Universidad Carlos III de Madrid

Máster en Ingeniería Matemática

Departamento de Estadística

Trabajo presentado en cumplimiento parcial de los requisitos para obtener el grado de Máster en Ingeniería Matemática

Modalidad-A

Economía de la Atención: Maximizando el Valor para el Usuario Mediante el Índice de Whittle

Autor: German Mendoza Villacorta

Tutor: José Niño Mora

Leganés, 2016

Índice general

Agradecimientos

Quiero agradecer al Dr. José Niño Mora, por asesorarme en la elección, elaboración y desarrollo del presente trabajo y sobre todo por la enorme paciencia que me brindó, así mismo a los profesores de los departamentos de Matemática y Estadística de la Universidad Carlos III de Madrid, por las sesiones de clase impartidas durante el Máster.

Al Programa Nacional de Becas y Crédito Educativo (Pronabec) del Ministerio de Educación (Minedu), institución de mi país, por la beca que me concedió para poder realizar los estudios de Máster en Ingeniería Matemática en esta prestigiosa universidad y así poder desarrollarme no solo profesionalmente, sino también en otras áreas de mi vida. Esta experiencia definitivamente me marcó en muchos sentidos, me llevo gratos recuerdos, anécdotas y vivencias.

A mi familia y amigos de mi país, porque a pesar de la distancia que nos separaba, siempre estuvieron presentes conmigo dándome ánimos y apoyo moral, nombrarlos a todos tomaría más espacio que esta pequeña redacción, también a las amistades que pude tener en este paso por Europa.

A Jehová el todopoderoso, porque el amor nunca dejará de ser, pero el conocimiento y la ciencia acabarán. El hombre es sabio en su propio entendimiento, pero Dios pesa los espíritus.

Gracias.

DEDICATORIA

- *A Santiago Honorio mi amado padre por el apoyo y la amistad que siempre me brindo.*
- *A Marcelina mi adorada madre por su paciencia , sus consejos y su amor incondicional.*
- *A mis hermanos Lin, Andres y Jan por su comprensión, su apoyo y estar siempre a mi lado.*

Resumen

En ambientes ricos en información como es el entorno Web, la competencia por la atención de los usuarios conduce a una inundación de contenidos, por lo que la gente a menudo encuentra difícil separar las piezas más relevantes y útiles. En este contexto trataremos el concepto de *economía de la atención*, donde la atención humana es considerada como un recurso escaso y valioso.

En lugar de dejar a los usuarios enfrentarse con esta sobrecarga de información, los proveedores de contenido a menudo intentan captar su atención y beneficiarlo con lo que deciden mostrarle.

Con esto en mente planteamos el problema de optimizar el valor de la información presentada al usuario como un problema de asignación de prioridades dinámico y estocástico, el cual se puede modelar como un problema *Multi-Armed Restless Bandit* en el marco de los procesos de decisión Markovianos con una estructura particular.

Para aproximar la solución de este problema introduciremos un mecanismo automático que generará la información más relevante, la cual será presentada al usuario cuya atención a priori es limitada y se realizará a través de un algoritmo computacionalmente eficiente. Este algoritmo asigna el valor de un índice a cada elemento de la información, el cual determinará si es o no favorable presentarlo al usuario en un momento dado.

Palabras Clave: Economía de la atención.

Multi-Armed Restless Bandit.

Índice de Whittle.

Índice de Productividad Marginal.

Cota de Whittle.

Abstract

In environments rich with information such as the Web and the digital media, the competition for the attention of users leads to an overflow of content, which often leads to difficulties in picking out the relevant and useful pieces. In this context we shall discuss the concept of "economy of attention", where human attention is considered as a scarce and valuable resource.

Instead of letting the users deal with this overload of information, the providers often try to capture their attention and benefit form it with an optimal choice.

With this in mind we pose the problem of optimizing the value of the information presented to the users as a problem of dynamic and stochastic allocation of priorities, which can be modeled as a "Multi-Armed Restless Bandit"problem in the context of decision Markov processes with a particular structure.

To approximate the solution of this problem we introduce an automatic mechanism that generates the most relevant information which is presented to the user whose attention a priori is limited and is carried out by a computationally efficient algorithm. This algorithm assigns the value of an index to each item of information which determines whether or not it is favorable to present the item to the user at a given time.

Keywords: Economy of Attention.

Multi-Armed Restless Bandit.

Index Whittle.

Marginal Productivity Index.

Whittle's Bound.

Capítulo 1

Introducción

1.1. Descripción del Problema: Economía de la Atención

La popularidad de la Web y los medios digitales han dado lugar a una impresionante abundancia de información, motores de búsqueda como Google o Yahoo, billones de sitios web, publicidad dirigida y de fácil acceso a contenido digital, nos proporcionan miles de maneras de satisfacer nuestras necesidades informativas y de entretenimiento más complejos.

La economía de la atención se trata de un concepto que describe la realidad de la saturación de información a la que ha llegado la sociedad actual y que ha tomado forma con las aportaciones de diferentes autores. Según sus teorías, hoy nos vemos "atacados" en todo momento por más estímulos y datos de los que podemos asimilar. Por lo tanto, más que el contenido de la información que ofrecen los emisores de los mensajes, lo verdaderamente importante ha pasado a ser la porción de atención que estos son capaces de captar en sus destinatarios. Es decir, dado que hay abundancia de información y una gran accesibilidad a la misma, la atención humana se ha convertido en un bien tan escaso como valioso.

Esta teoría adquiere una importancia capital tanto para los profesionales de la comunicación, como para las empresas que luchan por captar la atención del público en general.

El primero que escribió sobre este tema fue Herbert Alexander Simon, economista, politólogo y teórico de ciencias sociales, premio nobel de economía en 1978 y premio Turing en 1975. Simon no utilizó el término "economía de la atención", pero en Simon (1971, p. 40- 41) señalo que:

> *"In an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it ".*

El objetivo de los gestores de información y publicidad de una empresa es ahora captar la atención del usuario, beneficiándolo con la información que deciden mostrarle, lo cual se traducirá también en un beneficio para la empresa.

Pensando en un entorno Web, en lugar de dejar al consumidor enfrentarse con esta sobrecarga de distracción, los proveedores de contenido a menudo intentan presentar primero los enlaces más destacados de su inventario teniendo en cuenta el espacio de visualización disponible en un dispositivo. Motores de búsqueda como Yahoo no exhiben todos sus resultados de búsqueda en una sola página Web, en lugar de eso los priorizan y los muestran en páginas consecutivas cuyo valor de información se supone cada

vez menor para el usuario. Google incluso ha implementado una función llamada "voy a tener suerte", el cual dado un conjunto de términos de búsqueda elige automáticamente el "mejor" resultado para el usuario.

Este enfoque adolece de dos problemas descritos en [1]Huberman y Wu (2008). El primero se encuentra con el proveedor de contenidos, quien tiene que decidir qué priorizar para obtener la atención del usuario. Esta decisión puede ser tomada en base a criterios objetivos (popularidad del sitio, número de recomendaciones para el software, prominencia de noticias,etc.) o alguna regla heurística que el proveedor de contenidos desarrolla. En cualquier caso, no es claro que tales procedimientos maximicen el valor para el usuario. Por ejemplo, mientras un algoritmo como PageRank inserta en la primera página de un resultado de búsqueda los enlaces más ligados a los términos buscados, otros enlaces en otras páginas frecuentemente contienen información valiosa incipiente que no está disponible para el usuario. Ver Huberman y Wu (2006).

El segundo problema se deriva del número de enlaces que el usuario puede atender en un intervalo de tiempo dado. Hay evidencia empírica que en una búsqueda de resultados un usuario raramente visite páginas más allá de la primera, por lo tanto es menos probable que un usuario visite las últimas páginas de un motor de búsqueda. Ver Cho y Roy (2004). Este comportamiento tiende a reforzar la posición de liderazgo de aquellos enlaces "top" y además incrementa su popularidad, que a su vez perjudica a los nuevos contenidos no muy conocidos todavía. Por lo tanto, es más fácil para un enlace mantenerse en la lista "top" y difícil para otro del fondo salir a la superficie, a pesar de que este último a menudo puede ser más valioso. Es importante romper este refuerzo de distorsión alentando a los usuarios a explorar más elementos, lo que aumentará en promedio el valor que ellos obtienen. Ver Pandey y cols. (2005).

En el trabajo de Huberman y Wu (2008) se propone una solución a los problemas mencionados, esto a través de una indexación de los estados de un enlace a través de métodos *multi-armed resstless bandit*, con la finalidad de mostrar al enlace cuando su estado tenga el mayor índice. Sin embargo, no desarrolla el caso de múltiples enlaces en competencia por ser mostrados, no hace experimentos computacionales y tampoco aproxima el objetivo de maximización.

1.2. Objetivos, Procedimiento y Estudios Aportados

Objetivos

En el presente trabajo intentamos solucionar las dificultades presentadas en la sección anterior a través de la configuración de un mecanismo que maximice el beneficio para el usuario que busca información en la Web. Extenderemos los trabajos realizados por Huberman y Wu (2008); Servia-Rodríguez, Huberman, y Asur (2014). Haremos esto formulando el problema de optimización de la información que se obtiene de la exploración de páginas Web o cualquier otro contenido digital como un problema de asignación de prioridades dinámico y estocástico para una serie de proyectos (enlaces en nuestro caso) en competencia por ser mostrados. Dicha Formulación cae en el marco del problema *multi-armed restless bandit*, el cual es un proceso de decisión Markoviano con una estructura particular.

En Papadimitriou y Tsitsiklis (1999) se observó que el problema *multi-armed restless bandit* es en general computacionalmente intratable. Para aproximar la solución de nuestro problema usaremos el enfoque introducido en Whittle (1988), el cual usa los llamados *índices de Whittle* que son asignados a

[1]El Profesor Bernardo A. Huberman dirige el Laboratorio de computación social de los Laboratorios Hewlett-Packard en Palo Alto (USA), entre otras áreas investiga la economía de la atención en la Web.

cada enlace en competencia, para luego usar la *política-índice de Whittle*: "escoger los enlaces con índice de Whittle actual más grande".

El índice de Whittle se define de manera implícita. Por ello usaremos un algoritmo computacionalmente eficiente llamado algoritmo *"adaptive-greedy"* (AG) introducido en Niño-Mora (2001, 2002) para producir el *índice de productividad marginal* (IPM), el cual bajo ciertos supuestos que se cumplen en el modelo de nuestro trabajo coincide con el índice de Whittle .

Nuestros objetivos son:

- Analizar el grado de suboptimalidad de la política-índice de Whittle en nuestro problema (descrito y desarrollado en el capítulo 4). Haremos esto comparando el rendimiento de la política-índice de Whittle con la llamada *cota de Whittle* (valor numérico próximo al objetivo original de maximización de la información para el usuario).
- Comparar el rendimiento de la política-índice de Whittle con el rendimiento de la política-índice Greedy en nuestro problema; analizar si hay mejoras en el rendimiento y ver si esta es sustancial.

Estudios Aportados

En los trabajos realizados por Huberman y Wu (2008); Servia-Rodríguez, Huberman, y Asur (2014), se abordó el caso de indexación de un proyecto (enlace) y se realizó a través de un algoritmo introducido en Niño-Mora (2001), el cual necesita supuestos (condiciones de PCL-indexabilidad) para asegurar que produce el índice de Whittle. Tales trabajos no tomaron en cuenta dichos supuestos.

Por otro lado, tampoco analizaron el caso multiproyecto. Con la finalidad de mejorar y extender los resultados de los trabajos antes mencionados, en el capítulo 4 desarrollamos lo siguiente: abordamos el caso de indexación de un proyecto, trabajamos el caso de indexación multiproyecto, calculamos la cota de Whittle y la comparamos con el rendimiento de la política-índice de Whittle, finalmente estudiamos la llamada *conjetura de Whittle*.

Procedimiento

La presente memoria se realiza en la **Modalidad-A**, en ese sentido se desarrolla de la siguiente manera:

- **Primera parte**:
  El problema que intentamos solucionar y que describimos en el capítulo 4, está formulado en el contexto de *economía de la atención*. Por ello, en la sección 1.1 describimos el concepto de economía de la atención y nos enfocamos en entornos Web, específicamente en el trabajo realizado por Huberman y Wu (2008), el cual es nuestro punto de partida y que además de mejorar extendemos en la presente memoria.

