the evaluation of the results of a fuzzy unsupervised classification based on a fuzzy reference framework.

1.3.4. Publications associated.

- **Review paper:**

  Rojas, K., Gómez, D., Rodríguez, J.T., Montero, J., Valdivia, A., Paiva, F. Development of child's home environment indexes based on consistent families of aggregation operators with prioritized hierarchical information. Fuzzy Sets and Systems: Special Issue EUROFUSE 2011. (in second revision) (2013).

- **Published paper:**

  Rojas, K., Gómez, D., Rodríguez, J.T., Montero, J. Strictly stable families of aggregation operators. Fuzzy Sets and Systems.
  http://dx.doi.org/10.1016/j.fss.2012.12.010 (2013).

  Rojas, K., Gómez, D., Rodríguez, J.T., Montero, J. Stability in aggregation operators. In Greco, S. et al. (Eds.) Advances in Computational Intelligence. Springer. 299, 317-325 (2012).

  Rojas, K., Gómez, D., Rodríguez, J.T., Montero, J. Some properties of consistency in the families of aggregation operators. In B. de Baets (Ed.) et al., Eurofuse 2011, AISC. Springer. 107, 169-176 (2011).

- **Conference contributions:**

2010

Karina Rojas, Javier Montero: Medición del desarrollo infantil temprano: un enfoque contextual, *FuzzyMAD* 2010, Madrid, España.

2011

K. Rojas, D. Gómez, J. Montero, J.T. Rodríguez: Algunas propiedades de consistencia de las familias de operadores, *FuzzyMAD* 2011, Madrid, España.

K. Rojas, D. Gómez, J.T. Rodríguez, J. Montero: Some properties of consistency in the families of aggregation functions, EUROFUSE 2011, Regua, Portugal.

2012

K. Rojas, D. Gómez, J. Montero: Algunas propiedades de consistencia de las familias de operadores de agregación, *ESTYLF* 2012, Valladolid, España.

D. Gómez, K. Rojas, J. Montero, J.T. Rodríguez: Stability in aggregation operators, *IPMU* 2012, Catania, Italia.

K. Rojas, D. Gómez, J. Montero: Construcción de índices psicosociales basados en familias consistentes de op. de agregación sobre información con estructura jerárquica., *FuzzyMAD* 2012, Madrid, España.

Capítulo 2

Preliminares.

2.1. Definiciones y métodos para la construcción de índices de desarrollo infantil.

En esta sección, indagamos respecto de algunos enfoques, etapas y componentes del desarrollo del niño, así como también, las condicionantes que intervienen en el logro del desarrollo infantil dadas por las características del entorno del hogar del niño. Primero resaltamos la importancia de la temática en términos de impacto en política pública, además de la interrelación que existe entre el estudio del desarrollo infantil y del entorno del hogar del niño, para luego acercarnos a una definición de desarrollo infantil temprano, de entorno del hogar del niño y métodos de construcción de índices psicosociales.

El desarrollo humano corresponde a un fenómeno de la realidad que se manifiesta a lo largo del tiempo en un individuo en consonancia con su entorno. Todos los aspectos del desarrollo humano son afectados por su entorno y por la acumulación de experiencias desde el prenatal hasta la infancia temprana, siendo el entorno temprano del hogar por tanto, el mejor predictor de las habilidades cognitivas y no cognitivas [49].

En efecto, durante la primera infancia (entre el primer día de vida y el primer día de último ciclo de educación infantil), la calidad de la estimulación sensorial tiene un efecto directo sobre la estructura y función del cerebro, período en el cual el cerebro se desarrolla muy rápidamente. Shonkoff y Phillips en [79] lo confirman al mencionar que durante la infancia temprana, el desarrollo avanza muy rápido, siendo considerado como un período formativo, en tanto el soporte de prácticamente todos los sistemas del organismo humano, desde la más pequeña célula a la capacidad para las relaciones íntimas, son construidos durante este período de edad. Siendo esta etapa por lo tanto, la mejor en términos del impacto que pueden llegar a producir las intervenciones socio-educativas, ya que corresponde a la de mayor desarrollo de capacidades y habilidades de un sujeto, en la cual el ambiente social y educativo en que vive, influye en toda su vida escolar.

Es por esto que las políticas públicas dirigidas a la población en primera infancia son tan relevantes, y como consecuencia de esto, la importancia de los métodos de evaluación de resultados de dichas políticas. Sin embargo, se considera necesario tomar en cuenta que tales intervenciones no actúan de la misma manera en todos los individuos [83], en tanto el efecto y el sujeto interactúan produciendo distinto grado de incidencia de la intervención en función de otros factores, tales como; habilidades de entrada, variables del contexto y variables socio-demográficas.

En [49] se demuestra que las condicionantes familiares están asociadas al desarrollo del niño, como también la condición socioeconómica, sin embargo, algunos estudios muestran una considerable variabilidad entre los sujetos dentro de una misma clase [72], lo que sugiere la existencia de condicionantes familiares que son transversales a la clase socioeconómica de la familia y que inciden en el desarrollo infantil, evidenciándose la necesidad de diferenciar los entornos de los niños también respecto de la dinámica familiar en términos de conductas, actitudes y organización en torno al fomento del desarrollo infantil.

Frente a esto, parece lógico y metodológicamente correcto considerar el entorno del hogar de la población en edad temprana al momento de estudiar y comparar su nivel de desarrollo, en tanto el desarrollo óptimo de un niño requiere un ambiente que garantice la satisfacción de todas las necesidades físicas básicas y las disposiciones para el cuidado de la salud y la seguridad [10], constituyéndose entonces, en un factor diferenciador. De aquí el frecuente uso de indicadores de entorno del hogar del niño en investigación educativa, asi como también en planificación y evaluación de programas educativos. Por lo tanto, el acercamiento a una definición, medición y clasificación del entorno temprano del hogar del niño, surge como un paso necesario a la hora de estudiar el desarrollo infantil temprano.

2.1.1. Desarrollo infantil temprano.

En psicología del desarrollo se entienden como determinantes del desarrollo a un conjunto de *factores* considerados como necesarios para que este proceso se produzca. Estos factores se pueden clasificar en: *biológicos*, *sociales* y *psicológicos*. Según I. S. Kon [54], las distintas concepciones del desarrollo pueden clasificarse en función del orden de importancia de cada uno de estos factores, a saber; *concepción biologicista o biogenética*, *concepción sociologicista o sociogenética* y *concepción psicogenética o centrada en la persona*.

La concepción biologicista o biogenética releva el desarrollo de lo biológico, entendido como lo hereditario y lo congénito, y muy especialmente, como proceso de maduración del organismo y, en particular, del cerebro humano. Mientras que la concepción sociologicista o sociogenética, releva el medio como factor que determina esencialmente el desarrollo psicológico, el cual no siempre es entendido en su carácter socio-histórico, sino fundamentalmente como *ambiente* físico. Finalmente, la concepción psicogenética o enfoque centrado en la persona, sitúa en un primer plano a los factores subjetivos, es decir, a lo propiamente psicológico, en términos de las principales funciones, procesos psíquicos y contenidos personológicos, que caracterizan el desarrollo de la subjetividad humana, en sus distintas etapas.

Vygotski analiza dos principales concepciones sobre el desarrollo: "*podemos reducir todas las teorías del desarrollo infantil a dos concepciones fundamentales. Según una de ellas, el desarrollo no es más que la realización, el cambio y la combinación de las capacidades innatas. No surge nada nuevo a excepción del crecimiento, despliegue y reagrupación de los elementos dados desde el principio. Para la otra concepción, el desarrollo es un proceso continuo de auto movimiento, que se distingue, en primer lugar, por la permanente aparición y formación de lo nuevo, no existente en estadios anteriores. Ese punto de vista sabe captar en el desarrollo algo esencial para la comprensión dialéctica del proceso.*" [89]

El desarrollo ocurre según Vygotski mediante el proceso de interiorización. El niño, en su actividad conjunta con el adulto o con sus iguales, asimila los procedimientos de realización de la actividad y aquellos correspondientes para utilizar determinados medios, orientados a dirigir su propio comportamiento. En esta interacción surgen los procesos ínter-psíquicos. Luego estos procedimientos asimilados al inicio en forma externa se transforman y se convierten en procesos internos, intra psíquicos. Asi mismo, Vygotski apunta: "*...en el desarrollo del niño, lo que es posible lograr al final y como resultado del proceso de desarrollo, está presente ya en el entorno desde el principio mismo. Y no sólo está presente en el entorno desde el principio mismo, sino que influye desde el mismo principio del desarrollo del niño.* (1935)". El entorno visto desde la perspectiva vygotskiana "*... constituye la fuente de todos los rasgos específicos del niño, y si la forma ideal apropiada no está presente en el entorno, dejará de desarrollarse en el niño la actividad, la característica o el rasgo correspondiente* [90]".

Respecto de algunos hitos del desarrollo, en [79] se presentan tres ámbitos entre los muchos logros que caracterizan el desarrollo durante la infancia temprana:

- La auto-regulación, incluyendo aprender a regular las propias emociones, comportamientos y atención.
- El desarrollo temprano del lenguaje, el razonamiento y la resolución de problemas.
- Aprender a relacionarse bien con otros niños, la confianza, el amor y la resolución constructiva de conflictos.

Desde aqui se relevan tanto habilidades cognitivas como no-cognitivas, en la cual ambos tipos de habilidades son influenciadas por el ambiente familiar, siendo formadas y moldeadas en diferentes etapas del ciclo de vida [26]. Las primeras son medibles mediante pruebas de aptitudes, mientras que las segundas son más difíciles de medir, sin embargo, desempeñan un papel igualmente importante en el logro del desarrollo infantil temprano.

Un instrumento comúnmente utilizado para la medición del desarrollo infantil es el *Inventario de desarrollo Battelle*. Este inventario fue elaborado en el año 1984 por un grupo de profesionales de diversos campos, y la adaptación española se realizó el año 1996. Uno de los propósitos fundamentales con el que se creó este inventario, fue el de poder proporcionar información sobre los puntos fuertes y débiles en diversas áreas del desarrollo del niño, a fin de facilitar la elaboración de programas de intervención individualizados. Contiene una batería de ítems cuyo fin es evaluar las habilidades fundamentales del desarrollo en niños menores de 8 años basándose en el comportamiento, se aplica de manera individual y está estandarizado. Está compuesto por 341 ítems que se valoran en tres categorías ordinales divididos en 5 áreas, que a su vez se dividen en sub-áreas, tal como se muestra a continuación:

Cuadro 2.1: Componentes del desarrollo medido por el inventario Battelle.

Dimensión	Sub-dimensión
Personal/Social.	Interacción con el adulto
	Expresión de sentimientos/afecto
	Autoconcepto
	Interacción con los compañeros
	Colaboración
	Rol social
Adaptativa.	Atención
	Comida
	Vestido
	Responsabilidad personal
	Aseo
Motora.	Control muscular
	Coordinación corporal
	Locomoción
	Motricidad fina
	Motricidad perceptiva
Comunicación.	Receptiva
	Expresiva
Cognitiva.	Discriminación perceptiva
	Memoria
	Razonamiento y habilidades escolares
	Desarrollo conceptual

La recogida de la información se realiza en base a tres métodos; la observación del comportamiento del niño durante sus actividades normales en su entorno, la respuesta del niño frente a un estímulo, y la información entregada por los padres del niño. Ahora bien, como las habilidades se van desarrollando a distintas edades, se estable la edad de medición de cada habilidad de manera empírica en base a la edad en que el 75 % de los niños logran las pruebas que miden dicha habilidad.

Respecto de la puntuación, si el niño es capaz de demostrar cada habilidad de manera regular, se le da una puntuación de 2, si la habilidad es emergente, el niño recibe una puntuación de 1, y si el niño es incapaz de demostrar la habilidad, se da una puntuación de 0. Estas puntuaciones se suman para determinar la puntuación bruta de cada sub-dimensión, los que son nuevamente sumados para determinar la puntuación de cada dimensión, que finalmente sumados entregan la puntuación total. Las puntuaciones son llevados a una escala de 0 a 100, y posteriormente es posible transformarlos en las puntuaciones usualmente utilizados en psicometría, a saber; z, T, CI y ECN (el manual contiene tablas de transformación para cada sub-dimensión y edad). Las puntuaciones de cada dimensión y el puntaje total son transformados a una escala con media 50 y desviación estándar de 15, conformando los cocientes de desarrollo y el rango percentil. Por lo tanto, el cociente de desarrollo obtenido en cada dimensión representa la capacidad general del niño en el área correspondiente, y el rango percentil representa una medida general de desarrollo del niño.

La estandarización del test se realizó utilizando unas muestras representativas de 2,500 niños menores de 8 años, y la consistencia interna del test se evaluó con la fórmula Spearman-Brown [82], revelando una confiabilidad entre un 0,95 y 0,99 en la puntuación total, y una confiabilidad no menor a 0,8 en cada uno de los cocientes de desarrollo, salvo en adaptación que resultó un poco menor. Sin embargo, a nivel de sub$-$dimensión, la confiabilidad cae muy por debajo del recomendado 0,8 para algunas edades. Esto muestra que para interpretar la puntuación obtenido desde el inventario de desarrollo, se comparan los resultados con una tabla construida en base a un grupo de referencia o normativo, que corresponde al comportamiento estándar o esperado de la población. Sin embargo, estas normas no son universales, en tanto dependen de las características de la población evaluada, y más aún, si esta población corresponde a niños en primera infancia, tales características influyen directamente en sus resultados de desarrollo. Frente a esta situación, se considera adecuado construir una medida de desarrollo infantil temprano en la cual el grupo de referencia sobre el cual se comparan los resultados, considere las características del entorno familiar del niño evaluado, y de esta forma, su nivel de desarrollo represente la proporción del potencial de desarrollo logrado por el niño, más que la comparación con un niño "*estándar*".

2.1.2. Entorno del hogar del niño.

Si consideramos al desarrollo infantil como un proceso dinámico y multifactorial que se produce en consonancia con el entorno inmediato del niño, es necesario estudiar las características del entorno del hogar del niño al momento de interpretar sus logros en desarrollo. Según Heckman J. (Premio Nobel de Economía año 2000) [4], un niño está afectado por el ambiente en el que vive, ya que éste impacta desde el desarrollo neuronal del cerebro hasta su capacidad de empatía. Así, mejores ambientes familiares a edad temprana son mejores predictores de las habilidades cognitivas y no cognitivas de un ser humano.

Frente a esto, la calidad de las prácticas parentales y de manera más global, el entorno familiar en su conjunto, desempeñan un papel clave en el desarrollo de las dimensiones cognitivas y no cognitivas del ser humano en sus distintas etapas de crecimiento. Al respecto, hoy en día existe la preocupación por el estudio cualitativo de la estimulación cognitiva disponible en los entornos sociales más significativos en que se desenvuelven los niños [12]. Esto se traduce concretamente en el análisis de indicadores de la calidad de los entornos familiares y de los centros de cuidado infantil, y su incidencia en diversas dimensiones del desarrollo ([34]; [35]). Relacionado con lo anterior, existe además un interés en la comparación de la influencia entre los distintos entornos y en describir cómo las diferencias en el cuidado de los niños se relacionan con su desarrollo social emocional e intelectual ([34]; [9].

Los estudios sobre la relación entre el contexto familiar y el desarrollo infantil han utilizado diversas aproximaciones para conceptualizar y analizar la calidad de los entornos. En la actualidad, el enfoque predominante consiste en la utilización del inventario *Home obervation measurement of the enviroment (HOME)* como instrumento para la evaluación de la estimulación cognitiva disponible en el entorno familiar. El uso de escalas como el HOME [10], que combina el uso de entrevistas semi estructuradas con pautas de observación de la interacción del niño con su madre [57], supone un avance y un complemento a las medidas de desarrollo cognitivo basadas en el uso aislado de evaluaciones de coeficiente intelectual como predictor del desarrollo, y a la conceptualización del ambiente familiar únicamente en términos de sus características estructurales a través de indicadores de nivel socio económico ([85]; [3]; [71]).

La aproximación propuesta por el HOME, confiere mayor relevancia a un enfoque ecológico de los sistemas familiares y educativos [85], que considera la interdependencia y mutua influencia entre los contextos más significativos para el niño [11]. Sin embargo, algunos autores (ej. [57]) proponen la reevaluación de las sub escalas del HOME para determinar aquellos ítems que conformen nuevas escalas con significado conceptual, en este sentido se propone privilegiar la agrupación de indicadores basados en su conectividad teórica y no en el análisis factorial estadístico, técnica a través de la cual se configuraron las escalas actuales.

El entorno familiar y las pautas de crianza que organizan dicho entorno en función del cuidado y promoción del desarrollo cognitivo, físico y socio-afectivo de los niños en sus primeros años de vida resultan fundamentales. Dentro de los factores que se han estudiado de este entorno, es posible identificar dos grandes grupos, aquellos vinculados con las condiciones materiales donde ocurre la crianza del niño, y aquellos que dicen relación con las pautas de interacción para el desarrollo. Ambos grupos se presentan y observan de manera conjunta en las prácticas familiares.

A continuación se presentan principales factores del entorno familiar que se han estudiado como relevantes en el desarrollo de la primera infancia. Para efectos de esta definición se tomó como referencia inicial los inventarios Home (*Observation for measurement of the environment*), considerando de manera específica las escalas de 0 a tres años, y de 4 a 6 años [57].

- *Entorno físico*: En este factor se consideran aspectos vinculados con la calidad de la vivienda en términos de la protección, salud y seguridad del niño, así como con las pautas de organización de la rutina de crianza en relación con el uso de dicho espacio.
- *Recursos materiales de estimulación*: Diversidad y calidad de los recursos materiales que promueven el desarrollo físico, cognitivo y socioemocional. Un aspecto central es la pertinencia a la edad y nivel de desarrollo que tenga el niño, es decir, que la respuesta o reacción demandada no sea demasiado simple ni demasiado desafiante (Smith, Landry y Swank, 2000 en [80]). Se pueden distinguir: recursos lúdicos y recursos didácticos para el aprendizaje (estimulación intelectual).
- *Pautas de estimulación del lenguaje*: Comunicación verbal entre el niño y el responsable del cuidado que promueve el desarrollo del lenguaje. Este factor resulta clave en el desarrollo socio-cognitivo, y está en estrecha relación con el siguiente.
- *Pautas de estimulación intelectual*: Considera principalmente las acciones que promueven el desarrollo cognitivo y habilidades intelectuales que permiten la generación de conocimiento futuro. Por ejemplo: aprendizaje de los colores; aprendizaje

de géneros discursivos (canciones, poesías, etc.); aprendizaje de relaciones espaciales; aprendizaje de los números; conocimiento de letras y desarrollo de la conciencia fonológica.

- *Afecto y calidez en el cuidado*: Calidad de las interacciones comunicativas y afectivas promovidas por el responsable de la crianza. En la primera etapa de la infancia (de 0 a 3 años) este factor está relacionado con las prácticas de apego especialmente vinculadas con la figura materna, luego, en la etapa entre 4 y 6 años las interacciones comunicativas mediadas por el lenguaje verbal toma más relevancia y se evalúa en relación al estilo afectivo y de contención emocional que adopta. Algunos autores además, precisan la importancia que tiene la cálida recepción y aceptación de las necesidades e intereses del niño con respuestas que son inmediatas y contingentes a las señales que exhiben, se consideran como comportamiento que da cuenta de una capacidad de respuesta. De esta manera, los niños experimentan que sus necesidades serán satisfechas de una manera confiable [55]. En este factor, se tiende a identificar indicadores asociados a la interacción física, de la interacción verbal, como ocurre en el inventario del HOME.
- *Normas y disciplina*: Este factor está asociado con pautas que establecen los límites de la conducta de los niños, transmitiendo con ello los patrones culturales éticos, morales y de protección de la infancia. En este factor se pueden considerar: modelado y estimulación de la Madurez social, donde se socializa al niño en las normas y límites de comportamiento social aceptado (respeto de autoridades, expresión de sentimientos y emociones negativas de manera "*aceptable*", horarios y rutinas marcadas, etc.), y prácticas educativas frente a una falta infantil, asociada a la actitud asumida por adultos responsables en relación al castigo.
- *Experiencias de estimulación social*: En este factor se presentan las pautas de crianza que promueven la interacción de niños con un entorno social más allá del núcleo familiar, por medio de la oferta de una diversidad de actividades fuera del hogar.

2.2. Enfoques y métodos de construcción de índices psicosociales.

En la sección anterior, nos acercamos a una definición de desarrollo infantil, intentando describir muy brevemente etapas y componentes que lo configuran, así como también, las principales condicionantes que inciden en el logro del desarrollo infantil. Desde este último aspecto, se describieron los principales factores del entorno del hogar del niño que se han estudiado como relevantes en el desarrollo de la primera infancia. Ahora bien, una vez se tiene claridad respecto de los conceptos que se intentan medir, es necesario definir el procedimiento que se utilizará en la construcción de los índices de dichos conceptos. En esta sección, se presentan algunos enfoques y métodos de construcción de índices aplicados en ciencias sociales y del comportamiento.

La cuestión de los indicadores en ciencias sociales, constituye un problema fundamental en el proceso de generar explicaciones y conocimientos sobre fenómenos que se dan en el acontecer social. Expresa un problema epistemológico que se refiere tanto a la relación sujeto/objeto como a la construcción de lo concreto/abstracto, a la relación entre la reconstrucción empírica de la realidad y la teoría, que debe resolverse en el ámbito del modo en que el sujeto la piensa. Fundamentalmente porque todo conocimiento brota del sujeto [51].

El objetivo que se persigue ante la construcción de indicadores es *describir características, comportamientos o fenómenos de la realidad, a través de la evolución de una variable o el establecimiento de una relación entre variables, la que comparada con períodos anteriores, productos similares o una meta o compromiso, permita evaluar el desempeño y su evolución en el tiempo* [63], implicando por tanto, la disposición de una adecuada representación del conocimiento a través de información contenida en un conjunto de datos, así como también, un procedimiento orientado a disminuir la dimensionalidad de dicho conjunto, y el establecimiento de una referencia que sirva de estándar normativo para interpretar los resultados del indicador.

Los indicadores son una representación empírica de un objeto de estudio, y tiene que ver con la necesidad de cuantificar o medir un fenómeno [47]. En el proceso de construcción de indicadores, se definen unidades de medida en base a contenidos teóricos del objeto de estudio, siendo por tanto, una traducción que va desde la construcción teórica de un concepto que representa cierta cualidad del objeto, hacia la cuantificación del concepto a través de un indicador. Una estrategia comúnmente utilizada es dividir dicho concepto en dimensiones específicas de la representación original, y encontrar indicadores para cada una de estas dimensiones, que en conjunto constituyen por agregación, el concepto global a través de un índice. No obstante, no existe ningún límite interpretativo al proceso de especificación: toda dimensión puede dar lugar a nuevas dimensiones, pero es fácil intuir que un análisis excesivamente refinado puede conducir a la aceptación de dimensiones redundantes [56].

Una cuestión destacada en relación con los indicadores en ciencias sociales, es que no existen procedimientos definidos para su construcción, requiriéndose por tanto, una construcción add hoc a cada objeto de estudio. Por lo tanto, y dado que en materia de construcción de indicadores existe tal apertura, podemos encontrar en la literatura una diversidad enorme de clasificaciones y características de indicadores: cuantitativos y cualitativos; objetivos y subjetivos; simples y complejos; absolutos y relativos; autónomos e independientes; descriptivos y analíticos; internos y externos, etcétera [47]. No obstante, la dimensión del indicador corresponde a una clasificación transversal, distinguiendo entre los indicadores unidimensionales y multidimensionales.

Como se ha dicho anteriormente, uno de los desafíos planteados en esta memoria es el de la construcción de índices e indicadores multidimensionales, en tanto están compuestos por información contenida en más de una variable. Y centrados es este tipo de indicadores compuestos, el procedimiento más utilizado para su construcción consta de las siguientes tres etapas:

1. Selección de variables a utilizar.
2. Definición de un esquema de ponderación para cada variable.
3. Identificación de un umbral que otorga un juicio de valor a cada estado.

En relación a la segunda etapa, se han propuesto diferentes técnicas para la determinación de los pesos del conjunto de variables de entrada, algunos estudios aplican ponderaciones iguales a las variables ([86]; [69] y [68]), en otros, las variables se combinan usando ponderaciones determinadas por medio de un proceso de consultas con expertos, o también pueden ser aplicados de forma que reflejen la calidad de la información de las variables [76]. En ocasiones también se utiliza la valoración de los propios entrevistados, y en otros estudios, se desarrollan índices compuestos ponderando las variables sobre la base de sus frecuencias relativas, o basados en métodos estadísticos multivariados tales como los del *análisis de componentes principales.*

El análisis de componentes principales explica la estructura de varianzas y covarianzas de un conjunto de variables por medio de combinaciones lineales de ellas. Los objetivos perseguidos al utilizar este método estadístico son principalmente dos: reducir la dimensión del conjunto de datos, y permitir la interpretación del problema en la nueva dimensión ([36], [52], [67] y [53]). La estrategia comúnmente utilizada para aplicar este método consiste en cuantificar todas las variables, ya sea a través de juicio experto o de la técnica de escalamiento óptimo o cuantificación óptima. De esta forma, se asignan valores numéricos a los estados de las variables valoradas bajo un concepto abstracto. El procedimiento se encuentra descrito en [94], [78], [94]. Algunas aplicaciones a indicadores de condiciones de vida se encuentran en [22], [21], [23], [18] y [19]. Algunas propiedades del indicador han sido estudiadas en [18] ([20]).

Frente a este procedimiento, se evidencian al menos dos problemas, la representación numérica de los conceptos y la agregación lineal de los mismos.

En primer lugar, las variables que intentan representar conceptos abstractos suelen ser codificadas en base a una valoración lineal de sus estados, es decir, la distancia entre dos estados cualesquiera es la misma, lo que usualmente se aleja de su definición teórica, además de implicar un tratamiento numérico que no rescata la distinta naturaleza de las variables involucradas.

En segundo lugar, respecto de la agregación de la información, la principal dificultad tiene relación con el usual comportamiento no lineal de las variables asociadas a estudios relativos a las ciencias sociales y del comportamiento, en tanto dicha técnica intenta describir las características principales de un conjunto de datos multivariados, identificando nuevos ejes incorrelados, a fin de representar los datos mediante combinaciones lineales de las variables originales. De esta forma, el análisis rescata la asociación lineal entre las variables relacionadas y las cuantifica combinándolas a través de una media ponderada, en la cual sus pesos se definen en función de los valores propios de la matriz de correlación o de covarianzas.

2.3. Lógica borrosa.

2.3.1. Representación borrosa del conocimiento.

Frente al estudio del desarrollo infantil temprano, del mismo modo que en otras áreas de las ciencias sociales y del comportamiento, surge la dificultad de conseguir una adecuada representación del conocimiento que se manifiesta en la realidad. El primer paso es sistematizar los conceptos, en el que intentamos descifrar y registrar la realidad a partir de una percepción o una respuesta observable de los sujetos, su colectivo y su entorno. Pero estas respuestas no suelen corresponder a hechos concretos, aislados y fácilmente medibles, sino que a un sistema que se arma, se modifica y se desarma constantemente en función de un conjunto de factores interrelacionados que visualizamos como respuestas con un importante grado de vaguedad. La incertidumbre dada por la vaguedad que lleva consigo la percepción, la actitud o una situación particular, provoca que no nos parezca natural medir este tipo de conceptos a través de variables absolutas como las que se utilizan usualmente en matemáticas, evidenciando la existencia de un tipo de incertidumbre distinta a la que se estudia en probabilidad. En este escenario, parece ser necesario incorporar la imprecisión para representar la información más que considerarla como un factor de distorsión.

Zadeh en 1965 [95], interesado en proveer un lenguaje formal para las ciencias sociales y del comportamiento, propuso un modelo matemático que ayuda a formalizar la vaguedad y graduación de los conceptos lingüísticos a través de *conjuntos borrosos* [88]. Los conceptos que aparecen en los estudios de desarrollo infantil y de entorno familiar, son representados por *variables lingüísticas*, que se caracterizan por tomar valores del lenguaje natural. Con ellas, es posible precisar su significado por medio de conjuntos borrosos asociados, que se constituyen al recibir un valor de verdad graduado y perteneciente al intervalo unitario desde su *función de pertenencia*, representando el grado en que la etiqueta asociada a la *variable borrosa* se hace presente en la realidad del individuo a través del discurso.

Simplemente por poner un ejemplo del tipo de variables que surgen en estudios relacionados con el desarrollo infantil temprano, veamos lo que pasa con el *High-risk index* [70] utilizado para medir el nivel de riesgo psicosocial de los niños. Uno de los conceptos considerados en este índice es la *condición económica* de la familia del niño, que es medido a través de las variables *ingreso familiar*, *trabajo precario e inestable del padre* y *recepción de subsidios*. Si bien el ingreso familiar y los subsidios recibidos son variables objetivas, la calidad del trabajo del padre tiene un alto nivel de subjetividad, y por tanto, hay que hacer un esfuerzo para diferenciar un trabajo precario e inestable de uno que no lo es, siendo prácticamente imposible imaginar una distinción nítida. Es más natural ver un continuo entre ambos polos que caracterizan la calidad de un trabajo, implicando por tanto, que los conjuntos asociados a los estados de esta variable no sean excluyentes, más bien que se solapen. Las claves parecen ser: límites borrosos y un continuo entre dos polos, donde estos dos polos corresponden a los valores de verdad del paradigma tradicional, que son la base de la delimitación nítida que se suele utilizar en la definición de conceptos sociales y del comportamiento.

Un conjunto borroso A contiene observaciones acompañadas por sus respectivos estados de verdad frente a un predicado, por ejemplo, los estados frente a *trabajo precario* de la variable lingüística *tipo de trabajo.* A diferencia del enfoque tradicional, este conjunto no tiene una frontera clara, siendo la función de pertenencia la encargada de definir la transición entre pertenecer y no pertenecer a cada una de las categorías, y por consiguiente, el estado de cada observación frente a una característica, entregando de esta forma, una representación más natural del conocimiento, con límites borrosos y una transición continua entre ambos estados.

2.3.2. Conceptos fundamentales.

Definición 2.1. (Conjunto borroso). Sea X un espacio de objetos no vacío con un elemento genérico de X que denotamos por x.

Un *conjunto borroso* o clase borrosa A en X es una función μ_A de X llamada *función de pertenencia* que asocia a cada punto de X un número real en el intervalo $[0, 1]$. El valor de $\mu_A(x)$ representa el grado de pertenencia de x a A. Así, cuanto más cerca esté $\mu_A(x)$ de la unidad, más alto es el grado de pertenencia de x a A. Claramente la función de pertenencia es la extensión de la función característica clásica de un subconjunto de X. [95]

Definición 2.2. (α-corte). Sea A conjunto borroso en el universo X. Definimos *α-corte* de A y lo denotamos por A^{α} al siguiente conjunto $\{x \in X : \mu_A(x) \geq \alpha\}$.

Definición 2.3. (Conjunto vacío). Un conjunto borroso A se dice que es *vacío* si y sólo si su función de pertenencia es idénticamente 0 en X, es decir, $\mu_A(x) = 0 \ \forall x \in X$.

Definición 2.4. (Igualdad entre conjuntos). Se dice que dos conjuntos borrosos A y B son *iguales* si y sólo si $\mu_A(x) = \mu_B(x) \ \forall x \in X$.

Estas definiciones ponen de manifiesto que la teoría de conjuntos borrosos parte de la teoría clásica de conjuntos, asociando incertidumbre a un conjunto al incorporar una *función de pertenencia* a él, que está definida como un número real perteneciente al intervalo unitario. Un elemento x pertenece a un conjunto A con un grado de pertenencia $\mu_A(x)$ que puede variar entre 0 y 1, es decir, $\mu_A : U \to [0, 1]$ con U como el universo de discurso, y se representa como un conjunto de pares ordenados de x con su respectivo valor de pertenencia al conjunto como: $A = \{(x, \mu_A(x))/x \in U\}$. Por lo tanto, una variable puede ser caracterizada por diferentes *valores lingüísticos*, cada uno de los cuales representa un *conjunto borroso.*

Para la definición de las funciones de pertenencia se utilizan algunas familias de formas estándar, en tanto se asocian intuitivamente con el significado lingüístico de las etiquetas más utilizadas. Estas son la función de tipo trapezoidal, triangular, S, exponencial y de tipo π, que se describen a continuación.

- *Función trapezoidal*: requiere la definición de cuatro parámetros, y permite definir un conjunto borroso con pocos datos, además de calcular su valor de pertenencia con pocos cálculos.

$$S(u;a,b,c,d) = \begin{cases} 0 & \text{si } u < a \\ \dfrac{u-a}{b-a} & \text{si } a \leq u \leq b \\ 1 & \text{si } b \leq u \leq c \\ \dfrac{d-u}{d-c} & \text{si } c \leq u \leq d \\ 0 & \text{si } u > d \end{cases}$$

- *Función triangular*: requiere la definición de tres parámetros, y es adecuada para modelar un rango de valores estrecho en torno a b.

$$T(u;a,b,c) = \begin{cases} 0 & \text{si } u < a \\ \dfrac{u-a}{b-a} & \text{si } a \leq u \leq b \\ \dfrac{c-u}{c-b} & \text{si } b \leq u \leq c \\ 0 & \text{si } u > c \end{cases}$$

- *Función de tipo* π: requiere de dos parámetros, tiene forma de campana y es adecuada para modelar valores en torno a c, siendo posible ser definida mediante expresiones exponenciales o cuadráticas como la campana da Gauss.

$$\pi(u;a,b) = \begin{cases} S(u;c-b,c-b/2,c) & \text{si } u \leq c \\ 1 - S(u;c-b,c-b/2,c) & \text{si } u \geq c \end{cases}$$

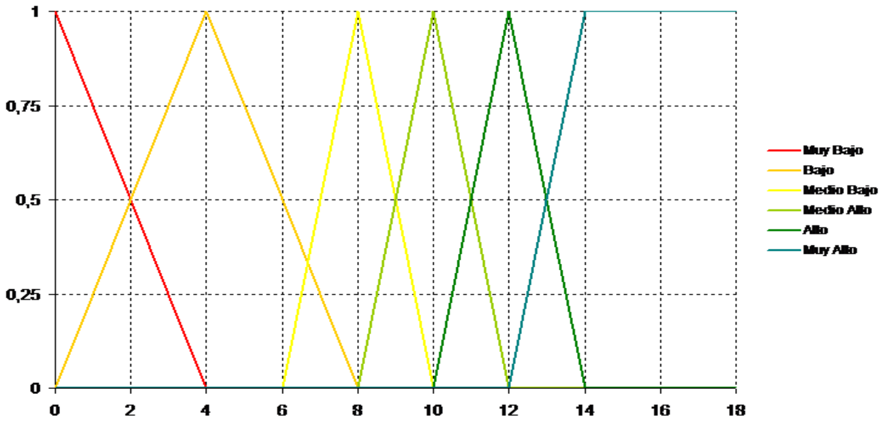
Si bien esta representación de los conceptos abstractos parece ser más adecuada frente a variables cualitativas, en muchas situaciones las variables con las que trabajamos están originalmente definidas, donde los valores numéricos han sido asignados a términos de lenguaje natural de forma lineal y nítida, con el fin de permitir el análisis estadístico posterior. Frente a este escenario, parece lógico tratar de mantener o más bien recuperar la riqueza de la información presente en los discursos cualitativos, y centrarse en una representación más adecuada de los conceptos cuando los valores numéricos han sido asignados.

En el ámbito de las ciencias sociales y del comportamiento, se suelen generar índices en base a factores o indicadores que en su conjunto representan un concepto no medible directamente en términos cuantitativos. Para su construcción, se requiere una discusión teórica adecuada a fin de proporcionar una definición conceptual del constructo, y relacionar estos conceptos con los indicadores seleccionados [62]. Una estrategia útil para establecer la coherencia entre las definiciones operativas y las conceptualizaciones teóricas de la construcción social que se está investigando, es una técnica utilizada en investigación social llamada *juicio experto* [81]. Esta técnica también puede ser utilizada en la definición de los conjuntos borrosos asociados a las variables lingüísticas.

