

## Reseñas de trabajos de grado de pregrado del Programa de Matemáticas de la Universidad Nacional de Colombia – Sede Bogotá presentados en el período 2021-II

Omar Duque Gómez<sup>a</sup>

### 1. Practicas laborales-Contrato de aprendizaje en Banco Caja Social

Estudiante **Edwin Alexander Acero Cuesta\***

Directora *Margaret Johanna Garzón\*\**

Emails \*eaaceroc@unal.edu.co, \*\*mjgarzonm@unal.edu.co

RESUMEN. El presente Trabajo de Grado expone el proceso de práctica llevado a cabo en el Banco Caja Social en el área de Cobranzas, con el fin de realizar un estudio de el comportamiento de la cartera. Este trabajo se enfocó principalmente en la cartera caída, que es la de mayor interés para el área ya que se necesita que aumenten los pagos por parte de los deudores y disminuir la cartera que entra en mora. El estudio realizado está centrado en identificar, mediante un análisis estadístico, los períodos donde se presentan las caídas significativas y se analiza qué tipo de población es la que causa este comportamiento. Los procedimientos a realizar son un análisis descriptivo del portafolio, para así identificar patrones en indicadores de caída a cierre de cada mes, encontrar el segmento de población que mas tiende a no pagar su obligación al momento de llegar su fecha de pago y, adicionalmente, realización de regresiones para conocer qué variables son las que tienen mayor relación con la caída para que el Banco genere una mejor gestión al momento de cobrar y esté preparado para los tiempos donde se presentan estos períodos de aumento de cartera en mora.

---

<sup>1</sup>Coordinador del Programa Curricular de Matemáticas, Universidad Nacional de Colombia, Bogotá, Colombia

<sup>a</sup>oduqueg@unal.edu.co

## 2. Counterfactual Explanations of Symptoms Decreases in Schizophrenia using Digital Phenotyping

Estudiante **Juan Sebastián Cañas Silva\***

Director *Francisco Gómez\*\**, Codirector *Omar Costilla-Reyes\*\*\**

Emails \*jscanass@unal.edu.co, \*\*fagomezj@unal.edu.co,

\*\*\*omar.costillareyes@manchester.ac.uk

RESUMEN. La práctica clínica actual está saturada por una gran demanda de servicios sanitarios y los escasos recursos disponibles. Los nuevos paradigmas de datos de salud impulsados con herramientas de ciencia de datos posibilitan una mejora en el flujo de trabajo clínico en etapas cruciales del seguimiento a pacientes. En este trabajo, proponemos un sistema de aprendizaje automático capaz de predecir, detectar y explicar cambios individuales en los síntomas de pacientes con esquizofrenia, mediante el uso de datos de fenotipado digital. Pronosticamos síntomas de pacientes con un error por debajo de 10%, detectamos deterioros en los síntomas utilizando algoritmos de punto de cambio, y exploramos explicaciones contrafactuales como una alternativa en un escenario simulado de monitoreo continuo en salud. Esta propuesta va en línea con las aproximaciones hacia los sistemas integrados de apoyo a la toma de decisión en medicina personalizada.

## 3. Bredon Cohomology and Mackey Machine over Finite Groups

Estudiante **Jairo Esteban Castro Mora\***

Director *Andrés Ángel\*\**, Codirector *Edwin Becerra\*\*\**

Emails \*jaecastro@unal.edu.co, \*\*ja.angel908@uniandes.edu.co,

\*\*\*eabecerrah@unal.edu.co

RESUMEN. En este documento intentamos descomponer los grupos de cohomología de Bredon [1] para  $G$ -espacios utilizando herramientas de teoría de la representación. Comenzamos introduciendo todo el vocabulario necesario para los  $G$ -espacios, incluyendo los mapas  $G$ -equivariantes y los  $G$ -CW complejos, luego definimos la cohomología de Bredon ( $G$ -equivariante) para estos espacios. También introducimos la terminología básica de la Teoría de la Representación y las herramientas necesarias, específicamente los funtores de restricción e inducción que son fundamentales para la Teoría de Mackey [3]. Finalmente, aplicamos estas herramientas utilizando representaciones proyectivas, de forma análoga al trabajo realizado por Ángel, Becerra y Velásquez para la teoría  $K$ -equivariante [2].