  El trabajo de Servia-Rodríguez, Huberman, y Asur (2014) usa conceptos desarrollados en Huberman y Wu (2008) y lo aplica a contenido en Twitter, por ello también puede ser mejorado y extendido.
- **Segunda parte**:
  En el capítulo 2; presentamos y explicamos el problema *multi-armed restless bandit*, el cual es una extensión del problema *multi-armed bandit* y que usaremos para modelar nuestro problema.

En el capítulo 3, calculamos la llamada *cota de Whittle* y presentamos una metodología para calcular el índice de Whittle, el cual generaremos a través del algoritmo *adaptive-greedy* (AG).

En el capítulo 4, Presentamos, modelamos y aproximamos la solución de nuestro problema, el cual básicamente busca priorizar enlaces para ser mostrados al usuario de una página Web, con el objetivo de maximizar el beneficio (en el sentido de enlaces mejor valorados) a largo plazo para dicho usuario.

- **Tercera parte**:
  En el capítulo 5; presentamos las conclusiones y una posible extensión de nuestro trabajo.

Usaremos tipo de letra cursiva para referirnos a los problemas *multi-armed bandit* y *multi-armed restless bandit*, debido a que no tienen traducción definida al castellano.

1.3. Marco de los Procesos de Decisión Markovianos (PDM)

En un problema de asignación de recursos dinámico y estocástico, el controlador puede influir con sus acciones en la evolución futura del sistema subyacente en varios puntos en el tiempo. En tal proceso de decisión secuencial existen recompensas (o costos) incurridos en el tiempo que dependen de las medidas adoptadas y la forma en que evoluciona el sistema. El objetivo del controlador puede ser maximizar la recompensa total esperada o minimizar el costo total esperado durante un determinado horizonte temporal. Si el horizonte es infinito, entonces es conveniente el uso de un descuento o promedio a largo plazo con la finalidad de tener un objetivo finito y valioso. Ver Puterman (1994).

Cuando la información necesaria para predecir la evolución futura de un sistema está contenida en el *estado* actual del sistema y depende de la acción actual, llamaremos a tal decisión secuencial un *Proceso de decisión Markoviano* (PDM). Los PDMs tienen un gran poder de modelado, el cual puede proveer resultados sobre la existencia y estructura de buenas políticas, así como también métodos para el cálculo de políticas óptimas. Por lo tanto, es natural que sean utilizados en una gran variedad de aplicaciones en áreas que incluyen ingeniería de sistemas, investigación operativa, economía y probabilidad aplicada.

En sistemas dinámicos estocásticos, típicamente no es posible tener información sobre los estados futuros al momento de tomar una acción, es por eso que las acciones no pueden basarse en tales estados. Por lo tanto, un concepto de solución útil para un PDM es de una *política no anticipativa* (o *política dependiente de la historia*), la cual se define como un conjunto de reglas que especifican las acciones que deben tomarse en cada punto de decisión en el tiempo y para cada posible estado del sistema, utilizando solamente la información actual y pasada.

Una política debe responder la siguiente pregunta: ¿Qué acción se debe tomar en un tiempo dado, si el sistema está en un determinado estado? Las *políticas estacionarias* son de gran interés, una política es estacionaria si la respuesta a la pregunta que acabamos de exponer no depende del punto en el tiempo (es decir, que es de tiempo homogéneo) y sólo depende de su estado actual. Tal política es apropiada para tomar acciones en un PDM evolucionando en el tiempo, esto debido a que los PDMs son de la naturaleza de Markov.

El avance de los PDMs fue hecha por un enfoque, que ahora se llama programación dinámica, desarrollada por Richard Bellman en la década de 1950. Ver, por ejemplo, Bellman (1957). La idea de la programación dinámica se basa en el principio de optimalidad: en cualquier punto en el tiempo una

política óptima debe prescribir una acción que optimiza la suma de la recompensa inmediata y la esperanza de la recompensa total obtenida si la política óptima se aplica desde el punto subsiguiente en el tiempo. El concepto matemático asociado al principio de optimalidad es el de las ecuaciones de optimalidad de programación dinámica, llamadas *ecuaciones de Bellman*. Para horizonte infinito las ecuaciones de Bellman se simplifican de manera que son independientes del tiempo; de hecho, el valor objetivo óptimo es único.

La programación dinámica proporciona un marco teórico coherente para el estudio de PDMs. Como tal, conduce a varios resultados teóricos generales, incluyendo la condición necesaria y suficiente para la optimalidad de una política estacionaria en algunos casos generales. Por ejemplo, implica que para los PDMs de estados y acciones finitos existe una política óptima que es estacionaria, determinista e independiente del estado inicial. Desde el punto de vista práctico, es notable que el enfoque de programación dinámica reduce la optimización sobre la secuencia de decisiones en diferentes puntos en el tiempo a una secuencia de optimizaciones de parámetros en cada punto del tiempo, por lo tanto, puede reducir significativamente la complejidad del problema.

Sin embargo, para muchos problemas esto puede no ser suficiente para hacer manejable su solución. Un dificultad típica que surge en su uso, es que las recurrencias de programación dinámica pueden ser muchas (o un número infinito) para permitir el cálculo real. El tamaño de la formulación de programación dinámica crece exponencialmente con el número de estados, lo cual es conocido como la *maldición de la dimensionalidad*. Aquí surge la necesidad de otros enfoques que proporcionen políticas implementables en la práctica y cuyo grado de suboptimalidad sea pequeño. Una de las alternativas es la formulación de *programación lineal* (LP) de las ecuaciones de Bellman. Sin embargo, la formulación LP, como tal, no ayuda a hacer frente a la maldición de la dimensionalidad. La sección 3.1 discute una relajación de Lagrange, el cual es un enfoque que se adoptara en este trabajo, ayudando a descomponer problemas complejos con estructura especial, con el fin de obtener soluciones subóptimas de buen rendimiento.

En este trabajo nos centraremos en PDMs con espacio de estados y acciones finitos, evolucionando en periodos de tiempo discreto, con horizonte infinito y criterio de optimización descontado.

1.4. Políticas Basadas en Índices de Prioridad

Debido a la maldición de la dimensionalidad, problemas asignación de recursos dinámica y estocástica, se analizan y resuelven normalmente por técnicas ad-hoc. Además, si el objetivo es hallar buenas políticas (no necesariamente óptimas) basadas en prioridades, mucha personas puede haber propuesto reglas de prioridad ad-hoc.

Quizás, la clase más natural y simple de políticas de asignación de prioridad está basada en el uso de *índices de prioridad*. Por lo tanto, si uno necesita priorizar dinámicamente múltiples proyectos a través del tiempo, un *índice* es definido para cada proyecto como función de su estado. El resultado es una *política índice*, la cual da prioridad al número de proyectos requeridos con índice actual más grande en un periodo de tiempo dado.

Un ejemplo de política índice es la *política Moipe*: el índice de un proyecto en un determinado estado es su recompensa actual. Por lo tanto, da prioridad al número de proyectos requeridos con mayor recompensa actual. Esta política sólo mira el presente sin tener en cuenta que a largo plazo esto puede no ser adecuado.

El problema que es objeto de este trabajo y que será desarrollado en el capítulo 4 es un problema de asignación de prioridades dinámico y estocástico (tenemos que priorizar enlaces para ser mostrados al usuario de una página Web), para aproximar su solución nos centraremos en el uso de la *política-índice de Whittle*, la cual será explicada en el siguiente capítulo.

Observación: Trade-off Fundamental en Procesos de Decisión Markovianos.

El trade-off fundamental en procesos de decisión Markovianos es encontrar un balance óptimo entre recompensas a corto plazo y recompensas a largo plazo, en este sentido, para un factor de descuento β ($0 < \beta < 1$) de las recompensas futuras ganadas, se tiene que:

- $\beta \approx 0$: Da más peso en las decisiones a las recompensas presentes, es decir políticas que planifican a corto plazo, entre ellas la política Greedy o Miope.
- $\beta \approx 1$: Da más peso en las decisiones a las recompensas futuras, es decir políticas óptimas que planifican a largo plazo.

Capítulo 2

Problema *Multi-Armed Restless Bandit*

En este capítulo revisaremos procesos de decisión Markovianos con una estructura particular, los cuales caen en el marco del problema *multi-armed restless bandit*, el cual modela problemas de asignación de recursos dinámica y estocástica.

El problema *multi-armed restless bandit* es una extensión del problema clásico, llamado problema *multi-armed bandit*. Por ello, primero en la sección 2.1 presentamos y explicamos el problema *multi-armed bandit*. A continuación en las secciones 2.2 abordamos el problema *multi-armed restless bandit*. Finalmente en la sección 2.4 describimos algunas características y el desarrollo histórico de las políticas índice que dan solución o aproximan la solución de los problemas mencionados.

A manera de ilustración podemos decir lo siguiente: el nombre *one-armed bandit* en inglés, es el de una maquina tragamonedas con un solo brazo que uno puede encontrar en casinos. Darle este nombre a este tipo de problemas no es casualidad, dado que la idea básica de un problema *multi-armed bandit* son máquinas tragamonedas donde uno tiene que pagar para tirar de una palanca ganando o perdiendo con una cierta probabilidad. Supongamos ahora que tenemos muchas de estas máquinas tragamonedas y una decisión que tomar sobre cuáles de ellas tirar a continuación. ¿Cuáles deben ser tiradas para maximizar la ganancia esperada de todos los tirones? Resulta que podemos construir un índice que se puede utilizar para elegir el mejor *one-armed bandit* a tirar para maximizar nuestras ganancias, este se llama el *índice de Gittins* y ha abierto los problemas *multi-armed bandit* como toda una nueva área de investigación.

2.1. Problema *Multi-Armed Bandit*

Podemos formular el problema *multi-armed bandit* como el problema que se ocupa de como asignar dinámicamente un recurso escaso entre varios proyectos estocásticos alternativos. Ver J. Gittins (1989)
El problema *Multi-Armed Bandit* se describe de la siguiente manera.

Sea una cartera de N proyectos, sea también $X_n(t) = i \in \mathbb{X}_n$, donde $X_n(t)$ es el estado del n-ésimo proyecto en el periodo de tiempo $t = 0, 1, 2, \ldots$ y $\mathbb{X}_n$ su espacio de estados finito. En cada periodo de tiempo t el controlador o manager debe decidir sobre la asignación de un recurso escaso, al proyecto que se le asigna el recurso escaso se le llama *proyecto activo*. Si un proyecto es seleccionado activo, inmediatamente genera una recompensa $R_n(X_n(t))$ y cambia al estado $X_n(t+1) = j \in \mathbb{X}_n$ de manera Markoviana con probabilidad de transición $p_n(i, j)$; los proyectos no activos permanecen congelados, es decir, no generan recompensa ni producen ningún cambio de estado.

El objetivo del manager es maximizar una función relacionada con las recompensas futuras, la cual es típicamente tomada como una esperanza, el objetivo de maximización más relevante y también más investigado es maximizar la *esperanza de la ganancia total descontada* dada por:

$$E_{i^0}^{\pi}\left[\sum_{t=0}^{\infty}\sum_{n=1}^{N} R_n\left(X_n(t)\right)\beta^t\right],$$

donde $i^0 = (X_n(0))_{n=1}^N$ es el vector de estados iniciales y $0 < \beta < 1$ un factor de descuento. La optimización se hace sobre un conjunto de *políticas admisibles* $\pi \in \Pi$, las cuales son aquellas que en cada periodo de tiempo t seleccionan proyectos como activos, basándose solamente en el vector de estados actuales $(X_n(t))_{n=1}^N$.

Observación: En el problema *multi-armed restless bandit* definido a continuación, los proyectos no activos pueden cambiar de estado o generar recompensa. De esta manera podemos ver a dicho problema como una exención del problema *multi-armed bandit*.