Otro importante concepto utilizado en lógica borrosa es el de *variable lingüística*, que corresponde a la quíntupla $(x, T(x), U, G, M)$, en el cual la variable x tiene asociado un conjunto de términos o *valores lingüísticos* $T(x)$ que pertenecen al *universo de discurso* U, que está constituido por el conjunto o rango de valores que puede tomar x. Los valores lingüísticos se generan a través de la regla sintáctica G, y están asociadas a *conjuntos borrosos* determinados por la regla semántica M, que asocia un significado a cada valor. Sobre esta base, es posible representar el conocimiento mediante variables lingüísticas y cálculos entre ellas, enlazando predicados -que representan una característica definida mediante un conjunto borroso- a través de *conectores* y *modificadores.*

A modo de ejemplo, como menciona Ramey y Smith en [70] , donde define el *High-risk index*, la condición educativa familiar incide en el nivel de riesgo psicosocial del niño, favoreciéndolo cuando el entorno presenta un alto nivel educativo. Si bien el indicador que se utiliza corresponde a los años de estudio aprobados por los padres, la transición entre un año de estudio y otro no es nítida en la realidad. Frente a esto, parece ser más adecuado utilizar conjuntos borrosos asociados a los predicados *Muy alto, Alto,* *Medio alto, Medio bajo, Bajo, Muy bajo*, que actuarán sobre el universo de discurso representado en x, cuya valoración se determina mediante la función de pertenencia $P_k : x \to [0,1]$ definida para $k \in \{$Muy alto, Alto, Medio alto, Medio bajo, Bajo, Muy bajo$\}$ como lo muestra la siguiente gráfica:

Figura 2.1: Nivel educativo de los padres.



Ahora bien, para el tratamiento de problemas en ambiente borroso es imprescindible el uso de operaciones básicas como por ejemplo la negación, la unión y la intersección. Sin embargo, en contexto borroso pueden existir diversas funciones para cada uno de estos tipos de operadores, existiendo apertura frente a la definición de *operadores borrosos*. Es por esto que es más común referirse a familias de operadores, siendo el tipo más común el dado por las *t-normas*, *t-conormas* y los *operadores de agregación*.

Zadeh, sugirió la utilización de $Min(\varphi, \psi)$ y el $Max(\varphi, \psi)$ como extensión natural de las operaciones intersección y unión en conjuntos nítidos respectivamente. Estos constituyen un caso particular de las *t-normas* y *t-conormas*, y juegan un papel esencial en los procesos de agregación de conjuntos borrosos. A continuación se presentan las conjunciones y disyunciones que son la base de la definición de la intersección y unión de conjuntos borrosos.

Definición 2.5. (T-norma). Una función $T : [0,1]^2 \rightarrow [0,1]$ que verifica:

- $T(1,x) = x \ \forall x \in [0,1]$
- $T(x,y) = T(y,x) \ \forall x, y \in [0,1]$
- $T(x,T(y,z)) = T(T(x,y),z) \forall x, y, z \in [0,1]$
- $T(x,y) \leq T(u,v) \ si \ 0 \leq x \leq u \leq 1, \ 0 \leq y \leq\leq v \leq 1$

Se denomina norma triangular o *t-norma*

Definición 2.6. (T-conorma). Una función $S : [0,1]^2 \rightarrow [0,1]$ que verifica:

- $S(0,x) = x \ \forall x \in [0,1]$
- $S(x,y) = T(y,x) \ \forall x, y \in [0,1]$
- $S(x,S(y,z)) = S(S(x,y),z) \forall x, y, z \in [0,1]$
- $S(x,y) \leq S(u,v) \ si \ 0 \leq x \leq u \leq 1, \ 0 \leq y \leq\leq v \leq 1$

Se denomina conorma triangular o *t-conorma*

A continuación se presentan algunos ejemplos de t-normas y t-conormas.

1. Zadeh.

   $T_1(x,y) = Min\{x,y\}$

   $S_1(x,y) = Max\{x,y\}$

2. Probabilística.

   $T_2(x,y) = x \cdot y$

   $S_2(x,y) = x + y - x \cdot y$

3. Lukasiewicz.

   $T_3(x, y) = Max\{+y - 1, 0\}$

   $S_3(x, y) = Min\{x + y, 1\}$

2.4. Proceso de agregación.

En la sección anterior, se presentó un enfoque que ayuda a representar el conocimiento de una manera más flexible, y por tanto, acorde a una realidad psicosocial. En esta sección, nos referiremos a la construcción de índices en base a información con incertidumbre, enfocado como un proceso de agregación.

La agregación de la información contenida en n variables para ser transformada en un valor escalar, define un *proceso de agregación*. El objetivo que se persigue ante un problema de agregación es simplificar la información a través de la reducción de dimensionalidad del conjunto de datos de entrada, y de esta forma, facilitar el análisis y la toma de decisiones. De aquí la importancia de la robustez del proceso de agregación en términos de los resultados obtenidos.

Un proceso de agregación está constituido por un conjunto de datos de entrada organizado en $n - tuplas$ $(x_1, x_2, ..., x_n)$, por la definición de funciones $A_n, n \in \mathbb{N}$ que agregarán la información, y por datos de salida dados por $A_n(x_1, x_2, ..., x_n)$ para cada $n \in \mathbb{N}$. De aquí se deduce que tanto las características del conjunto de datos de entrada, como la definición de los operadores que generará la información agregada para cada n, inciden individual y conjuntamente en el proceso de agregación en términos de la robustez de los resultados obtenidos, y por tanto, en la consistencia del proceso.

A continuación se presentan algunas definiciones y propiedades asociadas a un proceso de agregación que serán utilizadas en este trabajo.

2.4.1. Operadores de agregación.

Definición 2.7. Un operador de agregación corresponde a una función real definida en el intervalo unitario, cuyas $n \geq 2$ variables del conjunto de datos de entrada dan lugar a un valor escalar.

$$A_n : [0,1]^n \rightarrow [0,1], \;\; n \in \mathbb{N} \tag{2.1}$$

(2.1) es no decreciente en cada una de sus variables, es decir, si $x < y$ entonces $A_n(x) < A_n(y) \; \forall x, y \in [0,1]^n$, y satisface las siguientes condiciones de borde:

$$A_n(0, ..., 0) = 0 \;\; y \;\; A_n(1, ..., 1) = 1 \tag{2.2}$$

A continuación se presentan las propiedades esenciales que han sido estudiadas en relación con los operadores de agregación.

- **Monotonicidad estricta.** Un operador de agregación $A_n : [0,1]^n \rightarrow [0,1]$ diremos que es monótono si:

$$x_1 \leq y_1, ..., x_n \leq y_n \;\; \rightarrow \;\; A_n(x_1, ..., x_n) \leq A_n(y_1, ..., y_n) \tag{2.3}$$

- **Idempotencia.** Un operador de agregación $A_n : [0,1]^n \to [0,1]$ diremos que es idempotente si dado $x \in [0,1]$:

$$A_n(x,...,x) = x \tag{2.4}$$

- **Elemento neutro.** Sea $A_n : [0,1]^n \to [0,1]$ un operador de agregación, diremos que $e \in [0,1]$ es un elemento neutro de A_n si para cada $t \in [0,1]$ en cualquier posición, se tiene que:

$$A_n(e,e,...,e,t,e,...e,e) = t \tag{2.5}$$

- **Elemento absorbente.** Sea $A_n : [0,1]^n \to [0,1]$ un operador de agregación, diremos que $a \in [0,1]$ es un aniquilador de A_n si $\forall x_1, x_2, ..., x_n \in [0,1]$ con $x_i = a$ para algún $i \in \{1,2,...,n\}$, se tiene que:

$$A_n(x_1,x_2,...,x_n) = a \tag{2.6}$$

- **Invarianza.** Un operador de agrgación $A_n : [0,1]^n \to [0,1]$ diremos que satisface la propiedad de invarianza ante transformaciones si $\forall \lambda \in [-1,1]$ y $\forall (x_1,x_2,..,x_n) \in [0,1]^n$ se cumple que:

$$A_n(x_1+\lambda,..,x_n+\lambda) = A_n(x_1,x_2,..,x_n)+\lambda, \tag{2.7}$$

siempre que $(x_1+\lambda,..,x_n+\lambda) \in [0,1]^n$ *y* $A_n(x_1,..,x_n)+\lambda \in [0,1]$.

- **Homogeneidad.** Un operador de agrgación $A_n : [0,1]^n \to [0,1]$ diremos que es homogéneo si $\forall \lambda \in [0,1]$ y $\forall (x_1,x_2,..,x_n) \in [0,1]^n$ se cumple que:

$$A_n(\lambda x_1,..,\lambda x_n) = \lambda A_n(x_1,x_2,..,x_n) \tag{2.8}$$

- **Simetría**. Un operador de agregación $A_n : [0,1]^n \rightarrow [0,1]$ diremos que es simétrico si $\forall x_1, x_2, ..., x_n \ \in [0,1]$:

  $$A_n(x_1, x_2, ..., x_n) = A_n(x_{p(1)}, x_{p(2)}, ..., x_{p(n)}) \tag{2.9}$$

  para toda permutación $p = (p(1), p(2), ..., p(n))$ de $(1, 2, ..., n)$ con $n \geq 2$.

Para finalizar esta sección, nos gustaría hacer hincapié en la importancia que va a tener en esta memoria y en general en los procesos de agregación el concepto de simetría de un operador de agregación. El concepto de ***simetría*** permite diferenciar entre aquellos operadores de agregación que trabajan con conjuntos de datos estructurados o sin estructurar, cuestión muy importante a la hora de definir un proceso consistente de agregación. En nuestra opinión, no todos los operadores de agregación son adecuados para un conjunto de datos determinado. Por ejemplo, si el conjunto de datos de entrada tiene una estructura inherente, los operadores que se utilicen en la agregación de la información deben estar acorde a esta estructura, y por tanto parecería razonable pensar que el operador de agregación fuese no simétrico ya que la importancia relativa de sus ítems incide en el resultado del proceso de agregación. Por el contrario, si lo datos no tienen una estructura determinada, la posición de los ítem que deben agregarse carece de importancia y para esta situación sería más razonable pensar en operadores de agregación simétricos.

A continuación se presentan algunos operadores de agregación comúnmente utilizados.

Definición nominal	Definición operativa
Mínimo	$Min_n(x_1, ..., x_n) = Min\{x_1, ..., x_n\} = x_{(1)}$
Máximo	$Max_n(x_1, ..., x_n) = Max\{x_1, ..., x_n\} = x_{(n)}$
Mediana	$Md_n(x_1, ..., x_n) = \begin{cases} x_{(n/2)} & \text{si n es par} \\ x_{((n+1)/2)} & \text{si n es impar} \end{cases}$
Media aritmética	$M_n(x_1, ..., x_n) = \sum_{i=1}^{n} \frac{x_i}{n}$
Media geométrica	$G_n(x_1, ..., x_n) = \left(\prod_{i=1}^{n} x_i \right)^{1/n}$
Media armónica	$H_n(x_1, ..., x_n) = \frac{n}{\sum_{i=1}^{n} 1/x_i}$
Productorio	$P_n(x_1, ..., x_n) = \prod_{i=1}^{n} x_i$
Op.binario inductivo hacia adelante	$A_n^f(x_1, ..., x_n) = A_2(..(A_2(A_2(x_1, x_2)..), x_{n-1})$
Op.binario inductivo hacia atrás	$A_n^b(x_1, ..., x_n) = A_2(A_1, A_2(..., A_2(x_{n-2}, x_{n-1})))$
Media ponderada	$W_n(x_1, ..., x_n) = \sum_{i=1}^{n} x_i \cdot w_i$
OWA	$O_n(x_1, ..., x_n) = \sum_{i=1}^{n} x_{(i)} \cdot w_i$

Ahora bien, cuando a lo largo de un proceso de agregación se producen cambios de cardinalidad en el conjunto de datos de entrada, ya sea por la incorporación de nuevas variables, o por la existencia de una o más variables sin información, dicho proceso ya no involucra sólo un operador de agregación A_n, sino que a una colección de operadores. En este contexto, si el número de variables del conjunto de datos de entrada es desconocido, se considera A_n $\forall n$, lo que se traduce en una *familia de operadores de agregación*.

2.4.2. Familia de operadores de agregación.

Definición 2.8. [15] Dado un operador de agregación A_n *con* $n \in \mathbb{N}$, la *familia de operadores de agregación* asociada está constituida por el conjunto de operadores $A_2, A_3, ...A_n$ $\forall n \in \mathbb{N}$ representado por:

$$\{A_n\}_n : \bigcup_{n \in \mathbb{N}} [0,1]^n \rightarrow [0,1] \tag{2.10}$$

(2.10) también es llamada *función de agregación extendida (EAF)* por otros autores [45, 16].

Si bien esta definición no condiciona la existencia de una relación entre los operadores de agregación que la componen, la utilización de algunos tipos de operadores acerca a la idea de una familia como un todo, en tanto tales operadores se encuentran relacionados y utilizan una fórmula genérica a lo largo del proceso de agregación [5], como es el caso de los operadores recursivos. El problema de agregación considerado como un proceso se hace más claro cuando se utiliza este tipo de operadores, en tanto la idea que está a la base es una regla de agregación que se basa en una aplicación iterativa de operadores binarios a partir de las agregaciones previas. Por lo tanto, dada la relación que existe entre tales operadores de agregación para cada n, estos definen una agregación con coherencia a lo largo del proceso, que además debe ser acorde con las características de los datos de entrada en términos de orden y existencia o no de prioridades.

Formalmente, la recursividad en operadores de agregación fue introducida en [28] en el contexto de operadores OWA. Posteriormente y siguiendo la línea desarrollada en [28], en [2, 29, 40, 13] se estudia el concepto de recursividad y derivados en un contexto más general que permiten establecer ciertas nociones de consistencia en operadores de agregación. A continuación se introduce la noción de recursividad que requiere en primer lugar de la definición de una *regla de orden*.

Definición 2.9. (ver [2, 29, 40] para más detalles).

Denotemos por $\pi_n(x_1, \ldots, x_n) = (x_{\pi_n(1)}, \ldots, x_{\pi_n(n)})$. Una *regla de orden* π es una familia consistente de permutaciones $\{\pi_n\}_{n \geq 2}$ tal que such para cualquier colección finita de numeros, cada extra item x_{n+1} es asignado manteniendo la posición relativa previa de sus items, es decir

$\pi_{n+1}(x_1, \ldots, x_n, x_{n+1})$ es igual a $(x_{\pi_n(1)}, \ldots, x_{\pi_n(j-1)}, x_{n+1}, x_{\pi_n(j)}, \ldots x_{\pi_n(n)})$ para algún $j \in \{1, 2, \ldots, n+1\}$.

Definición 2.10. (ver [2, 29, 40] para más detalles). Una familia de operadores de agregación $\{A_n : [0,1]^n \longrightarrow [0,1]\}_{n>1}$ es recursiva por la izquierda si existe una familia de operadores de agregación binario $\{L_n : [0,1]^2 \longrightarrow [0,1]\}_{n>1}$ tal que $A_2(x_1, x_2) = L_2(x_{\pi(1)}; x_{\pi(2)})$ y

$A_n(x_1, x_2, \ldots, x_n)$ es igual a $L_n(A_{n-1}(x_{\pi(1)}, \ldots, x_{\pi(n-1)}; x_{\pi(n)})$ para todo $n \geq 2$,

donde π es una regla de orden.

De manera similar, es posible definir las reglas recursivas por la derecha de la siguiente manera.

Definición 2.11. (ver [2, 29, 40] para más detalles). Una familia de operadores de agregación $\{A_n : [0,1]^n \longrightarrow [0,1]\}_{n>1}$ es recursiva por la derecha si existe una familia de operadores de agregación binario $\{R_n : [0,1]^2 \longrightarrow [0,1]\}_{n>1}$ tal que $A_2(x_1, x_2) = R_2(x_{\pi(1)}; x_{\pi(2)})$ y

$A_n(x_1, x_2, \ldots, x_n)$ es igual a $R_n(x_1, A_{n-1}(x_{\pi(2)}, \ldots, x_{\pi(n-1)})$ para todo $n \geq 2$,

donde π es una regla de orden.

De las definiciones previas, es posible introducir el concepto de LR recursividad de la siguiente manera.

Definición 2.12 (ver [2, 29, 40] para más detalles)**.** . Una familia de operadores de agregación $\{A_n : [0,1]^n \longrightarrow [0,1]\}_{n>1}$ es recursiva por la derecha y por la izquierda simultaneamente si existen dos familias de operadores de agregación binario $\{R_n : [0,1]^2 \longrightarrow [0,1]\}_{n>1}$ and $\{L_n : [0,1]^2 \longrightarrow [0,1]\}_{n>1}$ tales que se satisfacen las condiciones de recursividad por la derecha y por la izquierda simultaneamente.

Un caso particular de las definiciones anteriores se da cuando el operador de agregación binario es siempre el mismo y la recursividad es desde la izquierda. Esta definición de recursividad puede verse en [15, 16], en la cual se dice que una familia $\{A_n\}_{n\in N}$ es recursiva cuando se satisface lo siguiente:

$$A_n(x_1, x_2, .., x_n) = A_2(A_{n-1}(x_1, x_2, .., x_{n-1}), x_n).$$

Las propiedades presentadas en la sección anterior, corresponden a características asociadas a un operador de agregación A_n para un n en particular. Si bien también se han estudiado dichas propiedades para una FAO, dado que la definición (2.10) no exige ninguna restricción que relacione los operadores de agregación que componen una misma familia, el cumplimiento de dichas propiedades por parte de $\{A_n\}_{n\in\mathbb{N}}$ se traduce en que cada operador que la constituye las satisfaga, independientemente si los $A_2, A_3, ..., A_n$ se encuentran o no relacionados. Esta extensión en el cumplimiento de propiedades por parte de una FAO se muestra en [5, 45].

A continuación se presentan otras propiedades que han sido estudiadas en relación a las familias de operadores de agregación, en las que se establecen condiciones entre los diferentes miembros de la familia.

- **Asociatividad.** Una familia de operadores de agregación $\{A_n\}_{n\in\mathbb{N}}$ diremos que es asociativa si $\forall n, m \in \mathbb{N}$ y $\forall x_1, x_2, ..., x_n, y_1, y_2, ..., y_m \ \in [0,1]$:

$$A_{n+m}(x_1, ..., x_n, y_1, ..., y_m) = A_{n+m}(A_n(x_1, ..., x_n), A_m(y_1, ..., y_m)) \tag{2.11}$$

- **Descompensación.** Una familia de operadores de agregación $\{A_n\}_{n\in\mathbb{N}}$ diremos que cumple la propiedad de descompensación si $\forall n, m = 1, 2, ...,$ $\forall x \in [0,1]^m$ y $\forall y \in [0,1]^n$ se cumple que:

$$A_{m+n}(x_1, .., x_m, y_1, .., y_n) \quad = \quad A_{n+m}(\underbrace{A_m(x_1, .., x_m), ..., A_m(x_1, .., x_m)}_{m\ veces}, y_1, .., y_n) \tag{2.12}$$

- **Bisimetría.** Una familia de operadores de agregación $\{A_n\}_{n\in\mathbb{N}}$ diremos que es bisimétrica si $\forall n, m = 1, 2, ...$ y $\forall x \in [0,1]^{mn}$ se cumple que:

$$\begin{aligned} A_{mn}(x_1, .., x_{mn}) &= A_m(A_n(x_{11}, .., x_{1n}), ..., A_n(x_{m1}, .., x_{mn})) \\ &= A_n(A_m(x_{11}, .., x_{m1}), ..., A_m(x_{1n}, .., x_{mn})) \end{aligned} \tag{2.13}$$

- **Self-identity.** Una familia de operadores de agregación $\{A_n\}_{n\in\mathbb{N}}$ diremos que satisface la propiedad self-identity si $\forall n \in \mathbb{N}$ y $\forall x_1, x_2, ..., x_n \ \in [0,1]$:

$$A_n(x_1, ..., x_n) = A_{n+1}(x_1, ..., x_n, A_n(x_1, ..., x_n)) \tag{2.14}$$

Algunas propiedades han sido estudiadas y relacionadas a fin de acercarse a la generación de resultados agregados consistentes. En [93] se releva la propiedad asociativa en tanto entrega consistencia al proceso de agregación. No obstante, tal consistencia también puede ser provista si una FAO no es asociativa, pero cumple las propiedades de idempotencia, monoticidad y self-identity. Siguiendo en esta línea, en [5] se presentan las propiedades de continuidad y estabilidad para una familia de operadores de agregación.

- **Continuidad.** Una familia de operadores de agregación $\{A_n\}_{n\in\mathbb{N}}$ diremos que es continua si $\forall n \in \mathbb{N}$ y $\forall X \ e \ Y \ \in [0,1]^n$, con $X \leq Y$ y $n \ \geq 2$:

$$\exists Z \in [0,1]^n \ con \ X \leq Z \leq Y : \ A_n(Z) = c, \tag{2.15}$$

  donde $c \in [A_n(x_1, ..., x_n), A_n(y_1, ..., y_n)]$

- **Continuidad Lipschitziana.** Una familia de operadores de agregación $\{A_n\}_{n\in\mathbb{N}}$ diremos que es lipschitz continua si $\forall n \in \mathbb{N}$ y $\forall X \ e \ Y \ \in [0,1]^n$, con $n \ \geq 2$:

$$\exists M > 0 : \ |[A_n(x_1, ..., x_n) - A_n(y_1, ..., y_n)]| \leq M \ d(X, Y), \tag{2.16}$$

  donde $d(X, Y)$ es una distancia entre $X \ e \ Y$

2.4.3. Consistencia y estabilidad en procesos de agregación.

La información agregada es el resultado del proceso de agregación, y por tanto, a través de ella es posible evaluar la robustez del proceso respecto de la *estabilidad* de sus resultados ante cambios de cardinalidad del conjunto de datos de entrada.

Como se señala en [15], en términos generales, la estabilidad de un modelo matemático para problemas aplicados se refiere a "*pequeños errores deentrada*" no se traducen en "*grandes errores de salida*". La propiedad de estabilidad para una función de agregación A_n ha sido definida de una manera similar a la condición Lipschitz, en el cual pequeños cambios en el vector x no pueden producir grandes cambios en $A_n(x)$.

La estabilidad corresponde a un concepto que ha sido estudiado en otros trabajos referidos a operadores de agregación. Por ejemplo, en [5] se propone una definición de *p-stability* para un operador de agregación A_n, extendiendo la definición de estabilidad a una familia de operadores de agregación, siendo considerada *p-stable* si todos los operadores que la definen son *p-stable* , es decir, si $\forall\ n \in \mathbb{N}$ la función $n - ary$ satisface la propiedad, del mismo modo, como se extiende un conjunto de propiedades presentadas en [45].

Definición 2.13. [15] Sea $\{A_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ una familia de operadores de agregación.

- Diremos que $\{A_n\}_{n\in\mathbb{N}}$ es n-p-stable si $\forall p \in [1,\infty)$ y $\forall x, y \in [0,1]^n$ *con* $x \neq y$:

$$\|A_n\|_p = sup\left(\frac{|A_n(x) - A_n(y)|}{\|x - y\|_p}\right) \leq 1 \tag{2.17}$$

- Diremos que $\{A_n\}_{n \in \mathbb{N}}$ es p-stable si $\forall p \in [1, \infty)$:

$$\|\{A_n\}_{n\in\mathbb{N}}\|_p \quad = \quad sup\,(\|A_n\|_p, n \in \mathbb{N}) = 1, \tag{2.18}$$

  i.e. si $\{A_n\}_{n\in\mathbb{N}}$ $es\ n - p - stable$ $\forall n \in \mathbb{N}$

Bajo esta definición de estabilidad para una familia de operadores de agregación, es posible definir una FAO p-stable usando por ejemplo el mínimo si n es par y el máximo si n es impar, en tanto en el punto (2.18) de la definición sólo se requiere que A_n sea p-stable $\forall n \in \mathbb{N}$, y en este ejemplo, ambos operadores son n-p-stable. Sin embargo, es evidente que el proceso de agregación asociado a esta familia no será robusto, no siendo posible garantizar un comportamiento estable en términos de los resultados del proceso cuando se producen cambios de cardinalidad. Por lo tanto, esta definición de estabilidad no asegura la existencia de un concepto unificador entre los miembros de una FAO que se traduzca en algún nivel de estabilidad de los resultados.

Esta idea de estabilidad también ha sido rescatada para datos ordenados de izquierda a derecha bajo la propiedad de *self-identity* definida por Yager en [93], en tanto asegura que el resultado de agregar n ítems coincide con el de $n + 1$ ítems cuando el último elemento incorporado corresponde a la agregación de los anteriores.

Definición 2.14. [93] Sea $\{A_n : [0,1]^n \rightarrow [0,1], n \in \mathbb{N}\}$ una familia de operadores de agregación. Diremos que la familia $\{A_n\}_{n\in\mathbb{N}}$ satisface la propiedad self-identity si $\forall n \in \mathbb{N}$ y $\forall x_1, \ldots x_n \in [0,1]$, se cumple:

$$A_n(x_1, x_2, ...x_{n-1}, A_{n-1}(x_1, x_2..., x_{n-1})) = A_{n-1}(x_1, x_2, ..., x_{n-1}) \tag{2.19}$$

Es importante notar que si la familia A_n no es simétrica, es decir, si existe un n ante el cual el operador de agregación A_n no es simétrico, entonces la posición del nuevo elemento es relevante en el resultado final del proceso de agregación, y entonces esta definición no es aplicable. Por ejemplo, si analizamos la propiedad self-identity en las familias de operadores de agregación compuesta por operadores binarios con extensión inductiva hacia atrás $\{A_n^b, n \in \mathbb{N}\}$, y con extensión inductiva hacia adelante $\{A_n^f, n \in \mathbb{N}\}$ [15] definidos para $n > 2$ como:

$$A_n^b = L_2\left(x_1, L_2(\ldots, L_2(x_{n-1}, x_n)\ldots)\right.$$
$$A_n^f = L_2\left(\ldots, (L_2(L_2(x_1, x_2), x_3)), \ldots, x_n\right)$$

donde L_2 es un operador de agregación binario.

Se observa que $\{A_n^f, n \in \mathbb{N}\}$ satisface la propiedad self-identity si L_2 es idempotente, mientras que $\{A_n^b, n \in \mathbb{N}\}$ no la satisface, en tanto el orden en la cual se agrega la información es inverso (de derecha a izquierda).

$$\begin{aligned} A_n^f(x_1, ..., x_{n-1}, A_{n-1}^f(x_1, ..., x_{n-1})) &= A_2(..(A_2(A_2(x_1, x_2)..), x_{n-1}), A_{n-1}^f) \\ &= A_2(A_{n-1}^f, A_{n-1}^f) \\ &= A_{n-1}^f(x_1, ..., x_{n-1}) \end{aligned}$$

A partir de esta observación, a fin de permitir la aplicación de esta propiedad en operadores no simétricos, además de la distinción entre distintos niveles de estabilidad, en el capítulo siguiente se propone una definición de estabilidad que extiende la noción de self-identity para conjuntos de datos ordenados linealmente en ambos sentidos y para datos estructurados lineal o jerárquicamente.

2.5. Sistemas de clasificación borrosa no supervisada.

En el contexto de construcción de índices de entorno del hogar del niño enfocado desde un proceso de agregación, en la sección anterior se presentaron las definiciones de operadores de agregación y de familia de operadores de agregación, además de algunas de sus propiedades, profundizando en aquellas asociadas a la consistencia de un proceso, con el fin de asegurar robustez de sus resultados. Una vez ha sido agregada la información relativa al entorno del hogar del niño mediante un proceso de agregación determinado, es necesario identificar una clasificación borrosa del conjunto de datos en función de dichos índices de entorno. De esta manera, las clases borrosas de la partición podrán ser utilizadas como población normativa de los niños que presentan un nivel de pertenencia significativo a ellas, insumo necesario para la construcción de índices contextuales de desarrollo infantil temprano.

Frente a un problema de clasificación, el supuesto que está a la base es la existencia de clases que conforman una partición de un conjunto de datos determinado. El propósito que se persigue es estructurar la información que puede extraerse desde este conjunto de datos, a fin de modelar la pertenencia de los elementos a tales clases. El resultado es básicamente un conjunto de criterios que debe satisfacer un elemento para pertenecer a una u otra clase, para el cual se requiere definir medidas de similitud y pertenencia a dichas clases, asi como también indicadores que aseguren un nivel de calidad del sistema de clasificación.

Si las clases se encuentran previamente definidas en el conjunto de datos, entonces se trata de una clasificación supervisada. En tal caso el objetivo es de tipo predictivo, en el cual se estudian las características asociadas a pertenecer a cada clase, a fin de estimar la pertenencia de observaciones externas al conjunto de datos. Mientras que si el conjunto de clases se encuentra previamente definido en base a estudios anteriores o supuestos teóricos, no siendo parte del conjunto de datos, se trata de una clasificación no supervisada. En este caso el objetivo es de tipo descriptivo, en el cual se intenta organizar el conjunto de datos en grupos basado en similaridades entre sus elementos.

En este trabajo nos centramos en sistemas de clasificación no supervisada. El método más utilizado bajo un problema de este tipo corresponde al *análisis de conglomerado*, para el cual existen diversos algoritmos que buscan patrones de comportamiento que están a la base del conjunto de datos. En [84], se muestra una clasificación de algoritmos de conglomerados respecto de tres criterios de agrupación; *dirección*, *pertenencia* y *estructura*. La dirección del proceso se refiere a cómo se generan las clases, de un modo aglomerativo (de abajo hacia arriba) o de un modo divisivo (de arriba hacia abajo), es decir, las clases se definen en base a la adición de elementos, o en base a particiones del conjunto de datos. La pertenencia se refiere a la medida que se utiliza para indicar la pertenencia de los elementos a las clases, pudiendo ser una medida nítida, borrosa o probabilística, es decir, de tipo booleana, en base a grados de pertenencia o asociada a una distribución de probabilidad. Y por último, la estructura de la partición, es decir, las clases de una partición son independientes (sin estructura), en el cual el usuario define *a priori* la cantidad de particiones que se van a realizar sobre el conjunto de datos, o las posibles particiones del conjunto de datos definen una estructura jerárquica o de otro tipo. A esta clasificación de algoritmos de conglomerados le incorporamos un cuarto criterio, el *tipo de partición*, que puede ser determinista o borrosa, es decir, las clases resultantes son excluyentes o admiten un nivel de solapamiento entre ellas.

La noción de partición borrosa fue introducida por Ruspini en [77], suponiendo que dada una familia discreta de clases C , se asume que para cada elemento $x \in X$ se verifica que:

$$\sum_{c \in C} \mu_c(x) = 1$$

Por tanto, cada elemento tiene un grado de pertenencia a cada clase, donde el total de la pertenencia se distribuye entre todas las clases, generalizando así el concepto de partición nítida, donde la función de pertenencia $\mu_c(x) \in \{0,1\}\ \forall x, c$. De ésta manera, cada elemento estará en una y solo una clase, por tanto $\forall x \in X, \exists c \in C : \mu_c(x) = 1, \mu_k(x) = 0\ \forall k \neq c$. Esta generalización se considera deseable en muchas situaciones, sin embargo, en la práctica, para una modelización borrosa sigue siendo muy restrictivo [1].

Considerando el objetivo de investigación y la naturaleza borrosa de la información analizada, en este trabajo nos centramos en algoritmos en la cual la medida de pertenencia de los elementos a las clases es gradual y la partición es borrosa y estructurada. A continuación, se profundiza brevemente respecto de conceptos básicos, tipos de algoritmos y medidas de calidad de un sistema de clasificación no supervisada en contexto borroso.

2.5.1. Análisis de conglomerados en contexto borroso.

En términos generales, el objetivo del análisis de conglomerado es generar una partición de un conjunto de datos determinado, constituyendo clases, grupos o subconjuntos, en la cual se busca que los elementos con mayor grado de pertenencia a una cada clase, tengan un mayor grado de similitud, al mismo tiempo que dos elementos con alto grado de pertenencia a clases distintas, tengan un menor grado de similitud. Desde aquí se evidencia la necesidad de definir una medida de similaridad que determine el nivel de cercanía o lejanía entre vectores que contienen la información del conjunto de datos, medida que debe considerar el tipo y significado de los valores de cada variable de dicho conjunto.

Un conjunto de datos puede contener variables real valoradas o variables en el cual su valoración representa una clase abstracta. Además, puede revelar distintos tipos de agrupaciones, de diferentes dimensiones, formas y densidad, donde las clases pueden estar completamente desconectadas, o con algún nivel de solapamiento. Toda esta variedad de posibilidades depende del objetivo de investigación y de las características del conjunto de datos de entrada, frente a las cuales se han desarrollado variadas técnicas de análisis de conglomerado.

Otra clasificación de algoritmos de conglomerado es propuesta en [50], a partir de las siguientes familias no disjuntas:

- *Técnicas heurísticas o incompletas.*

  Referidas a aquellas en la cual se utilizan técnicas de reducción de dimensionalidad del conjunto de datos, a fin de obtener una representación gráfica de a lo más tres dimensiones.

- *Técnicas nítidas deterministas.*

  Técnicas que generan clases excluyentes, por tanto la intersección de las clases que componen la partición del conjunto de datos es vacía.

- *Técnicas nítidas solapadas.*

  Técnicas que generan clases solapadas o no excluyentes, en tanto cada elemento puede pertenecer a más de una clase.

- *Técnicas probabilísticas.*

  Referidas a aquellas técnicas en la cual se determina una probabilidad de pertenencia de los elementos a cada clase.

- *Técnicas posibilistas.*

  Aquellas que definen un grado de pertenencia de los objetos a las clases.

- *Técnicas jerárquicas.*

  Técnicas que en base a reglas de combinación y división de clases, van generando paso a paso particiones del conjunto de datos, en la cual las clases tienen cada vez menos elementos y por tanto la partición cada vez más clases.

- *Técnicas basadas en función objetivo.*

  Referidas a aquellas que en base a una función objetivo o evaluación, se asigna a cada partición una calidad o un error que debe ser optimizado, siendo la partición mejor evaluada la solución ideal.

- *Técnicas de estimación.*

  Técnicas que utilizan algoritmos de optimización en el cual la generación de la partición y la estimación de los parámetros se realiza en base a ecuaciones heurísticas.

Si bien esta clasificación mezcla los criterios de dirección, pertenencia, estructura y tipo de partición definidos anteriormente, incorpora otro aspecto relacionado con la técnica que está a la base del proceso, ya sea de reducción de dimensionalidad o de optimización de criterios de calidad o del error. Frente a esta clasificación de técnicas de análisis de conglomerados, es posible centrar este trabajo en técnicas posibilistas en base a función objetivo o evaluación, en tanto una partición del conjunto de datos en base a la evaluación del sistema de clasificación mediante indicadores de calidad, en el cual la pertenencia a las clases es gradual, parece ser más coherente con la naturaleza borrosa de la información y con el contexto psicosocial de la investigación.

Focalizados en técnicas borrosas o posibilistas según [50], a continuación profundizamos en algunas definiciones y algoritmos.

La información de entrada es representada por un conjunto finito de puntos $X = \{x_1, x_2, ..., x_n\} \subseteq \mathbb{R}^s$, que corresponden a s características de cada uno de los $n \geq 2$ elementos del conjunto de datos. El objetivo es particionar este conjunto en $c \geq 2$ clases de elementos con similares características, las que son representadas por los patrones o centros de clases $V = \{v_1, v_2, ..., v_c\} \subseteq \mathbb{R}^s$ que *a priori* son desconocidos. La pertenencia del i-ésimo elemento a la k-ésima clase es representada por u_{ik}, donde $i \in \{1, 2, ..., n\}$ y $k \in \{1, 2, .., c\}$, constituyendo la matriz de pertenencia $U \in [0, 1]^{c \cdot n}$.