**Palabras clave:** Cohomología Equivariante, Representaciones Proyectivas, Máquina de Mackey.

## Referencias

- [1] Glen E. Bredon, *Equivariant Cohomology Theories*, Springer Berlin Heidelberg, <https://doi.org/10.1007/bfb0082690>, 1967.
- [2] Andrés Angel and Edward Becerra and Mario Velásquez, *Proper actions and decompositions in equivariant K-theory*, 2020, arXiv:2003.09777.
- [3] George W. Mackey, *Unitary representations of group extensions. I*, Acta Mathematica, Institut Mittag-Leffler, Vol. 99, pp. 265 – 311, 1958, <https://doi.org/10.1007/BF02392428>.

### 4. Aproximación numérica de ecuaciones diferenciales parciales con la librería MFEM

Estudiante **Felipe Cruz Vasquez\***

Director *Boyan Lazarov\*\**, Codirector *Juan Galvis\*\**

Emails \*fcruzv@unal.edu.co, \*\*lazarov2@llnl.gov, \*\*\*jgalvisa@unal.edu.co

RESUMEN. Revisamos la formulación de elementos finitos para los espacios de elementos finitos de Lagrange, Raviart-Thomas y Taylor-Hood. Solucionamos la ecuación de Laplace en su formulación de primer y segundo orden, y comparamos las soluciones obtenidas con los espacios de elementos finitos de Lagrange y Raviart-Thomas al cambiar el orden de las funciones base y el nivel de refinamiento de la malla. Finalmente, resolvemos las ecuaciones de Navier-Stokes en un dominio bidimensional, donde la solución es un estado estable, y en un dominio tridimensional, donde el sistema presenta un comportamiento turbulento. Todos los experimentos numéricos se realizan utilizando la librería MFEM, la cual es también estudiada.

### 5. Clasificación de información por agrupamiento de datos: El algoritmo de k-medias

Estudiante **Víctor Andrés De La Hoz Luengas\***

Director *Jorge Mauricio Ruiz Vera\*\**

Emails \*vadela@unal.edu.co, \*\*jmruizv@unal.edu.co

RESUMEN. Una de las mayores necesidades estadísticas que surgen hoy dada la gran cantidad de información que se está generando, es la de poder clasificar la información en grupos que permitan identificar información relevante de los datos. Uno de los métodos más conocidos y tradicionales es el algoritmo de k-medias, que permite dar una solución al problema de optimización que surge de buscar la mejor agrupación de los datos. Sea  $\mathcal{K} = \{K_1, \dots, K_j\}$  un agrupamiento de  $\mathcal{X}$ , y  $\cdot$  una norma sobre  $\mathcal{X}$ . A cada  $K_j$  se le asigna un centro de grupo, llamado centroide definido por:  $c_j = \underset{x \in \mathbb{R}^q}{\operatorname{arg\,min}} \sum_{x_i \in K_j} x - x_i$ . Como se busca que la suma de las distancias de

los elementos de cada grupo a su respectivo centro sea mínima, entonces se introduce la función:  $f(\mathcal{K}) = \sum_1^j \sum_{x_i \in K_j} c_j - x_i$ . Así el problema de optimización para agrupamiento es:  $\mathcal{K} \in \mathcal{P}(\mathcal{X}) f(\mathcal{K})$ . Una vez explicado el algoritmo y demostrada su convergencia, se presenta una implementación en Python con la creación de una clase k-medias para finalmente mostrar cómo podría aplicarse en una imagen y sus ventajas en términos de almacenamiento.

## 6. Soluciones algorítmicas de sistemas semi-algebraicos con una aplicación al análisis de redes químicas pseudoquirales

Estudiante **Daniel Darío Fula Argüello\***

Director *Juan Andrés Montoya Argüello\*\**

Emails \*ddfulaa@unal.edu.co, \*\*jamontoyaa@unal.edu.co

RESUMEN. Se pretende estudiar la solución algorítmica de sistemas de ecuaciones y desigualdades polinomiales, que en la literatura se conocen como sistemas semi-algebraicos y en concreto se quiere mostrar una aplicación de eso al “análisis” de redes químicas pseudoquirales en donde se buscan condiciones para el rompimiento de simetría.