2.2. Problema *Multi-Armed Restless Bandit*

Sea en un proyecto X cuya evolución sobre periodos de tiempo $t = 0, 1, 2, \ldots$ es controlado por un manager, quien decide al inicio de cada periodo si el proyecto debe ser *activo* (trabajar) o *pasivo* (no trabajar) durante el periodo. Si al inicio del periodo t el proyecto ocupa el estado $X(t) = i \in \mathbb{X}$ (donde $\mathbb{X}$ denota el espacio de estados del proyecto y es finito) es activado, es decir se toma la *acción activa* $a(t) = 1$, esta produce una recompensa inmediata $R(i, 1)$ y se mueve al estado $X(t + 1) = j \in \mathbb{X}$ de una manera Makoviana con probabilidad de transición $P(i, j|1)$. Si, por otro lado, el proyecto es pasivo, es decir se toma la *acción pasiva* $a(t) = 0$, esta produce una recompensa inmediata $R(i, 0)$ y su estado se mueve a $X(t + 1) = k \in \mathbb{X}$ de una manera Markoviana con probabilidad $P(i, k|0)$.

El término *multi-armed restless bandit*, es usado como una metáfora para *cartera de proyectos*, pensado como una colección de $N \geq 2$ proyectos, cada uno con la estructura del párrafo anterior, de los cuales el manager tiene que escoger a lo sumo $M \leq N$ para activar en cada periodo de tiempo. Incorporaremos el índice del proyecto $n = 1, \ldots, N$ en lo que sigue para denotar un proyecto simple, escribiremos, por ejemplo $X_n,\ R_n(i, a),\ p_n(i, j|a),\ X_n(t)$ y $a_n(t)$, con significado obvio. En este ajuste, el manager de la cartera de proyectos observa al inicio de cada periodo t el *estado conjunto* $\mathbf{X}(t) = (X_n(t))_{n=1}^N$ y toma una *acción conjunta* $\mathbf{a}(t) = (a_n(t))_{n=1}^N$, la cual debe ser basada en la historia de estados y acciones conjuntas, además satisfacer que

$$\sum_{n=1}^{N} a_n(t) = M. \tag{2.1}$$

La elección de la acción conjunta es basada en la adopción de una *política de planificación* π, la cual se tomara de una clase resultante de *políticas de planificación admisibles* $\Pi(M)$, tal clase consiste de *políticas no anticipativas*. La transición de estados de la cartera está dada por la transición de estados de los proyectos individuales.

Un supuesto importante es que la transición de estados de un proyecto es estocásticamente independiente de la transición de estados de los otros proyectos. En cuanto a la recompensa conjunta, se supone que es aditiva a través de los proyectos.

El problema de horizonte infinito es encontrar una política de planificación admisible π^*, si lo hay, la cual indique que proyectos activar en cada periodo de tiempo (donde recompensas futuras son descontadas geométricamente con factor $0 < \beta < 1$), el cual maximiza la esperanza de la recompensa total descontada ganada. Formularemos el problema bajo el *criterio de descuento* β como:

$$\max_{\pi \in \Pi(M)} E_{i_0}^{\pi} \left[\sum_{t=0}^{\infty} \sum_{n=1}^{N} R_n \left(X_n(t), a_n(t) \right) \beta^t \right] \tag{2.2}$$

Donde $E_{i_0}^{\pi} [\cdot]$ denota la esperanza bajo la política de planificación π, condicionado a que el estado inicial de la cartera de proyectos es $\mathbf{X}(0) = i^0 = \left(i_n^0\right)_{n=1}^{N}$.

Aplicaciones:

Los problema *restless bandit* proporcionan un poderoso marco de modelamiento, como ilustramos en los siguientes ejemplos de aplicaciones.

Ensayos Clínicos (Whittle 1988). En este caso, los proyectos corresponden a los tratamientos médicos, el estado de un proyecto representa un estado de conocimiento sobre la eficacia del tratamiento correspondiente, operar el proyecto corresponde a probar el tratamiento. Si, por ejemplo, el virus que los tratamientos están intentando combatir está mutando, el conocimiento de la efectividad de cada tratamiento cambia si se prueba o no.

Vigilancia de una Aeronave (Whittle 1988). Supongamos m aeronaves intentando rastrear la posición de n submarinos enemigos, donde $m < n$. Por lo tanto, si cada submarino es monitoreado solamente por una aeronave, entonces cada aeronave tiene que cambiar de tarea de tiempo en tiempo. Consideraremos este como un caso de la operación de exactamente m proyectos de n, en que exactamente m submarinos de n están bajo vigilancia en un tiempo dado. El problema es asignar esta vigilancia. Este problema es *multi-armed restless bandit* en el sentido más literal. Mientras un submarino está bajo observación, ganamos información sobre su posición, etc. Pero sino, usualmente perdemos información, porque el submarino ciertamente tomara acción evasiva impredecible.

Economía de la atención (Huberman & Wu 2008). Este modelo es objeto de desarrollo del presente trabajo y lo abordaremos con más detalle en el capítulo 4.

Para más ejemplos ver Niño-Mora (2007).

2.3. Política de Indexación

Política Índice que Soluciona el Problema *Multi-Armed Bandit*

El problema *multi-armed bandit* originalmente descritos por Robbins (1952), es un modelo de un controlador optimizando sus decisiones y adquiriendo conocimiento al mismo tiempo. A pesar de ser un problema de optimización dinámica estocástica de planteamiento simple, su solución había sido un problema abierto por un tiempo considerablemente largo. Es debido a Gittins y Jones que el problema *Multi-Armed Bandit* ha sido solucionado. Ver J. Gittins y Jones (1974). La esencia de la solución fue la definición del *índice de Gittins* como una función del proyecto y su estado, interpretados como el precio justo para la activación del proyecto en Whittle (2002). El índice de Gittins en el estado x del proyecto k, se puede expresar como:

$$\nu_k(x) = \max_{\tau>0} \frac{E\left\{\sum_{t=0}^{\tau-1} \beta^t r_k(t)\left(x_k(t)\right)|x_k(0)=x\right\}}{E\left\{\sum_{t=0}^{\tau-1} \beta^t |x_k(0)=x\right\}},$$

donde la maximización es sobre todos los tiempos de parada $\tau \in T \cup \{\infty\}$ tal que, $\tau > 0$. En otras palabras el índice de Gittins define un valor asociado a cada estado de un proyecto como la recompensa esperada media alcanzable máxima (i.e, la recompensa esperada por unidad de tiempo descontado esperado), teniendo en cuenta que empezamos a trabajar en este proyecto desde el primer periodo. Note que estas cantidades son independientes del tiempo y dependen únicamente de la información relativa al proyecto k.

Gittins mostró como calcular su índice en los proyectos (usando soluciones de algunos problemas de parada óptima). Además, el significado del índice de Gittins es que puede ser calculado para cada proyecto por separado, es decir, es independiente de los otros proyectos.

Una característica interesante de los problemas *multi-armed bandit* es la optimalidad de la *política-índice de Gittins* (también llamada *regla de índice*) obtenido por Gittins y presentado en J. Gittins y Jones (1974) y J. C. Gittins (1979). Dado que los proyectos están compitiendo por un recurso escaso (para ser activados), hacemos corresponder a cada uno de ellos un índice de asignación dinámico (el índice de Gittins) y luego aplicamos la política-índice de Gittins, definida como sigue: "Asignar el recurso escaso al proyecto con índice de Gittins actual más alto". La solución del problema *multi-armed bandit* como política es dada por la *política-índice de Gittins*, para más detalles ver J. C. Gittins (1979).

Gittins dio una prueba de la optimalidad de la política-índice de Gittins difícil de seguir. Como se observó en Varaiya y cols. (1985), la optimalidad de la política-índice de Gittins se implicó de dos características del problema *multi-armed bandit*: que los proyectos no activos no cambian de estado y que no contribuyen ninguna recompensa. Además, la dinámica Markoviana solamente es útil en que permite un cálculo simple de los índices de Gittins.

Whittle (1988) presentó una prueba por medio de la solución explicita de la formulación de programación dinámica del problema, en Weber (1992) se introdujo una prueba breve y casi verbal, lo que proporciono una mejor comprensión del problema, como Whittle (2002) escribe, merece ser caracterizado por el término "prueba de Dios".

Una prueba más fue dada en Bertsimas y Niño-Mora (1996), donde ellos usan una formulación de programación lineal del problema y muestran por el teorema de dualidad que la política-índice de Gittins es óptima, este enfoque resulto ser muy útil para analizar y solucionar problemas más generales, tales

como los problemas *multi-armed restless bandit* en los cuales los proyectos no activos pueden cambiar de estado o generar recompensa.

Un ejemplo clásico de optimalidad de una política índice es la $c\mu$*-rule* para el problema de secuenciación de empleo. Ver Stidham Jr (2002). En este problema, los empleos con costo de mantenimiento lineal c_k y tiempo medio de servicio μ_k^{-1} deben ser programados para atender en un solo servidor de tal manera que la esperanza total del costo de mantenimiento se minimice. La $c\mu$*-rule* prescribe programar los empleos como sigue: " Asignar el servidor a un trabajo no realizado de mayor índice $c_k\mu_k$". Note, sin embargo, que tal asignación de índices es *estático*, en oposición al índice de Gittins, el cual es dinámico (dependiendo del estado actual del proyecto).

Política Índice para Aproximar la Solución del Problema *Multi-Armed Restless Bandit*

Una cuestión central en la literatura concerniente a extensiones del problema *multi-armed bandit* es identificar índices calculables de forma eficiente que hacen que una política índice tenga un buen rendimiento. A no ser que tengamos la suerte de encontrar una clase particular de proyectos con la estructura especial del problema *multi-armed bandit*. En general solo podemos esperar una forma de optimalidad asintótica de una política índice para el problema *multi-armed restless bandit*, como fue mostrado en Weber y Weiss (1990).

Encontrar la solución óptima del problema *multi-armed restless bandit* es en general computacionalmente intratable. Ver Papadimitriou y Tsitsiklis (1999). Nosotros usaremos el enfoque introducido por Whittle (1988), quien fue el primero en proponer una solución aproximada de este problema a través de la solución de una relajación por métodos lagrangianos. La relajación de Whittle fue remplazar la restricción de operar un número fijo de proyectos en cada instante de tiempo (un número infinito de restricciones) por una sola restricción (activar el número de proyectos requeridos *en promedio*), entonces usando un multiplicador de Lagrange, tal restricción puede ser dualizado he incluido en el objetivo. Esto permite simplificar la solución del problema *multi-armed restless bandit* a solucionar problemas *one-armed bandit* simples, desarrollaremos estas ideas en la sección 3.1.

Whittle (1988) además propone un índice, conocido desde entonces como el *índice de Whittle* (que en el caso de bandits clásicos coincide con el índice de Gittins), el cual es definido de forma implícita para cada proyecto como función de su estado y es usado por la *política-índice de Whittle*: "asignar el recurso escaso a los proyectos requeridos con mayor índice de Whittle actual". Tal política es típicamente subóptima, por ello nuestro interés en aplicarla a nuestro modelo de economía de la atención.

Whittle también observo que la existencia de sus índices no es garantizada para todos los *restless bandit*: solamente para aquellos que satisfacen la llamada propiedad de *indexabilidad*. El afirmó en Whittle (1988):

> . . . one would very much like to have simple sufficient conditions for indexability; at the moment, none are known.

Tales situaciones observadas por Whittle llevaron a Niño Mora a abordar estos y otros temas sobre indexabilidad de proyectos *restless bandit*. Niño-Mora (2001, 2002, 2006b) introduce métodos para determinar a priori si un modelo *restless bandit* es indexable y dio un algoritmo llamado *adaptive-greedy* (AG) basado en Klimov (1975) para calcular los llamados *índices de productividad marginal* (IPM).

El IPM mide la productividad marginal del trabajo en cada estado del proyecto y se basa en objetivos

alcanzados por la *teoría de productividad marginal* en economía, desarrollado por varios autores a finales de siglo diecinueve. Véase por ejemplo, el trabajo clásico de Clark (1902).

El IPM además de ser bastante manejable y aplicable, un creciente cuerpo de evidencias computacionales indica que su política índice suele exhibir un rendimiento cercano al óptimo y supera a políticas de enfoques alternativos.

Capítulo 3

Asignación Dinámica de Prioridades Mediante el Índice de Whittle

La primera sección de este capítulo la dedicamos a aproximar el objetivo de optimización del problema *multi-armed restless bandit*, haremos esto a través de la solución de una relajación por métodos lagrangianos. El valor del objetivo de optimización de dicha relajación es conocido como la *cota de Whittle* y es usada para compararla con el rendimiento de políticas que aproximan el objetivo de optimización del problema *multi-armed restless bandit* original. En la sección 4.6, comparamos el rendimiento de la política-índice de Whittle, la política Greedy y la cota de Whittle para nuestro problema específico en el contexto de economía de la atención.