Bajo estas definiciones, básicamente todas las técnicas borrosas centradas en medidas de distancia entre los elementos y los patrones de clase, se basan en la minimización de una función objetivo dado por:

$$J(U,V;X)=\sum_{i=1}^{n}\sum_{k=1}^{c}u_{ik}^{m}\cdot d_{ik}^{2} \tag{2.20}$$

donde $m \in \mathbb{R}_{>1}$ corresponde al parámetro de borrosidad, que representa el grado en el cual los elementos que pertenecen a una clase inciden en el valor que toma el patrón de dicha clase, y d_{ik} es la medida de distancia entre el i-ésimo elemento y el k-ésimo patrón de clase.

El algoritmo utilizado para minimizar la función (2.20) es el que otorga la variedad en la técnica de análisis de conglomerado ([50]). La técnica borrosa más conocida para la generación de conglomerados es *fuzzy c-means* (FCM). La primera versión de Duda y Hart en 1973 ([32]) corresponde a una partición de clases nítidas, sobre la cual Dunn, en 1974 realiza una versión borrosa [33], la que es complementada por Bezdek en 1981 al introducir el parámetro de borrosidad m ([7]). Este algoritmo genera c clases de similar tamaño representadas por un patrón o centro de la clase, en la cual la medida de distancia entre los elementos y el patrón de cada clase corresponde a la distancia euclídea.

La búsqueda de patrones del algoritmo FCM se basa formalmente en:

Definición 2.15. (Patrones del algoritmo FCM)
Sea $s \in \mathbb{N},\ D := \mathbb{R}^s,\ X = \{x_1, x_2, ..., x_n\} \subseteq D,\ C := \mathbb{R}^s,\ c \in \mathbb{N},\ P_c(C)$ el conjunto de todos los subconjuntos de C con cardinalidad c, la función objetivo J con $m \in \mathbb{R}_{>1}$, y

$d : D \times C \to \mathbb{R}$, $(x, p) \to \|x - p\|$ la medida de distancia entre un elemento de X y un patrón de clase.

Si J es minimizado respecto de todas las posibles particiones de X, donde $K = \{k_1, k_2, ..., k_c\} \in P_c(C)$ es el conjunto de patrones de clases y $f(x_i)(k_j) = u_{ij}$ la pertenencia de x_i a la clase k_j con $f : XxK \to [0, 1]$ como función de pertenencia, entonces se satisface:

$$k_i = \frac{\sum\limits_{i=1}^{n} u_{ij}^m \cdot x_i}{\sum\limits_{i=1}^{n} u_{ij}^m}$$

Cuando $m = 1$ el algoritmo FCM genera clases excluyentes, en la cual cada elemento del conjunto de datos sólo pertenece a una clase, no correspondiendo por tanto, a un análisis de conglomerado borroso.

Se han desarrollado otros algoritmos alternativos al FCM, donde la principal diferencia se centra en la medida de distancia. Uno de ellos es el algoritmo *Gustafson-Kessel*, utilizado para clases de distinto tamaño y forma elipsoidal ([46]), otro es el algoritmo *Gath-Geva*, que sigue el mismo objetivo que el Gustafson-Kessel pero en base a estimadores estadísticos ([38]), además de otros algoritmos que corresponden a sus respectivas versiones simplificadas.

El algoritmo Gustafson-Kessel incorpora una matriz simétrica cuyo determinante es uno, que está dada por: $A_i = \sqrt[p]{det(S_i)} S_i^{-1}$, donde $S_i = \sum\limits_{j=1}^{n} u_{ij}^m (x_j - v_i)(x_j - v_i)^T$, y en la cual $\|x\|_A := \sqrt{x^T \ A \ X}$. De esta forma, el algoritmo puede generar clases de distinto tamaño y forma, sin embargo, la definición de la matriz también requiere un conocimiento *a priori* de la partición. Por otro lado, el algoritmo Gath-Geva es una extensión del anterior, en tanto también toma en cuenta el tamaño y la densidad de las clases de la partición, pero no se basa en una función objetivo, sino que en un estimador estadístico. Gath-Geva interpreta al conjunto de datos como realizaciones de variables aleatorias con distribución Normal, en la cual cada una de las c distribuciones Normales $N_i, i \in \mathbb{N}_{\leq c}$ tiene un valor esperado v_i y una matriz de covarianzas A_i, que genera un dato x_j con probabilidad *a priori* P_i, para más detalle ver [50].

2.5.2. Evaluación del sistema de clasificación.

Tanto el algoritmo FCM como sus extensiones, requieren la definición previa de la cantidad de clases de la partición y de la medida de distancia, además de la inicialización de los patrones de clase. Una de las principales desventajas radica en que el número de clases es *a priori* desconocido. Frente a este problema, se han desarrollado algunas estrategias para seleccionar la partición más adecuada de un conjunto de datos determinado. Una de ellas es evaluar la calidad de particiones realizadas con distinto número de clases, a fin de identificar el mejor sistema de clasificación en base a nuevas funciones objetivo. Estos indicadores evaluativos o las funciones objetivo pueden estar orientados a evaluar la calidad de la partición, como también de cada clase de una partición determinada, ambos entregan un número óptimo de clases para un conjunto de datos determinado.

Bajo un conjunto de datos de entrada es posible generar una gran cantidad de diferentes clases, que dependerá de la cantidad de variables de entrada $X = \{x_1, x_2, ..., x_n\} \subseteq \mathbb{R}^s$ y de la cantidad de valores que puedan tomar tales variables. Si los posibles valores que puede tomar x_i $\forall i \in \{1, 2, ..., n\}$ pertenecen al conjunto dado por $T = \{t_1, t_2, ..., t_m\}$, entonces la cantidad de particiones que se pueden generar corresponde a $(m^s \cdot s)$. Frente a esta gran variedad de posibilidades, es necesario definir un proceso de evaluación de dichas particiones que entregue la solución ideal bajo ciertas condiciones de calidad.

Existen algunos indicadores de validación global que evalúan el sistema de clasificación completo, y otros que son de validación local, en tanto evalúan la calidad de cada clase de una partición determinada. A continuación se presentan algunas medidas de validación encontradas en [50]:

Medidas de validación global:

- *Coeficiente de partición (PC)*: Sea $f : XxK \to [0,1]$ una partición borrosa del conjunto de datos $X = (x_1, x_2, ..., x_n)$, en la cual $f(x_i)(k_j) = u_{ij} \in [0,1]$, $\forall x \in X \ y \ k \in K$.

  El coeficiente de partición PC de la partición f es definido por:

  $$PC(f) = \frac{\sum\limits_{x \in X} \sum\limits_{k \in K} f^2(x)(k)}{|X|}$$

  PC está acotado por $\frac{1}{|K|} \leq PC(f) \leq 1$, tomando valores cercanos a 1 a medida que la partición va siendo más nítida, y cercanos a 0 a medida que los grados de pertenencia de cada observación a cada clase van siendo más símiles, por tanto, es una medida de solapamiento entre clases.

- *Entropía de la partición (EP)*: Sea $f : XxK \to [0,1]$ una partición borrosa del conjunto de datos $X = (x_1, x_2, ..., x_n)$, en la cual $f(x_i)(k_j) = u_{ij} \in [0,1], \forall x \in X \ y \ k \in K$.

  La entropía de la partición f PE es definida por:

  $$PE(f) = -\frac{\sum\limits_{x \in X} \sum\limits_{k \in K} f(x)(k) ln(f(x)(k))}{|X|}$$

  PE está acotado por $0 \leq PE(f) \leq ln(|K|)$, tomando valores cercanos a 0 a medida que la partición va siendo más nítida, por tanto, con menos nivel de solapamiento entre clases.

- *Índice de separación (S)*: Sea $f : XxK \to [0,1]$ una partición borrosa del conjunto de datos $X = (x_1, x_2, ..., x_n)$, en la cual $f(x_i)(k_j) = u_{ij} \in [0,1],\ \forall x \in X\ y\ k \in K$, y d es la función distancia.

El índice de separación S de la partición f es definida por:

$$S(f) = \frac{\sum\limits_{x \in X} \sum\limits_{k \in K} f^2(x)(k)\ d^2(x,k)}{|K|\ min\{d^2(k,l):\ k,l \in K, k \neq 1\}}$$

El objetivo es indicar lo separadas que están la clases, evitando que dos elementos que pertenecen a clases distintas tengan una distancia similar a la de dos elementos de una misma clase, siendo por tanto, valores bajos del índice de separación, indicador de una buena partición.

S está acotado por $S \leq \frac{1}{D_1^2}$, que corresponde al valor que toma el índice cuando la partición es nítida, donde D_1 es el índice de separación de una partición nítida dado por:

$$D_1(f) = \min_{i \in K} \left\{ \min_{j \in K/\{i\}} \left\{ \frac{d(A_i, A_j)}{\max\limits_{k \in K} diam(A_k)} \right\} \right\}$$

donde $A_k := f^{-1}(k)$, $diam(A_k) = max\,\{d(x_i, x_j) : x_i, x_j \in A_k\}$ y $d(A_i, A_j) = min\,\{d(x_i, x_j) : x_i \in A_i, x_j \in A_j\}$.

- *Hipervolumen borroso (FHV)*: Sea $f : XxK \rightarrow [0,1]$ una partición borrosa del conjunto de datos $X = (x_1, x_2, ..., x_n)$ realizada mediante el algoritmo de *Gath-Geva*, en la cual $f(x_i)(k_j) = u_{ij} \in [0,1],\ \forall x \in X\ y\ k \in K.$

  Se define el hipervolumen borroso FHV por:

$$FHV(f) = \sum_{i=1}^{c} \sqrt{det(A_i)}$$

- *Densidad promedio de la partición (APD)*: Sea $f : XxK \rightarrow [0,1]$ una partición borrosa del conjunto de datos $X = (x_1, x_2, ..., x_n)$ realizada mediante el algoritmo de *Gath-Geva*, en la cual $f(x_i)(k_j) = u_{ij} \in [0,1], \ \forall x \in X \ y \ k \in K$.

Se define la densidad promedio de la partición APD por:

$$APD(f) = \frac{1}{c}\sum_{i=1}^{c}\frac{S_i}{\sqrt{det(A_i)}}$$

donde el número de elementos por clases es:

$S_i = \sum_{j \in Y_i} f(x_j)(k_i)$ y $Y_i = j \in \mathbb{N}_{\leq n} \ /(x_j - v_i)^T \ A_i^{-1}(x_j - v_i) < 1$

- *Densidad de la partición (PD)*: Sea $f : XxK \rightarrow [0,1]$ una partición borrosa del conjunto de datos $X = (x_1, x_2, ..., x_n)$ realizada mediante el algoritmo de *Gath-Geva*, en la cual $f(x_i)(k_j) = u_{ij} \in [0,1], \ \forall x \in X \ y \ k \in K$.

  La densidad de la partición f PD se define por:

$$PD(f) = \frac{\sum_{i=1}^{c} S_i}{FHV(f)}$$

Es posible notar que las medidas de evaluación de un sistema de clasificación borroso no supervisado presentada, sólo se enfocan a optimizar un criterio de calidad, la redundancia o solapamiento entre clases, entregando mejor valoración a aquella partición en la cual el nivel de solapamiento es menor. A nuestro juicio, la redundancia es un criterio necesario pero no suficiente, en tanto es importante evaluarla conjuntamente con otros criterios, por ejemplo, que todos los elementos del conjunto de datos queden representados por al menos una clase con un nivel mínimo de pertenencia. A modo de ejemplo, supongamos que se tiene el conjunto de clases $C = \{c_1, c_2\}$ y el conjunto de datos $X \in \{x_1, x_2, x_3\}$, donde $x_1 = (1,0), \ x_2 = (0,1), \ x_3 = (0,0)$, esta partición es bien evaluada por las medidas presentadas a pesar que x_3 no queda representado por la partición C. Llevándolo al extremo, dichas medidas también entregan una buena evaluación si se tiene la partición dada por $C = \{c_1, c_2, c_3\}$ en la cual $x_1 = (1,0,0), \ x_2 = (0,1,0), \ x_3 = (0,0,1)$, siendo evidentemente ilógica una solución como esta para un sistema de clasificación de tres elementos.

Medidas de validación local:

- *Relación de compatibilidad en clases GK*: Sea $p \in \mathbb{N}$, A matriz simétrica y definida positiva, y $C := \mathbb{R}^p \ x \ \{A \in \mathbb{R}^{pxp} : det(A) = 1\}$. Para cada $k_i = (v_i, A_i) \in C$, sea λ_i el mayor autovalor de A_i, y e_i el autovalor normalizado de A_i asociado con su menor autovalor. Para cada $\gamma = (\gamma_1, \gamma_2, \gamma_3) \in \mathbb{R}^3$ se define la relación de compatibilidad CCM $\doteq_\gamma \subseteq \ C \ x \ C$ por:

  $$\forall k_1, k_2 \in C : k_1 \doteq_\gamma k_2 \ ssi \ \text{ se cumplen las tres siguientes condiciones:}$$

  $$|e_1 e_2^T| \geq \gamma_1$$

  $$\left| \frac{e_1 + e_2}{2} \frac{(v_1 - v_2)^T}{\|v_1 - v_2\|} \right| \leq \gamma_2$$

  $$\|v_1 - v_2\| \leq \gamma_3 \left(\sqrt{\lambda_1} + \sqrt{\lambda_2}\right)$$

- *Densidad del contorno (SD)*: Sea $f : X \to K$ una partición del conjunto de datos $X = (x_1, x_2, ..., x_n)$ y Δ_{ij} la distancia entre $x_j \in X$ y el patrón de clase $k_j \in K$. Se define la densidad de contorno SD_i de la clase k_i en la partición f por:

  $$SD_i(f) = \frac{\sum\limits_{x_j \in S_i} f(x_j)(k_i)}{\sqrt{det(A_i)}}$$

  donde $S_i = \{x_j \in X : \Delta_{ij} \hat{A}_i^{-1} \Delta_{ij} < 1\}$ y A_i es la matriz de covarianzas del contorno borroso.

Nuevamente vemos que los indicadores presentados se enfocan a un criterio de calidad, el solapamiento entre clases. Además, algunos de ellos fueron ideados para ser utilizados sólo bajo ciertos algoritmos de clasificación (FHV, APD y PD), limitando por tanto su ámbito de aplicación.

Frente a este vacío de indicadores de calidad de una clasificación borrosa no supervisada en la cual no necesariamente se verifica una partición Ruspini (2.5), en [1] se presentan las medidas de ***relevancia***, ***redundancia*** y ***cobertura***. Estas medidas pueden constituir un sistema para evaluar la calidad de las distintas particiones posibles de generar desde un conjunto de datos de entrada determinado. La primera corresponde a una medida de calidad de la partición, mientras que las otras son medidas referidas a cada clase de una partición determinada.

- *Relevancia*: Una clase es relevante cuando al prescindir de ella como componente de una partición determinada, al menos uno de los elementos queda indeterminado, en caso contrario será irrelevante. Sin embargo, a pesar de ser una característica de cada clase, esta medida se evalúa sobre una familia de clases C, en tanto es parte de un sistema en la cual se espera que $\forall c \in C,\ \exists x \in X:\ \mu_c(x) > 0$. El nivel de relevancia de la clase c es gradual y está dado por $\phi\{\mu_c(x) : x \in X\}$, entonces, diremos que c es relevante si *para cada* $x \in X$, se cumple:

  $$\phi\{\mu_c(x) : c \in C\} \text{ es significativamente mayor que } \phi\{\mu_k(x) : k \in C, k \neq c\}.$$

  Y por el contrario, diremos que c es irrelevante si *para todos* los $x \in X$, se cumple:

  $$\phi\{\mu_c(x) : c \in C\} \text{ no es significativamente mayor que } \phi\{\mu_k(x) : k \in C, k \neq c\}.$$

- *Redundancia*: La redundancia de una partición puede ser entendida como el nivel de solapamiento entre las clases que la componen. Dicho grado de solapamiento entre las clases $c, k \in C$ *con* $c \neq k$ sobre el elemento x, lo representamos por $\varphi(\mu_c(x), \mu_k(x))$.

- *Cobertura*: La cobertura se refiere a que todos los elemento $x \in X$ deben estar contenidos en alguna de las clases $c \in C$ de la partición, y por tanto, la cobertura dada por $\phi\{\mu_c(x) : \ c \in \ C\}$, será mayor mientras mayor sea su grado de pertenencia. Por lo tanto, $\forall x \in X, \ \exists c \in C : \ \mu_c(x) > 0$.

Recordemos los ejemplos anteriores, se tenía el conjunto de clases $C = \{c_1, c_2\}$ y el conjunto de datos $X \in \{x_1, x_2, x_3\}$, donde $x_1 = (1, 0), \ x_2 = (0, 1), \ x_3 = (0, 0)$, esta partición entregará buenos resultados bajo el criterio de redundancia en tanto no hay solapamiento, pero no será bien evaluada bajo el de cobertura, ya que un elemento no queda representado por ninguna clase. Respecto del otro ejemplo, se tenía la partición dada por $C = \{c_1, c_2, c_3\}$ en la cual $x_1 = (1, 0, 0), \ x_2 = (0, 1, 0), \ x_3 = (0, 0, 1)$, esta partición será bien evaluada en términos de redundancia y cobertura, pero claramente existe al menos una clase no relevante.

Capítulo 3

Consistencia y estabilidad en familias de operadores de agregación.

Bajo el objetivo de generar índices robustos, a continuación se estudian los conceptos de consistencia y estabilidad en familias de operadores de agregación. En este sentido, se presentan propiedades que definen a una familia *consistente* en términos de la *estabilidad* de los resultados del proceso de agregación en el que participan. Estas propiedades consideran la estructura del conjunto de datos de entrada, siendo definidas para datos sin una estructura inherente, como también para datos que presentan un orden lineal con o sin prioridades asociadas a tal ordenamiento. Finalmente, se extienden dichas propiedades de estabilidad a conjuntos de datos jerárquicamente priorizados, que corresponde a la estructura que comúnmente presentan los datos sobre los cuales se generan índices psicosociales.

Tal como ha sido mencionado en el capítulo anterior, un *operador de agregación* A_n agrega la información de un conjunto de ítems $x_1, \ldots, x_n$ en $[0, 1] \subseteq [+\infty, -\infty]$, produciendo un valor agregado en $[0,1]$ [15]. Correspondiendo entonces, a una función real de n variables definida en el intervalo unitario, monótona no decreciente y que satisface las condiciones de borde descritas en (2.2). De esta forma, la secuencia $A_2, A_3, ..., A_n \ \forall \ n \in \mathbb{N}$, corresponde a la *familia de operadores de agregación* dada por $\{A_n : [0,1]^n \rightarrow [0,1], \ n \in \mathbb{N}\}$.

El proceso de agregación dado por la generación de dicho valor agregado $A_n(x_1, x_2, ..., x_n)$, persigue el objetivo de reducir la dimensionalidad del conjunto de datos de entrada. Sin embargo, el cumplimiento de este objetivo se ve afectado por el *problema de dimensionalidad*, en tanto a lo largo del proceso de agregación se pueden producir diversos cambios de cardinalidad en los datos, ante lo cual, cada vez que se produce un cambio, es necesario actualizar el valor agregado a partir del nuevo conjunto de datos, no existiendo restricciones que garanticen que el nuevo operador de agregación utilizado en esta actualización, mantenga la consistencia lógica respecto de la agregación previa.

Por ejemplo, un proceso de agregación puede estar dado por el mínimo para $n = 2$, el máximo para $n = 3$, la media geométrica para $n = 4$ y la mediana para $n = 5$, definición que evidentemente no generará resultados estables. Frente a esto, parece lógico esperar que los operadores que componen una misma familia deban tener algún tipo de relación, y así, la secuencia $A(2)$, $A(3)$, $A(4), \ldots$ sería consistente, en tanto no estaría compuesta por operadores desconectados, y el proceso de agregación seguiría siendo el mismo ante cualquier cambio en la dimensión n de los datos. Por lo tanto, a fin de asegurar un proceso de agregación robusto ante cambios de dimensionalidad, es necesario definir alguna propiedad que asegure un nivel de estabilidad en los resultados de dicho proceso.

Se han estudiado muchas propiedades que representan características deseables que debe tener un operador de agregación A_n para un n conocido, a saber; simetría, monoticidad, idempotencia, entre otras. Si bien, estas propiedades han sido extendidas a las familias de operadores de agregación $\{A_n\}_{n \in \mathbb{N}}$, esto se ha traducido en el cumplimiento para cada n, en tanto la definición de familia no exige la existencia de una relación entre los operadores $A_2, A_3, ..., A_n$ que los unifiquen bajo un concepto común, y que por tanto, resguarde la consistencia de una FAO como un todo ([2, 27, 28, 29, 40]). La noción de consistencia utilizada en este sentido, que exige la existencia de una relación entre los miembros de una misma FAO, es muy amplia, siendo más general que el concepto de recursividad, en tanto algunos operadores no recursivos, como por ejemplo la mediana, satisfacen la idea de consistencia. Esta idea responde por ejemplo, al problema de pérdida de datos, ya que cuando se tiene un conjunto de datos en la cual existen elementos que no contienen información en una o más variables, la agregación para tales elementos no es posible hacerla mediante el operador definido para el resto de los elementos que contienen toda la información, en tanto tiene una dimensión menor. En este caso, es necesario definir un nuevo operador de menor dimensión, pero que mantenga la unicidad conceptual, requiriéndose por tanto, establecer alguna restricción que resguarde la consistencia del proceso de agregación, por ejemplo, mediante una definición recursiva de operadores, o también, mediante el cumplimiento de alguna propiedad de estabilidad definida para una secuencia de operadores que componen una familias de operadores de agregación.

En la sección anterior, se presentaron algunas propiedades que persiguen la idea de generar resultados consistentes. Una de estas propiedades es la asociatividad, otra menos restrictiva es la propiedad self-identity, que en base a operadores idempotentes, asegura resultados consistentes bajo un conjunto de datos desestructurado o con un orden lineal de izquierda a derecha. Orientados al mismo fin, también se presentaron las propiedades de continuidad y estabilidad para familias de operadores de agregación, no obstante, el cumplimiento de dichas propiedades se traduce en el cumplimiento para cada operador que compone la familia, no existiendo por tanto, restricciones respecto a la unidad conceptual del proceso.

Frente a este vacío, y bajo el objetivo de entender las propiedades que definen a una familia consistente de operadores en términos de los resultados del proceso de agregación, pero además, que considere las características del conjunto de datos de entrada, a continuación se estudia la relación que debiera existir entre los operadores de una misma FAO, a fin de asegurar una estabilidad mínima ante cambios de cardinalidad del conjunto de datos de entrada.

Los desarrollos que a continuación se presentan, se dividen en dos partes; estabilidad en familias de operadores de agregación aplicadas sobre un conjunto de datos de entrada no estructurado, y estabilidad bajo un conjunto de datos de entrada estructurado. Las FAOs que se utilizan sobre conjuntos de datos no estructurados son simétricas, mientras que el ámbito de aplicación de las familias no simétricas se restringe a conjuntos de datos estructurados, que en este trabajo se centra en estructuras con orden lineal y con orden jerárquico.

A lo largo de las siguientes secciones y para cada tipo de conjunto de datos de entrada, se van incorporando definiciones de estabilidad cada vez menos restrictivas, comenzando con la *estabilidad estricta*, para luego seguir con la *estabilidad asintóticamente estricta*, la *estabilidad estricta casi segura* y finalmente la *inestabilidad*.

3.1. Estabilidad en FAOs bajo conjuntos de datos no estructurados.

Cuando el conjunto de datos de entrada no tiene estructura, necesariamente el resultado del proceso de agregación no varía si se intercambian los elementos de dicho conjunto, por tanto se trata de una familia simétrica $\{A_n\}_{n \in \mathbb{N}}$, es decir, A_n es un operador de agregación simétrico $\forall n \in \mathbb{N}$.

La propiedad de *estabilidad estricta* para una *FAO* restringe a que un operador de una familia estrictamente estable definido para n items, no difiera de un operador definido para $n-1$ ítems cuando el último elemento adicionado corresponde a la agregación del resto de los ítems. De esta forma, la estabilidad estricta entrega condiciones a considerar para mantener la consistencia lógica entre operadores de una misma familia, traduciéndose en que la robustez del proceso de agregación asociado no sea afectada cuando se producen cambios de cardinalidad en los datos.

No obstante, es importante considerar que el cumplimiento de esta propiedad tiene sentido cuando el valor resultante de la agregación de los $n-1$ ítems puede considerarse como otro elemento del mismo conjunto de datos de entrada. Por ejemplo, supongamos que se tiene un conjunto de datos en el intervalo unitario $\{x_1, ..., x_{n-1}\}$, que representan el nivel de riqueza de $n-1$ sujetos, e interesa agregarlos para estimar la *desigualdad existente en la población* a través del nivel de acumulación de la riqueza. Una posibilidad para agregar esta información sería utilizar el conocido índice de concentración (de un bien) de Gini o *Coeficiente de Gini*. Este índice toma valores en el intervalo $[0, 1]$, donde el valor 0 representa la mayor igualdad entre la riqueza de los sujetos (todos tienen la misma riqueza por lo que la concentración del bien es nula), y el valor 1 representa la desigualdad extrema (máxima concentración), es decir, sólo un sujeto acumula toda la riqueza de la población. En este caso, el conjunto de datos de entrada contiene la información de la variable lingüística *riqueza del sujeto*, mientras que el valor agregado intenta medir el *nivel de acumulación de la riqueza de la población*, que representa una característica de todo el conjunto de datos de entrada y no de cada uno de sus elementos, por lo que no tiene sentido evaluar la estabilidad en los términos que la estamos estudiando. Ahora bien, si se tiene un conjunto de datos $\{y_1, ..., y_{k-1}\}$ donde $y_i \in [0, 1]$ con $i \in \{1, 2, ..., k-1\}$, que representan $k-1$ características socioeconómicas de una población determinada, por ejemplo el índice de Gini, el ingreso *per cápita*, el promedio de años de estudio de la población, entre otras, e interesa agregar esta información para disponer de un índice socioeconómico de dicha población. Esta vez sí tiene sentido evaluar el cumplimiento de la propiedad de estabilidad estricta de la familia de operadores de agregación utilizada, en tanto $A_{k-1}(y_1, ..., y_{k-1})$ puede considerarse como otro elemento del conjunto de datos de entrada, correspondiendo entonces, a una "*$confirmación$*" del nivel socioeconómico de dicha población.

Definición 3.1. Sea $\{A_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ una familia de operadores de agregación. Diremos que $\forall n > 2$ y para todo conjunto de datos no estructurado $\{x_n\}_{n\in\mathbb{N}}$ en $[0,1]$, $\{A_n\}_{n\in\mathbb{N}}$ es una familia estrictamente estable si satisface:

$$A_n(x_1, x_2, ...x_{n-1}, A_{n-1}(x_1, x_2..., x_{n-1})) = A_{n-1}(x_1, x_2, ..., x_{n-1}) \tag{3.1}$$

La propiedad 3.1 coincide con la propiedad *self-identity* definida por Yager [93], por lo tanto, todas las familias que satisfacen la propiedad self-identity, serán familias simétricas *estrictamente estables.*

Es importante notar que dado que el conjunto de datos de entrada no tiene una estructura inherente, no importan la posición que tome $A_{n-1}(x_1, ..., x_{n-1})$ en A_n, en tanto no afecta el cumplimiento de la propiedad de estabilidad estricta dada por 3.1, pudiendo ubicarse tanto en la primera como en la i-ésima posición.

A continuación se prueba el nivel de estabilidad estricta de algunas de las *FAO's* simétricas más conocidas.

Proposición 3.1. *La FAO definida por el mínimo* $\{Min_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$*, es una familia estrictamente estable.*

Demostración:

$$\begin{aligned} Min_n(x_1,...x_{n-1},Min_{n-1}(x_1,...,x_{n-1})) &= Min_n(x_1,...x_{n-1},x_{(1)}) \\ &= x_{(1)} \\ &= Min_{n-1}(x_1,...,x_{n-1}) \end{aligned}$$

Proposición 3.2. *La FAO definida por el máximo* $\{Max_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$, *es una familia estrictamente estable.*

Demostración:

$$\begin{aligned} Max_n(x_1,...x_{n-1},Max_{n-1}(x_1,...,x_{n-1})) &= Max_n(x_1,...x_{n-1},x_{(n-1)}) \\ &= x_{(n-1)} \\ &= Max_{n-1}(x_1,...,x_{n-1}) \end{aligned}$$

Proposición 3.3. *La FAO definida por la mediana* $\{Md_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$, *es una familia estrictamente estable.*

Demostración:

Si $n-1$ es impar:

$$\begin{aligned} Md_n(x_1,...x_{n-1},Md_{n-1}(x_1,...,x_{n-1})) &= Md_n\left(x_1,...x_{n-1},x_{(\frac{n+1}{2})}\right) \\ &= x_{(\frac{n+1}{2})} \\ &= Md_{n-1}(x_1,...,x_{n-1}) \end{aligned}$$

Si $n-1$ es par:

$$\begin{aligned}
Md_n(x_1,...x_{n-1},Md_{n-1}(x_1,...,x_{n-1})) &= Md_n\left(x_1,...x_{n-1},\frac{x_{(\frac{n}{2})}+x_{(\frac{n}{2}+1)}}{2}\right)\\
&= \frac{x_{(\frac{n}{2})}+x_{(\frac{n}{2}+1)}}{2}\\
&= Md_{n-1}(x_1,...,x_{n-1})
\end{aligned}$$

Proposición 3.4. *La FAO definida por la media aritmética* $\{M_n:[0,1]^n\to[0,1],n\in\mathbb{N}\}$*, es una familia estrictamente estable.*

Demostración:

$$\begin{aligned}
M_n(x_1,...x_{n-1},M_{n-1}(x_1,...,x_{n-1})) &= \frac{1}{n}\left(\sum_{i=1}^{n-1}x_i+\frac{1}{n-1}\sum_{i=1}^{n-1}x_i\right)\\
&= \frac{1}{n}\left(\frac{(n-1)\sum\limits_{i=1}^{n-1}x_i+\sum\limits_{i=1}^{n-1}x_i}{n-1}\right)\\
&= \frac{1}{n(n-1)}\left(\sum_{i=1}^{n-1}x_i(n-1+1)\right)\\
&= M_{n-1}(x_1,...,x_{n-1})
\end{aligned}$$

Proposición 3.5. *La FAO definida por la media geométrica* $\{G_n:[0,1]^n\to[0,1],n\in\mathbb{N}\}$*, es una familia estrictamente estable..*

Demostración:

$$G_n(x_1,...x_{n-1},G_{n-1}(x_1,...,x_{n-1})) = \left(\prod_{i=1}^{n-1}x_i\left(\prod_{i=1}^{n-1}x_i\right)^{\frac{1}{n-1}}\right)^{\frac{1}{n}}$$

$$
\begin{aligned}
&= \left(\left(\prod_{i=1}^{n-1} x_i\right)^{\frac{n}{n-1}}\right)^{\frac{1}{n}} \\
&= G_{n-1}(x_1, ..., x_{n-1})
\end{aligned}
$$

Proposición 3.6. *La FAO definida por la media armónica* $\{H_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$*, es una familia estrictamente estable.*

Demostración:

$$
\begin{aligned}
H_n(x_1, ...x_{n-1}, H_{n-1}(x_1, ..., x_{n-1})) &= \frac{n}{\sum\limits_{i=1}^{n-1} \frac{1}{x_i} + \frac{1}{\frac{n-1}{\sum\limits_{i=1}^{n-1} \frac{1}{x_i}}}} \\
&= \frac{n}{\sum\limits_{i=1}^{n-1} \frac{1}{x_i} + \frac{\sum\limits_{i=1}^{n-1} \frac{1}{x_i}}{n-1}} \\
&= \frac{n(n-1)}{(n-1)\sum\limits_{i=1}^{n-1} \frac{1}{x_i} + \sum\limits_{i=1}^{n-1} \frac{1}{x_i}} \\
&= \frac{n(n-1)}{n \sum\limits_{i=1}^{n-1} \frac{1}{x_i}} \\
&= H_{n-1}(x_1, ..., x_{n-1})
\end{aligned}
$$

Ahora bien, si analizamos la familia de operadores de agregación dada por el producto, podemos ver que esta familia no es estrictamente estable en términos absolutos, ya que si se considera la secuencia de datos $(x_1, x_2, \dots) = (1/2, 1, 1, \dots)$, vemos que $P_{n-1} = P_{n-1}(x_1, ..., x_{n-1}) = 1/2$, pero $P_n = P_n(x_1, \dots, x_{n-1}, P_{n-1}) = (P_{n-1})^2 = 1/4 \ \ \forall \ n > 2$, por tanto la diferencia $\left|Dif_r(P_n, P_{n-1})\right| = |P_n - P_{n-1}|$ no es cero. Sin embargo, veremos a continuación que esta familia sí tiene un comportamiento consistente en términos asintóticos.

A fin de analizar el comportamiento de esta familia de operadores de agregación, se realizó un ejercicio de simulación donde se estudia su velocidad de convergencia, permitiendo así identificar el tamaño mínimo del conjunto de datos sobre el cual se puede garantizar estabilidad estricta del proceso de agregación bajo cierto nivel de tolerancia. Luego, se estudia la proporción sobre la cual se satisface la estabilidad estricta en el límite, dada por $Pr\ [\lim_{n\to+\infty} \left|Dif_r(A_n, A_{n-1})\right| < \varepsilon\ ]$, para cada n y nivel de tolerancia $\varepsilon \in \{10^{-1}, 10^{-2}, 10^{-3}, ..., 10^{-30}\}$.

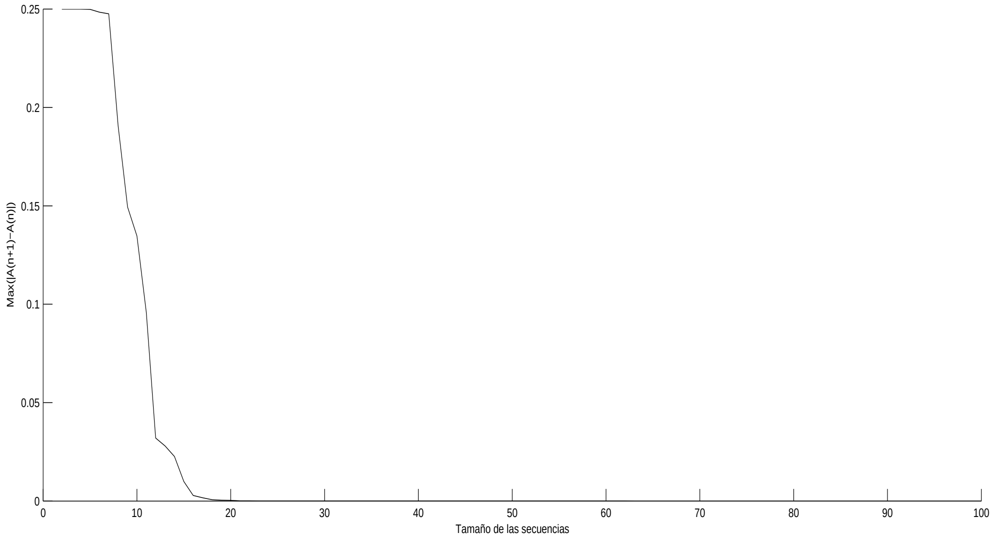
En el siguiente ejercicio de simulación, se calcula la diferencia dada por:

$$\left|Dif_r(A_n, A_{n-1})\right| = \left|A_n(x_1, ...x_{n-1}, A_{n-1}(x_1, .., x_{n-1})) - A_{n-1}(x_1, .., x_{n-1})\right|$$

para 1,000 secuencias $(x_1, \ldots, x_{n-1})$ de variables aleatorias independientes con distribución $U[0,1]$, y para los siguientes tamaños de las secuencias de datos: $\{10, 20, 30, ..., 50000\}$. A fin de estudiar el comportamiento de estas diferencias a medida que n aumenta, se calcula el valor máximo de $\left|Dif_r(A_n, A_{n-1})\right|$ para cada n.