### Planteamiento del problema

#### Problema general

Estudio de las soluciones algorítmicas de sistemas de ecuaciones y desigualdades polinomiales

#### Problema específico

Desarrollo de algoritmos para calcular los valores de los parámetros de redes químicas que pueden dar lugar a rompimiento de simetría y bifurcaciones Turing

#### Estado del arte

No es necesario ahondar en la importancia del estudio de sistemas de ecuaciones y desigualdades polinomiales dado que este es uno de los problemas centrales de las matemáticas desde sus orígenes y es el problema fundacional de la geometría algebraica real. Por otra parte, nos gustaría profundizar un poco más en la importancia de las aplicaciones en química. Específicamente estamos interesados en amplificadores quirales, que son unos estados que en principio corresponden a inestabilidades que pudieron haber estado presentes en la tierra prebiótica y que pudieron haber dado lugar al rompimiento de simetría quiral que precede al origen de la vida. Dichos procesos químicos pueden ser modelados parcialmente por estructuras combinatorias que dan lugar a sistemas de ecuaciones diferenciales no lineales que normalmente no podemos resolver en forma cerrada, lo que nos obliga a realizar un análisis cualitativo o análisis de estabilidad. Tal análisis cualitativo consiste en determinar, de la mejor manera posible, cuáles son los equilibrios químicos del sistema y clasificarlos de

acuerdo con diferentes criterios de estabilidad que pueden ser importantes en el análisis. Considerando que los equilibrios químicos están definidos por ecuaciones polinomiales que dependen de la estructura combinatoria de la red y que las condiciones de estabilidad que se estudian típicamente están definidas por desigualdades polinomiales, el problema se reduce a solucionar un sistema semi-algebraico. No obstante, dicho sistema puede ser muy complejo y es necesario tener algoritmos para hallar su solución computacionalmente.

### Objetivos

#### Generales

- Estudiar los principios de la teoría de la geometría algebraica real
- Estudiar algunos algoritmos para resolver sistemas semi-algebraicos

#### Específicos

- Mostrar una breve aplicación del algoritmo de Collins al análisis de modelos de redes químicas en las que se buscan condiciones de rompimiento de simetría
- Implementar una herramienta computacional que permita hacer el análisis de modelos de redes químicas

## 7. Modelo de sensibilidad de la estructura a término de las tasas de interés utilizando la TIR en deuda corporativa

Estudiante **Angie Eloisa Llanos Culma\***

Director *Oscar Javier López Alfonso\*\**

Emails \*aellanos@unal.edu.co, \*\*ojlopeza@unal.edu.co

RESUMEN. Este proyecto busca encontrar correlación histórica entre las curvas de rendimientos de bonos tasa fija Colombiana y Americana en deuda corporativa, con el fin de establecer y/o mejorar la estrategia de inversión para 5 fondos de inversión colectiva cuya estrategia se basa en instrumentos de Renta Fija en Davivienda Corredores (Interés, Multiescala, Renta Fija Largo Plazo, Balanceado Activo y Sintético).

Inicialmente se utilizan dos categorías de clasificación para los bonos americanos, la primera es Sector Económico y la segunda está dada por la Calificación Crediticia realizada por Standard & Poor's (S&P) Global Ratings. De donde, se seleccionan los bonos del Sector Económico Financiero y Calificación Crediticia BBB+, dado que es la categoría que presenta mayor frecuencia (759 de 8430 bonos). Posteriormente, se implementa muestreo aleatorio estratificado para tomar una muestra de tamaño óptimo de bonos americanos en pos de obtener el mejor ajuste de la estructura a término de tasas de interés para la Tasa Interna de Retorno (TIR), estimada con la metodología de Nelson y Siegel.

Para el caso Colombiano, se utilizan los parámetros betas suministrados por el proveedor de precios Precia para construir curvas cero cupón, valorar los bonos emitidos a la fecha, estimar una Tasa Interna de Retorno para esta valoración y de nuevo utilizar la metodología de Nelson y Siegel para estimar la curva de rendimientos de TIR. posterior a esto se toman curvas históricas y se analiza el movimiento de la curva americana respecto a la colombiana con el fin de encontrar tal correlación. Finalmente se da una idea de cómo se puede implementar en el negocio para tomar decisiones estratégicas tanto de compra como de venta de los productos financieros que se tranzan.