Por otro lado, Estamos interesados en aplicar la política-índice de Whittle en el modelo de economía de la atención que es objeto de este trabajo, en este sentido la segunda sección de este capítulo definimos el concepto de *indexabilidad* e *índice de Whittle*, dicho índice es definido de manera implícita y nuestro interés es calcularlo de una manera eficiente, haremos esto siguiendo las ideas de Niño-Mora (2001, 2002, 2006a, 2006b), donde se introduce y desarrolla el concepto de *índice de productividad marginal* (IPM) como aquel producido por el algoritmo *adaptive-greedy* (AG) y que bajo ciertos supuestos (condiciones de PCL-indexabilidad) coincide con el índice de Whittle.

En el modelo de economía de la atención, estudiado en este trabajo, se cumplen los supuestos necesarios para que el IPM coincida con el índice de Whittle, por ello usaremos el algoritmo AG para producir el índice de Whittle.

3.1. Relajación de Whittle, Relajacíon Lagrangiana y Descomposición

Siguiendo las ideas de Whittle (1988), podemos encontrar una cota superior $W(i^0)$ del objetivo de optimización del problema *multi-armed restless bandit* solucionando el siguiente problema (que es [1]equivalente a una relajación. Ver Bertsimas y Niño-Mora (2000)), donde asumimos que cada proyecto es controlado por separado, como si estuviera solo, bajo su propia política $\pi_n \in \Pi_n$, $i^0 = \left\{i_n^0\right\}_{n=1}^N$ es el vector de estados iniciales de la cartera de proyectos y relajando la restricción (2.1), la cual opera exactamente M proyectos en cada periodo de tiempo a operar M proyectos en promedio en cada periodo de tiempo:

[1]Un supuesto importante para esta equivalencia, es que la transición de estados de un proyecto es estocásticamente independiente de la transición de estados de los otros proyectos.

$$W(i^0) = \max \sum_{n=0}^{N} E_{i_n^0}^{\pi_n} \left[\sum_{t=0}^{\infty} R_n(X_n(t), a_n(t)) \beta^t \right] \tag{3.1}$$
$$\text{sujeto a: } \pi_n \in \Pi_n \quad n = 1, 2, \ldots, N$$
$$\sum_{n=1}^{N} E_{i_n^0}^{\pi_n} \left[\sum_{t=0}^{\infty} a_n(t) \beta^t \right] = \frac{M}{1-\beta}.$$

Note que $X_n(t)$ y $a_n(t)$ denotan el estado y la acción procesada por el n-ésimo proyecto bajo la política π_n empezando del estado i_n^0, además Π_n es la clase admisible de políticas sólo para el n-ésimo proyecto.

Relajación Lagrangiana y Descomposición

Para solucionar el problema (3.1) usamos un enfoque lagrangiano. En ese sentido, asignando un multiplicador $\lambda \in \mathbb{R}$ a la restricción de igualdad se produce la relajación lagrangiana

$$L(\lambda, i^0) = \frac{M\lambda}{1-\beta} + \max \sum_{n=0}^{N} E_{i_n^0}^{\pi_n} \left[\sum_{t=0}^{\infty} \left\{ R_n(X_n(t), a_n(t)) - \lambda a_n(t) \right\} \beta^t \right]$$
$$\text{sujeto a: } \pi_n \in \Pi_n \quad n = 1, 2, \ldots, N$$

entonces

$$L(\lambda, i^0) = \frac{M\lambda}{1-\beta} + \sum_{n=0}^{N} \max_{\pi_n \in \Pi_n} E_{i_n^0}^{\pi_n} \left[\sum_{t=0}^{\infty} \left\{ R_n(X_n(t), a_n(t)) - \lambda a_n(t) \right\} \beta^t \right].$$

Definiendo $L_n(\lambda, i_n^0)$ para $n = 1, \ldots, N$, como:

$$L_n(\lambda, i_n^0) = \max_{\pi_n \in \Pi_n} E_{i_n^0}^{\pi_n} \left[\sum_{t=0}^{\infty} \left\{ R_n(X_n(t), a_n(t)) - \lambda a_n(t) \right\} \beta^t \right], \tag{3.2}$$

tenemos que

$$L(\lambda, i^0) = \frac{\lambda M}{1-\beta} + \sum_{n=1}^{N} L_n(\lambda, i_n^0). \tag{3.3}$$

Para cada número natural n, fijo pero cualquiera entre 1 y N, $L_n(\lambda, i_n^0)$ es llamado [2]problema *one-armed restless bandit* y sus ecuaciones de Bellman asociadas son:

$$L_n(\lambda, i) = \max_{a_n \in \{0,1\}} \left\{ \{R_n(i, a_n) - \lambda a_n\} + \beta \sum_{j \in \mathbb{X}_n} P_n(i, j | a_n) L_n(\lambda, j) \right\} \quad , i \in \mathbb{X}_n \tag{3.4}$$

[2]Whittle definió su índice, conocido como el *índice de Whittle*, en base a la solución del problema *one-armed restless bandit*. Sin embargo, nosotros definiremos el índice de Whittle en la siguiente sección siguiendo las ideas de Niño-Mora (2001, 2007).

donde $\mathbb{X}_n$ es el espacio de estados del n-ésimo proyecto. Podemos solucionar las ecuaciones de Bellman (3.4) a través de su formulación de programación lineal

$$\begin{cases} \text{mín} \sum_{j\in\mathbb{X}_n} P_0(j)L_n(\lambda,j) \\ \text{sujeto a:} \\ x(i,a_n):\ L_n(\lambda,i) - \beta \sum_{j\in\mathbb{X}_n} P_n(i,j|a_n)L_n(\lambda,j) \geq R_n(i,a_n) - \lambda a_n\ , i \in \mathbb{X}_n \ \text{ y } \ a_n \in \{0,1\}\,, \end{cases} \tag{3.5}$$

donde, para todo $j \in \mathbb{X}_n$ el estado inicial es tomado de acuerdo a las siguientes probabilidades $P_0(j) = P\left\{i_n^0 = j\right\}$.

La solución del problema (3.5) está garantizada por el siguiente teorema de la teoría de programación dinámica.

Teorema 3.1 *Si para todo $j \in \mathbb{X}_n$ tenemos que $P_0(j) = P\{X_n(0) = j\} > 0$, entonces el problema (3.5) tiene solución única $\{L_n(\lambda,j) : j \in \mathbb{X}_n\}$.*

De esta manera obtenemos el valor de $L_n(\lambda, i_n^0)$, además como n era fijo pero cualquiera entre 1 y N, también obtenemos el valor de $L(\lambda, i^0)$ dado en la ecuación (3.3).

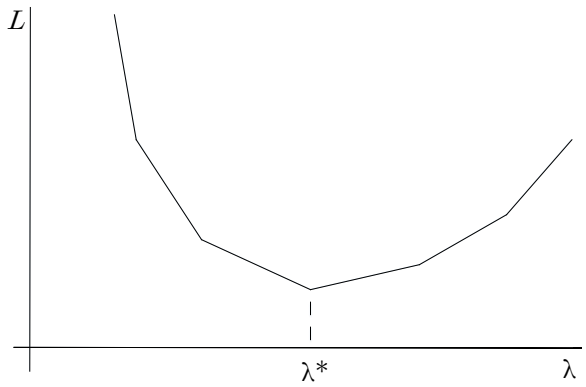

Figura 3.1: Función $L(\lambda, i^0)$

Cota de Whittle:

De la ecuación (3.2) tenemos que:

$$\begin{aligned} L_n(\lambda, i_n^0) &= \max_{\pi_n\in\Pi_n} E_{i_n^0}^{\pi_n}\left[\sum_{t=0}^{\infty}\left\{R_n(X_n(t),a_n(t)) - \lambda a_n(t)\right\}\beta^t\right] \\ &= \max_{\pi_n\in\Pi_n} E_{i_n^0}^{\pi_n}\left[\sum_{t=0}^{\infty} R_n(X_n(t),a_n(t))\beta^t\right] - \lambda E_{i_n^0}^{\pi_n}\left[\sum_{t=0}^{\infty} a_n(t)\beta^t\right] \\ &= \max_{\pi_n\in\Pi_n} F_{i_n^0}^{\pi_n} - \lambda G_{i_n^0}^{\pi_n}, \end{aligned}$$

donde los números $F_{i_n^0}^{\pi_n}$ y $G_{i_n^0}^{\pi_n}$ son las *medidas de recompensa y trabajo* respectivamente, que definiremos en la sección 3.2. De la última igualdad y la ecuación (3.3) tenemos que la función $L(\lambda, i^0)$ es una función convexa lineal a trozos que alcanza su mínimo en algún valor de $\lambda \in \mathbb{R}$ y cuya gráfica es como la de la Figura 3.1.

Por otro lado, como $L(\lambda, i^0)$ es una relajación lagragiana de $W(i^0)$, entonces existe un $\lambda^* \in \mathbb{R}$ tal que

$$L(\lambda, i^0) \geq W(i^0) = L(\lambda^*, i^0) = \min_{\lambda \in \mathbb{R}} L(\lambda, i^0).$$

El valor de $W(i^0)$ es llamado *Cota de Whittle* y es una aproximación el objetivo de optimización del problema original (2.2).

3.2. Indexación *Restless Bandit*: Teoría y Cálculo

En esta sección tomamos los conceptos introducidos en Niño-Mora (2001), revisados y explicados en Niño-Mora (2007), para presentar una metodología de *indexación* del problema *multi-armed restless bandit*, nos referiremos a *proyecto* como a uno tomado de la cartera de proyectos que definen al problema *multi-armed restless bandit*.

3.2.1. Indexabilidad y el Índice de Whittle

Dado un proyecto, evaluaremos una política π por medio de dos medidas. La primera es la *medida de recompensa*

$$F_{i^0}^{\pi} := E_{i^0}^{\pi}\left[\sum_{t=0}^{\infty} R\left(X(t), a(t)\right)\beta^t\right],$$

dando la esperanza del valor total descontado de recompensa ganada sobre un horizonte infinito, empezando en el estado i_0. La segunda medida se refiere a los gastos del recurso asociado. Así, si se gastan $Q(j, a)$ unidades de trabajo al tomar la acción a en el estado j, usaremos la *medida de trabajo*

$$G_{i^0}^{\pi} := E_{i^0}^{\pi}\left[\sum_{t=0}^{\infty} Q\left(X(t), a(t)\right)\beta^t\right],$$

dando la correspondiente esperanza del valor total descontado del trabajo gastado sobre un horizonte infinito, empezando en el estado i_0. Asumiremos que el parámetro de trabajo gastado satisface que $Q(i, a) = a$.

Nos referiremos a las correspondientes medidas f^{π} y g^{π} como *medidas de recompensa y trabajo promedio* respectivamente, obtenidas por tomar aleatoriamente el estado inicial de acuerdo a una función de masa de probabilidad arbitraria positiva $p_i > 0$ con $i \in \mathbb{X}$, es decir $f^{\pi} := \sum_{i \in \mathbb{X}} p_i F_i^{\pi}$ y $g^{\pi} := \sum_{i \in \mathbb{X}} p_i G_i^{\pi}$.

Supongamos que el trabajo tiene un *coste de activación* ν y consideremos el problema *ν-coste de activación*

$$\max_{\pi \in \Pi} f^{\pi} - \nu g^{\pi}, \tag{3.6}$$

el cual consiste en encontrar una política admisible de operación del proyecto que maximiza el valor de la recompensa ganada menos el coste de activación de trabajo incurrido. Usaremos (3.6) como un *problema de calibración* destinado a medir el valor de ganancia marginal producido por el trabajo en cada estado del proyecto.