La *Figura 3.1* muestra que los valores agregados de las secuencias de n y $n-1$ elementos, tienen mayor similitud a medida que n aumenta, donde la diferencia máxima de $|A_n - A_{n-1}|$ es 0 para cualquier secuencia de tamaño mayor a 20. En cuanto a la convergencia de la FAO del producto, se observa en *Tabla 3.1* que la velocidad de convergencia es bastante rápida, en tanto una secuencia de tamaño $n = 100$ ya garantiza resultados estables bajo cualquier nivel de tolerancia. La Figura 3.2 representa los resultados de la *Tabla 3.1.*

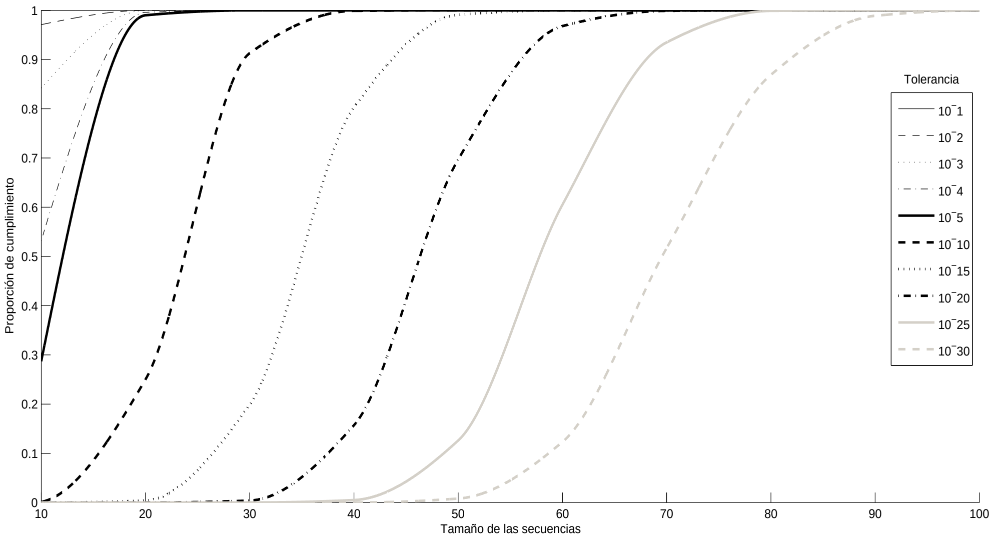
Figura 3.1: Estabilidad de la FAO del producto a medida que n aumenta.



Nota: La gráfica sólo muestra tamaños muestrales menores a 100.

Figura 3.2: Velocidad de convergencia de la familia del producto

Resultados de simulación de $Pr[\lim_{n\rightarrow+\infty} \left|Dif_r(P_n, P_{n-1})\right| < Tolerancia]$



Cuadro 3.1: Resultados de simulación de $Pr[\lim_{n\to+\infty} \left|Dif_r(P_n, P_{n-1})\right| < Tolerancia]$ para la FAO del producto.

Tamaño de	Nivel de tolerancia									
las secuencias	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-10}	10^{-15}	10^{-20}	10^{-25}	10^{-30}
10	1.000	0.979	0.867	0.53	0.327	0.002	0.000	0.000	0.000	0.000
20	1.000	1.000	1.000	0.998	0.978	0.239	0.005	0.000	0.000	0.000
30	1.000	1.000	1.000	1.000	1.000	0.906	0.216	0.004	0.000	0.000
40	1.000	1.000	1.000	1.000	1.000	1.000	0.797	0.181	0.010	0.000
50	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.693	0.134	0.004
60	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.965	0.585	0.121
70	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.934	0.492
80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.888
90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.989
100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Los resultados del ejercicio de simulación evidencian que a pesar de no satisfacer las propiedades de estabilidad estricta en términos absolutos, la *FAO* del producto muestra un comportamiento consistente. Frente a esto, y a fin de extender el enfoque propuesto a otras familias consistentes, se definen dos propiedades que corresponden a versiones menos restrictivas del mismo concepto de estabilidad estricta.

Definición 3.2. Sea $\{A_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ una familia de operadores de agregación. Diremos que $\forall n > 2$ y para todo conjunto de datos no estructurado $\{x_n\}_{n\in\mathbb{N}}$ en $[0,1]$, $\{A_n\}_{n\in\mathbb{N}}$ es una familia asintóticamente estrictamente estable si:

$$\lim_{n\to+\infty} \left| A_n(x_1,...x_{n-1}, A_{n-1}(x_1,...,x_{n-1})) - A_{n-1}(x_1,...,x_{n-1}) \right| = 0$$

Si bien, la propiedad de estabilidad asintóticamente estricta es menos restrictiva que la estabilidad estricta en términos absolutos, no recoge el tipo de consistencia de algunas familias de operadores de agregación, por ejemplo, veamos lo que ocurre con la familia del producto.

Dada la diferencia:

$$Dif_r(P_n, P_{n-1}) = \prod_{i=1}^{n-1} x_i \prod_{i=1}^{n-1} x_i - \prod_{i=1}^{n-1} x_i = \prod_{i=1}^{n-1} x_i \left(\prod_{i=1}^{n-1} x_i - 1 \right),$$

para aquellas sucesiones $\{x_n\}_{n\in\mathbb{N}}$ tal que $\prod_{i=1}^{n} x_n \overset{n\to\infty}{\longrightarrow} c \in (0,1)$, implica que $x_n \overset{n\to\infty}{\longrightarrow} 1$, pero la implicancia opuesta no es cierta, en tanto si $\exists k \;/\; x_k = 0$ entonces $\prod_{n=1}^{\infty} x_n = 0$.

Lo anterior lleva a suponer que aquellas sucesiones que provocan que la familia dada por el productor no sea estrictamente estable en términos asintóticos, constituyen un reducido subconjunto F del conjunto S de todas las sucesiones en $[0,1]$. Por lo tanto, la posibilidad de reunir una colección de ítems que potencialmente lleven al producto a un comportamiento inestable, es esperable que sea pequeño, siendo posible formalizar este argumento en términos de probabilidades.

Sea $S = \{\{x_n\}_{n\in\mathbb{N}} : x_i \in [0,1] \; \forall i\}$ un conjunto de sucesiones en $[0,1]$, y sea $\{A_n\}_{n\in\mathbb{N}}$ una FAO involucrada en el proceso de agregación de la sucesión $s_n = (x_1, \ldots, x_n) \in S$. Si se considera el experimento dado por "*observar la estabilidad de* $\{A_n(s_n)\}_{n\in\mathbb{N}}$", o en otras palabras, "*observar la distancia entre* $A_n(s_{n-1}, A_{n-1}(s_{n-1}))$ *y* $A_{n-1}(s_{n-1})$" (para una colección de datos ordenados de izquierda a derecha). El espacio muestral

E asociado, está dado por el conjunto de los posibles niveles de estabilidad estricta de $\{A_n(s_n)\}_{n\in\mathbb{N}}$. Si la cardinalidad de S tiende a infinito, es posible obtener la probabilidad que el evento dado por "*reunir una sucesión s_n que lleve a que $\{A_n(s_n)\}_{n\in\mathbb{N}}$ tenga un comportamiento estrictamente estable*", es decir,

$$\mathbb{P}[\lim_{n\to+\infty} |A_n(s_{n-1}, A_{n-1}(s_{n-1})) - A_{n-1}(s_{n-1})| = 0 \ ].$$

Si se asume que el valor de los ítems $(x_1, ...x_n)$ son variables aleatorias independientes que siguen la distribución $U([0,1])$, entonces es posible introducir una medida de probabilidad sobre el conjunto de sucesiones S mediante la función de distribución de la probabilidad conjunta:

$$\mathbb{P}(a_1 \leq X_1 \leq b_1, ..., a_n \leq X_n \leq b_n, ...) = \prod_{i=1}^{\infty} \mathbb{P}(a_i \leq X_i \leq b_i) = \prod_{i=1}^{\infty}(b_i - a_i),$$

donde $a_i, b_i \in [0,1]$ $\forall i$. Entonces, si por ejemplo la probabilidad dada para una sucesión $\{x_n\}_{n\in\mathbb{N}}$ es cercana a cero, y la probabilidad del conjunto de todas las sucesiones tal que $x_i \in [0,1/2]$ $\forall i \leq N$ y $x_i \in [0,1]$ $\forall i > N$, para un N determinado es $(1/2)^N$, entonces es posible ver que el conjunto dado por $F = \left\{ \{x_n\}_{n\in\mathbb{N}} / \prod_{n=1}^{\infty} x_n \overset{n\to\infty}{\longrightarrow} c \in (0,1) \right\}$, en la cual la familia del productorio deja de ser estrictamente estable, tiene probabilidad cero. Esto nos lleva a la siguiente definición:

Definición 3.3. Sea $\{A_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ una familia de operadores de agregación. Diremos que $\forall n > 2$ y para todo conjunto de datos no estructurado $\{x_n\}_{n\in\mathbb{N}}$ en $[0,1]$, $\{A_n\}_{n\in\mathbb{N}}$ es una familia casi seguro estrictamente estable si:

$$\mathbb{P}[\lim_{n\to+\infty} |A_n(x_1, ...x_{n-1}, A_{n-1}(x_1, ..., x_{n-1})) - A_{n-1}(x_1, ..., x_{n-1})| = 0 \ ] = 1, \forall x_i \sim U(0,1)$$

A continuación se prueba el nivel de estabilidad estricta en probabilidad de las FAO's más conocidas.

Proposición 3.7. *La FAO definida por el productorio* $\{P_n\}_{n\in\mathbb{N}}$ *es una familia casi seguro estrictamente estable.*

Demostración.

Tal como fue mencionado anteriormente, sea F el conjunto de sucesiones en la cual la FAO del productorio no es estrictamente estable, y donde $F \subset S1 = \{(x_n)_{n\in\mathbb{N}}/x_n \to 1\}$. Notar que para cada $\varepsilon > 0$, $S1$ puede ser particionado en los subconjuntos $C_n = \big\{\{x_k\}_{k\in\mathbb{N}}/\ \big|x_k - 1\big| < \ \varepsilon,\ \forall k \geq n\big\},\ n \in \mathbb{N}$. Considerando $C_n \bigcap C_m = \emptyset$ si $n \neq m$, se cumple que:

$$\mathbb{P}(F) \leq \mathbb{P}(S1) = \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} C_n\right) = \sum_{n=1}^{\infty} P(C_n) \leq \sum_{n=1}^{\infty}\prod_{k=n}^{\infty} \varepsilon = 0,$$

y entonces, es posible decir que la probabilidad que la familia dada por el productorio no tenga un comportamiento estrictamente estable es cero. Por lo tanto, esta familia satisface la noción de estabilidad estricta en una versión más débil, que puede ser caracterizada en términos de convergencia en probabilidad a una FAO estrictamente estable.

Proposición 3.8. *La FAO definida por* $\{Q_n\}_{n\in\mathbb{N}}$ *es una familia casi seguro estrictamente estable.*

Demostración.

Similar a la demostración anterior.

Finalmente, la definición de inestabilidad se deduce desde las definiciones de los distintos niveles de estabilidad estricta presentados, ya que tanto una *FAO* que satisface la propiedad de estabilidad asintóticamente estricta, como una *FAO* que safistace la propiedad de estabilidad estricta casi seguro, convergen a una *FAO* estrictamente estable. Siendo las *FAO's* estrictamente estables en el límite un caso particular de las familias estrictamente estables en probabilidad.

Proposición 3.9. *Sea* $\{A_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ *una familia de operadores de agregación, entonces:*

1. *Si la familia* $\{A_n\}_{n\in\mathbb{N}}$ *satisface la propiedad de estabilidad estricta, entonces también satisface la propiedad de estabilidad asintóticamente estricta.*
2. *Si la familia* $\{A_n\}_{n\in\mathbb{N}}$ *satisface la propiedad de estabilidad asintóticamente estricta, entonces también satisface la propiedad de estabilidad casi seguro estricta.*

Por lo tanto, si una *FAO* no es estrictamente estable en probabilidad, entonces no satisface ninguno de los tres niveles de estabilidad y en este caso, será una familia inestable.

Definición 3.4. Sea $\{A_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ una familia de operadores de agregación. Entonces, diremos que:

Satisface la propiedad de inestabilidad, si la familia no es casi seguro estrictamente estable, y entonces será una familia inestable.

A continuación se prueba la inestabilidad de una familia de operadores de agregación muy utilizada, los operadores ordered weighted averaging (OWA).

Proposición 3.10. *La FAO definida por operadores OWA* $\{O_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ *es una familia inestable.*

Demostración:

A fin de probar que $\{O_n\}_{n\in\mathbb{N}}$ es una FAO inestable, se considera la familia $\{IA_n\}_{n\in\mathbb{N}}$ definida como:

$$IA_n(x_1,\ldots,x_n) = \begin{cases} \text{Max}(x_1,\ldots,x_n) & \text{si n es impar} \\ \text{Min}(x_1,\ldots,x_n) & \text{si n es par} \end{cases}$$

Esta familia puede ser considerada como un caso particular de la familia dada por operadores OWA, si se consideran los pesos $w_n = (1,0,0,...)$ cuando n es impar y $w_n = (0,...,0,1)$ cuando n es par. Para probar que $\{IA_n\}_{n\in\mathbb{N}}$ es inestable, se verá que dado $\varepsilon > 0$, es posible encontrar una familia de suseciones R_ε con probabilidad positiva, talque para cualquier $\{x_n\}_{n\in\mathbb{N}} \in R_\varepsilon$, existe $n_0 \geq 2$ donde:

$$\left|IA_n(x_1,\ldots,x_{n-1},IA_{n-1}(x_1,\ldots,x_{n-1})) - IA_{n-1}(x_1,\ldots,x_{n-1})\right| \geq \epsilon, \text{ si } n \geq n_0.$$

Efectivamente, dado $\varepsilon \in (0,1)$, sea

$$R_\varepsilon = \left\{\{x_n\}_{n\in\mathbb{N}},\ x_1 \in \left[0,\frac{1}{2}-\frac{\varepsilon}{2}\right],\ x_2 \in \left[\frac{1}{2}+\frac{\varepsilon}{2},1\right],\ x_k \in [0,1]\ \forall k \geq 3\right\}$$

un conjunto de sucesiones en la cual el primer elemento pertenece a $\left[0,\frac{1}{2}-\frac{\varepsilon}{2}\right]$, el segundo a $\left[\frac{1}{2}+\frac{\varepsilon}{2},1\right]$, y no se impone ninguna otra restricción.

Ahora, es fácil ver que $\mathbb{P}(R_\varepsilon) = \left(\frac{1}{2}-\frac{\varepsilon}{2}\right)^2 > 0$ y también que para cualquier $\{x_n\}_{n\in\mathbb{N}} \in R_\varepsilon$, esto es

$$\left|IA_n(x_1,\ldots,x_{n-1},IA_{n-1}(x_1,\ldots,x_{n-1})) - IA_{n-1}(x_1,\ldots,x_{n-1})\right| \geq \varepsilon \text{ si } n \geq 2,$$

por tanto, la familia $\{IA_n\}_n$, que puede ser vista como un caso particular de la familia de OWA's, es inestable.

3.2. Estabilidad en FAO's bajo conjuntos de datos estructurados.

En esta sección se estudian las propiedades de estabilidad estricta bajo un conjunto de datos de entrada que tiene una estructura inherente. En este caso, el resultado del proceso de agregación varía si se intercambian los elementos del conjunto de datos, por lo tanto, las *FAO's* que se utilizan son familias no simétricas $\{A_n, n \in \mathbb{N}\}$, es decir, aquellas en la cual $\exists n \in \mathbb{N}$ talque A_n es un operador de agregación no simétrico.

En las secciones siguientes se definen propiedades que establecen restricciones para asegurar un nivel de robustez en el proceso de agregación en la cual la información de entrada se encuentra estructurada. Específicamente, nos enfocamos en conjuntos de datos con un ordenamiento lineal, y con un orden jerárquico (ambos asociados o no a una estructura de prioridades).

Del mismo modo como se presentó la estabilidad para datos no estructurados, a lo largo de la sección y para cada tipo de estructura analizada, se incorporan las definiciones de estabilidad para niveles cada vez menos restrictivos, comenzando con la *estabilidad estricta*, para luego seguir con la *estabilidad asintóticamente estricta*, la *estabilidad estricta casi segura* y finalmente la *inestabilidad*.

3.2.1. Datos linealmente ordenados.

A fin de ilustrar la importancia del ordenamiento de los datos frente a un proceso de agregación, estudiamos la familia de operadores binarios idempotentes con extensión inductiva hacia adelante dada por $\{A_n^f : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ y la *FAO* con extensión inductiva hacia atrás dada por $\{A_n^b : [0,1]^n \to [0,1], n \in \mathbb{N}\}$. Es posible ver que

$\{A_n^f\}_n$ satisface la estabilidad estricta definida en (3.1), sin embargo, dado que su agrupación iterativa obedece a un ordenamiento lineal desde la izquierda hacia la derecha dado por $A_2(..(A_2(A_2(x_1,x_2)..),x_{n-1})$, no es posible esperar resultados consistentes en el otro sentido, de derecha a izquierda. Análogamente, $\{A_n^b\}_n$ no satisface la propiedad de estabilidad estricta tal cual como la hemos definido, en tanto su agrupación iterativa obedece a un ordenamiento lineal con una orientación opuesta, que está dada por $A_2(A_1,A_2(...,A_2(x_{n-2},x_{n-1})))$.

Frente a esta situación, podemos ver que cuando el conjunto de datos de entrada tiene una estructura lineal inherente, es necesario identificar la orientación de dicho ordenamiento, y diferenciar si la propiedad de estabilidad estricta es satisfecha desde la derecha o desde la izquierda. Lo que nos lleva a complementar las definiciones de estabilidad estricta para datos linealmente ordenados desde la izquierda.

Definición 3.5. Sea $\{A_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ una familia de operadores de agregación y supongamos que los datos tienen una estructura de orden lineal inherente desde la derecha. Diremos que:

1. $\{A_n\}_{n\in\mathbb{N}}$ es *consistente* desde un punto de vista estable estricto cuando sea una familia R-estrictamente estable, es decir

$$A_n(x_1,x_2,...x_{n-1},A_{n-1}(x_1,x_2...,x_{n-1})) = A_{n-1}(x_1,x_2,...,x_{n-1})$$

$\forall n > 2$ y para todo conjunto de datos linealmente ordenado desde la derecha$\{x_n\}_{n\in\mathbb{N}}$ en $[0,1]$

Definición 3.6. Sea $\{A_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ una familia de operadores de agregación y supongamos que los datos tienen una estructura de orden lineal desde la izquierda. Diremos que:

1. $\{A_n\}_{n\in\mathbb{N}}$ es *consistente* desde un punto de vista estable estricto cuando sea una familia L-estrictamente estable, es decir

$$A_n(A_{n-1}(x_1, x_2, ..., x_{n-1}), x_1, x_2, ...x_{n-1}) = A_{n-1}(x_1, x_2, ..., x_{n-1}) \tag{3.2}$$

   $\forall n > 2$ y para todo conjunto de datos linealmente ordenado desde la izquierda $\{x_n\}_{n\in\mathbb{N}}$ en $[0,1]$

A continuación vemos la demostración de estabilidad estricta de las familias de operadores de agregación recursivas que ejemplificamos anteriormente.

Proposición 3.11. *La familia de operadores binarios idempotentes con extensión inductiva hacia adelante* $\{A_n^f : [0,1]^n \to [0,1], n \in \mathbb{N}\}$*, es una familia R-estrictamente estable (es decir sería consistente para una familia de datos linealmente ordenado desde la derecha).*

Demostración:

$$\begin{aligned} A_n^f(x_1, ...x_{n-1}, A_{n-1}^f(x_1, ..., x_{n-1})) &= A_2(..(A_2(A_2(x_1, x_2)..), x_{n-1}), A_{n-1}^f) \\ &= A_2(A_{n-1}^f, A_{n-1}^f) \\ &= A_{n-1}^f(x_1, ..., x_{n-1}) \end{aligned}$$

la última ecuación sólo es satisfecha si A_2 es un operador idempotente.

Pero el cumplimiento de la estabilidad estricta de esta FAO se da sólo en un sentido, ya que en general $\{A_n^f\}_n$ no es una familia estrictamente estable desde la izquierda. Considere por ejemplo $n = 3$:

$A_3^f(A_2^f(x_1, x_2), x_1, x_2)$ debiera ser equivalente con $A_2^f(x_1, x_2)$, pero

$A_3^f(A_2^f(x_1, x_2), x_1, x_2) = A_2\,(A_2(A_2(x_1, x_2), x_1), x_2)$, en general no es equivalente a $A_2(x_1, x_2)$.

Proposición 3.12. *La familia de operadores binarios idempotentes con extensión inductiva hacia atrás* $\{A_n^b : [0,1]^n \to [0,1], n \in \mathbb{N}\}$*, es una familia L-estrictamente estable.*

Demostración:

$$\begin{aligned} A_n^b(A_{n-1}^b, x_2, ..., x_n) &= A_2(A_{n-1}^b, A_2(..., A_2(x_{n-2}, x_{n-1}))) \\ &= A_2(A_{n-1}^f, A_{n-1}^b) \\ &= A_{n-1}^b(x_1, ...x_{n-1}) \end{aligned}$$

la última ecuación también se cumple si A_2 es un operador idempotente.

En la misma línea analítica anterior, es posible concluir que en general, la familia $\{A_n^b\}_n$ no es una familia *R-estrictamente estable.*

Análogamente a la definición más ampliada de estabilidad, que considera la estabilidad estricta desde la izquierda (3.2), se definen las propiedades de estabilidad asintóticamente estricta, estricta casi seguro e inestabilidad para un conjunto de datos linealmente ordenado.

Definición 3.7. Sea $\{A_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ una familia de operadores de agregación y supongamos que los datos tienen una estructura de orden lineal desde la izquierda. Diremos que:

1. $\{A_n\}_{n\in\mathbb{N}}$ es *consistente* desde un punto de vista asintóticamente estable cuando

sea una familia asintóticamente L-estrictamente estable, es decir:

$$\lim_{n\to+\infty} \big|A_n(A_{n-1}(x_1,...,x_{n-1}),x_1,...x_{n-1}) - A_{n-1}(x_1,...,x_{n-1})\big| = 0$$

para todo conjunto de datos linealmente ordenado desde la izquierda $\{x_n\}_{n\in\mathbb{N}}$ en $[0,1]$

Definición 3.8. Sea $\{A_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ una familia de operadores de agregación y supongamos que los datos tienen una estructura de orden lineal desde la derecha. Diremos que:

1. $\{A_n\}_{n\in\mathbb{N}}$ es *consistente* desde un punto de vista asintóticamente estable cuando sea una familia asintóticamente R-estrictamente estable, es decir:

$$\lim_{n\to+\infty} \big|A_n(x_1,...x_{n-1},A_{n-1}(x_1,...,x_{n-1})) - A_{n-1}(x_1,...,x_{n-1})\big| = 0$$

para todo conjunto de datos linealmente ordenado desde la derecha $\{x_n\}_{n\in\mathbb{N}}$ en $[0,1]$

Definición 3.9. Sea $\{A_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ una familia de operadores de agregación y supongamos que los datos tienen una estructura de orden lineal desde la izquierda. Diremos que:

1. $\{A_n\}_{n\in\mathbb{N}}$ es *consistente* desde un punto de vista estable casi seguro cuando sea una familia casi seguro L-estrictamente estable, es decir

$$\mathbb{P}[\lim_{n\to+\infty}[\ |A_n(A_{n-1}(x_1,...,x_{n-1}),x_1,...x_{n-1}) - A_{n-1}(x_1,...,x_{n-1})| = 0\ ] = 1,$$

para todo conjunto de datos linealmente ordenado desde la izquierda $\{x_n\}_{n\in\mathbb{N}}$ en $[0,1]$ $\forall x_i \sim U(0,1)$

Definición 3.10. Sea $\{A_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ una familia de operadores de agregación y supongamos que los datos tienen una estructura de orden lineal desde la derecha. Diremos que:

1. $\{A_n\}_{n\in\mathbb{N}}$ es *consistente* desde un punto de vista estable casi seguro cuando sea una familia casi seguro R-estrictamente estable, es decir

$$\mathbb{P}[\lim_{n\to+\infty} |A_n(x_1,...x_{n-1}, A_{n-1}(x_1,...,x_{n-1})) - A_{n-1}(x_1,...,x_{n-1})| = 0\ ] = 1,$$

para todo conjunto de datos linealmente ordenado desde la derecha $\{x_n\}_{n\in\mathbb{N}}$ en $[0,1]$, $\forall x_i \sim U(0,1)$

Definición 3.11. Sea $\{A_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ una familia de operadores de agregación. Entonces, diremos que:

1. Satisface la propiedad de R-inestabilidad si la familia no es casi seguro R-estrictamente estable, y entonces será una familia R-inestable.
2. Satisface la propiedad de L-inestabilidad si la familia no es casi segura L-estictamente estable, y entonces será una familia L-inestable.

Otras *FAOs* que suelen satisfacer la estabilidad estricta sólo en un sentido, son las familias de operadores de agregación ponderados. Estas familias tienen una definición de prioridades que está asociada a la posición relativa de los elementos del conjunto de datos de entrada, razón por la cual se consideran familias inestables en general, en tanto su estabilidad depende de la definición de la secuencia de pesos. A continuación se prueba la inestabilidad de algunas de estas *FAO's*.

Proposición 3.13. *La FAO definida por la media ponderada $\{W_n\}_{n\in\mathbb{N}}$ es una familia inestable.*

Demostración.

Sea la secuencia de pesos dada por:

$$w_i^n = \begin{cases} \dfrac{1}{2(n-2)} & \text{if } 1 < i < n \\ 1/4 & \text{if } i = 1 \\ 1/4 & \text{if } i = n \end{cases}$$

nótese que $\sum\limits_{i=1}^{n} w_i^n = 1 \ \forall n$, y entonces:

$$Dif_r(W_n, W_{n-1}) = \sum_{i=1}^{n-1} \left(w_i^n - (1 - w_n^n) w_i^{n-1} \right) x_i =$$
$$\frac{1}{16} x_1 + \sum_{i=2}^{n-2} \frac{n-6}{8n^2 - 40n + 48} x_i - \frac{3n-2}{16n-32} x_{n-1}$$

Asumiendo que todos los valores de x_2,...,x_{n-1} han sido generados por variables aleatorias independientes con distribución $U[0,1]$, en el límite se tiene que $\sum\limits_{i=2}^{n-2} \frac{n-6}{8n^2-40n+48} x_i \in [0, 1/8]$ y $\frac{3n-2}{16n-32} x_{n-1} \in [0, 3/16]$. Por lo tanto, si $\varepsilon = 0{,}01$ para n grande, se tiene que:

$$\left| Dif_r(W_n, W_{n-1}) \right| \ \geq \ \frac{1}{16} \left| x_1 - 3x_{n-1} \right| > \varepsilon = 0{,}01$$

se cumple si y solo si $1 \geq x_1 > 3x_{n-1} + 0{,}16$ o $0 \leq x_1 < 3x_{n-1} - 0{,}16$.

Como cualquier valor $x_{n-1} \in [0,1]$ pertenece a un intervalo con longitud positiva para x_1, en términos de medida de probabilidad se tiene que:

$$P\left(\left\{ (x_n)_{n \in N} \subset [0,1] \ / \lim_{n \to \infty} \left| Dif_r(W_n, W_{n-1}) \right| > 0{,}01 \right\} \right) > 0,$$

y por lo tanto, la FAO de medias ponderadas no es una familia casi seguro estrictamente estable. Además, desde:

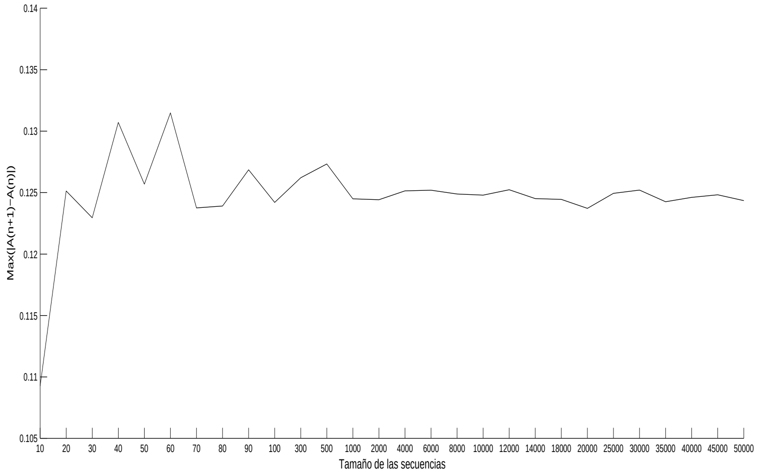
$$
\begin{aligned}
Dif_l(W_n, W_{n-1}) = \sum_{i=1}^{n-1} \left(w_{i+1}^n - (1 - w_1^n) w_i^{n-1} \right) x_i = \\
\frac{2-3n}{16n-32} x_1 + \sum_{i=2}^{n-2} \frac{n-6}{8n^2-40n+48} x_i + \frac{1}{16} x_{n-1},
\end{aligned}
$$

por lo tanto, la FAO dada por la media ponderada es inestable en general.

Del mismo modo como visualizamos el comportamiento de la familia de operadores de agregación dada por el productorio, a fin de visualizar el comportamiento inestable de esta FAO, a continuación se muestran los resultados de otro ejercicio de simulación realizado con la familia recién utilizada en la demostración de inestabilidad de la media ponderada.

Figura 3.3: Inestabilidad de la familia de medias ponderadas cuyos pesos se definen por

$$w_i^n = \frac{1}{(n-2)} \ si \ 1 \leq i \leq n \ \ y \ \ w_i^n = \frac{1}{4} \ si \ i = n \ o \ i = 1.$$



La Figura 3.3 muestra claramente este comportamiento, en tanto los resultados de la agregación de n y $n-1$ elementos no convergen a medida que n aumenta. Además, se evidencia que la proporción de ser estable no converge a 1, como puede verse en la Tabla 3.2.1.

Cuadro 3.2: Simulación de $Pr[\lim_{n\to+\infty} \left|Dif_r(W_n, W_{n-1})\right| < Tolerancia]$ para la FAO de medias ponderadas inestable.

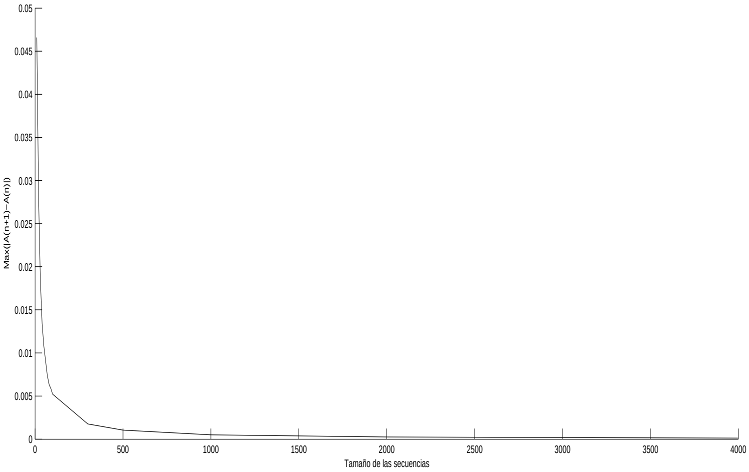
Size of the	Nivel de tolerancia				
sequences	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
10	1.000	0.157	0.012	0.002	0.000
20	0.986	0.123	0.006	0.001	0.000
30	0.978	0.133	0.009	0.000	0.000
40	0.965	0.120	0.015	0.000	0.000
50	0.957	0.117	0.012	0.001	0.000
60	0.960	0.117	0.014	0.000	0.000
70	0.959	0.123	0.008	0.001	0.000
80	0.971	0.121	0.014	0.003	0.000
90	0.965	0.121	0.015	0.003	0.000
100	0.950	0.109	0.009	0.000	0.000
300	0.956	0.101	0.007	0.002	0.000
500	0.935	0.097	0.014	0.001	0.000
1000	0.956	0.110	0.014	0.002	0.000
2000	0.953	0.137	0.006	0.001	0.000
4000	0.924	0.121	0.007	0.001	0.000
6000	0.947	0.105	0.008	0.000	0.000
8000	0.938	0.109	0.015	0.000	0.000
10000	0.941	0.109	0.013	0.001	0.000
12000	0.948	0.100	0.014	0.001	0.000
14000	0.962	0.115	0.008	0.000	0.000
20000	0.953	0.103	0.005	0.001	0.000
30000	0.958	0.098	0.013	0.003	0.000
40000	0.961	0.099	0.013	0.000	0.000
50000	0.943	0.123	0.009	0.002	0.000

Con esto se ilustra la inestabilidad de la familia de operadores de agregación dada por la media ponderara $\{W_n\}_{n \in \mathbb{N}}$. Sin embargo, a continuación veremos que existen familias de la media ponderada que sí satisfacen algún nivel de estabilidad estricta, como es el caso de la familia *asintóticamente R-estrictamente estable*, en la cual sus pesos $w^n \in [0,1]^n$ están definidos por:

$$w_i^n = \begin{cases} \dfrac{1}{(n+1)} & \text{si } 1 \leq i \leq n-1 \\ \dfrac{2}{(n+1)} & \text{si } i = n \end{cases}$$

Figura 3.4: Estabilidad asintótica de la FAO de medias ponderadas cuyos pesos se definen por

$$w_i^n = \frac{1}{(n+1)} \ si \ 1 \leq i \leq n-1 \ \ y \ \ w_i^n = \frac{2}{(n+1)} \ si \ i = n.$$



En la Figura 3.2.1 se observa que esta vez los valores agregados de las secuencias de n y $n-1$ ítems sí tienen mayor similitud a medida que n aumenta. Sin embargo, converge con una velocidad menor que la FAO del productorio. Para este ejemplo, es posible garantizar resultados estables con un nivel de tolerancia no menor a 10^{-3}, para una secuencia de datos cuyo tamaño es no menor a $n = 100$ (Figura 3.2.1).

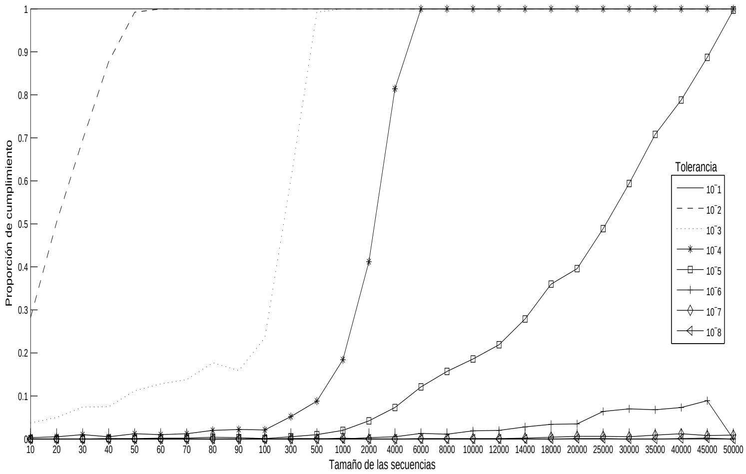
Cuadro 3.3: Simulación de $Pr[\lim\limits_{n\to+\infty} \left|Dif_r(W_n, W_{n-1})\right| < Tolerancia]$ para la familia de medias ponderadas.