## 8. Agrupamiento no paramétrico de datos genómicos

Estudiante **Gabriel Lozano\***

Director *Michael Levine\*\**, Codirector *Francisco Gómez\*\*\** En colaboración con *Nadia Atallah*

Emails \*golozanop@unal.edu.co, \*\*mlevins@purdue.edu, \*\*\*fagomezj@unal.edu.co

RESUMEN. La identificación de agrupamientos de genes coexpresados en datos del transcriptoma es una tarea difícil. Hay dos tipos principales de algoritmos para realizar este agrupamiento: basados en distancia o basados en modelos. Los algoritmos basados en distancia normalmente usan una métrica entre dos pares de puntos en los datos y los agrupan en grupos similares. Los algoritmos basados en modelos se basan en la teoría de los modelos mixtos, esta teoría ayuda a la interpretabilidad de los resultados que se obtienen. Una deficiencia de los modelos mixtos es la dificultad que encuentran para ajustar datos de encontrar una distribución multivariada apropiada de acuerdo a los datos que se manejan. En este proyecto final de grado se propone usar métodos no paramétricos como **npEM** y **npMSL** para solucionar este problema. De esta manera no se especifica la distribución de la muestra biológica y se hace la tarea de agrupamiento más sencilla. En nuestros resultados con datos reales mostramos que nuestros métodos generan más agrupamientos biológicamente significativos que los propuestos por algoritmos basados en modelos paramétricos.

### Introducción

El estudio del transcriptoma nos permite entender las funciones de una célula y de las proteínas que sintetiza. El Secuenciación del Transcriptoma Entero para Clonación al Azar (RNA-seq) es una técnica que nos permite entender los cambios en el transcriptoma de los diferentes tipos de células. En RNA-seq se contabilizan las ocurrencias de genes particulares en el transcriptoma, de esta manera se tiene que células contienen más ocurrencia de uno u otro gen. El correcto agrupamiento de los datos de RNA-seq nos permite encontrar genes que se relacionan con el mismo

proceso biológico, por tanto se pueden desarrollar drogas enfocadas en atacar genes específicos que son característicos por ser problemáticos.

Hay múltiples problemas con encontrar un agrupamiento “correcto” de genes. En general no hay un agrupamiento correcto del cual se pueda aprender y por tanto calificar los algoritmos que se usan para resolver el problema. Encontrar una métrica para el buen agrupamiento de genes es una tarea difícil en si misma, en este trabajo usamos técnicas de enriquecimiento funcional que nos permiten determinar la cantidad de clusters con significancia biológica.

Otro problema esta relacionado con los supuestos del modelo de agrupamiento tanto para los modelos basados en distancias como los basados en modelos mixtos. Para los modelos basados en distancia la elección de la función de distancia impone supuestos en la forma de los clusters, por ejemplo, en la métrica euclidiana los clusters van a ser de forma esférica. Para el caso de los modelos mixtos hay que elegir una distribución multivariada que sigan los clusters. Esto impone un comportamiento específico sobre los clusters de acuerdo a la elección hecha.

El tercer problema esta relacionado a los datos de RNA-seq, que son en general difíciles de tratar. Los datos suelen ser asimétricos, sobre-dispersos (la varianza es más grande que la media) y hay una correlación alta entre la longitud de los genes y las ocurrencias del mismo.

Los métodos actuales para tratar el problema de agrupamiento suelen usar modelos mixtos paramétricos. Modelos basados en distribuciones de Poisson [1] (Rau 2015), o en Poisson-log [2] (Silva 2019) muestran como se crean modelos específicos y poco intuitivos para la elección de distribución. En este trabajo de grado proponemos una manera alternativa de modelar usando métodos no paramétricos que nos permiten no especificar la distribución que se usa para los clusters, y permitir que la distribución que siguen sea más flexible y acomodada a los datos.

Los algoritmos que se proponen usar son inspirados el en algoritmo de expectación y maximización (EM), en su versión no paramétrica. Particularmente usamos el algoritmo No Paramétrico de Expectación y Maximización (**npEM**) y el algoritmo No Paramétrico de Maxima Similitud Suavizada (**npMSL**). En este trabajo nos encargamos de ilustrar el funcionamiento de ambos algoritmos y de mostrar su utilidad usando análisis de enriquecimiento funcional GO. Por último damos conclusiones y comentarios adicionales.