Dado que (3.6) es un proceso de decisión Markoviano descontado de acciones y estados finitos (estamos considerando que el conjunto de estados $\mathbb{X}$ del proyecto es finito), resultados estándar como los dados en Puterman (1994) aseguran la existencia de una política óptima que es: (i) estacionaria determinista (solo depende de los estados de manera determinista) y (ii) independiente del estado inicial. Es conveniente representar aquellas políticas por su *conjunto activo* $S \in \mathbb{X}$, el cual es el conjunto de estados donde se prescribe participar del proyecto (se toma la acción $a = 1$), nos referiremos a tal política como *política activa* S y escribiremos, por ejemplo, f^S y g^S. Podemos reducir el problema (3.6) a un *problema de optimización combinatoria* de encontrar un conjunto activo óptimo en la familia de todos los subconjuntos de $\mathbb{X}$, denotado por $2^{\mathbb{X}}$:

$$\max_{S \in 2^{\mathbb{X}}} f^S - \nu g^S.$$

Para todo coste de activación $\nu \in R$, la política óptima es caracterizada por la única solución de las ecuaciones de Bellman

$$\vartheta_i^*(\nu) = \max_{a \in \{0,1\}} R(i,a) - Q(i,a)\nu + \beta \sum_{j \in \mathbb{X}} p(i,j|a)\vartheta_j^*(\nu), \quad i \in \mathbb{X}, \tag{3.7}$$

donde $\vartheta_i^*(\nu)$ denota el valor óptimo de (3.6) empezando del estado i. Por lo tanto, existe un *conjunto activo óptimo* $S^*(\nu) \subseteq \mathbb{X}$ para (3.6), el cual es caracterizado en términos de (3.7) por

$$S^*(\nu) := \left\{ i \in \mathbb{X}^{\{0,1\}} : R(i,1) - \nu + \beta \sum_{j \in \mathbb{X}} p(i,j|1)\vartheta_j^*(\nu) \geq R(i,0) + \beta \sum_{j \in \mathbb{X}} p(i,j|0)\vartheta_j^*(\nu) \right\}.$$

Ahora, parece razonable que al menos en algunos modelos el conjunto activo $S^*(\nu)$ se expanda monótonamente de $\emptyset$ a todo el espacio de estados $\mathbb{X}$ cuando el coste de activación ν decrece de $+\infty$ a $-\infty$. Si tal es el caso, a cada estado i le asignamos un coste de activación crítico ν_i^* debajo del cual i entra en $S^*(\nu)$.

Definición 3.2 (Indexabilidad, Índice de Whittle) *Diremos que un proyecto es indexable, si para todo estado $i \in \mathbb{X}$ existe un único $\nu_i^* \in \mathbb{R}$, tal que para todo coste de activación $\nu \in \mathbb{R}$*

$$\vartheta_i^*(\nu) = R(i,1) - \nu + \beta \sum_{j \in \mathbb{X}} p(i,j|1)\vartheta_j^*(\nu) \Leftrightarrow \nu_i^* \geq \nu,$$

$$\vartheta_i^*(\nu) = R(i,0) + \beta \sum_{j \in \mathbb{X}} p(i,j|0)\vartheta_j^*(\nu) \Leftrightarrow \nu_i^* \leq \nu.$$

En tal caso diremos que

- $\nu^* = \{\nu_i^*\}_{i \in \mathbb{X}}$ *es el índice de Whittle del proyecto.*
- $\nu_i^* \in \mathbb{R}$ *es el índice de Whittle del estado i.*

En palabras, un proyecto es *indexable*, si para todo coste de activación:

- La acción activa es óptima en el estado i, si y solo si, ν_i^* es mayor o igual que el coste de activación.
- La acción pasiva es óptima en el estado i, si y solo si, ν_i^* es menor o igual que el coste de activación.
- ambas acciones son óptimas, si y solo si, ν_i^* es igual al coste de activación.

El concepto de indexabilidad fue introducido por Whittle (1988) bajo el criterio promedio a largo plazo en el caso $Q(i,a) = a$, en una formulación dada en términos de conjuntos pasivos optimales. Él también probó que hay proyectos que no son indexables. En Niño-Mora (2002) se introdujo la extensión de considerar funciones de consumo $Q(i,a)$ generales.

3.2.2. Algoritmo *Adaptive-Greedy* (AG) y Condiciones de PCL-indexabilidad

Los investigadores generalmente están interesados en establecer *analíticamente* si un modelo particular que surge de alguna aplicación es indexable bajo una serie de parámetros adecuados. Será útil tener condiciones suficientes y generales de indexabilidad que sean ampliamente aplicables. Niño-Mora (2001, 2002, 2006a) introduce, desarrolla y despliega por primera vez tales condiciones, junto con un algoritmo de índices llamado *adaptive-greedy* (AG) el cual producirá el llamado *índice de productividad marginal* (IPM). Revisaremos estas ideas a continuación.

Algoritmo *Adaptive-Greedy* (AG) y el Índice de Productividad Marginal (IPM)

Para un proyecto, dada una acción $a \in \{0,1\}$ y un conjunto activo $S \in \mathbb{X}$, denotaremos por $\langle a, S\rangle$ a la política que toma la acción a en el periodo inicial y luego adopta la política activa S. Además de las medidas de recompensa y trabajo presentadas antes, ahora vamos a definir la *medida de trabajo marginal*

$$w_i^S := G_i^{\langle 1,S\rangle} - G_i^{\langle 0,S\rangle} = 1 + \beta \sum_{j\in\mathbb{X}} \left(p(i,j|1) - p(i,j|0)\right) G_j^S,$$

y la *medida de recompensa marginal*

$$r_i^S := F_i^{\langle 1,S\rangle} - F_i^{\langle 0,S\rangle} = R(i,1) - R(i,0) + \beta \sum_{j\in\mathbb{X}} \left(p(i,j|1) - p(i,j|0)\right) F_j^S.$$

Nótese que w_i^S (respectivamente r_i^S) mide el incremento marginal en el trabajo empleado (respectivamente en el valor de la recompensa ganada) que resulta de trabajar en lugar de descansar en el periodo inicial a partir del estado i, a condición de que la política activa S se adopte después.

Además, si $w_i^S \neq 0$, se define la *medida de productividad marginal*

$$\nu_i^S := \frac{r_i^S}{w_i^S}.$$

También definimos el *índice de productividad marginal* (IMP) como aquel producido por el algoritmo *adaptive-greedy* (AG) mostrado en la tabla 3.1. En esencia, este algoritmo genera una familia de conjuntos anidados

$$\{S_0, S_1, \ldots, S_n\} \text{ , con } S_0 := \emptyset, S_n := \mathbb{X} \text{ y } S_k := \{i_1, \ldots, i_k\} \text{ , para } 1 \leq k \leq n,$$

donde, a cada conjunto S_k se le añade el estado $i_k \in \mathbb{X} \setminus S_{k-1}$ si este tiene la mayor medida de productividad marginal. La salida consiste en una cadena ordenada $i_1, \ldots, i_n$ de todos los estados del proyecto y con sus respectivos valores $v^*_{i_k}$, los cuales forman el *índice de productividad marginal* (IPM) $\left\{v^*_{i_k}\right\}_{k=1}^n$ del proyecto. El algoritmo está bien definido cuando las medidas de productividad marginal tienen denominador distinto de cero.

ALGORITMO AG

Output: $\left\{i_k, v^*_{i_k}\right\}_{k=1}^n$
$S_0 := \emptyset$
for $k := 1$ **to** n **do**
  pick $i_k \in \arg\max \left\{\nu_i^{S_k-1} : i \in \mathbb{X} \setminus S_{k-1}\right\}$
  $v^*_{i_k} := \nu_{i_k}^{S_{k-1}}$; $S_k := S_{k-1} \cup \{i_k\}$
end { for }

Tabla 3.1: Algoritmo AG.

A continuación usaremos el algoritmo AG para definir una cierta clase de proyectos. Note que el acrónimo "PCL" se refiere a *partial conservation laws* introducido en Niño-Mora (2001).

PCL-indexabilidad

Definición 3.3 (PCL-indexabilidad) *Diremos que un proyecto es PCL-indexable si satisface las siguientes condiciones:*

a) Trabajo marginal positivo: $w_i^S > 0$ *para* $i \in \mathbb{X}$, $S \in 2^{\mathbb{X}}$.

b) Cálculo de componentes del IPM monótono decreciente: Las componentes del IPM producidos por el algoritmo AG satisfacen

$$v^*_{i_1} \geq v^*_{i_2} \geq, \ldots, \geq v^*_{i_n}.$$

Note que la parte *a)* de la definición 3.3, asegura que el algoritmo está bien definido. El interés en la clase de proyectos con PCL-*indexabilidad* se basa en el siguiente resultado, demostrado en Niño-Mora (2001, Cor.2; 2002, Th. 6.3; 2006a, Th. 4.1) en entornos cada vez más generales.

Teorema 3.4 *Si un proyecto es PCL-indexable, entonces es indexable y su índice de Whittle es el IPM producido por el algoritmo AG.*

Para ejemplos de análisis detallados de modelos específicos donde se cumplen las condiciones de PCL-indexabilidad dadas en la definición 3.3 podemos referirnos a Niño-Mora (2002, 2006a, 2006b).

Proyecto de Doble Velocidad

Un proyecto es de doble velocidad si las probabilidades de transición de estados bajo las acciones activa y pasiva cumplen la siguiente relación.

$$P(i,j|0) = \begin{cases} \epsilon_i P(i,j|1) & i \neq j, \\ (1-\epsilon_i) + \epsilon_i P(i,i|1) & i = j. \end{cases}$$

donde $\epsilon_i \in (0,1)$.

Observación 3.5 *Se prueba en Glazebrook, Niño-Mora, y Ansell (2002), que un proyecto de doble velocidad es PCL-indexable. Por lo tanto, según teorema 3.4 es indexable y su índice de Whittle es el IPM que produce el algoritmo AG.*

Capítulo 4

Aplicación: Maximizando el Valor para para el Usuario en Ambientes Ricos en Información

El presente capítulo se desarrolla de la siguiente manera:
En la sección 4.1, formulamos nuestro problema de aplicación el cual está en el contexto de economía de la atención y que será modelado en el marco del problema *multi-armed restless bandit* de doble velocidad.
En las secciones 4.2-4.3, tratamos el caso de un solo enlace, según la observación 3.5, el enlace es indexable y podemos hallar el índice de Whittle usando el algoritmo AG. Por otro lado, haremos un análisis de sensibilidad para la prioridad de los estados al hacer variar el factor de descuento β.
En las secciones 4.4-4.5, tratamos el caso multiproyecto homogéneo y heterogéneo, comparamos los rendimientos de la política-índice de Whittle y la política-índice Greedy, para ello aproximaremos nuestro objetivo de maximización a través de una simulación de Montecarlo.
En la sección 4.6, calculamos la cota de Whittle y la comparamos con el rendimiento de la política-índice de Whittle.
En la sección 4.7, estudiamos la optimalidad asintótica del rendimiento de la política-índice de Whittle a través de la llamada *conjetura de Whittle*.

4.1. Formulación del Problema

Consideremos un sistema (por ejemplo el buscador Google o una página web) que quiere mostrar "N" enlaces (los enlaces serán los proyectos de nuestro problema *multi-armed restless bandit*) a un usuario, pero solamente puede mostrar M con $(M < N)$ en un periodo de tiempo $t = 0, 1, 2, \ldots$ dado, llamaremos a estos M enlaces la "lista top".

Parámetros para un enlace

Acciones: Las acciones al inicio de cada periodo de tiempo $t = 0, 1, 2, \ldots$ serán $a(t) = 1$ (acción activa) poner el enlace en la lista top y $a(t) = 0$ (acción pasiva) no poner el enlace en la lista top.

Estados: Si el sistema puede rastrear un cierto número de propiedades para cada enlace, tal como su reputación, historia o edad, etc. Diremos que el enlace está en un *estado* definido por aquellas propiedades. Al inicio de cada periodo de tiempo $t = 0, 1, 2, \ldots$ el *estado* del enlace será $X(t) = i$.

Dinámica de estados: Asumiremos que los estados de un enlace evolucionan de manera estocásticamente independiente a la evolución de los estados de los otros enlaces. En general el estado de un enlace puede cambiar con el paso del tiempo y con la acción tomada, usaremos la notación $P(i, j|a)$ como la probabilidad de que un enlace que está en el estado i pase al estado j al tomar la acción a.