Tamaño de las secuencias	Nivel de tolerancia							
	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}
10	1.000	0.289	0.029	0.003	0.000	0.000	0.000	0.000
20	1.000	0.478	0.064	0.004	0.000	0.000	0.000	0.000
30	1.000	0.673	0.067	0.008	0.000	0.000	0.000	0.000
40	1.000	0.875	0.081	0.003	0.001	0.000	0.000	0.000
50	1.000	0.992	0.109	0.017	0.002	0.000	0.000	0.000
60	1.000	1.000	0.123	0.015	0.001	0.000	0.000	0.000
70	1.000	1.000	0.135	0.017	0.001	0.000	0.000	0.000
80	1.000	1.000	0.166	0.008	0.001	0.000	0.000	0.000
90	1.000	1.000	0.196	0.021	0.003	0.000	0.000	0.000
100	1.000	1.000	0.172	0.011	0.000	0.000	0.000	0.000
300	1.000	1.000	0.600	0.054	0.004	0.000	0.000	0.000
500	1.000	1.000	0.990	0.092	0.009	0.002	0.000	0.000
1000	1.000	1.000	1.000	0.207	0.016	0.001	0.000	0.000
2000	1.000	1.000	1.000	0.403	0.043	0.003	0.000	0.000
4000	1.00	1.000	1.000	0.802	0.067	0.007	0.000	0.000
6000	1.000	1.000	1.000	1.000	0.122	0.007	0.003	0.000
8000	1.000	1.000	1.000	1.000	0.144	0.015	0.001	0.000
10000	1.000	1.000	1.000	1.000	0.220	0.019	0.001	0.000
14000	1.000	1.000	1.000	1.000	0.265	0.031	0.001	0.000
20000	1.000	1.000	1.000	1.000	0.389	0.036	0.003	0.000
30000	1.000	1.000	1.000	1.000	0.603	0.067	0.010	0.000
40000	1.000	1.000	1.000	1.000	0.821	0.087	0.003	0.000
50000	1.000	1.000	1.000	1.000	1.00	0.099	0.008	0.000

Nota: Para los tres últimos niveles de tolerancia, converge para n $2x10^6$, $2x10^7$ y $2x10^8$.

Figura 3.5: Velocidad de convergencia de la familia de medias ponderadas cuyos pesos se definen por

$$w_i^n = \frac{1}{(n+1)} \ si \ 1 \leq i \leq n-1 \ \ y \ \ w_i^n = \frac{2}{(n+1)} \ si \ i = n.$$



Frente a esta situación, parece ser necesario poder garantizar algún nivel de estabilidad para la familia de operadores de agregación dada por la media ponderada en función de la generación de la secuencia de sus pesos. A continuación se estudian las condiciones que deben imponerse sobre la secuencia de los pesos a fin de asegurar algún nivel de consistencia en el proceso de agregación.

Para cualquier cardinalidad n del conjunto de datos, usualmente los pesos estan dados por el siguiente vector $w^n = (w_1^n, ..., w_n^n) \in [0,1]^n$, tal que $\sum\limits_{i=1}^{n} w_i^n = 1$, y la media ponderada se define de la siguiente manera $W_n(x_1, \ldots, x_n) = \sum\limits_{i=1}^{n} x_i \cdot w_i$. Los pesos asociados a los elementos del conjunto de datos de entrada representan la *importancia* que tiene cada uno de estos elementos en el proceso de agregación. Es por esto que la media ponderada es uno de los operadores de agregación más utilizados en diferentes áreas: estadística, lógica borrosa, decisión multicriterio, entre otros, siendo la determinación de esta *importancia* definida por los pesos, uno de los problemas más estudiados en todas estas áreas del conocimiento. Ahora veremos la relación que debe existir entre dos vectores de pesos de dimensión r y s a fin de producir un proceso de agregación consistente.

Para ilustrar este objetivo, se presenta el siguiente ejemplo. Se tiene un problema de decisión multicriterio de cuatro criterios dados por C_1, C_2, C_3 *y* C_4. Un jurado evalúa las diferentes alternativas en base a estos cuatro criterios, y luego utiliza la media ponderada como regla de agregación. El objetivo no es decidir cómo debería ser el vector de pesos $w^4 = (w_1^4, w_2^4, w_3^4, w_4^4)$, sino que garantizar alguna estabilidad o consistencia en el proceso de agregación. Por ejemplo, si tenemos cuatro criterios, parece poco coherente evaluar a un candidato con $w^4 = (1/4, 1/4, 1/4, 1/4)$ para luego evaluarlo con $w^3 = (0{,}8, 0{,}2, 0)$ en caso que el criterio C_4 fuera descartado. Desde el punto de vista de la consistencia, este proceso de agregación no sería estable. El objetivo es entonces determinar la relación que debe existir entre estos vectores de criterios de diferentes dimensiones a fin de garantizar un proceso de agregación consistente.

Si se utiliza la usual secuencia de pesos dada por: $w_i^n = \frac{v_i}{\sum\limits_{i=1}^{n} v_i}, \forall i \in \mathbb{N}, v_i \in \mathbb{R}^+$, entonces la correspondiente familia de medias ponderadas $\{W_n\}_{n\in\mathbb{N}}$ es una familia *R-estrictamente estable*. Por lo tanto, si los pesos son construidos de manera que se preserve esta propiedad, se debe establecer alguna relación entre los pesos w^n y los pesos anteriores $w^{n-1}, w^{n-2}, \ldots$. Sin embargo, si se analiza la estabilidad estricta desde la izquierda, es posible ver que $\{W_n\}_{n\in\mathbb{N}}$ no es una familia *L-estrictamente estable*.

A continuación se presentan condiciones necesarias y suficientes que garantizan la estabilidad estricta y asintóticamente estricta de una FAO definida mediante la media ponderada.

Notar que para la *FAO* $\{W_n\}_{n\in\mathbb{N}}$ y sus pesos w^n *con* $n \in \mathbb{N}$, si:

$$Dif_r(W_n, W_{n-1}) = W_n(x_1, \ldots, x_{n-1}, W_{n-1}(x_1, \ldots, x_{n-1})) - W_{n-1}(x_1, \ldots, x_{n-1}),$$

entonces la propiedad de estabilidad estricta puede ser expresada como:

$$0 = Dif_r(W_n, W_{n-1}) = \sum_{i=1}^{n-1} (w_i^n - (1 - w_n^n)w_i^{n-1})\ x_i$$

para toda colección de datos $(x_1, \ldots, x_n) \in [0,1]^n$. Similarmente, si:

$$Dif_l(W_n, W_{n-1}) = W_n(W_{n-1}(x_1, \ldots, x_{n-1}), x_1, \ldots, x_{n-1}) - W_{n-1}(x_1, \ldots, x_{n-1}),$$

entonces la estabilidad estricta desde la izquierda es equivalente con:

$$0 = Dif_l(W_n, W_{n-1}) = \sum_{i=1}^{n-1} (w_{i+1}^n - (1 - w_1^n)w_i^{n-1})\ x_i$$

para toda colección de datos $(x_1, \ldots, x_n) \in [0,1]^n$. Esto se traduce en las siguientes proposiciones:

Proposición 3.14. *Sea* $w^n = (w_1^n, ..., w_n^n) \in [0,1]^n, n \in \mathbb{N}$ *una secuencia de pesos de la familia media ponderada* $\{W_n, n \in \mathbb{N}\}$ *tal que* $\sum_{i=1}^{n} w_i^n = 1$ *se satisface* $\forall n \geq 2$. *Entonces, la FAO* $\{W_n\}_{n\in\mathbb{N}}$ *es una familia R-estrictamente estable si y solo si la secuencia de pesos satisface:*

$$w_i^n = (1 - w_n^n) \cdot (w_i^{n-1})\ \forall n \in \mathbb{N}. \tag{3.3}$$

Demostración:

Directo desde análisis previos.

Observación 3.1. Observe que si la familia $\{W_n\}_{n\in\mathbb{N}}$ es R-strictamente estable, entonces cualquier vector de pesos $w^r = (w_1^r, \ldots, w_r^r)$ puede ser construido desde w^s, donde $r \leq s$. Por ejemplo, si $w^5 = (1/5, \ldots, 1/5)$, entonces se tiene que $w^r = (1/r, \ldots, 1/r)$ para cualquier $r \leq 5$. Además, para cualquier $r > 5$ el vector de pesos debe ser $w^r = (w, w, w, w, w, w_6^r, \ldots, w_r^r)$ para garantizar estabilidad estricta desde la derecha (i.e. los cinco primeros ítems deben cohincidir). Ahora bien, si $\{W_n\}_{n\in\mathbb{N}}$ es L-strictamente estable, entonces el vector de pesos $w^r = (w_1^r, \ldots, w_r^r)$ puede ser construido desde w^s, donde $r \leq s$. Por ejemplo, si $w^5 = (0{,}3, 0{,}2, 0{,}1, 0, 0{,}4)$, entonces se tiene que $w^4 = \left(\frac{0{,}2}{0{,}7}, \frac{0{,}1}{0{,}7}, \frac{0}{0{,}7}, \frac{0{,}4}{0{,}7}\right)$, $w^3 = \left(\frac{0{,}1}{0{,}5}, 0, \frac{0{,}4}{0{,}5}\right)$ y $w^2 = (0, 1)$.

Proposición 3.15. *Sea* $w^n = (w_1^n, ..., w_n^n) \in [0,1]^n, n \in \mathbb{N}$ *una secuencia de pesos de la FAO media ponderada* $\{W_n, n \in \mathbb{N}\}$ *tal que* $\sum_{i=1}^{n} w_i^n = 1$ *se satisface* $\forall n \geq 2$. *Entonces, la FAO* $\{W_n\}_{n\in\mathbb{N}}$ *es una familia L-estrcitamente estable si y sólo si la secuencia de pesos satisface:*

$$w_{i+1}^n = (1 - w_1^n) \cdot (w_i^{n-1}) \ \forall n \in \mathbb{N}. \tag{3.4}$$

Demostración:

Directo desde análisis previos.

Observación 3.2. De nuevo cabe remarcar que si la familia $\{W_n\}_{n\in\mathbb{N}}$ es L-estrictamente estable, entonces es posible construir un vector de pesos $w^r = (w_1^r, \ldots, w_r^r)$ cuando w^s es conocido, donde $r \leq s$. Por ejemplo si $w^5 = (1/5, \ldots, 1/5)$, entonces se tiene que $w^r = (1/r, \ldots, 1/r)$ para cualquier $r \leq 5$. Además, para cualquier $r > 5$, si se asume estabilidad estricta desde la izquierda entonces $w^r = (w_1^r, \ldots, w_{r-6}^r, w, w, w, w, w)$ (i.e. los cinco últimos items cohinciden).

De manera similar, a continuación se estudian las condiciones suficientes que garantizan estabilidad asintótica para una FAO definida por la media ponderada.

Sea $x_i \in [0,1]$, notar que para todo $n \in \mathbb{N}$:

$$\left|Dif_r(W_n, W_{n-1})\right| = \left| \sum_{i=1}^{n-1} \left(w_i^n - (1-w_n^n)w_i^{n-1}\right)\ x_i \ \right| \leq \ z(n)\cdot g(n) \ \leq \ (n-1)\cdot g(n)$$

donde $z(n) = \sharp\{i = 1, ..., n-1 \ / \ w_i^n - (1-w_n^n)w_i^{n-1} \neq 0\}$ y $g(n) = Max_{i<n}\{w_i^n - (1 - w_n^n)\ w_i^{n-1}\}$. Por lo tanto, si $\forall i$ las sucesiones $w_i^n - (1-w_n^n)\ w_i^{n-1}$ tienden a cero, es suficiente para imponer la condición $\lim\limits_{n\to+\infty} g(n)\cdot z(n) \ = \ 0$ a fin de garantizar que la FAO definida por la media ponderada resultante es una familia asintóticamente *R-estrictamente estable*. Un razonamiento similar se puede realizar para el caso de la estabilidad asintótica desde la izquierda.

A partir de esto se deducen los siguients resultados:

Proposición 3.16. *Sea $w^n = (w_1^n, ..., w_n^n) \in [0,1]^n, n \in \mathbb{N}$ una secuencia de pesos de la FAO definida por la media ponderada $\{W_n, n \in \mathbb{N}\}$ tal que $\sum\limits_{i=1}^{n} w_i^n = 1$ se satisface $\forall n$. Entonces, la familia $\{W_n\}_{n\in\mathbb{N}}$ es asintóticamente R-estrictamente estable si:*

$$\lim_{n\to+\infty} \left(w_i^n - (1-w_n^n)w_i^{n-1}\right) = 0, \quad \forall\ 1 \leq i \leq n-1 \ y \ \lim_{n\to+\infty} g(n)\cdot z(n) \ = \ 0 \ .$$

Demostración:

Directo desde análisis previos.

Proposición 3.17. *Sea* $w^n = (w_1^n, ..., w_n^n) \in [0,1]^n, n \in \mathbb{N}$ *una secuencia de pesos de la FAO definida por la media ponderada* $\{W_n, n \in \mathbb{N}\}$ *tal que* $\sum_{i=1}^{n} w_i^n = 1$ *se satisface* $\forall n$*. Entonces, la familia* $\{W_n\}_{n\in\mathbb{N}}$ *es asintóticamente L-estrictamente estable si:*

$$\lim_{n\to+\infty} \left(w_{i+1}^n - (1 - w_n^n)w_i^{n-1}\right) = 0, \quad \forall\ 1 \le i \le n-1 \ y \ \lim_{n\to+\infty} g(n) \cdot z(n) \ = \ 0 \ .$$

Demostración:

Directo desde análisis previos.

Nótese que el concepto de estabilidad estricta aplicado bajo un conjunto de datos linealmente estructurado, lo hemos definido suponiendo que el último o primer elemento que se incorpora al conjunto de datos, es la agregación del resto de los elementos (R o L estabilidad respectivamente). No obstante, estas propiedades se pueden extender suponiendo que tal elemento se incorpora a la posición i-ésima desde la derecha (i-R estabilidad) o a la posición j-ésima desde la izquierda (j-L estabilidad). Esta extensión ayuda resolver el problema de información faltante en alguna variable de un elemento del conjunto de datos, en tanto permite generar un operador de la misma familia pero de una dimensión menor, manteniendo la estabilidad de los resultados de dicho elemento.

3.2.2. Datos bi-direccionales.

El cumplimiento de las propiedades de estabilidad estricta en ambos sentidos, desde la derecha y desde la izquierda, sólo es posible para el caso de las *FAO's* simétricas, es decir, bajo datos no estructurados. No obstante, cuando tenemos un conjunto de datos linelamente ordenado, pudiera existir el caso en el cual a partir de un cierto $i \in \{2, 3, .., n\}$, los datos cambian de sentido en su orden lineal, ante lo cual, la consistencia del proceso de agregación generado a partir de la familia $\{A_n\}_{n\in\mathbb{N}}$, debe ser analizada bajo las propiedades de estabilidad estricta en ambos sentidos, es decir, bi-direccionalmente.

Definición 3.12. Sea $\{A_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ una familia de operadores de agregación y supongamos que los datos tienen una estructura de orden lineal bi-direccional Diremos que: $\{A_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ es una familia de operadores de agregación *consistente* desde un punto de vista estable estricto si dicha familia es LR estrictamente estable.

De manera análoga a las definiciones de estabilidad estricta bajo un conjunto de datos con orden lineal unidireccional, a continuación se define la estabilidad estricta asintótica y casi segura, además de la inestabilidad.

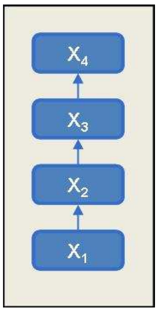
Definición 3.13. Sea $\{A_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ una familia de operadores de agregación y supongamos que los datos tienen una estructura de orden lineal bi-direccional Diremos que: $\{A_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ es una familia de operadores de agregación *consistente* desde un punto de vista asintóticamente estable estricto si dicha familia es LR asintóticamente estrictamente estable.

Definición 3.14. Sea $\{A_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ una familia de operadores de agregación y supongamos que los datos tienen una estructura de orden lineal bi-direccional Diremos que: $\{A_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ es una familia de operadores de agregación *consistente* desde un punto de vista casi seguro si dicha familia es LR asintóticamente casi seguro estrictamente estable.

3.2.3. Datos con estructura jerárquica priorizada.

La estructura de los datos sobre las cuales se construyen índices es con frecuencia de naturaleza jerárquica, y tiene un ordenamiento lineal asociado a prioridades. El caso más simple de estructura jerarquica corresponde a un conjunto de datos $\{x_n\}_{n\in\mathbb{N}}$ con orden lineal, en el cual sus itmes tienen definida una estructura de prioridades respecto de su posición relativa dada por $x_1 <_{prio} x_2 <_{prio} ... <_{prio} x_n$. Se puede considerar que este conjunto tiene una estructura jerárquica si el nivel de importancia de $x_{i/\ i\in\{1,...n\}}$ se define en función de su posición i-ésima, siendo entonces el elemento i menos prioritario que el $i+1$ y más que el elemento $i-1$, tal como se muestra en la *Figura 3.6.*

Figura 3.6: **Representación de un conjunto de datos priorizado y linealmente ordenado.**



Nota: $X_1 <_{prio} X_2 <_{prio} X_3 <_{prio} X_4$.

Este tipo de estructuras son muy usuales en variados problemas de agregación, en la cual el último elemento tiene mayor importancia que los primeros. En estas situaciones, parece aún más razonable suponer que si el último elemento es la agregación de los previos, entonces la agregación final debiera cohincidir con el resultado de la agregación previa, en tanto este último ítem representa una confirmación de la información que ya ha sido agregada. Por ejemplo, supongamos que con la información disponible hasta ayer, la probabilidad que llueva es 0,6, si la información de hoy es una confirmación de esto, entonces nuestra estimación no debiera variar.

Ahora bien, el razonamiento anterior no es evidente si se consideran las condiciones de estabilidad para datos con orden lineal pero en el sentido contrario, es decir, desde la izquierda. Si el conjunto de datos está linealmente ordenado y contenido en una estructura jerárquica, necesariamente los elementos que se encuentran en el primer nivel de la estructura tienen una mayor prioridad, y por tanto, esta igualdad $V_n(V_{n-1}(x_1, x_2 ..., x_{n-1}), x_1, x_2, ...x_{n-1}) = V_{n-1}(x_1, x_2, ..., x_{n-1})$ no será satisfecha necesariamente.

Por ejemplo, considere la *FAO* linealmente estructurada $\{V_n\}_{n\in\mathbb{N}}$ dada por:

$$V_n(x_1, \ldots, x_n) = \sum_{i\geq\frac{n+1}{2}} w_i x_i$$

Esta familia le otorga más importancia a los elementos que se encuentran en la parte superior de la estructura, eliminándo del proceso los elementos que se encuentra bajo la mitad de la estructura. Parece lógico que esta *FAO* se considere una familia estrictamente estable, a pesar que no satisfaga la propiedad desde la izquierda, en tanto agrega información sobre un conjunto de datos con una estructura jerárquica y linealmente ordenado desde la derecha, en el cual los últimos elementos tienen mayor prioridad que los primeros.

Teniendo en cuenta esto, la noción de estabilidad o consistencia para familias de operadores de agregación con estructura de datos lineal priorizada que cabría esperar es la noción de consistencia desde un punto de vista estable que hemos introducido para datos linealmente ordenados tanto desde un punto de vista estricto estable, asíntotico como casi seguro.

Por ejemplo, si la información a ser agregada tiene un orden lineal priorizado como se muestra en Figura 3.6. Entonces, la consistencia desde un punto de vista estricto estable sería:

$$V_n(x_1, x_2, ...x_{n-1}, V_{n-1}(x_1, x_2 ..., x_{n-1})) = V_{n-1}(x_1, x_2, ..., x_{n-1})$$

De manera análoga, la noción de consistencia que cabría esperar para dicha familia $\{V_n\}_n$ desde un punto de vista asintóticamente estable sería:

$$\lim_{n\to+\infty}\left|V_n(x_1,...x_{n-1},V_{n-1}(x_1,...,x_{n-1}))-V_{n-1}(x_1,...,x_{n-1})\right|=0$$

$\forall\{x_n\}_{n\in\mathbb{N}}$ en $[0,1]$.

Finalmente, la noción de consistencia que cabría esperar para dicha familia $\{V_n\}_n$ desde un punto de vista asintóticamente casi seguro sería:

$$\mathbb{P}[\lim_{n\to+\infty}\left|V_n(x_1,...x_{n-1},V_{n-1}(x_1,...,x_{n-1}))-V_{n-1}(x_1,...,x_{n-1})\right|=0]=1.$$

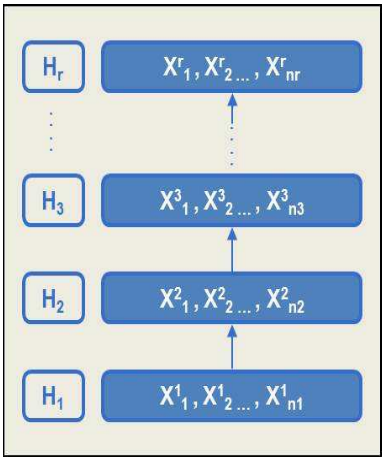
$\forall\{x_n\}_{n\in\mathbb{N}}$ en $[0,1]$ $\forall x_i\sim U(0,1)$.

Una *FAO* aplicada sobre un conjunto de datos linealmente priorizados $\{V_n\}_{n\in\mathbb{N}}$, debería satisfacer otras propiedades además de la estabilidad estricta, por ejemplo, no debiera ser simétrica como se ha comentado anteriormente. Además, en nuestra opinión, estas familias deberían dar mas relevancia a los ultimos elementos que a los primeros, entre otras propiedades.

La noción de estructura jerárquica de un conjunto de datos se utiliza para representar un conjunto en la cual existen dos posibles tipos de relación entre sus elementos: desestructurada y estruturada. Por lo tanto, la información de entrada contiene algunas preferencias o prioridades para un subconjunto de datos que deben ser previamente consideradas en el proceso de agregación.

En el conjunto de datos de entrada existen los conglomerados $H_1,\dots H_r$, en el cual es posible visualizar distintos niveles de preferencias o prioridades, en tanto H_r es más importante en el proceso de agregación que H_{r-1}, y más aún que H_1, existiendo un orden lineal entre los diferentes niveles $\{H_i\}_{i\in\{1,...r\}}$. Los elementos o variables de la data que están asociados a cada conglomerado son desestructurados, en tanto no se han definido niveles de preferencias *a priori* para ellos.

Figura 3.7: **Representación de un conjunto de datos con estructura jerárquica priorizada.**

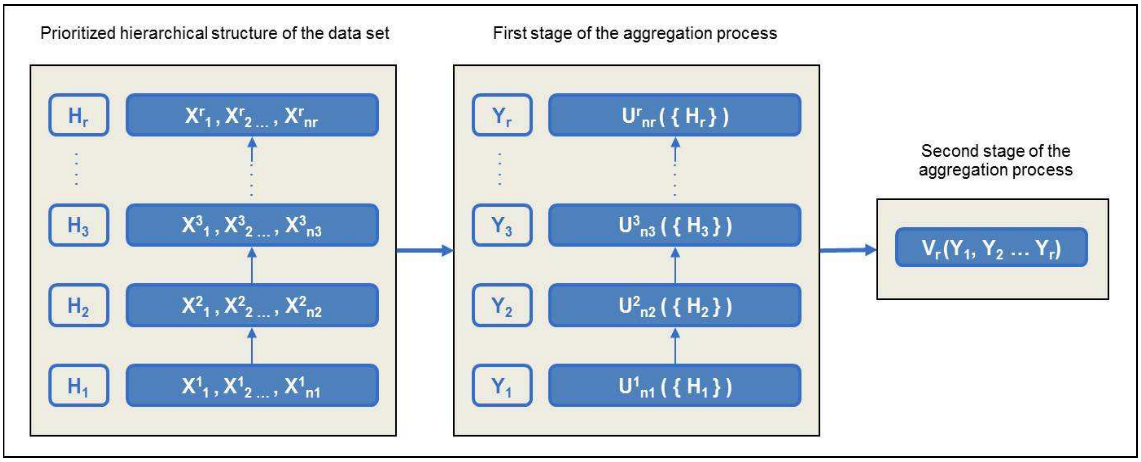


Definición 3.15. Sea X un conjunto de datos. X tiene una estructura jerárquica priorizada si es posible clasificar los elementos de X en r conglomerados $H_1, \ldots H_r$, satisfaciendo las siguientes condiciones:

- $\{H_1, \ldots H_r\}$ constituye una partición de X, i.e. $\bigcup_{i=1}^{r} H_i = X$, y $H_i \cap H_j = \emptyset$, si $i \neq j$.
- Los subconjuntos $H_1, \ldots H_r$ se definen en base a un orden lineal en el siguiente sentido: Dados dos elementos $x, y \in X$, si $x \in H_i$, e $y \in H_j$, con $i < j$, entonces el elemento y tiene más importancia en la agregación que el elemento x.
- Los elementos que pertenecen a la misma clase son agregados como en el caso desestructurado. Dos elementos $x, y \in X$, si $x, y \in H_i$ para $i \in \{1, \ldots, r\}$, tienen la misma importancia en el proceso de agregación.

Es posible considerar estructuras jerárquicas más complejas que la estructura propuesta en la definición 3.15, por ejemplo, con algunos niveles solapados. Sin embargo, este trabajo no se centra en este tipo de estructuras jerárquicas. Es importante enfatizar que si un proceso de agregación se realiza sobre un conjunto de datos de entrada con estructura jerárquica priorizada, el proceso consta de dos pasos. Primero, se agregan los elementos que pertenecen al mismo conglomerado, donde no se definen preferencias o prioridades, y por tanto la información se agrega de manera desestructurada. Y luego que esta agregación se ha realizado, la información agregada para cada conglomerado es agregada de acuerdo a su importancia en el proceso de agregación (ver un ejemplo de operador priorizado en [92]).

Figura 3.8: **Proceso de agregación de un conjunto de datos jerárquico priorizado.**



Esto nos lleva a la siguiente definición:

Definición 3.16. Sea $\{V_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ y $\{U_n^i : [0,1]^n \to [0,1], n, i \in \mathbb{N}\}$, familias de operadores de agregación que representan dos niveles (**v**ertical y **u**nstructured) en la agregación de un conjunto de datos priorizados jerárquicamente X. Se define una familia de operadores de agregación jerárquica priorizada $\{HIE_n\}_{n\in\mathbb{N}}$ compuesta por $\{V_n\}_{n\in\mathbb{N}}$, y $\{U_n^i\}_{n\in\mathbb{N}}$ para $i \in \mathbb{N}$ como:

$$HIE_{|X|}(X) = V_r \left(U^1_{|H_1|}(H_1), U^2_{|H_2|}(H_2), \ldots U^r_{|H_r|}(H_r) \right)$$

Notar que si $|H_i| = 1 \ \forall \ i = 1, \dots r$, se obtiene una familia de operadores de agregación linealmente estructurada. En el siguiente ejemplo se muestra que algunos operadores de agregación jerárquico priorizados, pueden ser vistos como un caso particular de la definición (3.16).

Ejemplo 3.1. La familia de operadores de agregación priorizados definida en [92] puede ser vista como un caso particular de $\{HIE_n\}_{n\in\mathbb{N}}$, donde la familia de operadores de agregación vertical $\{V_n\}_{n\in\mathbb{N}}$ se define como $V_n(y_1, \dots y_n) = \sum\limits_{i=1}^{n} w_i \prod\limits_{k=1}^{i} y_k \ \forall n \in \mathbb{N}$. Y la familia de operadores de agregación desestructurados $\forall l \ \{U_n^l\}_{n\in\mathbb{N}}$, corresponde a la familia definida por el mínimo $\{Min_n\}_{n\in\mathbb{N}} \ \forall l$.

Por lo tanto, para garantizar algún nivel de robustez en un proceso de agregación en el que está involucrado un conjunto de datos con estructura jerárquica priorizada como en el caso del ejemplo anterior, la definición de sus pesos debe satisfacer algunas propiedades similares a las descritas para datos linealmente ordenados. A continuación se extienden las condiciones para la secuencia de pesos dadas en (3.3) y (3.4), a fin de producir comportamientos estables bajo contextos jerárquicos.

Definición 3.17. Dada una familia de operadores de agregación jerárquica priorizada $\{HIE_n\}_{n\in\mathbb{N}}$, la cual está compuesta por $\{V_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$, y $\{U_n^i : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ para $i \in \mathbb{N}$. Diremos que $\{HIE_n\}_{n\in\mathbb{N}}$ es una familia *consistente* desde un punto de vista estrictamente estable para esa estructura de datos si y solo si $\{V_n\}_{n\in\mathbb{N}}$ es una familia R-estrictamente estable, y $\forall \ i \in \mathbb{N}$, la familia $\{U_n^i\}_{n\in\mathbb{N}}$ es una FAO simétrica y estrictamente estable.

Análogamente, a continuación se definen las nociones de estabilidad asintótica y casi segura para una familia de operadores de agregación jerárquica priorizada $\{HIE_n\}_n$.

Definición 3.18. Dada una familia de operadores de agregación jerárquica priorizada $\{HIE_n\}_{n\in\mathbb{N}}$, la cual está compuesta por $\{V_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$, y $\{U_n^i : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ para $i \in \mathbb{N}$. Diremos que $\{HIE_n\}_{n\in\mathbb{N}}$ es una familia *consistente* desde un punto de vista asintóticamente estable para esa estructura de datos si y sólo si $\{V_n\}_{n\in\mathbb{N}}$ es una familia R-asintoticamente estrictamente estable, y $\forall\ i \in \mathbb{N}$, la familia $\{U_n^i\}_{n\in\mathbb{N}}$ es una FAO simétrica y estrictamente estable.

Definición 3.19. Dada una familia de operadores de agregación jerárquica priorizada $\{HIE_n\}_{n\in\mathbb{N}}$, la cual está compuesta por $\{V_n : [0,1]^n \to [0,1], n \in \mathbb{N}\}$, y $\{U_n^i : [0,1]^n \to [0,1], n \in \mathbb{N}\}$ para $i \in \mathbb{N}$. Diremos que $\{HIE_n\}_{n\in\mathbb{N}}$ es una familia *consistente* desde un punto de vista casi seguro estrictamente estable si y sólo si $\{V_n\}_{n\in\mathbb{N}}$ es una familia casi seguro estrictamente R-estable, y $\forall\ i \in \mathbb{N}$, la familia $\{U_n^i\}_{n\in\mathbb{N}}$ es una FAO simétrica y estrictamente estable.

Retomando el *Ejemplo 3.1*, la siguiente proposición muestra que la consistencia desde un putno de vista de la estabilidad estricta del operador priorizado propuesto por Yager en [92] depende de la definición de la secuencia de los pesos de $\{V_n\}_{n\in\mathbb{N}}$, en tanto $\{Min_n\}_{n\in\mathbb{N}}$ es un FAO estrictamente estable y la estabilidad de $\{V_n\}_{n\in\mathbb{N}}$ depende de cómo se genere la secuencia de pesos.

Proposición 3.18. *Sea* $\{HIE_n\}_{n\in\mathbb{N}}$ *una* FAO *jerárquica priorizada.*

$$HIE_{|X|}(X) = V_r\left(U^1_{|H_1|}(H_1), U^2_{|H_2|}(H_2), \ldots, U^r_{|H_r|}(H_r)\right)$$

donde $\{U^i_{n_i} : [0,1]^{n_i} \to [0,1], n_i \in \mathbb{N}\}$ *y* $\{V_r : [0,1]^r \to [0,1], r \in \mathbb{N}\}$ *son definidas por:*

- $\{U^i_{n_i}\}_{n_i} = \{Min_{n_i}(H_i)\}$, *donde* H_i *es el i-ésimo elemento (desestructurado) de la partición jerárquica del conjunto de datos* X $\forall i \in \{1, 2, ..., r\}$
- $V_r(Y_1, Y_2, ..., Y_r) = \sum\limits_{j=1}^{r} w_j \prod\limits_{k=1}^{j} Y_k$

entonces, $\{HIE_n\}_{n\in\mathbb{N}}$ *es una* FAO *jerárquica priorizada consistente desde un punto de vista estrictamente estable si y sólo si la secuencia de pesos* $w^r = (w_1^r, w_2^r, ..., w_r^r)$ *de* $\{V_r\}_{r\in\mathbb{N}}$ *satisface* $w_k^r = (1 - w_r^r) \cdot (w_k^{r-1})\ \forall r \in \mathbb{N}$.

3.3. Construcción de índices de entorno del hogar de la población chilena en base a FAO's consistentes.

3.3.1. Descripción del contexto estudiado.

La Junta Nacional de Jardines Infantiles JUNJI, es una institución del Estado de Chile vinculado al Ministerio de Educación, cuya misión es "*Brindar educación inicial de calidad a niños y niñas menores de cuatro años en situación de vulnerabilidad, garantizando su desarrollo en igualdad de oportunidades, a través de la creación, promoción, supervisión y certificación de salas cuna y jardines infantiles administrados directamente o por terceros*".

A comienzos del mandato de la presidenta de la república de Chile Michelle Bachelet (2006-2010), se constituyó el *Consejo asesor presidencial* para la reforma de las políticas de infancia, a fin de "*...instalar un sistema de protección a la infancia destinado a igualar las oportunidades de desarrollo de los niños y niñas chilenos en sus primeros ocho años de vida, independiente de su origen social, género, o de la conformación de su hogar.*". Este Consejo elaboró un diagnóstico y generó una propuesta de políticas y programas que coordinen y orienten esfuerzos públicos y privados, a fin de asegurar a todas las niñas y niños de Chile un proceso de desarrollo pleno y equilibrado durante sus primeros años de vida.

En el informe realizado el año 2006 por dicho Consejo se concluye que la educación inicial en Chile no constituye un sistema con una institucionalidad única que tenga funciones claras respecto de estándares exigibles, de garantías de acceso y de mecanismos de financiamiento coordinados, entre otros elementos. Frente a este sistema ausente de políticas claras que asegure calidad en la prestación a una cobertura significativa, el Consejo propone un sistema que se fundamenta en el principio consagrado en la *Convención de derechos del niño* que señala que los derechos de niños y niñas son universales, y es responsabilidad de la sociedad en su conjunto garantizarlos. Por lo cual, tal sistema debe garantizar prestaciones a todos los niños a partir de sus necesidades, en base a relaciones de cooperación entre las familias, la comunidad organizada, la institucionalidad pública del ámbito territorial, la sociedad civil y el gobierno central, para el logro de metas universales de desarrollo infantil, y respetando las opciones y necesidades laborales y educacionales de las familias, apoyando de manera especial a niños cuyos padres y/o madres trabajan o estudian fuera del hogar, independientemente de la realidad socioeconómica, pero priorizando a las familias del quintil más pobre de la población.

Además del alineamiento político definido por el Consejo, la presidenta Michelle Bachelet, en discursos realizados en el contexto de la cuenta pública anual, comprometió que durante su período presidencial se ampliará la cobertura de educación pública para la primera infancia en 70,000 nuevos cupos en el nivel sala cuna (entre 3 meses y 1 año 11 meses de edad) y en 43,000 nuevos cupos en el nivel medio (entre 2 y 4 años de edad). Este compromiso fue clave para la posterior y gran ampliación de cobertura en educación pública hacia la primera infancia realizada en Chile, determinando con esto, la manera de educar a la población infantil chilena de menores ingresos, en base a una educación formal, a fin de facilitar el desarrollo económico de sus familias a través del fomento del trabajo de sus madres.

Esta política de ampliación de cobertura en educación pública hacia la primera infancia se basa en que *la generación de ingresos por parte del o de los adultos responsables de los niños, es una condición fundamental para el crecimiento, desarrollo y bienestar de los mismos* (Consejo asesor presidencial para la reforma de las políticas de infancia. 2006). Fue implementada de manera coordinada entre los distintos organismos involucrados en la educación hacia la primera infancia; JUNJI, INTEGRA y Ministerio de Educación, basándose en un trabajo territorial, y en una política focalizada dirigida a la población de menores ingresos, priorizando a hijos de madres que trabajan o estudian fuera del hogar.

Esta decisión política corresponde a uno de los posibles caminos con miras al fomento igualitario del desarrollo de la primera infancia, y a conciliar el cuidado materno de la población infantil con el desarrollo laboral de la mujer y económico de la familia, por lo que se evidenció la necesidad de evaluar su efectividad en términos del logro de desarrollo de los niños. Frente a esta necesidad, la JUNJI realizó un estudio longitudinal cuasi experimental caso-comparación orientado a explorar las condiciones familiares de los niños chilenos en condición de vulnerabilidad que puedan afectar en su desarrollo, al mismo tiempo de conocer los logros conseguidos por ellos en distintas etapas de su corta vida, y a través de esto, poder estimar la efectividad de la política de ampliación de cobertura en educación infantil, respecto del logro de desarrollo de los niños asistentes a un establecimiento de educación pública infantil.