## Métodos

Los modelos mixtos fueron introducidos originalmente por Pearson en 1894 [3] para analizar diferentes subespecies de cangrejo. En su momento Pearson usó un método de momentos pero no fue hasta la publicación del algoritmo **EM** [4] (Dempster 1997) que se popularizó su uso. La idea de

los modelos mixtos esta en dada una muestra aleatoria  $Y_1, \dots, Y_n$  *i.i.d.* de dimension  $d$  se considera la función de densidad  $f(y_i)$  tal que

$$f(y_i) = \sum_{z=1}^K \pi_z f_z(y_i)$$

donde  $0 < \pi_z < 1$  son los valores de la mezcla, y cumplen

$$\sum_{z=1}^K \pi_z = 1.$$

Luego decimos que  $f(y_i)$  es un modelo mixto de  $K$ -componentes.

Para el algoritmo **EM** se busca raíces de la función de similitud  $\frac{\partial}{\partial \Psi} L(\Psi) = 0$  donde

$$L(\Psi) = \prod_{i=0}^n f(y_i; \Psi) = \prod_{i=0}^n \sum_{z=1}^K \pi_z f_z(y_i; \theta_i)$$

donde  $\Psi$  representa los parámetros del modelo mixto. Algo usual es trabajar con el logaritmo de la función de similitud para poder transformar la productoria en una sumatoria. Además se trabaja con la función de similitud completa que evidencia cual es la probabilidad de que un elemento provenga de un cluster determinado. Se puede expresar como:

$$L_c(\Psi) = \sum_{i=0}^n \sum_{z=1}^K h_{iz} \{ \log(\pi_z) + \log(f_z(y_i; \theta_i)) \}$$

Para el método **npEM** se usa una función de densidad basada en densidad de kernel que asume independencia entre coordenadas, es decir,  $f_z(y_i) = \prod_{j=1}^d f_{jz}(y_{ij})$ . Luego la función de densidad puede ser expresada como

$$f(y) = \prod_{i=1}^n \sum_{z=1}^K \prod_{j=1}^d f_{jz}(y_{ij})$$

Para las iteraciones del algoritmo, se considera un paso **E** donde se ajustan los pesos  $h_{iz}$  usando

$$h_{iz} = \frac{\pi_z \prod_{j=1}^d f_{jz}(y_{ij})}{\sum_{z=1}^K \pi_z \prod_{j=1}^d f_{jz}(y_{ij})}$$

Luego en el paso **M** se usa actualiza el resto de parametros del modelo. Para actualizar  $\pi_z$  se usa  $\pi_z = \frac{1}{n} \sum_{i=0}^n h_{iz}$  luego para actualizar las funciones se usa

$$f_z(y) = \frac{1}{h} \frac{\sum_{j=1}^d \sum_{i=1}^n p_{iz} K\left(\frac{u-y_{ij}}{h}\right)}{\sum_{j=1}^d \sum_{i=1}^n h_{ij}}$$

Aquí se evidencia que el algoritmo no maximiza pues simplemente se tiene una fórmula no paramétrica para las funciones de densidad por cada dimensión. Es por esto que no clasifica como algoritmo **EM** sino que clasifica como un algoritmo similar a **EM**.

Para el algoritmo **npMSL** se precisa primero definir un operador suavizante no lineal:

$$\mathcal{N}f(x) := \exp \int K_h(x-u) \log(f(u)) du$$

donde  $K$  es una función de densidad de kernel definida en los reales con un producto de kernel en  $d$  dimensiones  $K(u) = \prod_{k=1}^d K(u_k)$  por tanto definimos la versión re-escalada  $K_h(u) = h^{-d} \prod_{k=1}^d h^{-1}u_k$ . A partir de estas definiciones podemos redefinir el paso **E** y el paso **M**. Para el paso **E** tomamos

$$h_{iz} = \frac{\pi_z \mathcal{N}f_z(y_i)}{\sum_{a=1}^K \pi_a \mathcal{N}f_a(y_i)}$$

luego actualizamos los valores  $\pi_i$  de la misma manera que en **npEM** y las funciones de densidad las recalculamos usando

$$f_{ik}(u) = \frac{1}{nh\pi_i} \sum_{j=1}^n h_{ij} K\left(\frac{u - y_{ik}}{h}\right)$$

Este nuevo algoritmo lo distinguimos como **npMSL**.