Recompensas: Consideraremos $R(i, a)$ como la recompensa total obtenida en un paso de tiempo al tomar la acción a cuando el enlace está en el estado i. Está recompensa puede depender de muchos factores, tales como el número total de usuarios eligiendo el enlace en un paso de tiempo o la calidad esperada del enlace. Dado que siempre podemos agrandar la definición de "estado" para incluir tales factores, al tomar la acción a la recompensa $R(i, a)$ es únicamente determinada por el estado i del enlace.

También haremos la suposición que poniendo un enlace en la lista top anima a más usuarios a usarlo, por lo tanto acelera su transición a un estado diferente, es decir, la probabilidad de transición hacia un estado diferente es mayor bajo la acción activa que bajo la acción pasiva ($P(i, j|1) > P(i, j|0)$, con $i \neq j$) y la probabilidad de permanecer en su estado bajo la acción activa será menor que bajo la acción pasiva ($P(i, i|1) < P(i, i|0)$), esta es una suposición de doble velocidad y puede ser escrita como

$$P(i, j|0) = \begin{cases} \epsilon_i P(i, j|1) & i \neq j, \\ (1 - \epsilon_i) + \epsilon_i P(i, i|1) & i = j. \end{cases}$$

donde $\epsilon_i \in (0, 1)$.

Nuestro objetivo es encontrar una política [1] π en el espacio Π de las políticas estacionarias que maximice la esperanza de la recompensa total [2] descontada ganada por el usuario. En un horizonte infinito esto es

$$\max_{\pi \in \Pi} E^{\pi}_{i^0} \left[\sum_{t=0}^{\infty} \sum_{n=1}^{N} R_n(X_n(t), a_n(t)) \beta^t \right], \tag{4.1}$$

donde $0 < \beta < 1$ es el factor de descuento futuro, $X_n(t)$, $a_n(t)$ y $R_n(X_n(t), a_n(t))$ son el estado, la acción y la recompensa procesada por n-ésimo proyecto (enlace) en el periodo de tiempo t, además condicionado a que el estado inicial de la cartera de proyectos (enlaces) sea $i^0 = (X_n(0))_{n=1}^N$ y también a que se aplique la acción activa a M enlaces en cada periodo de tiempo, es decir

$$\sum_{n=1}^{N} a_n(t) = M \quad , t = 0, 1, 2, \ldots$$

[1] Está política se traduce en el conjunto de enlaces mostrados en la lista top en cada periodo de tiempo.
[2] Total significa suma sobre todos los periodos de tiempo.

4.2. Ajustando los Parámetros de un Enlace para usar el Algoritmo AG

Estados:

Estamos considerando un sitio web que puede mostrar M enlaces de un total de N en un periodo de tiempo $t = 0, 1, 2, \ldots$, cada enlace puede tener la evaluación de estrellas entre 1 y 5, 5 estrellas denota la evaluación más alta y 1 la más baja, esta evaluación es agregada voluntariamente por los usuarios. Cada enlace tiene también un nivel de acceso entre 1 y 5, el cual indica el número de clicks en un paso de tiempo, donde 5 denota el mayor número de cliks recibidos, el nivel de acceso puede ser establecido de manera bastante arbitraria; en nuestro caso, nivel 1 cuando hay entre 1 y 1000 clicks, nivel 2 cuando hay entre 1001 y 2000 clicks, nivel 3 cuando hay entre 2001 y 3000 clicks, etc. Por tanto tenemos estados bidimensionales y pueden ser representados como un vector $(s, c) \in \{1, 2, 3, 4, 5\}^2$, donde s es la evaluación de estrellas y c es el nivel de acceso.

En adición a estos 25 estados hay un estado más, el estado "d", al cual llamaremos estado "desconocido", cada enlace empieza inicialmente en este estado ya que nunca ha sido accedido o evaluado. Asumiremos que ocasionalmente un enlace "muere" y si esto pasa es inmediatamente reemplazado por un nuevo enlace. Esto es equivalente a asumir que hay una pequeña probabilidad de transición de cada uno de los 25 estados al estado desconocido, lo cual implica empezar de nuevo, el estado "d" sirve como una fuente y pozo a la vez.

Enumeraremos los 25 estados bidimensionales (s, c) y el estado desconocido, de la siguiente manera:

$$y_k = \begin{cases} d & k = 1, \\ (1, k-1) & , k = 2, \ldots, 6, \\ (2, k-6) & , k = 7, \ldots, 11, \\ (3, k-11) & , k = 12, \ldots, 16, \\ (4, k-16) & , k = 17, \ldots, 21, \\ (5, k-21) & , k = 22, \ldots, 26. \end{cases}$$

Recompensas:

Mediremos las recompensas para el usuario en términos de una función de valor $V(s, c)$, definida como:

$$V(s, c) = \begin{cases} 0 & \text{en el estado desconocido,} \\ s.c & \text{en otro caso.} \end{cases}$$

Tomaremos $R(i, a)$ de la siguiente manera:

$$\begin{aligned} R(i, 1) = &\quad V(i), \\ R(i, 0) = &\quad 30\,\%V(i). \end{aligned}$$

Dinámica de Estados:

Tomaremos las probabilidades de transición de estado cuando se aplica la acción activa (poner el enlace en la lista top) de la siguiente manera:

$$
\begin{array}{lll}
P((s,c),(s+1,c)|1) = 0.1 & 1 \leq s \leq 4, & 1 \leq c \leq 5, \\
P((s,c),(s-1,c)|1) = 0.1 & 2 \leq s \leq 5, & 1 \leq c \leq 5, \\
P((s,c),(s,c+1)|1) = 0.2 & 1 \leq s \leq 5, & 1 \leq c \leq 4, \\
P((s,c),(s,c-1)|1) = 0.1 & 1 \leq s \leq 5, & 2 \leq c \leq 5, \\
P((s,c),(s+1,c-1)|1) = 0.1 & 1 \leq s \leq 4, & 2 \leq c \leq 5, \\
P((s,c),(s-1,c+1)|1) = 0.1 & 2 \leq s \leq 5, & 1 \leq c \leq 4, \\
P((s,c),(s+1,c+1)|1) = 0.2 & 1 \leq s \leq 4, & 1 \leq c \leq 4, \\
P((s,c),(s-1,c-1)|1) = 0.01 & 2 \leq s \leq 5, & 2 \leq c \leq 5, \\
P((s,c),d|1) = 0.01 & 1 \leq s \leq 5, & 1 \leq c \leq 5, \\
P(d,d|1) = 0.01 & 1 \leq s \leq 5, & 1 \leq c \leq 5, \\
P(d,(s,1)|1) = 0.2 & s = 1,3,4,5, & \\
P(d,(s,1)|1) = 0.19 & s = 2. &
\end{array}
$$

La probabilidad $P((s,c),d|1)$ será pequeña en relación a las probabilidades $P((s,c),j|1)$ con $j \neq d$, debido a que consideramos que un estado muere con una probabilidad pequeña.

Las probabilidades de transición no mostradas son cero salvo $P(i,i|1)$ que es igual a uno menos la suma de los demás valores de la i-ésima fila de la matriz estocástica que se genera.

Teniendo en cuenta la Figura 4.1, note que estamos considerando que un estado sólo tiene las siguientes alternativas de movimiento: moverse hacia los estados contiguos o moverse hacia el estado desconocido. Las demás opciones de movimiento se consideran con probabilidad cero, esto es debido a que consideramos el comportamiento natural de los usuarios de evaluar las características de un estado solo en más o menos un nivel o quedarse en el mismo nivel.

Note también la suposición de que el nivel de acceso tiende a crecer más que a decrecer, además que las características del estado tienden a crecer juntas, esto es debido al comportamiento natural del manager que muestra enlaces cada vez mejor valorados.

Para hallar las probabilidades correspondientes cuando el enlace no se pone en la lista top (acción pasiva) usaremos la doble velocidad con $\epsilon_i = 0.1$ para todo estado i. Observe que esto expresa el hecho de que mostrar un enlace en la lista top acelera su probabilidad de transición 10 veces con respecto a si el enlace no se muestra en la lista top. Los estados y probabilidades de transición son ilustrados en la Figura 4.1.

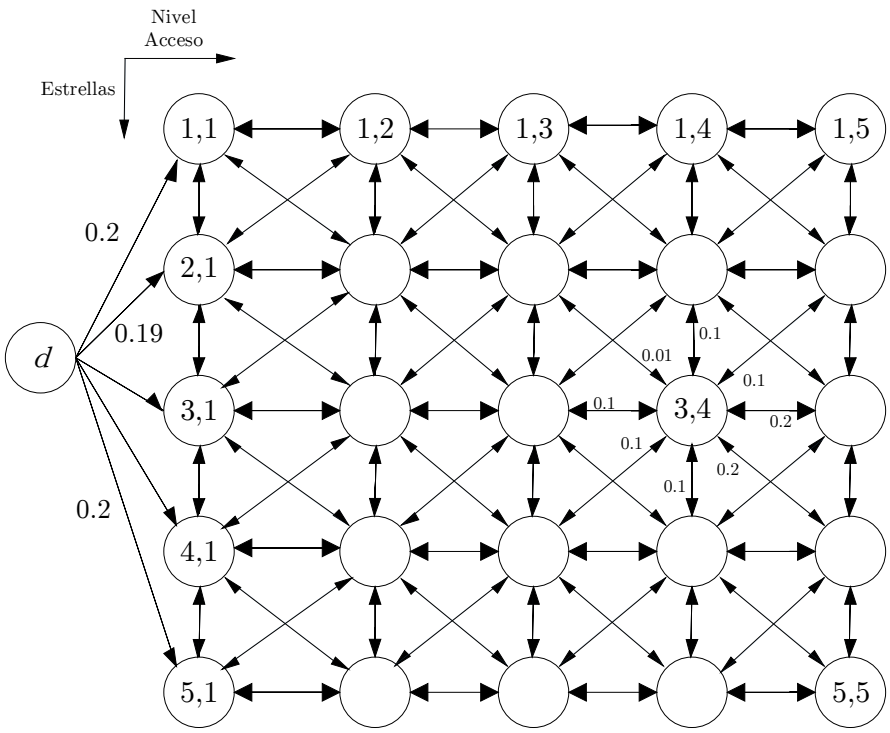

Figura 4.1: la figura muestra los 26 estados y las probabilidades de transición $P(i, j|1)$ entre ellas. El eje horizontal denota el nivel de acceso y el eje vertical la evaluación de estrellas. Lo que no muestra la figura es que cada estado transita al estado desconocido con una pequeña probabilidad 0.01, tampoco muestra la probabilidad de permanecer en el mismo estado.

Trabajo Gastado:

El trabajo gastado por el manager del proyecto al tomar la acción activa y pasiva en el estado i, serán tomados como $Q(i, 1) = 1$ y $Q(i, 0) = 0$ respectivamente.

4.3. Indexación de los estados de un Enlace Usando el Índice de Whittle

Tomando los parámetros de la sección anterior y usando el algoritmo AG podemos indexar los estados de un enlace con los índices de Whittle, note que en el algoritmo AG se va generando la medida de productividad marginal ν_i^S, esta medida depende del factor de descuento β. Por lo tanto, manteniendo fijos los parámetros del enlace, al hacer variar β el algoritmo AG produce distintos valores para los índices de Whittle.

Si por ejemplo, tomamos $\beta = 0.7$, el algoritmo AG nos devuelve el siguiente ordenamiento de los 26 estados y sus respectivos índices de Whittle, $\left\{i_k, \nu_{i_k}^*\right\}_{k=1}^{26}$:

- $S_{26} = \{i_1, \ldots, i_{26}\}$
  $= \{26, 25, 21, 20, 24, 16, 19, 15, 14, 23, 18, 11, 10, 13, 9, 17, 22, 5, 8, 12, 6, 4, 7, 3, 1, 2\}$

- $\{\nu^*_{i_1}, \ldots, \nu^*_{i_{26}}\} = \{16.6915, 14.4601, 14.0358, 13.3141, 11.6571, 11.0726, 11.0082, 10.7013,$
  $9.0981, 8.6127, 8.3771, 7.9111, 7.8803, 7.0154, 6.7919, 5.9074, 5.7042,$
  $5.3295, 5.3109, 5.0373, 4.9691, 4.6771, 3.8879, 3.7487, 3.7388, 2.7930\}.$

Además, podemos asignar una prioridad a cada estado, como se muestra en la Figura 4.2, donde tomaremos con mayor prioridad al estado con mayor índice de Whittle; es decir, 1 indica al estado con mayor índice de Whittle, 2 al estado con el siguiente mayor índice de Whittle, etc.