En este contexto se suscribe esta aplicación, a fin de proporcionar una medida contextualizada de desarrollo infantil temprano que entregue resultados más adecuados a la realidad de la población infantil chilena que se encuentra en condición de vulnerabilidad, y a través de esto, poder evaluar dicha política de una manera más estricta en términos metodológicos. A continuación se describe el origen y características metodológicas de los datos utilizados.

Se utiliza la información cuantitativa de un estudio longitudinal realizado por el *Centro de estudios de desarrollo y estimulación psicosocial (CEDEP)* durante los años comprendidos entre el 2007 y el 2009 bajo petición de la JUNJI. El propósito del estudio fue obtener información relevante, confiable y válida que ilumine la práctica institucional, a través del nivel de desarrollo/aprendizaje que logran los niños y niñas que habiendo ingresado a Sala cuna menor en 2007, continúen asistiendo a jardines infantiles de JUNJI, hasta el nivel Medio mayor en 2010.

El estudio se inscribe en un enfoque que plantea que el nivel de desarrollo/aprendizaje que expresan los niños depende del inter−juego entre las diferentes fuentes de influencia ambiental, que pueden ser favorecedoras u obstaculizadoras en diversos grados.

Respecto del tipo de estudio, es un estudio de cohorte prospectivo, de tres años de duración, longitudinal con medidas repetidas, cuasi experimental caso−comparación. La población objetivo se divide en grupo experimental y grupo de comparación. El grupo experimental corresponde a lactantes menores de 15 meses en mayo del año 2007 y que cursaban sala cuna menor a principios del año en 2007, y que continúan asistiendo durante los años 2008 y 2009. Y el grupo comparación a lactantes menores de 15 meses en mayo del año 2007 y que NO cursaban sala cuna menor a principios del año en 2007, y que continúan no asistiendo durante los años 2008 y 2009.

El muestreo fue probabilístico, estratificado y por conglomerados, considerando las regiones como estrato y los centros educativos como conglomerados para el grupo experimental. Mientras que para el grupo comparación, el muestreo fue aleatorio simple entre los asistentes a los centros de salud asociados a los centros educativos de la muestra, manteniendo la afijación proporcional del grupo experimental.

La distribución de la muestra a lo largo de los años es la siguiente:

Cuadro 3.4: Distribución de la muestra.

Grupo	2007	2008	2009
Experimental (escolarizados)	427	315	239
De comparación (no escolarizados)	184	114	65
Total	611	429	304

Como todo estudio longitudinal, la muestra inicial seleccionada el año 2007 sufre modificaciones a los largo de los años, ya que existen abandonos de participantes en el estudio debido a cambio de domicilio o pérdida de alguna característica esencial, como la incorporación a un establecimiento educacional por parte niños del grupo comparación, o abandono del establecimiento educacional por parte de niños del grupo experimental.

Respecto de la captura de la información, se realiza en base a dos instrumentos de medición:

- La adaptación española de De la Cruz y González del *Inventario de Desarrollo*

Battelle de Newborg, Stock y Wnek para medir el desarrollo/aprendizaje de los niños en las áreas motora, comunicación, cognitiva, personal-social y adaptativa.

- Pautas desarrolladas para la realización de entrevistas semi-estructuradas dirigidas a las madres o tutores de los niños.

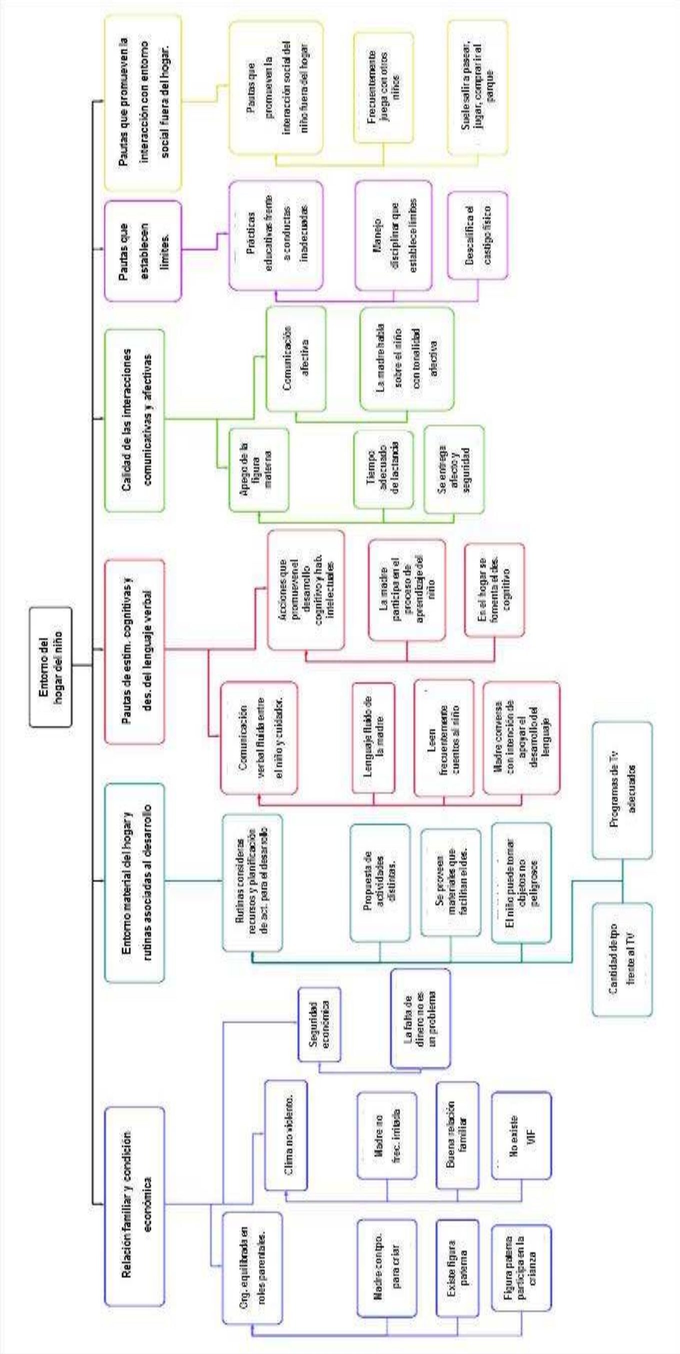
3.3.2. Representación borrosa del conocimiento.

La información disponible se extrajo de una muestra de niños chilenos en condición de vulnerabilidad, que a mayo del año 2007 tenían menos de 15 meses de vida. El conjunto de datos de entrada contiene 24 variables lingüísticas nítidas relativas al entorno del hogar del niño y 5 variables realvaloradas asociadas al logro del desarrollo infantil en cada componente del desarrollo.

La información del entorno del hogar fue generada el año 2008 a través de un proceso de codificación de transcripciones de entrevistas semi-estructuradas realizadas a las madres o tutores de los niños. Esto se tradujo en un conjunto de datos con variables lingüísticas nítidas, en la cual la distancia entre sus valores lingüísticos es la misma, no representando en todos los casos el significado teórico definido por expertos en desarrollo infantil. Y la información de logro de desarrollo se midió durante los años 2007, 2008 *y* 2009 a través del *Inventario de Desarrollo Battelle*, que entrega información de logro de los niños en las áreas motora, comunicación, cognitiva, personal-social y adaptativa.

Para modelar el conocimiento bajo un enfoque borroso, las variables lingüísticas nítidas relativas al entorno del hogar del niño se asociaron a conjuntos borrosos, teniendo en cuenta la opinión de los expertos y considerando que los valores asociados con cada variable lingüística están ordenados desde el caso óptimo al caso menos adecuado para el desarrollo infantil. Estos valores de verdad se definieron con el fin de representar su nivel de incidencia en el concepto asociado en cada componente del diagrama presentado en la Figura 3.9 (un ejemplo se presenta en la Figura 3.11). Así, la función de pertenencia de cada niño para cada variable representa el grado en el que el niño se encuentra en una condición adecuada para un lograr *buen desarrollo* frente a dicha variable.

Figura 3.9: **Conceptualización teórica del entorno del hogar del niño.**



El concepto de entorno del hogar del niño fue estudiado y definido por expertos en desarrollo infantil, siendo desagregado en seis constructos teóricos llamados *factores*. Cada factor se compone de sub-factores, y estos se descomponen a su vez en variables lingüísticas, que son priorizadas por los expertos de acuerdo a su impacto en el logro del desarrollo de los niños. Estos seis factores son desestructurados, ya que representan diferentes e incomparables aspectos del entorno del hogar del niño. Sin embargo, cada uno de ellos está compuesto por sub-factores estructurados jerárquicamente con un orden lineal que determinan el nivel de prioridad de los sub-factores sobre cada factor. De la misma manera, cada uno de estos sub-factores está compuesto por variables lingüísticas que también presentan un orden lineal asociado a un nivel de prioridad de las variables sobre cada sub-factor, como se muestra en el diagrama de la Figura 3.9.

La forma en que los sub-factores $S_1, \ldots, S_{10}$ se descomponen en las 24 variables, asi como también su orden de prioridad con estructura jerárquica se representa en la Figura 3.9. Teniendo en cuenta esta estructura, a partir del proceso de agregación de las 24 variables se obtienen los valores de los diferentes niños en cada sub-factor.

Una vez que los sub-factores $S_1, \ldots, S_{10}$ han sido construidos, los factores de $F_1, \ldots, F_6$ se descomponen en los 10 sub-factores, esta descomposición y su estructura jerárquica priorizada se muestra en la Figura 3.10.

Cada sub-factor fue generado a partir de la agregación priorizada de las variables lingüísticas borrosas que los constituyen, a modo de ejemplo, en la figura 3.11 se muestra la composición del cuarto sub-factor, correspondiente a "*Organización del ambiente físico y temporal*".

Figura 3.10: **Esquema jerárquico del conjunto de datos de entrada.**

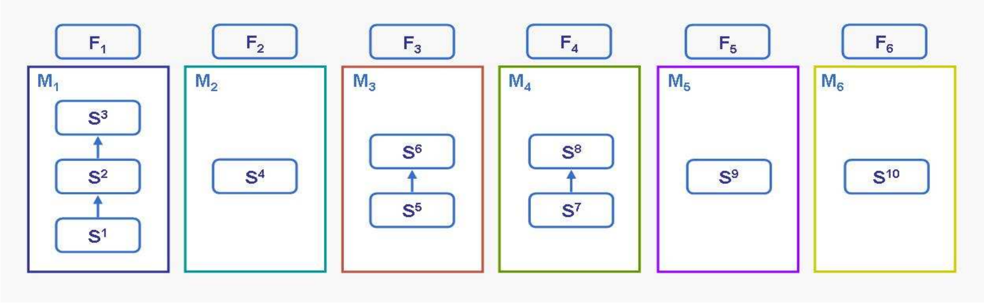
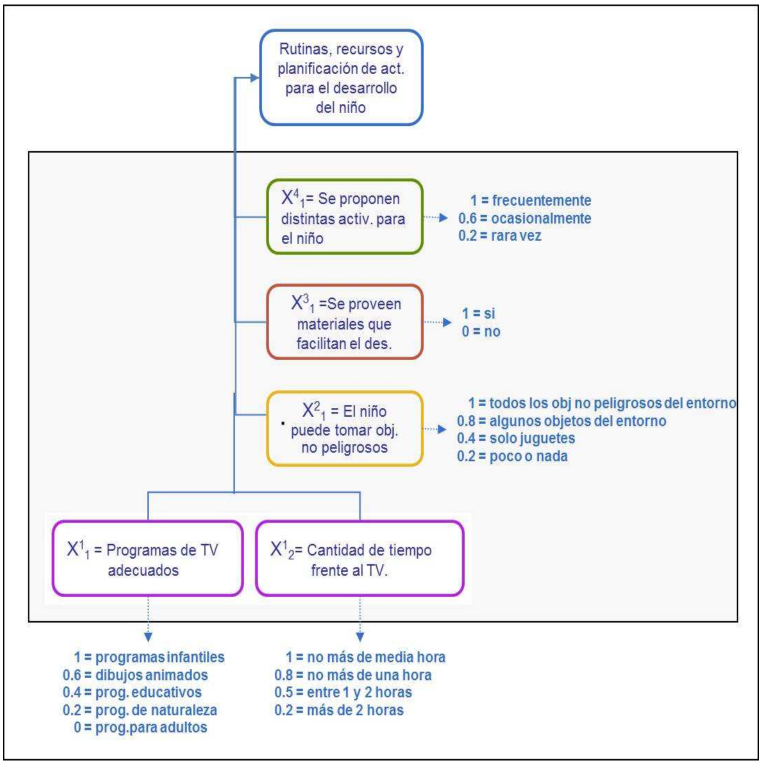


Figura 3.11: **Representación de "*Organización del ambiente físico y temporal*".**



3.3.3. Proceso de agregación.

Una vez que la estructura de los datos se ha establecido, se define la manera en que se calculan los índices correspondientes a los sub-factores y factores. Teniendo en cuenta que cada sub-factor de cada factor es una agregación de datos que presentan una estructura jerárquica priorizada, la familia $\{HIE_n\}_{n\in\mathbb{N}}$ para la construcción de los diferentes factores y sub$-$factores que aparecen en Figura 3.9 y 3.10 la definimos por:

Definición 3.20. Dada una estructura jerárquica priorizada de los datos X, la FAO jerárquica priorizada dada por $\{HIE_n\}_{n\in\mathbb{N}}$ se define como:

$$HIE_{|X|}(X) = V_r\left(U^1_{|H_1|}(H_1), U^2_{|H_2|}(H_2), \ldots, U^r_{|H_r|}(H_r)\right),$$

donde $\{U^i_{n_i} : [0,1]^{n_i} \to [0,1], n_i \in \mathbb{N}\}$ y $\{V_r : [0,1]^r \to [0,1], r \in \mathbb{N}\}$ están dadas por:

1. $\{U^i_{n_i}\}_{n_i} = \{M_{n_i}(H_i)\}\forall i \in \{1,2,...,r\}$ (media aritmética), con H_i el *i-ésimo* elemento desestructurado de la partición de X.
2. $V_r(U^1_{n_1}, U^2_{n_2}, ..., U^r_{n_r}) = \sum\limits_{j=1}^{r} l_j \prod\limits_{k=1}^{j} Y_k$, con $l_j = \dfrac{\alpha^{j-1}}{\sum\limits_{k=1}^{n_j} \alpha^{k-1}}$ $\forall j \in \{1,2,...,r\}$ y $\alpha \in [0,1]$

donde α representa un factor de "*descuento*" en la importancia de los diferentes niveles de la estructura jerárquica cuando se genera un sub-factor o factor determinado.

Notar que si $\alpha = 1$, entonces $l_j = 1/r$. De lo contrario, se da más importancia a los primeros elementos de esta agregación. Si $\alpha = 0$, sólo se consideran los elementos que aparecen en el nivel más alto de la estructura jerárquica.

Ejemplo 3.2. Sea X una estructura jerárquica con $H_1 = \{x_1, x_2\}$, $H_2 = \{x_3, x_4\}$, $H_3 = \{x_5\}$, y sea x un elemento (un niño específico) con grados de pertenencia para estas 5 variables dadas por (0,6, 0,7, 1, 1, 0,4). Si $alpha = 0,8$, la agregación de estos 5 valores es la siguiente:

$HIE_5(0{,}6, 0{,}7, 1, 1, 0{,}4) = 0{,}4\frac{1}{0{,}8+0{,}8^2} + (0{,}4x1)\frac{0{,}8}{0{,}8+0{,}8^2} + (0{,}4x1x0{,}65)\frac{0{,}8^2}{0{,}8+0{,}8^2}$

Como se explica en un capítulo anterior, la FAO de la media aritmética $\{M_n\}_{n\in\mathbb{N}}$ es estrictamente estable. Entonces, $\{HIE_n\}_{n\in\mathbb{N}}$ es una familia estrictamente estable si y sólo si $\{V_n\}_{n\in\mathbb{N}}$ es una familia estricta estable para datos linealmente estructurados. A raíz de la proposición

3.18 es posible demostrar la estabilidad de la familia $\{HIE_n\}_{n\in\mathbb{N}}$ propuesta en la Definición 3.20.

Proposición 3.19. *La familia* $\{HIE_n\}_{n\in\mathbb{N}}$ *propuesta en la Definición 3.20 es una familia estrictamente estable.*

Demostración: Teniendo en cuenta las consideraciones anteriores, sólo es necesario probar que $l_j^n - (1 - l_n^n)\ l_j^{n-1} = 0$, donde $l_j = \dfrac{\alpha^{j-1}}{\sum\limits_{k=1}^{n}\alpha^{k-1}}$ para cada sub-factor en particular, en tanto $\{V_r\}_{r\in\mathbb{N}}$ es una familia de medias ponderadas aplicada sobre un conjunto de datos con orden lineal hacia la derecha.

Y efectivamente:

$$\begin{aligned}
\frac{\alpha^{j-1}}{\sum\limits_{k=1}^{n}\alpha^{k-1}} - \left(1 - \frac{\alpha^{n-1}}{\sum\limits_{k=1}^{n}\alpha^{k-1}}\right)\cdot\frac{\alpha^{j-1}}{\sum\limits_{k=1}^{n-1}\alpha^{k-1}} &= \frac{\alpha^{j-1}\left[\sum\limits_{k=1}^{n-1}\alpha^{k-1} - \alpha^{k-1} - \alpha^{n-1}\right] + \alpha^{n-2+j}}{\sum\limits_{k=1}^{n}\alpha^{k-1}\sum\limits_{k=1}^{n-1}\alpha^{k-1}} \\
\frac{-\alpha^{j+n-2} + \alpha^{j+n-2}}{\sum\limits_{k=1}^{n}\alpha^{k-1}\sum\limits_{k=1}^{n-1}\alpha^{k-1}} &= 0
\end{aligned}$$

Por lo tanto, es posible construir índices de entorno del hogar del niño mediante un esquema anidado de estructuras jerárquicas priorizadas. Particularmente, se ha demostrado que cuando el conjunto de datos estructurado asociado se agrega mediante dos $FAO's$ estables (i.e., una familia priorizada con el fin de agregar los datos linealmente estructurados, y la media aritmética para agregar los datos desestructurados), entonces se obtiene un proceso de agregación jerárquico priorizado robusto, en tanto sus resultados son estables.

Sólo para mostrar un ejemplo del proceso de agregación, en la figura 3.11, asociada al sub-factor *Organización del ambiente físico y temporal*, vemos que los conjuntos de datos con estructura jerárquica están compuestas por los conglomerados $H_1 = \{X_1^1, X_2^1\}, H_2 = \{X_1^2\}, H_3 = \{X_1^3\}$ *y* $H_4 = \{X_1^4\}$. Por ejemplo, sea c un niño que tiene los valores 0,6, 0,8, 0,2, 0, 0,6 asociados a las variables X_1^1, X_2^1 , X_1^2, X_1^3 y X_1^4 respectivamente. Notar que x_1^3 es una variable dicotómica, por lo que no puede ser transformada en un conjunto borroso. Este tipo de variables se definen como una contribución en el grado de pertenencia de otra variable relacionada con el mismo sub$-$factor (tema que también fue definido por expertos). En este caso, cuando $x_1^3 = 1$, el grado de pertenencia de x_1^2 aumenta en 0,2 si su grado de pertenencia original era menos de 0,5, y se incrementa en 0,1 si su original grado de pertenencia era superior a 0,5. Así, el niño c tiene los nuevos valores $H_1 = \{0{,}6, 0{,}8\}, H_2 = \{0{,}2\}$ *y* $H_3 = \{0{,}6\}$ para el segundo sub$-$factor.

Los operadores de agregación priorizados y desestructurados (V_r y $U_{n_i}^i$ respectivamente) utilizados fueron presentados en la definición anterior. El nivel de importancia de las estructuras se representó por $\alpha = 0{,}7$ cuando se agregan cuatro elementos, y $\alpha = 0{,}8$ cuando se agregan dos o tres elementos. Finalmente, el valor agregado representa el grado de pertenencia que tiene el niño al conjunto borroso asociado con la variable lingüística "*adecuada organización del ambiente físico y temporal*" es $0,3548$.

En la Figura 3.9 se mostró que cada uno de los seis factores que representan el entorno del hogar del niño está compuesto por sub-factores priorizados, por tanto es posible construir cada factor y sub-factor de la misma manera como se ejemplificó anteriormente. Así, el índice de entorno del hogar del niño está compuesto por un conjunto de estructuras jerárquicas anidadas que se agregan a través de operadores priorizadas ponderados V_j con $j \in \{1, 2, ..., 16\}$, así como también, por operadores desestructurados $U_{n_i}^i$ con $i \in \{1, 2, ..., 23\}$, donde los V_j se utilizan para agregar los conjuntos de datos con orden lineal y los $U_{n_i}^i$ se utiliza para agregar los conjuntos de datos desestructurados.

Por lo tanto, el índice global puede ser definido por:

$$Index(X) = U_6(F_1, F_2, F_3, F_4, F_5, F_6), \text{ donde } F_k = HIE_{|M_k|}(M_k), \ \forall k \in \{1, 2, ..., 6\}$$

y a su vez $M_k = HIE_{|H_i|}(H_i)$, donde H_i es una partición desestructurada de X $\forall i \in \{1, 2, ..., 23\}$

3.3.4. Validación del proceso de agregación y resultados.

En base al procedimiento mostrado en la sección anterior, se generaron todos los factores y subfactores representados en el esquema 3.10, en la cual cada niño tiene asociado a un conjunto de valores de $x_1, ..., x_n$ que representan el grado de pertenencia a los conjuntos borrosos asociados con cada variable lingüística, información que es agregada a fin de valorar cada sub-índice para cada niño. La síntesis de los resultados estadísticos de dichos procesos de agregación se presentan en la tabla siguiente:

Con el fin de validar estos resultados, se comparan los índices de ambiente familiar del niño con la percepción de riesgo psicosocial de la familia registrada por el entrevistador que realizó las entrevistas en profundidad a las madres de los niños. Además, con el fin de validar el método propuesto, se comparan los resultados de dos índices obtenidos a través de procedimientos diferentes, pero ambos en base a la información asociada al factor " *adecuadas craacterísticas familiares y ocndición económica*". Un método corresponde a la propuesta de este trabajo, que utiliza conjuntos borrosos asociados a las variables lingüísticas y un proceso de agregación jerárquica priorizada, mientras que el otro se basa en el método aditivo utilizado por Ramey [71] para construir el $High - Risk\ Index$, índice utilizado para distinguir a los niños según el origen económico de la familia y el nivel de exposición a múltiples riesgos. El $High - Risk\ Index$ considera trece factores de riesgo asociados al hogar del niño, medidos a

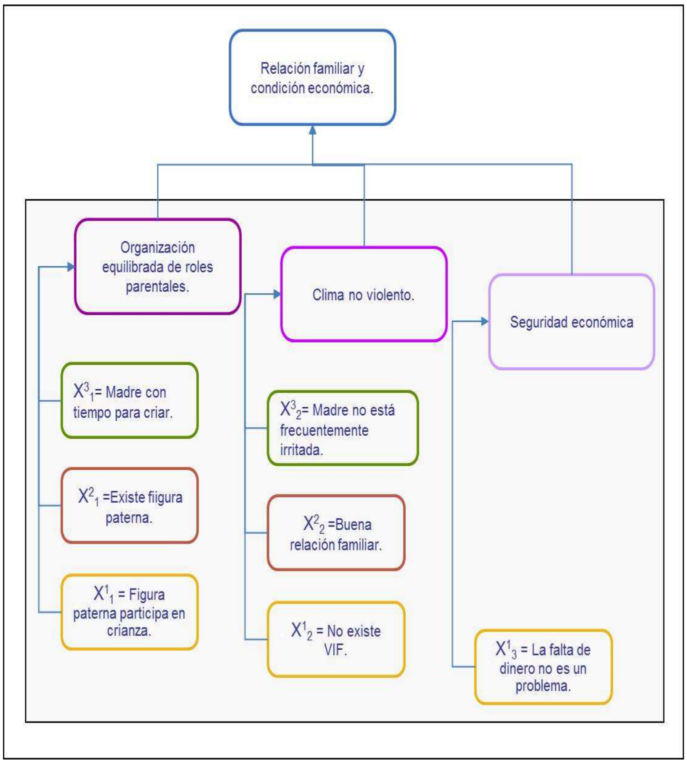
Cuadro 3.5: **Estadísticas generales de los índices de entorno del hogar del niño.**

Índices	**Media**	**Mediana**	**Desv.Est.**	**Min**	**Max**	**Rango Int.**
Adecuadas características familiares y condición económica	0.30	0.29	0.14	0.00	0.72	0.19
Adecuada organización del espacio físico y temporal	0.50	0.47	0.26	0.00	0.99	0.45
Presencia de patrones de estimulación cognitiva y desarrollo verbal	0.56	0.53	0.25	0.00	1.00	0.46
Calidad de las interacciones afectivas madre-hijo	0.81	0.88	0.16	0.00	1.00	0.12
Adecuadas pautas que establecen límites de la conducta del niño	0.59	0.63	0.29	0.04	1.00	0.48
Presencia de patrones que fomenta la interacción social fuera del entorno familiar	0.74	0.80	0.21	0.27	1.0	0.22

partir de variables relacionadas con el nivel socio-económico y educativo de la familia del niño, en la cual cada una de estas variables tiene un peso asociado que representa su contribución al indicador de riesgo del hogar del niño. La construcción del índice es aditivo, por lo que la mayor acumulación de factores hace más probable que el niño se encuentre en un hogar en riesgo.

A pesar que la construcción del segundo índice se basa en un método aditivo de la acumulación de factores de riesgo medidos por variables lingüísticas nítidas, para esta comparación, se han utilizado variables lingüísticas asociadas a conjuntos borrosos, ya que estamos interesados en validar el método y no la construcción teórica del entorno familiar del niño. Por lo tanto, el desarrollo de los índices que se utilizan en ambos métodos se basan en el mismo conjunto de datos, la misma conceptualización teórica presentada en la Figura 3.12 y el mismo modelado borroso de la información.

Figura 3.12: **Representación de ”*Relación familiar y condición económica*”.**



El riesgo psicosocial de la familia percibido por el entrevistador, se midió a una parte de la muestra mediante una variable ordinal de tres niveles que representan un riesgo familiar alto, medio y bajo, por lo que a fin de permitir el análisis comparativo con los índices, éstos se transformaron en el mismo tipo de variables. La proporción de riesgo obtenida por ambos índices es similar. Sin embargo, en la comparación de métodos de agregación existen algunos casos afectados por el *trade-off* que permite el método aditivo, cuestión que no es aceptable para este problema de agregación, ya que la definición conceptual del entorno familiar del niño definida por los expertos tiene una estructura jerárquica en la que las variables tienen diferentes niveles de importancia sobre el entorno del hogar. La siguiente tabla muestra las similitudes y diferencias entre las clasificaciones del entorno resultantes por cada tipo de método de construcción de índices.

Cuadro 3.6: **Índice aditivo vs índice priorizado de la calidad del entorno del hogar del niño.**

		Índice jerárquico priorizado			
		Precario entorno	Normal entorno	Adecuado entorno	Total
Índice aditivo	Entorno precario	89	30	**1** (Caso 1)	120
	Normal entorno	34	71	26	131
	Entorno adecuado	**1** (Caso 2)	27	100	128
Total		124	128	127	379

En la tabla se evidencian dos casos con una gran discordancia, los que son analizados a continuación. El comportamiento de los dos niños con mayor discordancia en el índice de "*Adecuadas características familiares y condición socioeconómica*" ha sido estudiado, analizándo las valoraciones de los sub−factores y de las variables que componen este factor. El sub-factor "*Composición de la familia y los roles parentales*" está compuesto por una estructura jerárquica priorizada de tres variables, donde la variable "*Tiempo disponible para el cuidado de la madre*" tiene la mayor importancia. El sub-factor "*Clima del hogar no violento*" está compuesto por la misma estructura, donde "*La madre no se percibe irritable o nerviosa*" tiene la mayor importancia.

La siguiente tabla muestra los grados de pertenencia de las variables borrosas asociadas a estos sub-factores, y las puntuaciones resultantes de cada clasificador.

En ambos casos, el método de agregación aditivo compensa la baja valoración de las variables más prioritarias con las otras variables que tienen menor nivel de importancia, mientras que el método de agregación jerárquica priorizada no produce este *trade-off*, ya que considera la

Cuadro 3.7: **Pertenencia a conjuntos borrosos asociados a "*Características familiares y condición socioeconómica*".**

Sub-factor	Variable	Caso 1	Caso 2
1. *Familia, composicion*	Tiempo que tiene la madre para criar	0.00	0.99
y roles parentales	Existe una figura paterna	0.00	0.00
	F. paterna involucrada en la crianza	0.00	1.00
Puntuación del sub-factor 1	con el método jerárquico priorizado	0.00	0.41
Puntuación del sub-factor 1	con el método aditivo	0.00	0.66
2. *Clima no violento*	Madre no está frecuentemente irritada	0.09	0.99
en el hogar	Buena relación familiar	0.71	0.99
	No hay violencia intrafamiliar	1.00	0.99
Puntuación del sub-factor 1	con el método jerárquico priorizado	0.07	0.98
Puntuación del sub-factor 1	con el método aditivo	0.60	0.99
3. *Estabilidad económica*	La falta de dinero no es un problema	0.52	0.72
Puntuación del sub-factor 1	con el método jerárquico priorizado	0.52	0.72
Puntuación del sub-factor 1	con el método aditivo	0.52	0.72
Puntuación del sub-factor 1	con el método jerárquico priorizado	0.00	0.37
Puntuación del sub-factor 1	con el método aditivo	0.33	0.81

estructura de prioridades al agregar la información.

Se observa que el método aditivo sobre-estima el *Caso* 1 en el sub-factor *Clima del hogar* *no violento*, en tanto no considera la baja valoracion de la variable "*La madre no se percibe irritable o nerviosa*" al momento de agregar la información, que es considerada de alta relevancia por los expertos. Lo mismo ocurre con el $Caso$ 2 en el sub-factor *Composición familiar y roles parentales* al no considerar la baja valoracion de la variable "*Existe figura paterna*".

Capítulo 4

Búsqueda de patrones basado en índices de calidad de sistemas de clasificación no supervisada.

Hasta ahora hemos visto cómo generar índices robustos asociados a distintos factores que componen un adecuado entorno del hogar del niño en cuanto a facilitar el logro de su desarrollo. Para esto, se definieron propiedades a tener en cuenta a la hora de definir las familias consistentes de operadores de agregación que participarán en el proceso de agregación de la información, las cuales dependerán de la estructura del conjunto de datos de entrada. Una vez la información de entorno del hogar del niño ha sido agregada en factores mediante familias consistentes de operadores de agregación, es necesario identificar patrones asociados a los distintos tipos de entorno del hogar presentes en la población infantil, bajo el objetivo de relacionar cada niño con una población normativa que comparta dichos condicionantes de logro de desarrollo medidos a través de los índices de entorno del hogar del niño.

En la sección 4 del capítulo 2 se explica brevemente el algoritmo *fuzzy c-means* y algunas de sus extensiones. Pudimos ver que todos los algoritmos presentados requieren la definición de la cantidad de clases de la partición, información que *a priori* es desconocida. Esto evidencia la necesidad de generar un sistema de evaluación de la calidad de distintas particiones, a fin de seleccionar la clasificación más adecuada al conjunto de datos de entrada y su respresentación teórica. En esta misma sección, se presentan algunas medidas de evaluación que han sido desarrolladas bajo este fin, donde vemos que las tres medidas de calidad global más utilizadas, a saber; *Coeficiente de partición*, *Entropía de la partición* y el *Índice de separación*, evalúan el nivel de solapamiento entre las clases de una partición, entregándole más valor a aquellas particiones en la cual el solapamiento es menor. Las otras medidas de evaluación global presentadas corresponden a evaluaciones de particiones realizadas bajo un algoritmo específico (Gath-Geva), teniendo por tanto, un alcance limitado. También se presentaron algunas medidas de calidad que evalúan el comportamiento de cada clase, uno se acota a una partición realizada bajo un algoritmo específico (Gustafson-Kessel), mientras que el otro evalúa la densidad del contorno de las clases, valorando las clases que tienen una menor cantidad de datos en su contorno.

Podemos ver que si bien se han desarrollados medidas que evalúan la calidad de una partición de clases borrosas, tanto a nivel global como local, estas se centran en la evaluación de sólo el criterio de solapamiento entre clases, dejando de lado otros aspectos igualmente relevantes, como la presencia de altos niveles de pertenencia a una clase, asi como también, el nivel de cobertura de las mismas. Además, estas medidas entregan una mejor evaluación a los modelos de clasificación no supervisada en la cual sus clases no se encuentran solapadas, tendiendo con esto, a valorar más los modelos cuyos resultados se acercan a una partición de clases nítidas.

Desde nuestro punto de vista, una partición que presenta un nivel de solapamiento entre sus clases borrosas no necesariamente es descartable, en tanto puede corresponder a una cualidad de la información que se extrae de los datos, además, el solapamiento no corresponde al único criterio de calidad relevante a evaluar en una partición. Las cualidades que se persiguen en una clasificación borrosa de un conjunto de datos son:

- Que todos los elementos queden representados en al menos una clase con un nivel de pertenencia mínimo.
- Que pocos elementos tengan similares y altos niveles de pertenencia a clases distintas.
- Que la mayor cantidad de elementos tengan un alto nivel de pertenencia a una clase.

Como se requiere evaluar tres criterios de calidad, es necesario generar un sistema de evaluación de calidad de particiones de clases borrosas. En este trabajo se estudiaron conjuntamente los criterios de *cobertura*, *redundancia* y *relevancia*, donde cada uno alude a los objetivos recién descritos y son generados a partir de definiciones de familias consistentes de operadores de agregación. A continuación se presentan tales medidas y las FAOs que las generan, en el contexto de búsqueda de patrones mediante restricciones de calidad sobre dichos criterios.

4.1. Búsqueda de patrones.

La evaluación de calidad de un conjunto particiones de X es aplicada sobre un conjunto finito de posibles clases borrosas que las determinan, las que son generadas a partir de sus correspondientes centros o patrones de clase. Estos patrones son definidos intencionadamente, a fin de cubrir los valores que pueden tomar las variables de entrada. A continuación se presentan algunos conceptos necesarios para la presentación del proceso de búsqueda de patrones definidos en [50].

Definición 4.1. (Espacio de análisis borroso). Sea $D \neq \emptyset$ un conjunto y R un conjunto de conjuntos. Llamaremos a D el espacio de datos y R el espacio de resultados. Entonces, entendemos por Espacio de análisis borroso a $A_{fuzzy}(D, R) := \{f/f : XxK \rightarrow [0, 1],\ X \subseteq D,\ K \in R\}$.

Definición 4.2. (Partición de clases borrosas). Sea $A_{fuzzy}(D, R)$ un espacio de análisis borroso. Un resultado de $f : XxK \rightarrow [0, 1] \in A_{fuzzy}(D, R)$ es llamado Partición de clases borrosas si $\forall k \in K : \sum\limits_{x \in X} f(x)(k) > 0$, donde $f(x)(k)$ corresponde al grado en que el elemento $x \in X$ es representado por la clase $k \in K$.

Para iniciar el proceso de búsqueda de la mejor partición de clases borrosas, se define un conjunto $G \subseteq [0, 1]^s$, donde $s \in \mathbb{N}$ corresponde a la cantidad de variables del conjunto de datos, que contiene los posibles centros de clase o patrones de clase de X. Este conjunto lo llamaremos *rejilla o cuadrícula*, y se define por $G = \{(g_1, ...g_s) : g_i \in \{0, 1/q, 2/q, ..., q/q\}, \forall i \in \mathbb{N}_{\leq s}\}$ con $q \in \mathbb{N}$. Los elementos de G corresponden a los centros de las clases borrosas que serán evaluadas a fin de seleccionar el subconjunto P de G que represente de manera más adecuada la segmentación "*natural*" del conjunto de datos X. La definición del valor q determina la cantidad de celdas de la rejilla o cuadrícula, y por tanto, la cantidad de clases a evaluar, correspondiendo a $(q + 1)^s$ clases para cada uno de los $s!$ ordenes lexicográficos de las coordenadas de los elementos de X.