### Escogencia de la cantidad de clusters

Los métodos no paramétricos propuestos tienen un parámetro oculto del que no se ha discutido: la cantidad de agrupamientos. Es por esta razón que normalmente se conocen como métodos semi-paramétricos. Para escoger este número se puede usar métodos paramétricos que tienen criterios de selección como AIC, BIC, ICL y CAIC. Para este trabajo se usó la librería de *mixtools* para ajustar modelos mixtos paramétricos gaussianos y escoger el número de agrupamientos adecuado.

Es importante resaltar que se necesita usar una técnica de inicialización “smallEM” que consiste en correr el algoritmo por pocas iteraciones desde muchos puntos iniciales y escoger el que tenga la mejor función de similitud.

### Análisis de datos reales

Se analizaron datos de RNA-seq usando métodos de **npEM** y **npMSL** después de hacer una transformación logarítmica a los datos. Los clusters generados se comparan con los generados por el método propuesto por Rau 2015 [1] basado en Poisson.

Los datos analizados corresponden a líneas celulares que han mostrado ser sensibles a agentes de quimioterapia, en particular C42 y LNCaP.

Estos datos se combinan con datos de líneas celulares identificadas como resistentes a agentes de quimioterapia, en particular MR49F y C42B. De los datos obtenidos se realiza un filtrado de genes que no ocurren en las líneas celulares. Y además se realiza una transformación logarítmica para controlar la asimetría de los datos. Para los genes que ocurrían solo en algunas líneas celulares y en las otras no se usó un método de relleno en el cual se tomó el mínimo valor obtenido y se recalculó las variables como una variable aleatoria entre cero y dicho mínimo. Cabe aclarar que dado que son ocurrencias siempre son no negativas.

No se aplicó la versión estándar de los algoritmos **npEM** y **npMSL**. Se aplicó la versión de bloques en la cual se tienen repeticiones de las medidas tomadas. Todas las líneas celulares tenían 3 repeticiones.

Para ilustrar nuestros resultados usamos diagramas lambda como en la Figura 1. En estos diagramas cada columna representa un cluster, y el grosor de la columna determina la proporción de genes que pertenecen a dicho cluster. La altura está dividida en fragmentos, cada fragmento representa una línea celular, hay 4 líneas celulares pero cada una tiene 3 repeticiones por tanto se muestran 12 fragmentaciones. La altura de cada fragmentación representa la proporción de ocurrencias en esa línea comparado a la suma de ocurrencias de las demás líneas para dicho cluster.

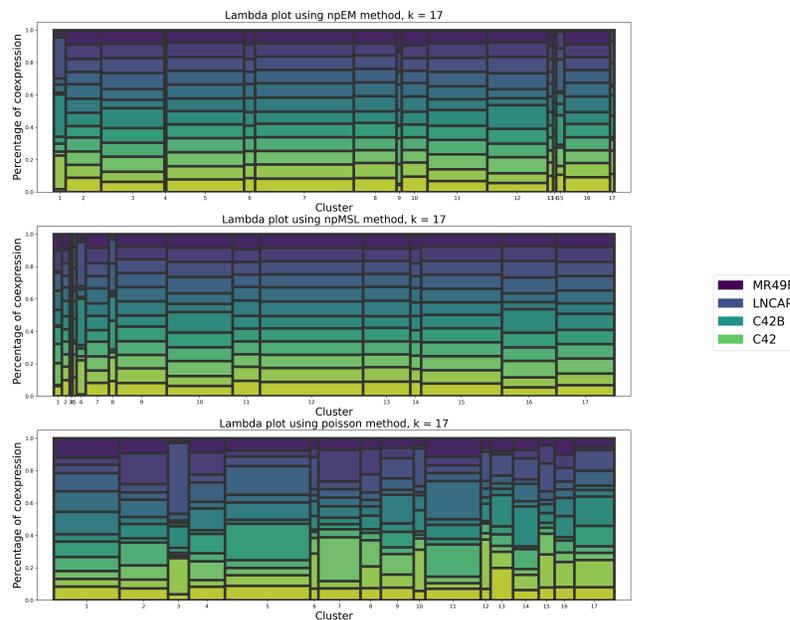


Figura 1: Diagrama lambda comparativo entre los métodos de Poisson, **npEM** y **npMSL**

Para determinar la significancia biológica de los clusters se usó ontología genética (GO) en cada uno de los clusters. De esta manera se evalúa que los clusters corresponden a codificaciones para proteínas que realizan procesos biológicos similares. Para el método de Poisson se identificaron solo 9 clusters biológicamente significativos. Para **npEM** se descubrieron 10 clusters significativos. Para el caso de **npMSL** se descubrieron 14 clusters significativos.