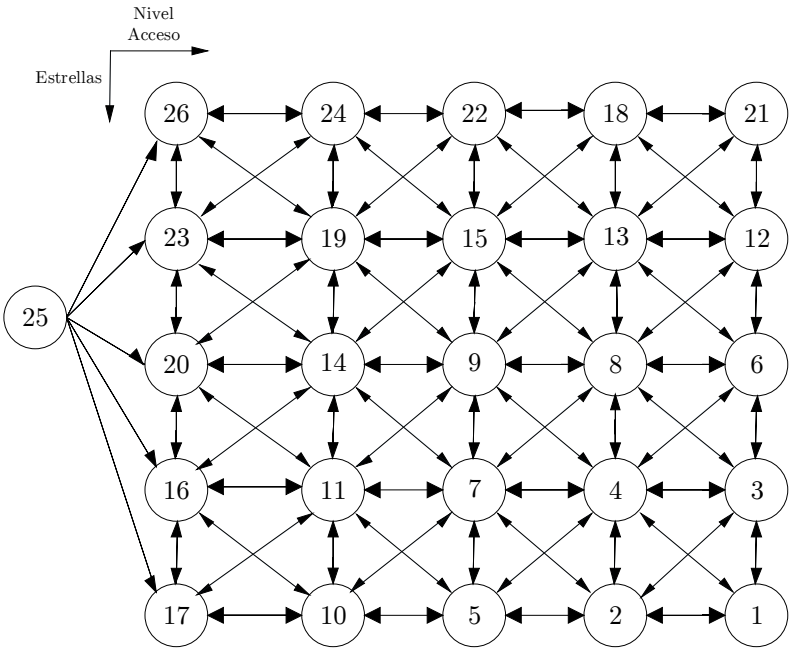

Figura 4.2: Ordenamiento de estados para $\beta = 0.7$.

Teniendo en cuenta la enumeración de los estados, podemos ver en la Figura 4.2 que el estado 26 tiene el mayor índice de Whittle y el estado 2 tiene el menor índice de Whittle, etc.

Análisis de sensibilidad:

Podemos hacer un análisis de sensibilidad para el ordenamiento de los 26 estados al hacer variar beta, como se muestra en la Tabla 4.1.

Los gráficos desde la Figura 4.3 hasta la Figura 4.10, nos ayudarán a visualizar la monotonía de la prioridad de los estados al mantener fijo un parámetro (nivel de acceso, número de estrellas) y aumentar el otro, para cada valor de beta.

β	Estados																									
0.05	26	25	21	20	24	16	19	15	23	11	14	18	10	13	9	22	6	17	5	8	12	4	7	3	2	1
0.1	26	25	21	20	24	16	19	15	23	11	14	18	10	13	9	22	6	17	5	8	12	4	7	3	2	1
0.2	26	25	21	20	24	16	19	15	23	11	14	18	10	13	9	22	6	17	5	8	12	4	7	3	2	1
0.3	26	25	21	20	24	16	19	15	23	11	14	18	10	13	9	22	6	17	5	8	12	4	7	3	2	1
0.5	26	25	21	20	24	16	19	15	14	23	11	18	10	13	9	22	17	6	5	8	12	4	7	3	1	2
0.7	26	25	21	20	24	16	19	15	14	23	18	11	10	13	9	17	22	5	8	12	6	4	7	3	1	2
0.9	26	25	20	21	19	24	15	16	14	18	23	10	13	9	17	11	22	12	8	5	1	4	7	6	3	2
0.95	20	19	25	26	15	21	24	14	18	13	16	23	17	10	9	12	22	1	8	5	11	4	7	3	6	2
0.98	1	17	12	18	13	14	19	22	9	4	8	7	5	15	20	3	23	10	2	24	6	25	11	16	21	26
0.99	1	17	12	7	4	3	13	22	2	8	5	18	9	14	19	10	6	23	15	11	20	24	16	25	21	26
0.999	1	2	7	3	12	4	17	8	5	22	6	13	9	18	10	14	23	11	19	15	24	16	20	25	21	26

Tabla 4.1: Se muestran los 26 estados ordenados de mayor a menor índice de Whittle para cada valor de β.

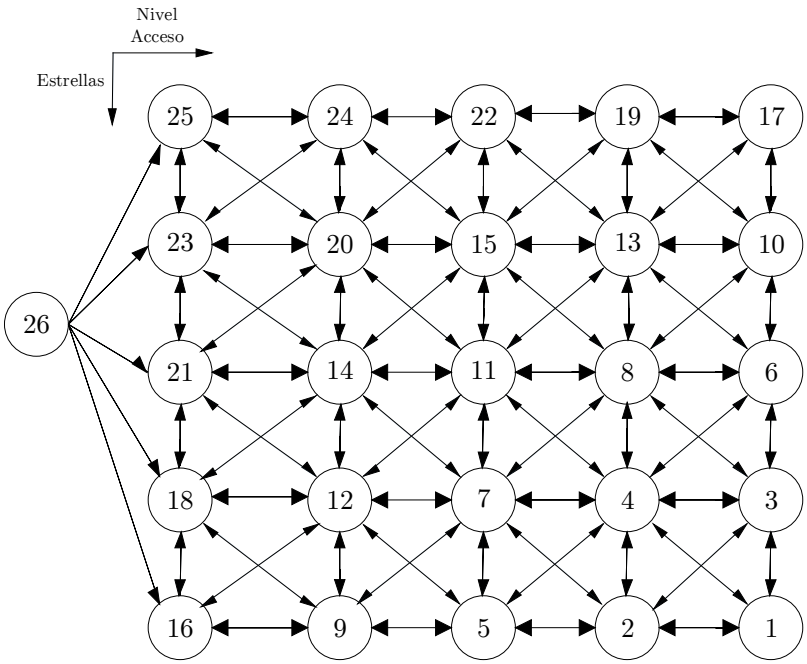

Figura 4.3: Ordenamiento de estados para $\beta = 0.05, 0.1, 0.2$ y 0.3.

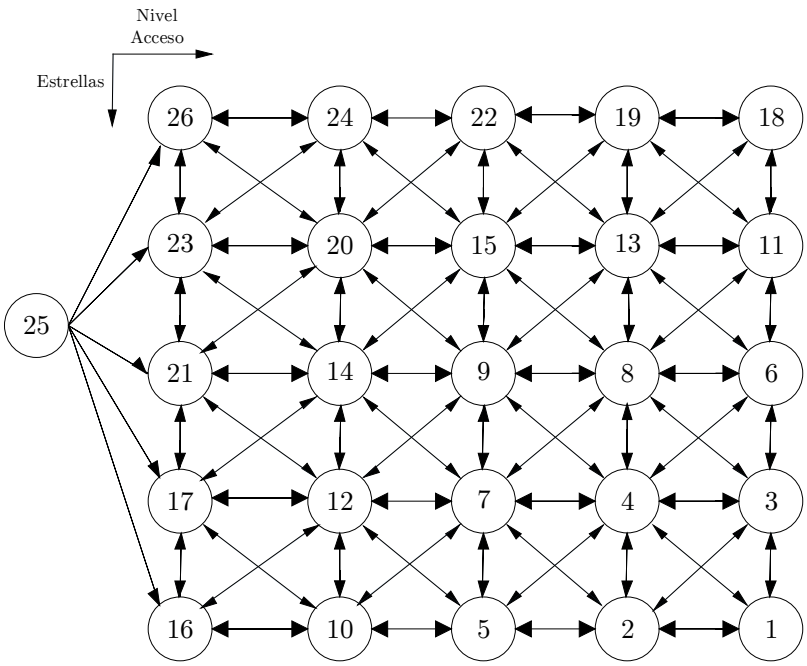

Figura 4.4: Ordenamiento de estados para $\beta = 0.5$.

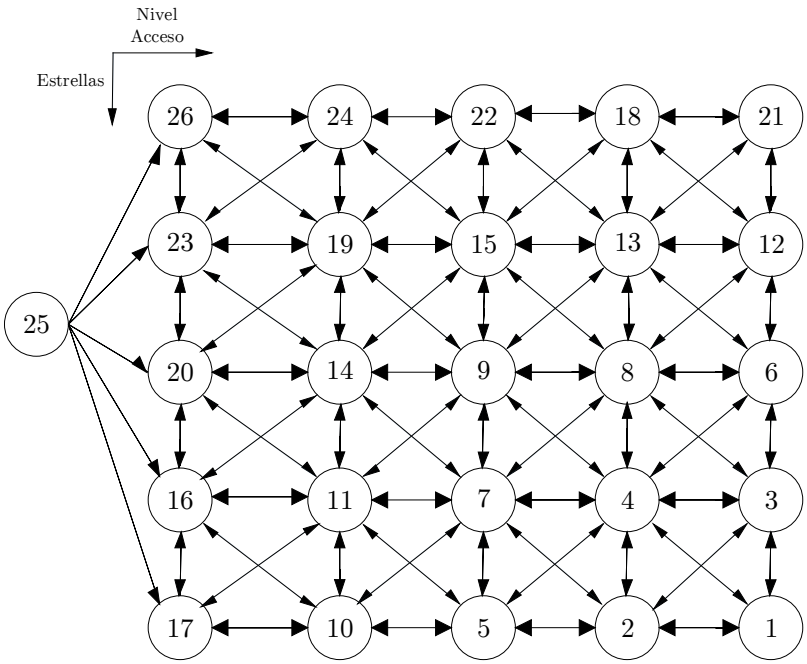

Figura 4.5: Ordenamiento de estados para $\beta = 0.7$.

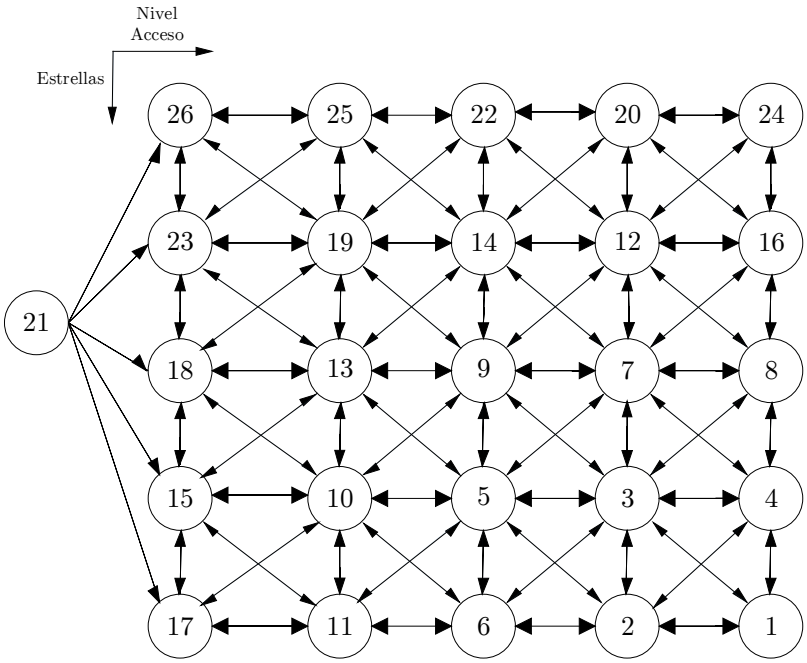

Figura 4.6: Ordenamiento de estados para $\beta = 0.9$.