Definición 4.3. (Patrón de clase borrosa). Sea $f : XxK \rightarrow [0, 1]$ una partición de clases borrosas del conjunto de datos $X = \{x_1, ..., x_n\} \subseteq [0, 1]^s,\ s \in \mathbb{N}$, dada por la pertenencia al conjunto de clases $K = \{K_1, ..., K_c\}$ con $c \in \{2, 3, ..., n - 1\}$. Llamaremos patrón de la clase $K_j\ con\ j \in \{1, 2, ..., c\}$ al punto $k \in G$ que determina una clase borrosa a partir de la función $f(x)(k) = 1 - d(x, k)$, donde d es una medida de distancia.

El conjunto de los posibles patrones de clase G son sometidos a evaluaciones de calidad, dicho proceso elimina aquellas clases que no cumplen determinadas restricciones de calidad, hasta terminar con el conjunto de clases borrosas que en su conjunto mejor representan la partición natural de los datos.

4.1.1. Índices de calidad.

El sistema de evaluación utiliza los criterios de calidad de *Relevencia, Redundancia* y *Cobertura* presentados en el Capítulo 2. A continuación se definen tales criterios de calidad en base a familias consistentes de operadores de agregación.

Consideremos la partición de clases borrosas del conjunto de datos $X = \{x_1, ..., x_n\} \subseteq [0,1]^s$ con $s \in \mathbb{N}$ dada por la pertenencia a las clases del conjunto $K = \{K_1, ..., K_c\}$ con $c \in \{2, 3, ..., n-1\}$, entonces, las medidas de relevancia, redundancia y cobertura se definen por:

- *Relevancia*: El nivel de relevancia de la clase K_j, con $j \in \{1, .., c\}$ es gradual y está dado por $\phi\{\mu_{K_j}(x) : x \in X\}$. Para esta medida utilizamos a la familia consistente de operadores de agregación dada por el máximo, de tal modo que el nivel de relevancia de una clase borrosa K_j está dado por:

  $$\phi(K_j) = Max_{K_j}\{\mu_{K_j}(x)\},\ \forall x \in X.$$

  Entonces, diremos que K_j es relevante si *para cada* $x \in X$, se cumple que:

  $$Max_{K_j}\{\mu_{K_j}(x)\} \text{ es significativamente mayor que}$$
  $$Max_{K_l}\{\mu_{K_l}(x)\}\ con\ K_j, K_l \in K, K_j \neq K_l.$$

  Y por el contrario, diremos que K_j es irrrelevante si *para todos* los $x \in X$, se cumple que:

  $$Max_{K_j}\{\mu_{K_j}(x)\} \text{ no es significativamente mayor que}$$
  $$Max_{K_l}\{\mu_{K_l}(x)\}\ con\ K_j, K_l \in K, K_j \neq K_l.$$

- *Redundancia*: La redundancia entre las clases $K_j, K_l \in K$ con $K_j \neq K_l$ se evalúa sobre cada elemento x, correspondiendo a la proporción de elementos de X que se encuentran con un nivel de solapamiento mínimo, y la representamos por $\varphi(\mu_{K_j}(x), \mu_{K_l}(x))$. Para esta medida utilizamos dos familias consistentes de operadores de agregación, el mínimo y la media aritmética, de tal modo que el nivel de redundancia entre dos clases K_j, K_l está dado por:

$$\varphi(\mu_{K_j}(x), \mu_{K_l}(x)) = \sum_{i=1}^{n} I_{(j,l)}(x_i) \ / \ n, \text{ donde}$$

$$I_{(j,l)}(x_i) = \begin{cases} 1 & \text{si } \mathrm{Min}_x\{\mu_{K_j}(x), \mu_{K_l}(x)\} \leq \beta_2, \text{ con } \beta_2 \in (0,1] \\ 0 & \text{si no} \end{cases}$$

- *Cobertura*: La cobertura se evalúa sobre cada elemento x, correspondiendo a la proporción de elementos de X que se encuentran cubiertos por la clase K_j, y está dada por $\phi\{\mu_{K_j}(x) : \ K_j \in \ K\}$. Para esta medida utilizamos dos familias consistentes de operadores de agregación, el máximo y la media aritmética, de tal modo que la cobertura de una clase borrosa K_j está dada por:

$$\phi(\mu_{K_j}(x)) = \sum_{i=1}^{n} I_{(K_j)}(x_i) \ / \ n, \text{ donde}$$

$$I_{(K_j)}(x) = \begin{cases} 1 & \text{si } \mathrm{Max}_x\{\mu_{K_j}(x)\} \geq \gamma_2 \text{ con } \gamma_2 \in (0,1] \\ 0 & \text{si no} \end{cases}$$

4.1.2. Restricciones de calidad.

Utilizando las medidas de calidad descritas en la sección anterior, se definen las siguientes tres restricciones sobre los criterios de relevancia, cobertura y redundancia:

- *Restricción 1*: Una clase c será relevante si:

$$\forall x \in X, \ Max_c\{\mu_c(x)\} > \alpha, \text{ con } \alpha \in (0,1].$$

- *Restricción 2*: Dos clases c, k no serán redundantes si:

  $$\forall x \in X, \sum_{i=1}^{n} I_{(c,k)}(x_i) \ / \ n \leq \beta, \ \ \textit{con} \ \beta \in (0, 1].$$

  y donde la función $I_{(c,k)}(x)$ se define por:

  $$I_{(c,k)}(x) = \begin{cases} 1 & \text{si } \mathrm{Min}_x\{\mu_c(x), \mu_k(x)\} \leq \beta_2, \ \text{con } \beta_2 \in (0, 1] \\ 0 & \text{si no} \end{cases}$$

- *Restricción 3*: Una clase c tiene una cobertura adecuada si:

  $$\forall x \in X, \sum_{i=1}^{n} I_{(c)}(x_i) \ / \ n \geq \gamma, \ \ \textit{con} \ \gamma \in (0, 1].$$

  y donde la función $I_{(c)}(x)$ se define por:

  $$I_{(c)}(x) = \begin{cases} 1 & \text{si } \mathrm{Max}_x\{\mu_c(x)\} \geq \gamma_2 \ \text{con } \gamma_2 \in (0, 1] \\ 0 & \text{si no} \end{cases}$$

En base a estas restricciones, primero se eliminan aquellas clases que no representan mínimamente a ningún elemento, en tanto tienen un nivel de pertenencia muy bajo. Luego, se eliminan aquellas clases que no son necesarias, en tanto existen otras que representan a los mismos elementos satisfaciendo mejor la primera restricción. Y finalmente, quedan fuera aquellas clases que tienen un bajo nivel de cubrimiento sobre el conjunto de datos X.

Para aplicar estas restricciones es necesario establecer los umbrales dados por $\alpha, \beta, \beta_2, \gamma, \gamma_2$, que pueden ser definidos luego de un análisis gráfico de la cobertura dada por $\sum_{i=1}^{n} I_{(c)}(x_i) \ / \ n$, para distintos niveles de γ_2.

Una vez seleccionada la partición de clases borrosas que mejor representa el conjunto de datos de entrada, es necesario describir el comportamiento a fin de evaluar su coherencia teórica. Para esto, se propone en primera instancia, interpretar los significados de los patrones de clase de la partición a partir de los valores que toman cada una de sus coordenadas, como también en base a su cobertura bajo distintos niveles de relevancia, y su posición en el conjunto de datos dado por el percentil en el cual se ubican. Finalmente, a fin de facilitar su interpretación, es posible analizar las asociaciones lineales entre los niveles de pertenencia a cada clase borrosa y los niveles de pertenencia a cada una de las variables representadas por las coordenadas de X.

4.2. Clasificación borrosa de la población infantil chilena según características del entorno del hogar.

4.2.1. Conjunto de datos de entrada.

Se tiene un conjunto de datos con información de seis índices que representan factores incidentes en que el niño tenga un adecuado entorno de hogar para lograr desarrollar todo su potencial. Estos factores son:

1. Adecuada relación familiar y condición económica.
2. Entorno material y rutinas que promueven el desarrollo.
3. Adecuadas pautas de estimulación cognitiva, intelectual y del lenguaje.
4. Buena calidad de las interacciones comunicativas y afectivas de la madre.
5. Adecuadas pautas que establecen límites de la conducta del niño.
6. Pautas que promueven la interacción del niño en un entorno social.

Al estudiar su comportamiento, se observó que si bien teóricamente estos factores son independientes y representan distintos conceptos, algunos de ellos se encuentran asociados. Por lo tanto, y en base a la opinión de expertos, se agregó el factor "*Entorno material y rutinas que promueven el desarrollo*" con "*Adecuadas pautas de estimulación cognitiva, intelectual y del lenguaje*", representando conjuntamente, a la presencia de pautas de estimulación del desarrollo del niño. Del mismo modo, se agregaron los factores "*Buena calidad de las interacciones comunicativas y afectivas de la madre*" y "*Pautas que promueven la interacción del niño en un entorno social*", representando conjuntamente, a la presencia de interacciones de calidad en el hogar y la promoción de interacciones con otros niños.

Cuadro 4.1: **Estadísticas generales de las variables de entrada.**

Índices	**Media**	**Mediana**	**Desv.Est.**	**Min**	**Max**	**Rango Int.**
Adecuadas características familiares y condición económica	0.30	0.29	0.14	0.00	0.72	0.19
Presencia de pautas de estimulación del desarrollo del niño	0.52	0.50	0.20	0.10	0.97	0.30
Interacciones de calidad	0.78	0.81	0.14	0.21	1.00	0.17
Adecuadas pautas que establecen límites de la conducta del niño	0.59	0.63	0.29	0.04	1.00	0.48

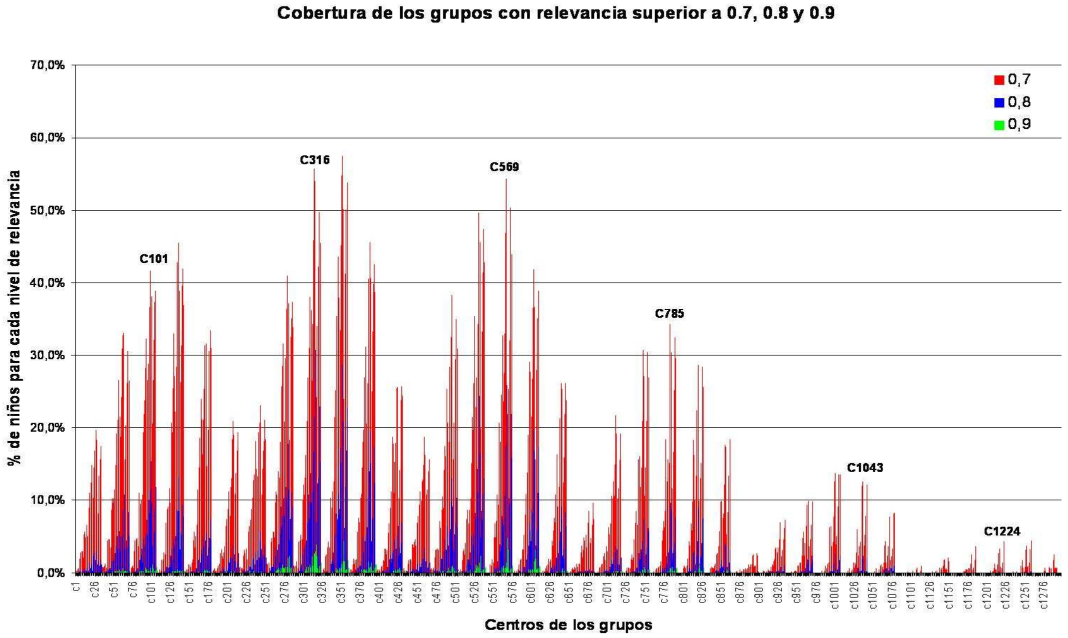
Índices	**Pc.5**	**Pc.10**	**Pc.25**	**Pc.50**	**Pc.75**	**Pc.90**	**Pc.95**
Adecuadas características familiares y condición económica	0.03	0.15	0.19	0.29	0.37	0.50	0.56
Presencia de pautas de estimulación del desarrollo del niño	0.22	0.26	0.36	0.50	0.67	0.81	0.87
Interacciones de calidad	0.50	0.56	0.70	0.81	0.86	0.94	0.95
Adecuadas pautas que establecen límites de la conducta del niño	0.08	0.09	0.32	0.63	0.80	0.99	0.99

4.2.2. Búsqueda de los mejores patrones y resultados.

El método utilizado supone que el conjunto de datos de entrada se encuentra altamente concentrado, dificultando la identificación de patrones. Tal como se señala en el capítulo anterior, se requiere como parámetros de entrada la amplitud de las celdas que constituyen la rejilla o cuadrículas para cada una de las variables de entrada. En este caso se tienen cuatro factores definidos en el intervalo unitario, y que teóricamente aportan conjuntamente en la determinación de la calidad del entorno del hogar del niño en términos de facilitar el desarrollo del mismo. Por lo tanto, los patrones corresponden a vectores de cuatro coordenadas que toma valores entre 0 y 1, en la cual cada una representa los distintos factores que fueron generados anteriormente mediante un proceso de agregación consistente de información borrosa estructurada jerárquicamente.

Se analizaron 1296 patrones de cuatro coordenadas para cada uno de las 24 posibles permutaciones entre las 4 variables de entrada, que corresponden a todos los posibles centros de las clases donde las coordenadas toman un valor en el intervalo unitario a distancia 0,2. Posteriormente, en base a las restricciones de calidad dadas por los indicadores de redundancia, relevancia y cobertura definidas anteriormentes, se identificaron los mejores patrones.

Para la selección de las mejores clases borrosas, primero se estableció la restricción de cobertura no menor al 20 % de elementos con relevancia superior a 0,70. A modo de ejemplo, la siguiente gráfica muestra la cobertura de los 1,296 centros ordenados bajo el criterio de orden dado por: $F1$, $F2-3$, $F4-6$ y finalmente $F5$, para las relevancias 0,7, 0,8 y 0,9.



El eje y corresponde al porcentaje de niños con un nivel de pertenencia mayor a 0,7, 0,8 y 0,9 a cada clase. Y el eje x corresponde a todos los posibles patrones de las clases, ordenados de izquierda a derecha según la presencia de los factores evaluados (0 indica nula presencia y 1 mayor presencia), considerando el siguiente orden: Factor 1, Factor $2-3$, Factor $4-6$, Factor 5. Por tanto, a lo largo del eje x se mantienen fijos los valores de las primeras coordenadas, y las últimas varían tomando los valores $0, 0{,}2, 0{,}4, 0{,}6, 0{,}8$ y 1. En la gráfica se observa que la muestra está desvalanceada hacia una relación familiar y condición económica menos adecuada para el desarrollo infantil (se aglutina a la izquierda), por lo que los dos grupos de la derecha tienen una cobertura baja, lo que tiene relación con el contexto de vulnerabilidad en el que se encuentran los niños de la muestra.

Otra condición utilizada, restringe a las clases que tienen mayor cobertura local, que en la gráfica están dados por:

Este procedimiento se realizó para cada uno de los 24 posibles criterios de ordenamiento de

Cuadro 4.2: **Mejores patrones para un sistema de clasificación de seis clases.**

Índices	**C101**	**C316**	**C569**	**C785**	**C1043**	**C1224**
Adecuadas características familiares y condición económica	0.00	0.20	0.40	0.60	0.80	1.00
Presencia de pautas de estimulación del desarrollo del niño	0.40	0.40	0.60	0.60	0.80	0.60
Interacciones de calidad	0.80	0.80	0.80	0.80	1.00	1.00
Adecuadas pautas que establecen límites de la conducta del niño	0.80	0.60	0.80	0.80	0.80	1.00

los cuatro factores, obteniéndose finalmente los siguientes mejores seis centros:

Cuadro 4.3: **Mejores patrones para un sistema de clasificación de seis clases.**

Índices	**Clase 1**	**Clase 2**	**Clase 3**	**Clase 4**	**Clase 5**	**Clase 6**
Adecuadas características familiares y condición económica	0.16	0.48	0.20	0.25	0.20	0.20
Presencia de pautas de estimulación del desarrollo del niño	0.62	0.60	0.32	0.58	0.60	0.40
Interacciones de calidad	0.83	0.80	0.80	0.60	0.80	0.82
Adecuadas pautas que establecen límites de la conducta del niño	0.82	0.80	0.62	0.80	0.00	0.31

En la tabla se observa que el factor que representa la calidad de las interacciones presenta una baja variación en los 6 patrones de clase (entre 0,6 y 0,83). Sin embargo, y tal como vimos anteriormente, el valor mínimo de ese factor corresponde a $0,21$ y el 50 % de la muestra tiene valores mayores a 0,81, por lo que estas pequeñas variaciones en el valor de este factor, sí representan un cambio considerando la muestra. A si mismo, el mayor valor del factor que representa la relación familiar y condición económica de la familia es de 0,48, lo que también se corresponde con la muestra, en tanto su valor promedio es 0,3.

En la siguiente tabla se presentan estadísticas de la redundancia entre cada par de clases. En ella se observa que la relevancia promedio se encuentra aproximadamente entre 0,5 y 0,65, siendo la tercera clase la que presenta mayores niveles de redundancia con más grupos. No obstante, aporta en disperción del factor relativo a las pautas que establecen límites de conducta del niño al tener asociado un valor medio (0,62), por lo que no se considera adecuado eliminarlo.

Cuadro 4.4: **Estadística del nivel de solapamiento de las clases.**

Pares de clases	**Máx.**	**Mín.**	**Media**
Clase 4 - Clase 5	$0,704$	$0,256$	$0,497$
Clase 2 - Clase 5	$0,708$	$0,256$	$0,494$
Clase 1 - Clase 5	$0,727$	$0,256$	$0,514$
Clase 2 - Clase 6	$0,768$	$0,299$	$0,568$
Clase 3 - Clase 5	$0,774$	$0,256$	$0,536$
Clase 4 - Clase 6	$0,777$	$0,323$	$0,574$
Clase 1 - Clase 6	$0,790$	$0,339$	$0,594$
Clase 2 - Clase 3	$0,834$	$0,299$	$0,619$
Clase 3 - Clase 4	$0,844$	$0,323$	$0,631$
Clase 5 - Clase 6	$0,848$	$0,256$	$0,568$
Clase 3 - Clase 6	$0,851$	$0,339$	$0,623$
Clase 1 - Clase 3	$0,859$	$0,372$	$0,654$
Clase 2 - Clase 4	$0,870$	$0,299$	$0,634$
Clase 1 - Clase 2	$0,883$	$0,299$	$0,652$
Clase 1 - Clase 4	$0,899$	$0,323$	$0,654$

4.2.3. Interpretación del sistema de clasificación.

- *Grupo* 1*: Pautas de crianza favorables para el desarrollo en entornos familiares precarios en su organización.*

  El patrón de este entorno familiar se caracteriza por poseer la relación familiar y percepción de las condiciones económicas más precaria de todos los grupos, en tanto el factor "Adecuada relación familiar y condición económica" toma el valor 0,16 en el intervalo unitario, y se ubica en el percentil 0,11 sobre el total de familias de la muestra. Sin embargo, este patrón muestra la mayor presencia de pautas de estimulación del desarrollo del niño, tomando el valor 0,62 en el factor "Presencia de pautas de estimulación de desarrollo del niño", y ubicándose en el percentil 0,71 de la muestra, además de interacciones de calidad, especialmente en relación al vínculo afectivo que promueve la madre, tomando el valor $0,83$ en el factor "Interacciones de calidad" y concentrándose en el percentil 0,54 de las familias estudiadas. Por último, las pautas que establecen límites de conducta a los niños es el factor más representativo del patrón de crianza en este grupo, tomando el valor 0,82 en el intervalo unitario y ubicándose en el percentil 0,84 del total de familias.

En términos descriptivos, sólo un 11 % de las familias presentan un patrón más precario en términos de relación familiar y condiciones económicas, sin embargo, si bien la figura paterna puede no estar asegurada, cuando existe, participa en la crianza, percibiéndose un clima no violento al interior de la familia. A pesar de tal precaria condición familiar, este patrón también se asocia con una estimulación cognitiva e iniciación en el lenguaje verbal del niño, en tanto la madre, además de ser evaluada con un lenguaje fluido, estimula el desarrollo del niño por medio de la lectura frecuente de cuentos, la conversación con su hijo y la participación en situaciones de aprendizaje, lo que se traduce en que sólo un tercio de la muestra se ubica en una mejor situación en términos de pautas de estimulación. Además, corresponde a un entorno familiar favorable para el desarrollo de los niños, con prácticas de promoción del desarrollo social de manera contenida y educativa, favoreciendo una socialización fuera del hogar por medio de una de las actividades más relevantes en el desarrollo durante la infancia, el juego con pares, además de presentar interacciones cotidianas que sostienen el cuidado y la crianza, y donde el clima de afecto y cuidado expresado por la madre en los primeros años de vida es clave. Finalmente, en este patrón se observa que la madre tiene un manejo disciplinar que establece límites claros, no involucra y descalifica el castigo físico, por lo que muestra un adecuado manejo de las normas y límites que guían la conducta de manera adecuada para el desarrollo del niño.

- *Grupo 2: Entornos familiares favorables para el desarrollo.*

  Este patrón de este entorno familiar se caracteriza por presentar la más alta valoración en el factor relación familiar y percepción de las condiciones económicas, ubicándose en el $0,48$ dentro del intervalo unitario y existendo sólo un 12 % de la muestra con mayor pertenencia al factor "Adecuada relación familiar y condición económica". Además de una alta valoración en la presencia de pautas de estimulación del desarrollo del niño, tomando un grado de pertenencia de 0,6, y ubicándose en el percentil 0,69 de la muestra. En relación al factor interacciones de calidad, en especial en relación al vínculo afectivo que promueve la madre, este patrón mantiene una situación favorable para el desarrollo de niños, con un grado de pertenencia de $0,83$ y ubicándose en el percentil 0,45 sobre el total de las familias. Por último, las adecuadas pautas que establecen límites de conducta a los niños es otra característica representativa de este patrón de crianza, valorizando en $0,82$ este factor y ubicándose en el percentil 0,78 de la muestra.

  En términos de la organización familiar y distribución de los roles parentales, este es el único grupo que presenta asociación con la disposición de tiempo de la madre para criar, además de presentar una figura paterna presente y participando en la crianza del niño. Por otra parte, el clima familiar es percibido como no violento y favorable para el desarrollo del niño, este es el único grupo asociado con una baja frecuencia de irritabilidad y nerviosismo de la madre, además de no asociarse con la presencia de violencia intrafamiliar. En cuanto a la seguridad económica, solo en este grupo la falta de dinero no es percibida como problema. Todo esto indica un entorno familiar con relaciones, clima y percepción de las condiciones económicas, que ofrece las condiciones necesarias para que el desarrollo de los niños ocurra de manera adecuada. Este patrón comprate con el grupo 1 las carcaterísticas de los otros factores, presentando mayor relevancia en variables relacionadas con la estimulación cognitiva y la iniciación en el lenguaje verbal que realizan los adultos responsables por medio de la lectura frecuente de cuentos, la conversación madre - hijo y la participación en situaciones de aprendizaje de los niños. Cabe destacar que este patrón es el que muestra mayor asociación con el lenguaje fluido de la madre, además de ser el único que se asocia positivamente con el adecuado tipo de programas de televisión visto por el niño. De manera relacionada, este patrón se asocia favorablemente con la frecuencia con que juega el niño con sus pares, es decir, la crianza favorece una socialización tanto dentro como fuera del hogar. Finalmente, se observa alta asociación con un manejo disciplinar de la madre que establece límites claros, la cual no involucra ni acepta el castigo físico,mostrando un adecuado manejo de las normas y límites que guían la conducta de manera adecuada para el desarrollo del niño.

- *Grupo* 3*: Entorno con carencias en las relaciones familiares y condiciones económicas y pocos recursos para la promoción y desarrollo de los niños.*

Este grupo o patrón de entorno familiar se caracteriza por ser uno de los más desfavorables, presentando baja valoración en casi todos los factores y variables. Se caracteriza por presentar la menor presencia de pautas de estimulación del desarrollo del niño, tomando el valor 0,32 en el factor ”presencia de pautas de estimulación del desarrollo del niño”, y ubicándose en el percentil 0,19 de la muestra. A esto se suma una relación familiar y percepción de las condiciones económicas precaria, en tanto el patrón toma el valor 0,28 en el factor asociado con una adecuada condición familiar y económica, ubicándose en el percentil 0,28 sobre el total de familias de la muestra. No obstante, presenta una adecuación moderada en las pautas que establecen límites de conducta a los niños, donde el patrón toma el valor 0,62 y se ubica prácticamente en la mitad de la muestra.

En términos descriptivos, este patrón muestra una condición precaria en la mayoría de los aspectos. Sin embargo, presenta una valoración media respecto de las pautas que establecen límites de conducta asociado con un adecuado manejo disciplinar de la madre en tanto establece límites sin castigo físico. Se observa una relación familiar y condición económica precaria, en tanto no tiene una figura paterna asegurada, además de una relación familiar que no se percibe como buena, donde la madre se percibe con frecuencia irritable o nerviosa y la falta de dinero sí se percibe como un problema. Tal vez la mayor complejidad respecto de este patrón, es la baja valoración en el factor pautas de estimulación del desarrollo, siendo un patrón de entorno que no ofrece actividades que estimulen y apoyen el desarrollo físico motor, cognitivo y lingûístico de los niños, con ambientes pobres y de escasos recursos sociales, cognitivos, y educativos para favorecer su adecuado desarrollo, en tanto la fluidez en el lenguaje verbal de la madre no es adecuada, asociado a esto, tanpoco se observan que la madre proponga conversaciones con el niño de modo de iniciarlo en el lenguaje, ni que participa en aprendizajes del niño.

- *Grupo* 4*: Precaria socialización fuera del hogar, pero con relevancia en el manejo disciplinar adecuado para el desarrollo del niño.*

El patrón de este entorno familiar se caracteriza por poseer una precaria socialización fuera del hogar con sus pares, en tanto presenta el menor grado de pertenecia al factor "Interacciones de calidad" (0,6) y se ubica en el percentil 0,12 de la muestra. No obstante, el patrón de este grupo está fuertemente organizado por el factor "Pautas que establecen límites de conducta", en tanto presenta una alta valoración (0,8) y se ubica en el percentil 0,78 de la muestra. Por otra parte, muestra una pertenencia moderada a los factores asociados con adecuadas condiciones y relaciones familiares (0,25), y presencia de pautas de estimulación del desarrollo (0,58), ubicándose en los percentiles 0,42 y 0,66 respectivamente.

En términos descriptivos, este patrón no promueve la calidad de las interacciones de los niños, por cuanto la afectividad en las formas maternas de vincularse se mantiene sin significatividad, y la variable relacionada con la socialización fuera del hogar con pares presenta una correlación negativa significativa. No obstante, la madre tiene manejo disciplinar con límites, descalificándo el castigo físico y no existiendo violencia intrafamiliar. En cuanto a la presencia de pautas de estimulación del desarrollo, solo la lectura de cuentos y la conversación de la madre con el niño están presentes en este grupo. Y respecto de las condiciones familiares, en términos generales existe una figura paterna que participa en la crianza.

- *Grupos* 5 *y* 6*: Dificultades para poner límites y potencial clima familiar violento.*

Estos grupos se caracterizan principalmente por presentar pautas de límites de conducta de gran precariedad, en tanto presenta un grado de pertenencia 0,27 al factor "Adecuadas pautas que establecen límites de conducta del niño", ubicándose en el percentil 0,16 de la muestra. Sumado a esto, presenta escasas pautas de estimulación además de desfavorables condiciones familiares, en tanto este grupo tiene un grado de pertenencia de 0,43 al factor asociado con la presencia de pautas de estimulación del desarrollo y de 0,2 al factor relativo a las adecuadas condiciones y relaciones familiares, comportandose de manera similar en estos aspectos al entorno 3, que corresponde al más precario y ubicándose el los percentiles 0,28 y 0,38 respectivamente.

En términos descriptivos, se observa que a este patrón pertenecen con mayor intensidad madres con falta de compentencias de crianza en cuanto a la disciplina y estimulación del desarrollo, que están al servicio de un desarrollo sano y potenciador de los niños. Además, muestra un entorno propenso a la presencia de violencia intrafamiliar, en tanto no existe una figuar paterna, las madres se encuentran frecuentemente irritadas o nerviosas, la falta de dinero es un problema, no descalifican el castigo físico y muestran un manejo violento de los límites.

Capítulo 5

Índice contextual basado en particiones borrosas.

5.1. Determinación del contexto.

En el capítulo 3 de esta memoria se han desarrollado técnicas para la generación de indices basados en familias consistentes de operadores de agregación que trabajan sobre conjuntos de datos con estructura jerárquica priorizada e información imprecisa. Una vez establecidos los indices para cada niño construidos de manera jerarquica, los factores y los sub-factores asociados a dichos indices, en el capítulo 4 de esta memoria se establece un procedimiento que da lugar a una clasificación borrosa no supervisada de todos los niños. Esta clasificación borrosa (no necesariamente una partición de Ruspini) establece una serie de entornos borrosos del conjunto de todos los niños X. Estos entornos serán de gran utilidad para la elaboración de los indices contextuales que tengan en cuenta los logros alcanzados por cada niño en base a su entorno y no a una población de referencia ajena. En este capítulo, se utiliza la clasificación borrosa generada a través de dicho sistema de clasificación, para construir índices contextuales.

Con el término *contextual* nos referimos a índices que contienen información del entorno del elemento que está siendo evaluado, bajo el supuesto que dicho entorno aporta información respecto de los factores que inciden positiva o negativamente en el logro medido por dicho índice.

Una clasificación borrosa del conjunto de todos los niños X en las clases $K = \{K_1, ..., K_c\}$, entrega información acerca del entorno de cada elemento $x \in X$ por medio del grado de pertenencia de cada elemento x a las diferentes clases del conjunto K. Denotaremos por $\mu_K(x)$ al vector c-dimensional $\mu_K(x) = (\mu_{K_1}(x), \ldots, \mu_{K_c}(x))$. De esta forma, $\forall x \in X, \mu_{K_j}(x) \in [0, 1]$ representa el grado en el cual el elemento x pertenece a la clase K_j, y por lo tanto, el vector de pertenencias de x $\mu_K(x)$ representa a su "*ubicación*" dentro de la partición de entornos. La siguiente definición establece el entorno contexto del cada elemento $x \in X$.

Definición 5.1. (Entorno Contexto Borroso). Dada una clasificación borrosa del conjunto de datos X, con conjunto de clases $K = \{K_1, ..., K_c\}$ y función de pertenencia $\mu_K(x) = (\mu_{K_1}(x), ..., \mu_{K_c}(x))$ $\forall x \in X$. Definimos contexto entorno asociado a un elemento x en base a un sistema de clasificación borroso, al conjunto borroso $E(x)$ construido a partir de sus alfa cortes como sigue:

$$E(x)_\alpha = \big\{y \in X : \phi\big[\varphi_2\left(\mu_{K_1}(x_i), \mu_{K_1}(y)\right), ..., \varphi_2\left(\mu_{K_c}(x_i), \mu_{K_c}(y)\right)\big] \geq \alpha\big\} \cup \{x\},$$

donde ϕ y φ son T normas y T co-normas respectivamente.

Es decir, para un nivel de exigencia α, diremos que un elemento $y \in Y$ esta en $E(x)$ si x e y pertenencen simultaneamente a alguna de las clases de K.

Notese que es posible contruir el conjunto borroso $E(x)$ en su manera estandar (es decir $E(x) = \big\{(y, \mu_{E(x)})y \in Y\big\}$) a partir de los alfa cortes del conjunto $E(x)$. De esta manera se tiene que para todo $x \in X$,

$$\mu_{E(x)}(y) = Max\,\{\alpha\ /\ y \in E(x)_\alpha\}$$

Para el caso particular en el que los operadores ϕ y φ sean el máximo y el mínimo respectivamente, se tiene que:

$$E_\alpha(x) = \Big\{ y \in X : Max_c \Big\{ Min_2 \big[\mu_{K_1}(x), \mu_{K_1}(y)\big], ..., Min_2 \big[\mu_{K_c}(x), \mu_{K_c}(y)\big] \Big\} \geq \alpha \Big\}$$

Una vez establecido el contexto de cada elemento x del conjunto de datos X, es posible definir índices contextuales que contengan la información de su entorno $E_\alpha(x)$.

Estos indices contextuales serán construidos teniendo en cuenta que el entorno es un concepto borroso, y por tanto serán funciones que dependerán de los diferentes alfa cortes asociados al entorno $E(x)$. En el capitulo 3 se han construido diferentes indices, factores, subfactores y puntuaciones para cada niño. Sin pérdida de generalidad, denotemos por $Z(x)$ al valor, puntuación o índice obtenido por el elemento $x \in X$.

La pregunta que tratamos de resolver en esta sección ahora es la de como construir a partir de este indice $Z(x)$ un índice contextual. Nótese que fijo $\alpha \in [0,1]$, el conjunto $E(x)_\alpha$ representa el conjunto nitido de elementos que se encuentran en el entorno de x a nivel α. Teniendo en cuenta esto, se proponen algunos indices contextuales:

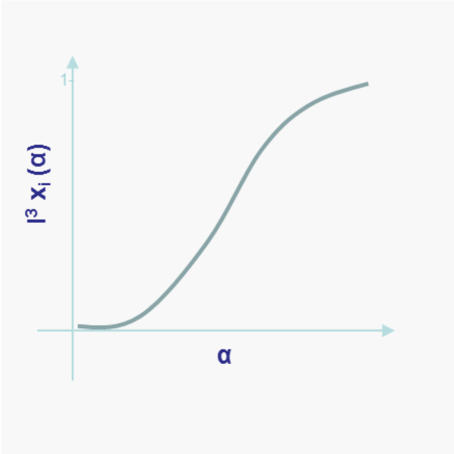
- $I_x^1(\alpha) = Pc_{\{Z(y), y \in E(x)_\alpha\}}(z(x))$, $\forall \alpha \in [0,1]$
  que corresponde a la ubicación porcentual que tiene el logro del elemento x, sobre el conjunto de puntuaciones obtenidos por su entorno $E_\alpha(x_i)$.

- $I_x^2(\alpha) = \frac{Z(x)}{Max\{Z(y), y \in E(x)_\alpha\}}$, $\forall \alpha \in [0,1]$
  que corresponde a la proporción de logro obtenido por x, suponiendo que el logro esperado es el máximo alcanzado por los elementos de su entorno $E_\alpha(x)$.

- $I_x^3(\alpha) = \frac{Z(x)}{Pc_{95}\{Z_y, y \in E(x)_\alpha\}}$, $\forall \alpha \in [0,1]$
  que corresponde a la proporción de logro obtenido por el elemento x, suponiendo que el logro esperado es el percentil 95 del conjunto de puntuaciones de su entorno $E_\alpha(x)$ definido por la puntuacion Z.

Por lo tanto, cada elemento x del conjunto de datos X, tiene asociado una función que representa el resultado del índice contextual bajo distintos niveles de exigencia sobre el nivel de pertenencia a las clases borrosas que respresentan su entorno. De esta manera, cuando el valor de α es pequeño, el logro de x_i es comparado prácticamemte con todo el conjunto X, pero a medida que aumenta el valor de α, el entorno de x va siendo cada vez más pequeño, y por tanto su logro es comparado con los elementos de X más similares respecto de su pertenencia a las clases de la clasificación borrosa dada por K.

Si por ejemplo utilizamos el índice $I_x^2(\alpha)$, la función asociada a un x con una puntuación promedio tendrá la siguiente forma:



Para valores pequeños de α, el entorno de x abarcará gran parte del conjunto X y por lo tanto, el nivel de exigencia sobre su logro será mayor, mientras que para valores grandes de α, el entorno será más reducido y su logro final dependerá de los elementos más símiles a él y por tanto, será menos exigente el baremo asociado. Ahora bien, si x es el elemento cuya puntuación es el máximo de todo el conjunto X, tal función será constante en 1, mientras que si es el mínimo, será constante en un valor cercano a 0.