### Discusión y Trabajo Futuro

En este trabajo de grado se muestra que los métodos no paramétricos de **npEM** y **npMSL** son útiles para el agrupamiento de datos genéticos, en particular aquellos relacionados con tratamiento de cáncer. Estos métodos también se desempeñan mejor en datos reales que modelos paramétricos en el estado del arte. Es importante notar que pese a obtener mejores resultados los algoritmos no paramétricos son computacionalmente caros. Se deja para trabajos futuros eliminar la hipótesis de independencia en las coordenadas en la función de densidad de los clusters.

### Agradecimientos

Esta investigación ha sido financiada por Purdue University Center for Cancer Research[P30CA082709], por IU Comprehensive Cancer Center [P30ca082709], y por Walther Cancer Foundation.

Agradecimientos especiales a la convocatoria de la ORI (Oficina de Relaciones Internacionales de la Universidad Nacional de Colombia) UREP-C. A Juan Diego Velasquez que se encarga de esta convocatoria en la Universidad de Purdue.

## Referencias

- [1] Andrea Rau, Cathy Maugis-Rabusseau, Marie-Laure Martin-Magniette and Gilles Celeux, *Co-expression analysis of high-throughput transcriptome sequencing data with Poisson mixture models*, Bioinformatics, Vol. 31, No. 9, pp. 1420–1427, Oxford University Press, 2015.
- [2] Anjali Silva, Steven J. Rothstein, Paul D. McNicholas and Sanjeena Subedi, *A multivariate Poisson-log normal mixture model for clustering transcriptome sequencing data*, BMC bioinformatics, Vol. 20, No. 1, pp. 1–11, BioMed Central, 2019.
- [3] Karl Pearson, *Contributions to the mathematical theory of evolution*, Philosophical Transactions of the Royal Society of London. A, Vol. 185, pp. 71 – 110, JSTOR, 1894.
- [4] Arthur P. Dempster, Nan M. Laird and Donald B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm*, Journal of the

Royal Statistical Society: Series B (Methodological), Vol. 39, No. 1, pp. 71 – 22, Wiley Online Library, 1977.

9. **Dungeon of the Weird Mage: Modelado de parámetros de juego para la aplicación de optimización y aprendizaje reforzado en el ajuste de dificultad dinámico**

Estudiante **Juan José Rueda Mejía\***

Director *Alvaro Mauricio Montenegro Díaz\*\**

Emails \*jjruedam@unal.edu.co, \*\*ammontenegrod@unal.edu.co

RESUMEN. Al jugar un videojuego, el usuario busca algún tipo de gratificación. Existen tantas formas de lograr esto en el diseño de juegos como jugadores, ya que los jugadores experimentan los videojuegos de una manera personal. Aunque el diseño de un juego no requiere que este tenga algún tipo de **dificultad** para que el usuario pueda experimentar esa gratificación, es bastante habitual que una característica que interesa a distintos tipos de jugadores sea un reto de mayor o menor nivel según sus gustos.

Esto hace que, una de las decisiones fundamentales de la planeación de un videojuego sea el definir un tipo de dificultad (o su ausencia). El problema, incluso con un nivel de dificultad definido, es que el jugador realmente experimente el reto como fue planeado y su **percepción de dificultad** sea la correcta.

La dificultad de un videojuego radica, como en cualquier otro tipo de tareas, en los conocimientos y habilidades que se tengan e influyan en la tarea. En esta medida no es equiparable la dificultad percibida entre personas, puesto que sus habilidades y conocimientos no son comparables, de hecho no podemos asumir tampoco que sea comparable la habilidad de una misma persona en momentos distintos del tiempo.