++ EJEMPLOS ++++

5.2. Construcción de índices contextuales de desarrollo infantil temprano.

Como vimos en la sección anterior, el término *contextual* se refiere a la ubicación de lo observado sobre un entorno específico. Respecto de la contrucción de índices, y bajo el supuesto que dicho entorno aporta información respecto de los factores que inciden en el logro medido por dicho índice, un *índice contextual* considera información del entorno del elemento evaluado. Llevado a la construcción de índices de desarrollo infantil temprano, el contexto corresponde al entorno del hogar del niño, en tanto constituye un factor diferenciador respecto del logro del desarrollo de un niño.

En materia de evaluación de políticas públicas, los indicadores se clasifican bajo cuatro aspectos no disjuntos, la medida, el nivel de intervención, la jerarquía y la calidad. El primero distingue entre indicadores cuantitativos (real valorados) o cualitativos (abstractos), mientras que el nivel de intervención se refiere a indicadores de impacto (efecto a corto, mediano y largo plazo), de resultado (efecto inmediato), de producto (bienes o servicios) y de proceso (seguimiento). La jerarquía por su parte, distingue entre indicadores de gestión (administrativos y operativos) y estratégicos (método y solución), y por último, los de calidad, entre indicadores de eficacia (logro de objetivos bajo un escenario ideal), eficiencia (logro de objetivos en el tiempo y costo planeado) y efectividad (logro de objetivos bajo un escenario real). Este último tipo de indicador, tiene el mismo objetivo que los índices contextuales a los que nos referimos, en tanto son medidas que consideran las condicionantes que afectan a la consecusión del logro.

En el *Capítulo 2*, se presentaron algunos enfoques acerca del desarrollo del niño. Vimos la conveniencia de considerar el entorno del hogar del niño al momento de estudiar y comparar su nivel de logro, en tanto constituye un factor diferenciador. Además, se presentan estudios que señalan la inadecuación de los tests tradicionales para revelar la capacidad de aprender en niños en los que factores no intelectuales son la causa de sus fallos. Siguiendo este mismo enfoque, vimos que se ha avanzado en la evaluación del *potencial de aprendizaje* en la línea de Vygotzki, en la cual su objetivo no es medir tan sólo la ejecución de los sujetos, sino su posibilidad de aprendizaje.

En este trabajo, se propone una medida de logro de desarrollo infantil visto como la proporción del *potencial de desarrollo* logrado por el niño al momento de la medición. Dicho potencial puede ser estimado a partir del máximo puntuación conseguido por niños que tienen similares condicionantes frente al logro de su desarrollo, de esta manera, se obtiene un índice de desarrollo contextualizado, donde el grupo de referencia o normativo comparte características del entorno del hogar del niño evaluado.

En el *Capítulo 3* fueron generados los índices de entorno del hogar del niño, en el *Capítulo 4* se realizó una clasificación borrosa de los niños en función de dichos índices de entorno familiar, y en la sección anterior de esta capítulo, se definió la manera de establecer el contexto de los niños a partir de dicha clasificación borrosa de entornos, y de la definición de un α-corte, constuyendo con esto, la población normativa de cada niño. Todo esto nos permite construir los índices contextuales de desarrollo infantil temprano.

Sea $Z = \{z_1, ..., z_n\} \in [0, 1]^5$ el conjunto de datos de n niños que contiene la información de sus puntuaciones de desarrollo en las áreas comunicación, cognición, adaptación, motora y personal/social. Y sea Z_{x_i} el conjunto de puntuaciones de los niños que pertenecen al entorno del i-ésimo niño. Entonces, tal como vimos en la sección anterior, el conexto o entorno del i-ésimo niño se establece mediante el siguiente conjunto:

$$E_\alpha(x_i) = \Big\{ y \in X : Max_c \Big\{ Min_2\big[\mu_{K_1}(x_i), \mu_{K_1}(y)\big], ..., Min_2\big[\mu_{K_c}(x_i), \mu_{K_c}(y)\big] \Big\} \geq \alpha \Big\}$$

donde $\alpha \in (0, Max_c\{\mu_{K_{j\in\{1,..,c\}}}(x_i)\})$, representa el nivel de exigencia sobre la pertenencia a las clases borrosas de la partición K de los elementos de X que constituyen el entorno de x_i.

El potencial de desarrollo del niño lo estimamos a partir del logro ubicado en el percentil 95 en una población con similares condiciones de entorno del hogar, definida por $E_\alpha(x_i)$ para un α determrinado. Constituyéndose de esta forma la referencia o baremo del i-ésimo niño mediante:

$$P_\alpha(z_i) = Max\,\{z_i, Pc_{95}\{Z_{x_i}\}\}$$

Finalmente, el índice contextual de desarrollo infantil temprano representará la proporción del potencial de desarrollo que el niño ha alcanzado al momento de la medición, y se define por:

$$\mathrm{I}_\alpha(z_i) = \ z_i \ / \ P_\alpha(z_i)$$

Capítulo 6

Conclusiones, contribuciones y futuras líneas de trabajo.

6.1. Conclusiones.

Las conclusiones de este trabajo se presentan en relación a los objetivos planteados:

Objetivo 1: Desarrollar un método de construcción de índices.

- *Formular el problema de construcción de índices como un problema de agregación en ambiente borroso.*

  La naturaleza imprecisa de los conceptos que se estudian en las ciencias sociales y del comportamiento, dificulta el tránsito entre el desarrollo teórico de los conceptos y la representación numérica de los mismos. En este trabajo, el problema de la representación es facilitado al formularlo desde la lógica borrosa, en tanto permite representar los conceptos a través de variables lingüísticas asociados a conjuntos borrosos, en la cual el valor de verdad de sus estados rescata la vaguedad de dichos conceptos.

Siguiendo con el mismo enfoque, la construcción de índices puede ser considerado como un proceso de agregación en contexto borroso, asegurando un nivel de robustez en los resultados, en tanto a través de familias consistentes de operadores de agregación, es posible relacionar un conjunto de variables lingüísticas borrosas a fin de obtener un valor agregado que asegure estabilidad ante posibles cambios de cardinalidad del conjunto de datos. Por ejemplo, ante el usual problema de pérdida de información en una o más variables de algunos elementos del conjunto de datos de entrada, es posible agregar la información de dichos elementos de menor dimensionalidad si el proceso de agregación se ha definido en base a familias consistentes de operadores de agregación. En este trabajo, se formula el problema de construcción de índices como un proceso de agregación en contexto borroso, asociando los estados de las variables lingüísticas a conjuntos borrosos y definiendo la estructura de prioridades de las mismas, para luego, garantizar la robustez de los resultados mediante la definición de propiedades que deben satisfaer las familias de operadores de agregación involucradas.

- *Representar de manera más precisa la información dada por el conjunto de datos de entrada, intentando rescatar las valoraciones reales de las variables lingüísticas frente al concepto medido.*

La definición de los valores de verdad de variables lingüísticas mediante la asociación a conjuntos borrosos, asi como también, el establecimiento de una relación de orden y prioridad entre ellas frente a ditintos niveles de incidencia sobre el concepto medido, permite acercarse de mejor manera a la realidad, en tanto entrega mayor flexibilidad y precisión al momento de re-establecer el conjunto de datos de entrada en función de la definición teórica y la información disponible. En este trabajo, se utiliza la técnica cualitativa *juicio experto* para la definición de un conjunto de datos de entrada en contexto borroso, lo cual permite determinar su estructura (desestructurada, lineal o jerárquica), asi como tambien, la definición de pertenencias a los estados de las variables lingüísticas.

- *Establecer restricciones que garanticen estabilidad en los resultados de un proceso de agregación, en concordancia con las características del conjunto de datos de entrada.*

  La actual definición de familia de operadores de agregación no condiciona la existencia de una relación entre sus miembros, permitiendo definiciones en base a un conjunto de operadores desconectados para distintos dimensiones del conjunto de datos, con lo cual no es posible asegurar consistencia en el proceso de agregación en el cual participa. En este trabajo, se establecen restricciones de consistencia sobre las familias de operadores de agregación, en el sentido que el resultado obtenido en base a un operador definido para n ítems, no difiera del resultado obtenido en base a un operador definido para $n-1$ ítems, cuando el elemento adicionado corresponde a la agregación de los elementos ya contenidos en el conjunto de datos. En base a esta idea, se definen distintos niveles de estabilidad estricta (absoluta, en el límite y en probabilidad), tanto para datos sin una estructura inherente, como para conjuntos de datos linealmente ordenados y con estructura jerárquica priorizada. De esta forma, dada una estructura determinada del conjunto de datos de entrada, es posible definir las familias de operadores de agregación que se utilizarán en el proceso de agregación en base al nivel de estabilidad que se requiere para asegurar resultados robustos.

- *Desarrollar un método de búsqueda de patrones en base a la evaluación de calidad de una clasificación no supervisada bajo ambiente borroso.*

Los métodos comúnmente utilizados en clasificación no supervisada, requieren la definición *a priori* de la cantidad de clases que tendrá la clasificación borrosa del conjunto de datos, cuestión que se ha resuelto en base a la definición de medidas que evalúan la calidad de las particiones para distinto número de clases. No obstante, estas medidas se suelen basar sólo en un criterio, tendiendo a valorar aquellas particiones borrosas con bajo nivel de solapamiento. En este trabajo, se establece un método de selección de patrones que determinarán clases borrosas de una partición, mediante el comportamiento conjunto de índices de relevancia, redundancia y cobertura del sistema de clasificación borroso no supervisado. Este método no requiere la definición anticipada de la cantidad de clases que tendrá la partición, ya que se evalúan todas aquellas clases posibles de generar en base a un conjunto de patrones en la cual sus coordenadas son definidas y acotadas sobre una rejilla o cuadrícula determinada. La definición de esta rejilla o cuadrícula debe cubrir la amplitud de los valores que toman las variables del conjunto de datos de entrada. Durante el proceso de búsqueda de patrones, se establecen restricciones de calidad en función de los indicadores mencionados.

Objetivo 2: Generar un modelo borroso del desarrollo infantil temprano aplicado a la realidad chilena.

- *Construir un modelo teórico del entorno del hogar del niño, en función de las condicionantes que inciden en el logro de su potencial de desarrollo.*

La codificación de la información cualitativa mediante variables lingüísticas nítidas, en la cual la distancia entre las etiquetas es la misma, se considera poco preciso cuando se trata de representar conceptos abstractos. En este trabajo, mediante la técnica *juicio experto*, se establecen nuevos valores de verdad y un orden de prioridad de las distintas variables lingüísticas disponibles, asi como también, la estructura del conjunto de datos de entrada. Las entrevistas semi-estructuradas realizadas a dos expertos en desarrollo infantil, permitieron estudiar el tema y el enfoque propuesto, entregando como resultado una estructura teórica del entorno del hogar de la población infantil chilena. La estructura resultante del conjunto de datos de entrada es jerárquica y priorizada, en tanto el concepto global es dividido en dimensiones desestructuradas, las que a su vez fueron divididas en sub-dimensiones con una estructura de prioridad lineal, en la cual cada una está constituida por variables lingüísticas que también tienen una estructura lineal de prioridades.

- *Generar índices de factores asociados a un entorno familiar que fomenta el desarrollo del niño.*

  El método para determinar los pesos de las variables lingüísticas para la construcción de índices de entorno del hogar del niño es un problema abierto, siendo el análisis de componentes principales la técnica comúnmente utilizada para estos fines. Sin embargo, su principal inconveniente es que se restringe a relaciones lineales entre sus variables, sin considerar las características del conjunto de datos de entrada. En este trabajo, se generan índices de entorno del hogar del niño mediante procesos de agregación definidos para una estructura jerárquica priorizada del conjunto de datos de entrada, en la cual se utilizan familias consistentes de operadores de agregación. Para la agregación de los datos sin una estructura inherente, se utilizó la familia estrictamente estable dada por la media aritmética. Mientras que para la agregación de los datos con una estructura lineal hacia la derecha, en la cual su ordenamiento está asociado a una prioridad, se utilizó la familia estrictamente estable por la derecha, dada por la media ponderada. Esta última familia está constituida por un tipo de *operador priorizado* definido por *Yager* ([92]), cuya característica fundamental es no permitir la compensación entre variables.

- *Identificar patrones representativos de los distintos entornos del hogar del niño.*

  La información que se utiliza en estudios de ciencias sociales y del comportamiento, suele presentar un nivel de concentración que dificulta la partición del conjunto de datos de entrada en clases que permitan identificar diferencias entre ellas. Nuestra temática no está ajena a esta dificultad, y frente a esto, en este trabajo se identificaron cuatro tipos de entornos del hogar del niño mediante un método de búsqueda de patrones en el cual se utilizan índices de calidad de las clases borrosas. Dicha búsqueda de patrones se realiza sobre un conjunto de posibles patrones definidos sobre una rejilla o cuadrícula, de manera que abarque lo mejor posible la amplitud de valores que pueden tomar las variables del conjunto de datos de entrada, para luego evaluar las clases que se generan a partir de dichos patrones en función de índies de relevancia, redundancia y cobertura. Por lo tanto, además de no requerir *a priori* la cantidad de clases de la partición, garantiza que las clases seleccionadas son las mejor evaluadas respeto de un conjunto de clases que cubren la amplitud de los valores que toman las variables de entrada.

6.2. Contribuciones.

Si bien el enfoque de este trabajo ya constituye una contribución en el desarrollo de índices bajo un escenario psicosocial, en tanto se le otorga importancia al contexto con el objetivo de relativizar las medidas de evaluación, claramente, el principal aporte se presenta en el tercer capítulo. En este capítulo, se evidencia la necesidad de establecer restricciones en la definición de ***familia de operadores de agregación (FAO)***, a fin se asegurar algún nivel de consistencia del proceso de agregación en el que participa, y de esta forma, garantizar resultados agregados robustos.

Frente a esto, se definen los conceptos de ***consistencia*** y ***estabilidad*** en el contexto de un problema de agregación, y se presentan definiciones de tres niveles de ***estabilidad estricta*** para una familia de operadores de agregación, en términos absolutos, en el límite y en probabilidad. Tales definiciones, consideran la naturaleza del conjunto de datos de entrada en términos de su estructura, focalizándonos en datos desestructurados, y en datos con estructura lineal y jerárquica priorizada. El análisis de estas propiedades permite conocer *a priori* el nivel de estabilidad que tendrán los resultados, luego de un proceso de agregación bajo determinada estructura del conjunto de datos.

Los tres niveles de estabilidad estricta de familias de operadores de agregación que son utilizadas para agregar datos con una estructura lineal, se han definido para datos cuyo orden lineal va desde la izquierda hacia la derecha como ***R-estrictamente estable***, y por tanto, el último elemento que se agrega es la agregación de los anteriores, y para aquellos que van desde la derecha hacia la izquierda como ***L-estrictamente estable***, y por tanto, el primer elemento es la agregación de los siguientes. Esto deja el camino para extender dichas definiciones estabilidad para el caso en que se agregue un elemento en la posición i-ésima desde la derecha o j-ésimo desde la izquierda, y de esta forma definir las propiedades ***iR-estrictamente estable*** y ***jL-estrictamente estable***.

Otra contribución de este trabajo se muestra en el cuarto capítulo, en el cual se utilizan familias consistentes de operadores de agregación para definir los criterios de ***relevancia***, ***redundancia*** y ***cobertura*** de las clases de una clasificación borrosa de un conjunto de datos. En base a la combinación de estos tres criterios, se genera un método de evaluación de calidad de un sistema de clasificación borrosa no supervisada.

Una de las principales ventajas de este método, es que **no exige la definición previa de la cantidad de clases que tendrá la partición**, en tanto evalúa la calidad de un conjunto de clases borrosas determinadas por los posibles patrones de clase de una partición, permitiéndo **seleccionar las mejores clases en función de criterios de calidad** . Este conjunto de posibles patrones se definen en una rejilla o cuadrícula, de tal manera que cubran los valores que toman los variables de entrada, considerando además, todas las permutaciones que existen entre dichas variables. Este proceso de búsqueda de buenos patrones, y por consiguiente, de una clasificación borrosa deseable de un conjunto de datos, va eliminando clases en la medida que no satisfacen las restricciones establecidas sobre los criterios de relevancia, redundancia y cobertura.

Finalmente, la contribución del quinto capítulo reúne las de los capítulos anteriores, en tanto se define el **contexto de un elemento a través de una clasificación borrosa realizada en base a familias consistentes de operadores de agregación**. Si bien se deja apertura respecto del tipo de índice contextual que se utilizará, a lo largo de este trabajo se establecen las pautas tanto para representar el conocimiento de una manera más precisa, construir índices contextuales mediante procesos de agregación robustos, que consideran la estructura de los datos que se agregan, y para establecer baremos relativos a cada elemento evaluado.

6.3. Futuras líneas de trabajo.

La representación numérica de los conceptos abstractos medidos mediante técnicas cualitativas, aún tiene muchos ámbitos de mejoramiento. En este trabajo, se utiliza información cualitativa que ha sido rescatada desde los discursos de las personas, sin embargo, tales discursos han sido "*cuantificados*" mediante variables lingüísticas cuyos estados o categorías son nítidas y presentan el mismo nivel de importancia ante el concepto medido. Frente a esto, se ha intentado "rescatar" la riqueza de los discursos mediante un trabajo conjunto con expertos en desarrollo infantil, asociando las variables lingüísticas a conjuntos borrosos que representan los distintos estados de dichas variables, estados que además tienen diferentes grados de importancia frente a su correspondiente variable lingüística. De esta forma, cada niño tiene asociado un grado de pertenencia a cada uno de los estados o categorías de las variables lingüísticas, aportando una mayor flexibilidad a la representación del concepto. Al respecto, una línea de trabajo futuro es contribuir al mejoramiento de la representación numérica de este tipo de conceptos, sobre la base de los discursos originales extraidos mediante técnicas cualitativas, y de esta forma, poder definir los estados de las variables lingüísticas y sus correspondientes grados de pertenencia desde la fuente original.

Las definiciones de estabilidad esticta en familias de operadores de agregación que son utilizadas sobre datos con una estructura lineal, se han definido en cada uno de los niveles para datos cuyo orden va desde la izquierda hacia la derecha como *R-estrictamente estable*, y por tanto, el último elemento que se agrega es la agregación de los anteriores. Mientras que para aquellos que van desde la derecha hacia la izquierda como *L-estrictamente estable*, y por tanto, el primer elemento es la agregación de los siguientes. Ciertamente, una línea de trabajo futuro es extender estas propiedades para el caso en que se agrege un elemento en la posición i-ésima desde la derecha o j-ésima desde la izquierda, y de esta forma definir las propiedades *iR-estrictamente estable* y *jL-estrictamente estable* para cada uno de los niveles de estabilidad (absoluta, asintótica y en probabilidad). Esta extensión ayuda resolver el problema de información faltante en alguna variable de un elemento del conjunto de datos, en tanto permite generar un operador de la misma familia pero de una dimensión menor, manteniendo la estabilidad de los resultados de dicho elemento.

En este trabajo, el entorno del hogar de cada niño se constituye a través del grado de pertenencia a cada uno de las clases borrosas de la partición del conjunto de datos, información que puede ser utilizada para establecer su población normativa en función de una medida de similitud entre entornos familiares. Esto lleva a plantearnos un par de futuras líneas de investigación, una orientada a definir una medida de similitud entre vectores en las cuales sus coordenadas representan grados de pertenencia, y que por tanto, tal medida debe considerar que cada coordenada está asociada a un concepto diferente, implicando relevar la posición del dato en el vector de coordenadas de pertenencias. Y otra que permita comparar los resultados obtenidos en base al método de clasificación borrosa no supervisada propuesto, con distintos métodos comúnmente utilizados, y bajo diferentes ámbitos de aplicación.

Bibliografía

[1] Amo, A., Gómez, D., Montero, J., Biging, V. Relevance and Redundancy in fuzzy clasification systems. Mathware and Soft Computing. 8, 203-216 (2001).

[2] Amo, A., Montero, J., Molina, E. Representation of consistent recursive rules. European Journal of Operational Research. 130, 29-53 (2001).

[3] Anders, Y., Rossbach, H-G., Weiner, S., Ebert, S., Kuger S., Lehrl S., von Maurice, J. Home and preschool learning environments and their relations to the development of early numeracy skills. Early Childhood Research Quarterly. 37(2), 231-244 (2012).

[4] Bahamondes, F., Becerra, M., Sánchez, D. Heckman, Behrman, Hertzman, Välimäki, Tremblay, Melhuish, Vega, Urzúa: Un compendio de publicaciones sobre primera infancia. Junta Nacional de Jardines Infantiles, JUNJI. Ministerio de Educación de Chile. Santiago de Chile (2009).

[5] Beliakov, G., Pradera, A. , Calvo, T. Aggregation Functions, a Guide to Practitioners. Springer-Verlag. Berlín (2007).

[6] Beltrán, M. Cinco vías de acceso a la realidad social. En García Ferrando, Ibañez y Alvira. El análisis de la realidad social. Métodos y técnicas de investigación. Alianza. Madrid (1986).

[7] Bezdek, J. Pattern recognition with fuzzy objective function algorithms. Kluwer Academic Publishers Norwell. MA, USA (1981).

[8] Boudon R., Lazarsfeld P. Metodología de las ciencias sociales. I. Conceptos e índices. Laia. Barcelona (1973).

[9] Bradley, R.H., Caldwell, B. M., Corwyn, R. F. The child care HOME Inventories: Assessing the quality of family child care homes. Early Childhood Research Quarterly. 18, 294-309 (2003).

[10] Bradley, R.H., Caldwell, B.M. The HOME inventory and family demographics. Developmental Psychology. 20, 315-320 (1984).

[11] Bronfenbrenner, U. La ecología del desarrollo humano. Paidos. Barcelona (1987).

[12] Burston A, Puckering C, Kearney E. At HOME in Scotland: validation of the home observation for measurement of the environment inventory. Child Care Health Dev. 31(5), 533-8 (2005).

[13] Bustince, H. , de Baets, B. , Fernández, J. , Mesiar, R. , Montero, J. A generalization of the migrativity property of aggregation functions. Information Sciences. 191, 76-85 (2012).

[14] Bustince, H., Barrenechea, E. ,Fernández, J. , Pagola, M. , Montero, J., Guerra, C. Aggregation of neighbourhood information by means of interval type 2 fuzzy relations.

[15] Calvo, T., Kolesarova, A., Komornikova, M., Mesiar, R. Aggregation operators, properties, classes and construction methods. In T. Calvo et al. (Eds.): Aggregation Operators New trends ans Aplications. Physica-Verlag. 3-104, Heidelberg (2002).

[16] Calvo, T., Mayor, G., Torrens, J., Suñer, J., Mas, M., Carbonell, M. Generation of weighting triangles associated with aggregation fuctions. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. 8 (4), 417-451 (2000).

[17] Carlsson, C.H. , Fuller, R. Fuzzy reaasoning in decision making and optimization. Springer-Verlag. Heidelberg (2002).

[18] Castaño, E. Diseño del sistema de identificación de beneficiarios de programas sociales, Informe Técnico. CONAM. Quito (1999).

[19] Castaño, E. Construcción de un indicador de Calidad de vida para Jamaica. BID. (2000).

[20] Castaño, E. El indicador multidimensional de condiciones de vida. Centro de estudios de opinión, Facultad de ciencias sociales y humana, Universidad de Antioquía. (2008).

[21] Castaño, E., Correa, C., Salazar B. La construcción de un indicador de Calidad de Vida para la ciudad de Medellín. Misión Social, DNP. DNP (1998).

[22] Castaño, E., Moreno, H. Metodología Estadística para la Selección de Variables del Sistema de Beneficiarios de Programas Sociales, SISBEN. Misión Social, DNP. DNP (1994).

[23] Castaño, E., Valencia, L.A. Indicador de Calidad de los Estratos para el Área Metropolitana de Medellín. Revista Lecturas de Economía. 50 (1999).

[24] Castiñeira, E., Calvo, T., Cubillo, S. A first approach to an axiomatic model of multi-measures. Proceedings of the 11th International Conference on Computational and Mathematical Methods in Science and Engineering, CMMSE. 1, 296-301 (2011).

[25] Creswell, J., Plano, V. Designing and conducting Mixed methods research. SAGE Publications. Los Angeles (2011).

[26] Cunha, F., Heckman, James J. Investing in our Young People. NBER Working Paper 16201. Department of Economics, University of Chicago (2006).

[27] Cutello, V., Montero, J. Hierarchical aggregation of OWA operators: basic measures and related computational problems. Uncertainty, Fuzzinesss and Knowledge-Based Systems. 3, 17-26 (1995).

[28] Cutello, V., Montero, J. Recursive families of OWA operators. Proceedings FUZZ-IEEE Conference. IEEE Press. 1137-1141, Piscataway (1994).

[29] Cutello, V., Montero, J. Recursive connective rules. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. 14, 3-20 (1999).

[30] De Baets B., Fodor J., Serodio C., Couto P., Melo-Pinto P. Eurofuse 2011: Workshop on Fuzzy Methods for Knowledge-Based Systems (Advances in Intelligent and Soft Computing). Springer-Verlag. Berlin and Heidelberg GmbH and Co. K (2011).

[31] Dubois, D. , Gottwald, S. , Hajek, P. , Kacprzyk, J. , Prade, H. Terminological difficulties in fuzzy set theory - the case of intuitionistic fuzzy sets. Fuzzy Sets and Systems. 156, 485-491 (2005).

[32] Duda, R.O., Hart, P.E. Pattern classification and scene analysis. Wiley. New York (1973).

[33] Dunn, J. C. Well separated clusters and optimal fuzzy partitions. Journal of cybernetic. 4: 95-104 (1974).

[34] Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH, DHHS. The NICHD Study of Early Child Care and Youth Development (SECCYD): Findings for Children up to Age 4 1/2 Years. U.S. Government Printing Office. Washington, DC (2006).

[35] Evans, J. Myers,R. Illfeld, E. Early Childhood Counts- A Programming Guide on Early Childhood Care for Development. . Washington, DC (2000).

[36] Filmer, D., Pritchett, L. Estimating wealth effect without expenditure data- or tears: An application to educational enrollments in states of India. Demography. 38, 115-132 (2001).

[37] Fung, L.W., Fu, K.S. An axiomatic approach to rational decision making in a fuzzy environment. In: L.A. Zadeh et at. (Eds.): Fuzzy Sets and their Applications to Cognitive and Decision Processes. Academic Press. 227-256, New York (1975).

[38] Gath, I., Geva, A.B. Unsupervised optimal fuzzy clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence. 7, 773-781 (1989).

[39] Gómez, D., Montero, J. Sequential aggregation. In M. González et al. (Eds.) Proceedings of the Fifth International Summer School on Aggregation Operators, AGOP. Universitat de les Illes Balears, Palma de Mallorca. 183-187 (2009).

[40] Gómez, D., Montero, J. A discussion of aggregation function. Kybernetika. 40, 107-120 (2004).

[41] Gómez, D., Montero, J. Determinig the accuracy in image supervised classification problems. EUSFLAT: Advances in Intelligent Systems Research. 342 - 349 (2011).

[42] Gómez, D., Montero, J., Rodríguez, J.T., Rojas, K. Stability in aggregation operators. In Greco, S. et al. (Eds.) Advances in Computational Intelligence. Springer. 299, 317-325 (2012).

[43] Gómez, D., Montero, J., Yánez, J. A coloring algorithm for image classification. Information Sciences. 176, 3645-3657 (2006).

[44] Gómez, D., Montero, J., Yánez, J, Poidomani, C. A graph coloring algorithm approach for image segmentation. Omega. 35:173–183 (2007).

[45] Grabisch, M., Marichal, J., Mesiar,R., Pap, E. Aggregation Functions. Encyclopedia of Mathematics and its Applications. Cambridge (2009).

[46] Gustafson, D.E. , Kessel, W.C. . Fuzzy clustering with a fuzzy covariance matrix. In Proc. IEEE CDC. . 761-766, San Diego, CA, USA. (1979).

[47] Gutiérrez, D. The construction of indicators as an epistemological problem. División de Ciencias Sociales, Universidad de Sonora. Sonora, México (2009).

[48] Heckman, J. Skill formation and the economics of investing in disadvantaged children. Science. 312 (2006).

[49] Heckman, J. Skill formation and the economics of investing in disadvantaged children. Science. 312 (2006).

[50] Höppner, F., Klawonn, F., Kruse, R., Runkler, T. Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition. Wiley IBM PC. Chichester (1999).

[51] Husserl, E. Meditaciones cartesianas. FCE. México (2005).

[52] Klasen, S. Measuring poverty and deprivation in South Africa. Review of Income and Wealth. . 46, 33-58 (2000).

[53] Kolenikov, S., Angeles, G. The Use of Discrete Data in PCA: Theory, Simulations, and Applications to Socioeconomic Indices. Working Paper of Measure. Carolina Population Centre, University of North Carolina (2004).

[54] Kon, I. S. Psicología de la edad juvenil. Editorial Pueblo y Educación. La Habana, Cuba (1990).

[55] Landry, S.H., Smith, K.E., Swank, P.R. Responsive parenting: Establishing early foundations for social, communication, and independent problem-solving skills. Developmental Psychology. 42:627-642 (2006).

[56] Lazarsfeld, P.F. Qualitative Methods. Free press. New York (1972).

[57] Linver, M., Brooks, G., Cabrera, N. The Home Observation for Measurement of the Environment (HOME) Inventory: The Derivation of Conceptually Designed Subscales. Parenting: Science and Practice. 4(2/3), 99-114 (2004).

[58] López, V., Garmendia,L., Montero, J., Resconi, G. Specification and computing states in fuzzy algorithms. Uncertainty, Fuzzinesss and Knowledge-Based Systems. 16, 301-336 (2008).

[59] López, V., Montero, J. Software engineering specification under fuzziness. Multiple-Valued Logic and Soft Computing. 15, 209-228 (2009).

[60] López, V., Montero, J., Rodriguez, J.T. Formal Specification and implementation of computational aggregation Functions. Computational Intelligence, Foundations and Applications Proceedings of the 9th International FLINS Conference. 523-528 (2010).

[61] Luce, R. Duncan. The Mathematics Used in Mathematical Psychology. American Mathematical Monthly. 71 (4), 364-78 (1964).

[62] MacKenzie, S. Podsakoff, P. Podsakoff, N. Construct measurement and validation procedures in mis and behavioral research: integrating new and existing techniques. MIS Quarterly. 35 (2), 293-334 (2011).

[63] Maldonado H., Sepúlveda C., Vargas A. et all. Guía para Diseño, Construcción e Interpretación de Indicadores.Estrategia para el Fortalecimiento Estadístico Territorial. (DANE. Bogotá (2010).

[64] Montero, J. A note on Fung-Fu‘s theorem. Fuzzy Sets and Systems. 13, 259-269 (1985).

[65] Montero, J., López, V., Gómez, D. The role of fuzziness in decision making. Studies in Fuzziness and Soft Computing. 215, 337-349 (2007).

[66] Montero, J., Gómez, D., Bustince, H. On the relevance of some families of fuzzy sets. Fuzzy Sets and Systems. 158, 2429-2442 (2007).

[67] Nagar, A. L., Basu, S. Weighting socio-economic indicators of human development (a latent variable approach). National Institute of Public Finance and Policy. New Delhi (2001).

[68] Nolan, B., Whelan, C.T. Resources, Deprivation and the Measurement of Poverty, The Economic and Social Research Institute. Working Paper. 21 (1991).

[69] PNUD. Human Development Report. . New York (1997).

[70] Ramey, C., Smith, B. Assessing the Intellectual Consequences of early intervention with High-Risk Infants. American Journal of Mental Deficiency. 8 (4), 318-324 (1976).

[71] Ramey, S. L., Ramey, C. T. Early childhood experiences and developmental competence. In S. Danziger, J. Waldfogel (Eds.), Securing the future: Investing in children from birth until college. Russell Sage Foundation. 122-152, New York (2000).

[72] Robert H. B., Caldwell, B. M. The HOME Inventory and Family Demographics. Developmental Psychology. 20 (2), 315-20 (1984).

[73] Rojas, K., Gómez, D., Rodríguez, J.T., Montero, J. Some properties of consistency in the families of aggregation operators. In B. de Baets (Ed.) et al., Eurofuse 2011, AISC. Springer. 107, 169-176 (2011).

[74] Rojas, K., Gómez, D., Rodríguez, J.T., Montero, J. Strictly stable families of aggregation operators. Fuzzy Sets and Systems. http://dx.doi.org/10.1016/j.fss.2012.12.010 (2013).

[75] Rojas, K., Gómez, D., Rodríguez, J.T., Montero, J., Valdivia, A., Paiva, F. Development of child's home environment indexes based on consistent families of aggregation operators with prioritized hierarchical information. Fuzzy Sets and Systems: Special Issue EUROFUSE 2011. (in second revision) (2013).

[76] Rowena, J., Smith, P., Goddard, M. Measuring Preformance: An Examination. Centre for health economics. The University of York (2004).

[77] Ruspini, E. H. A new approach to clustering. Information and control. 15, 239-253 (1978).

[78] Saporta, G. Multidimensional data analysis and quantification of categorical variables, en New Trends in Data Analysis and Applications. J. Janssen, J.F. Marcotorchino, J.M. Proth Eds., Elsevier Science Publishers B.V. North-Holland (1983).

[79] Shonkoff, J.P., Phillips, D. A. . From Neurons to Neighborhoods: The Science of Early Childhood Development. National Research Council and Institute of Medicine. Committee on Integrating the Science of Early Childhood Development. Board on Children, Youth, and Families, Commission on Behavioral and Social Sciences and Education. Washington, DC (2000).

[80] Sigelman, C., Rider, E. Life-span human development. Wadsworth Cenage Learning. Copyright 2008 by Marsha Ironsmith and Marion Eppler. Belmont CA (2009).

[81] Slocumb E., Cole F. A practical approach to content validation. Applied Nursing Research. 4, 192 - 195 (1991).

[82] Spearman, Charles, C. Correlation calculated from faulty data. British Journal of Psychology. 3, 271-295 (1910).

[83] Strasser, K. Evaluación de programas de intervención temprana. Expansiva. ISSN 0717-9987 (2006).

[84] Torra, V. Fuzzy c-means for fuzzy hierarchical clustering. FUZZ-IEEE 2005. p.646 - 651, Reno, Nevada (2005).

[85] Totsika, V., Sylva, K. The Home Observation for Measurement of the Environment revisited. Child and Adolescent Mental Health. 9 (1), 25-35 (2004).

[86] Townsend, P. Poverty in the United Kingdom, a Survey of Household Resources and Standards of Living. Penguin Books and Allen Lane. London (1979).

[87] Tzuriel, D. Dynamic Assessment is not Dynamic Testing. Issues in Education. 7(2), 237-250 (2001).

[88] Verkuilen, J. Assigning Membership in a Fuzzy Set Analysis. Sociological Methods and Research. 33 (4), 462-496 (2005).

[89] Vigotsky, L. S. El Problema de la Edad. En Problemas de la Psicología Infantil. Editorial Pedagógica. Moscú (1984).

[90] Vigotsky, L. S. El Problema del Entorno. En: Fundamentos de la Podología. Izdanie Instituto. Leningrado (1935).

[91] Yager, R.R. On ordered weighted averaging aggregation operators in multi-criteria decision making. IEEE Transactions on Systems, Man and Cybernetics. 18, 183-190 (1988).

[92] Yager, R.R. Prioritized aggregation operators. International Journal of Approximate Reasoning. 48:263-274 (2008).

[93] Yager, R.R., Rybalov, A. Nonconmutative self-identity aggregation. Fuzzy Sets and Systems. 85:73-82 (1997).

[94] Young, F.W., Takane, Y. y de Leeuw, J. The Principal Components of Mixed Measurement Level Multivariate Data: An Alternanting Least Squares Method with Optimal Scaling Features. Psychometrika. 43, 279-281 (1978).

[95] Zadeh, L. Fuzzy Sets. Information and control. 8, 338-53 (1965).