El Ajuste de dificultad dinámico (**DDA** por sus siglas en inglés), como presenta [1], planteado dentro del contexto de los videojuegos como solución teórica a los problemas descritos asociados a la dificultad. En el DDA se plantea que la experiencia óptima se alcanza cuando hay un nivel adecuado de dificultad que **no permite aburrimiento ni frustración**. A medida que el jugador aprende y modifica su comportamiento, la dificultad debe aumentar correspondientemente para mantener el reto en el nivel adecuado, de esta manera también se alcanza un **aprendizaje de la tarea de manera óptima**.

El trabajo realizado en “Dungeon of the Weird Mage: Aprendizaje Reforzado Aplicado al Ajuste de Dificultad Dinámica” [2] es un primer acercamiento al DDA con aprendizaje reforzado, el cual se vuelve a abordar en este trabajo de grado, **generalizando las técnicas de parametrización** que se usan en la caracterización de la percepción de dificultad del

usuario en el momento de aprender una tarea e identificación de características de control de dificultad, de manera que es aplicable en una gran número de videojuegos, además se formula su respectivo procesamiento e integración en un **sistema de aprendizaje reforzado que pueda moderar la dificultad** de forma efectiva, personalizada y gratificante, como se propone en el DDA.

Estas herramientas se implementan y experimentan en el contexto del videojuego “Dungeon of the Weird Mage”, que se desarrolla para la elaboración de este proyecto, con la intención de tener un ambiente controlado y flexible que permita tanto verificar como explorar todas las posibilidades del modelo.

## Referencias

- [1] Zohaid, M. (2018), “Dynamic difficulty adjustment (dda) in computer games: A review”, *Advances In Human-Computer Interaction* 2018, 1–12.
- [2] Jarma Montoya, O. (2021), “Dungeon of the weird mage: Aprendizaje reforzado aplicado al ajuste de dificultad dinámica”, Tesis de Pregrado, Universidad Nacional de Colombia, Facultad de Ciencias, Departamento Estadística, Bogotá.

### 10. Restricted directed column convex polyominoes and some statistics

Estudiante **Fabio Alejandro Velandia Sierra\***

Director *José Luis Ramírez Ramírez\*\**

Emails \*fvelandias@unal.edu.co, \*\*jlr Ramirez@unal.edu.co

RESUMEN. En este trabajo estudiamos algunas familias de poliominós y su relación con otros objetos combinatorios. Primero, introducimos los poliominós dirigidos y convexos por columnas (dcc) para definir los  $d$ -poliominós, una restricción particular que es estudiada detalladamente mediante descomposiciones simbólicas y funciones generatrices. Con ayuda de una conocida biyección entre caminos de Dyck no decrecientes y poliominós dcc, recuperamos resultados existentes para la familia de los caminos de Dyck  $d$  restringidos en términos de estadísticas (área, número de columnas, etc.) para  $d$ -poliominós. Posteriormente, estudiamos poliominós dcc con celdas hexagonales y celdas triangulares. Obtenemos sus funciones generatrices respecto a algunas estadísticas (área, número de columnas y altura) para establecer su relación con ciertas familias de composiciones de enteros. En particular, presentamos una correspondencia biyectiva entre poliominós hexagonales dcc y las composiciones de un

entero donde la parte 1 tiene tres posibles colores. A lo largo del trabajo empleamos el software *Wolfram Mathematica*, donde implementamos esta última biyección para generar poliomínos hexagonales.

## Bibliografía

- [1] E. Barucci, R. Pinzani y R. Sprugnoli, *Directed column-convex polyominoes by recurrence relations*, Lecture Notes in Computer Science, Springer, 1993
- [2] E. Barucci, F. Bertoli, A. Del Lungo y R. Pinzani, *The average height of directed column-convex polyominoes having square, hexagonal and triangular cells*, Mathematical and Computer Modelling, Volume 26, 2009.
- [3] M. Bousquet Mélou, *A method for various classes of column convex polygons*, Discrete Mathematics, 1996.
- [4] P. Flajolet y R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, Cambridge, 2009.
- [5] R. Flórez, J. L. Ramírez, F. A. Velandia, y D. Villamizar, *A refinement of Dyck paths: A combinatorial approach*, Discrete Mathematics, Algorithms and Applications, 2021.