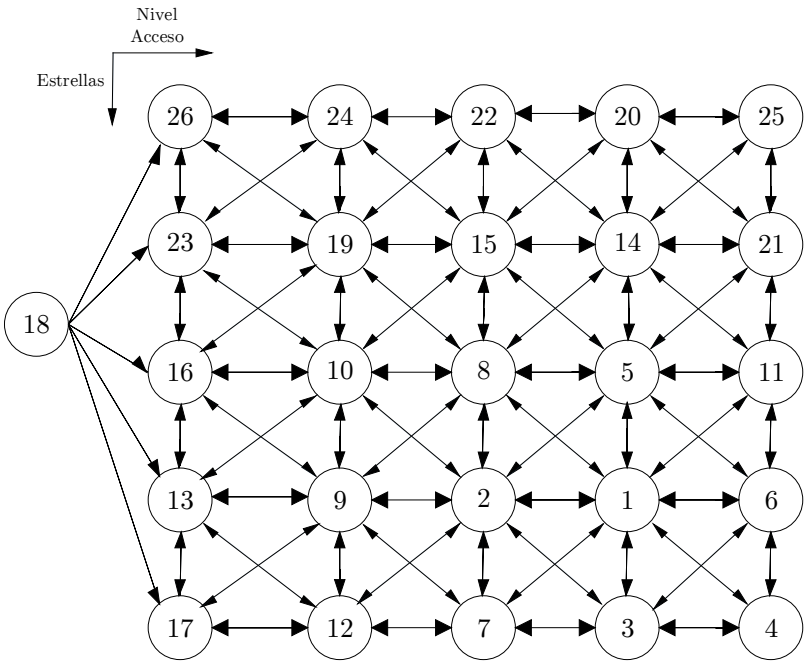

Figura 4.7: Ordenamiento de estados para $\beta = 0.95$.

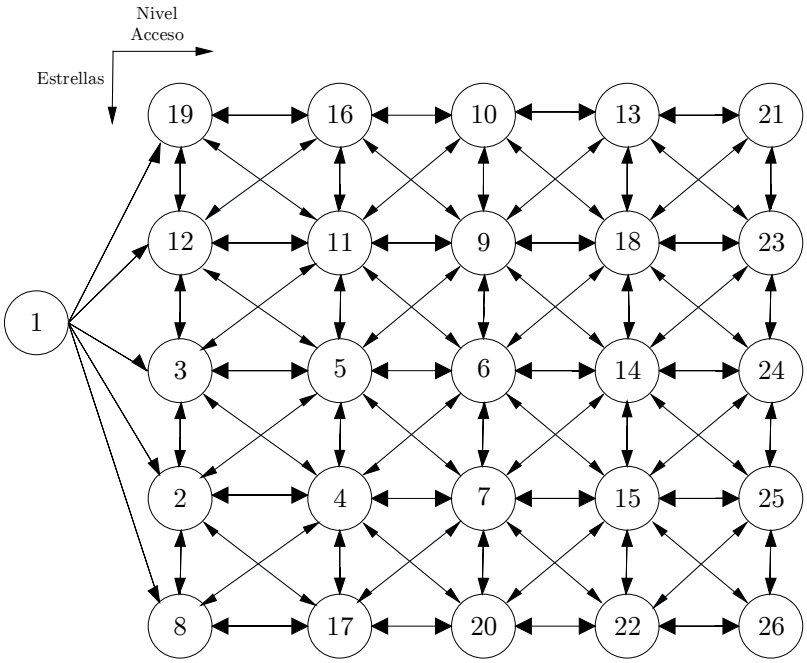

Figura 4.8: Ordenamiento de estados para $\beta = 0.98$.

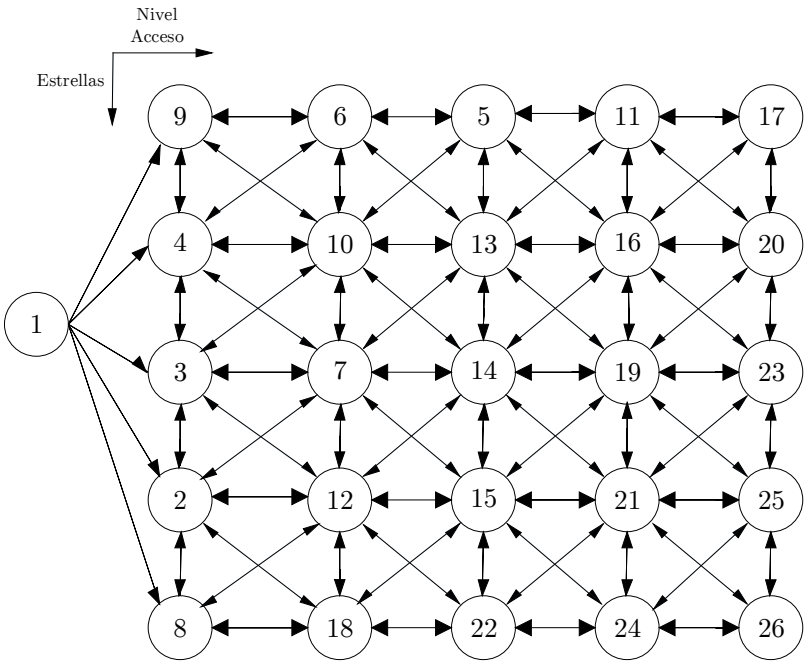

Figura 4.9: Ordenamiento de estados para $\beta = 0.99$.

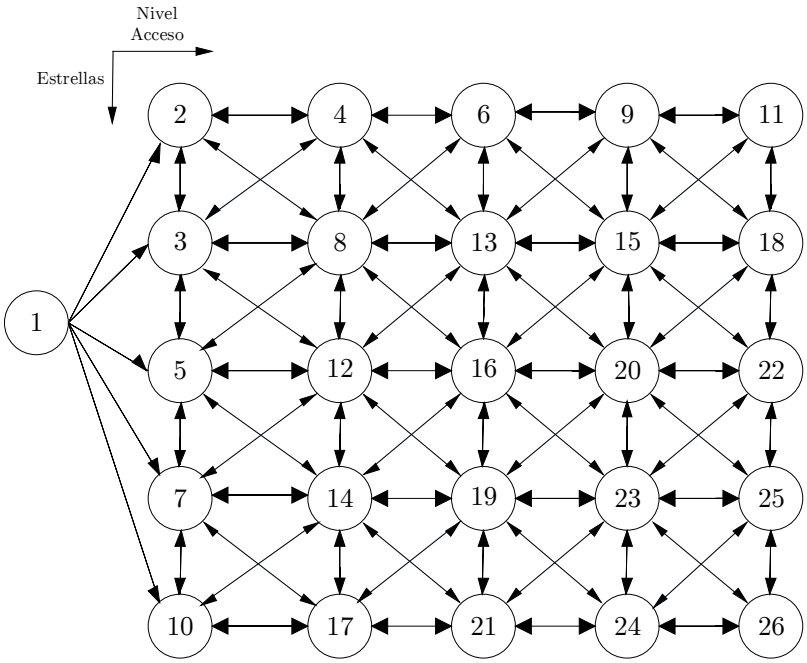

Figura 4.10: Ordenamiento de estados para $\beta = 0.999$.

Podemos observar de la Tabla 4.1 que para β desde 0.05 hasta 0.9 la mayor prioridad la tiene el estado 26 que es igual a $(5, 5)$, cuyas recompensas (activa y pasiva) son la mayores de todas y que la menor prioridad la tiene el estado 1 (estado desconocido), cuyas recompensas son las menores de todas. Por otro lado, para beta mayor o igual que 0.98 la situación se invierte, es decir, tiene mayor prioridad el estado 1 y menor prioridad el estado 26. Podemos concluir entonces que para β desde 0.05 hasta 0.9 se valora más la explotación que la exploración y que para β mayor o igual que 0.98 se valora más la exploración que la explotación.

Por otro lado, si mantenemos fijo un parámetro del estado y hacemos aumentar el otro, entonces para $\beta = 0.05, 0.1, 0.2$ y 0.3 podemos observar en la Figura 4.3 que la prioridad aumenta, lo cual es razonable, dado que en corto plazo estamos valorando más la explotación que la exploración. Sin embargo, para $\beta = 0.999$ podemos observar en la Figura 4.10 que la prioridad disminuye, lo cual no es trivial y es debido a que en el largo plazo estamos valorando más la exploración que la explotación.

4.4. Caso Multiproyecto Homogéneo

El caso multiproyecto consiste en escoger M enlaces de un total de N en cada periodo de tiempo, con la finalidad de maximizar la esperanza de la recompensa total descontada ganada por el usuario, como planteamos en la ecuación (4.1).

En esta sección trabajaremos el caso multiproyecto homogéneo, es decir, donde los N enlaces tienen la misma recompensa y la misma dinámica de estados (las definidas en la sección 4.2).

Escogeremos $M = 1$ enlace de un total de $N = 12$ enlaces, tomaremos $\beta = 0.99$ ya que nos interesa el caso de la planificación a largo plazo, el estado inicial de la cartera de proyectos es arbitrario, tomaremos

$$i^0 = \left\{i_n^0\right\}_{n=1}^{12} = \{20, 21, 4, 5, 7, 13, 8, 9, 26, 2, 12, 6\}.$$

Cálculo de la Esperanza de la Recompensa Total Descontada Ganada por el Usuario:

Compararemos el rendimiento de la *política-índice de Whittle* con el rendimiento de la *política-índice Greedy*, esta última política prioriza (toma acción activa) a los M enlaces cuyo estado tiene mayor medida de productividad marginal actual (cuando $\beta = 0$), también se dice que prioriza lo que es mejor aquí y ahora.

El problema original (4.1) es intratable debido a que tiene horizonte de planificación infinito, por ello la aproximaremos mediante una simulación de Montecarlo para la política-índice de Whittle y otra para la política-índice Greedy, está simulación de Montecarlo será el promedio de 100 simulaciones de la forma

$$\sum_{t=0}^{74000} \sum_{n=1}^{12} R_n(X_n(t), a_n(t))\beta^t,$$

como observamos, cada una con horizonte de planificación de 74000 periodos de tiempo.

En cada periodo de tiempo las transiciones de estado de los enlaces serán simulados como una variable aleatoria discreta usando la función *rand* de MATLAB.

Aproximaciones e Intervalos de Confianza:

Utilizaremos el siguiente intervalo al 98 % de confianza para la media con varianza desconocida:

$$\bar{x} \pm t_{(n-1),\frac{\alpha}{2}} \left(\frac{s}{\sqrt{n}} \right),$$

donde:

- n: Es el tamaño de la muestra $n = 100$.
- $\bar{x}$: Es la esperanza de la recompensa total descontada ganada por el usuario, con las 100 simulaciones.
- $t_{(n-1),\frac{\alpha}{2}}$: Es el percentil de la distribución t-Student al 98 % de confianza, es decir $\alpha = 2$.
- s: Es la desviación estándar muestral.

Usando Matlab (Anexo 1), tenemos que:

RecompensaMedia(P0,P1,R0,R1,beta,d,m,I0,n,sim): Devuelve $\bar{x}$ y los límites del intervalo confianza para la política-índice de Whittle.

RecompensaMediaGreedy(P0,P1,R0,R1,beta,n,I0,sim): Devuelve $\bar{x}$ y los límites del intervalo confianza para la política-índice Greedy.

Presentamos los resultados en la siguiente tabla:

	Límite Inferior	$\bar{x}$	Límite Superior
Política-índice de Whittle	6191.1	6285.9	6380.6
Política-índice Greedy	5078.7	5174	5269.3

Tabla 4.2: Media e intervalo de confianza para el caso multiproyecto homogéneo.

La variación porcentual del rendimiento de la política-índice de Whittle con respecto al rendimiento de la política-índice Greedy es

$$\frac{6285.9 - 5174}{6285.9} \times 100\,\% = 17.6\,\%.$$

Por lo tanto, tenemos que la política-índice de Whittle mejora a la política-índice Greedy en un 17.6 %, además como los intervalos de confianza son disjuntos se asegura la mejora.

4.5. Caso Multiproyecto Heterogéneo

En esta sección trabajaremos el caso multiproyecto heterogéneo, es decir, donde los enlaces tienen distinta dinámica estados, generaremos las dinámicas de estado aleatoriamente pero con las mismas alternativas de movimiento que en el caso homogéneo.

Tomaremos $M = 1$, $N = 12$, $\beta = 0.99$, $\{i_n^0\}_{n=1}^{12} = \{20, 21, 4, 5, 7, 13, 8, 9, 26, 2, 12, 6\}$, igual que en el caso homogéneo.

Cálculo de la Esperanza de la Recompensa Total Descontada Ganada por el Usuario:

Compararemos el rendimiento de la *política-índice de Whittle* con el rendimiento de la *política-índice Greedy* aproximando la esperanza de la recompensa total descontada ganada por el usuario a través de una simulación de Montecarlo como lo hicimos para el caso homogéneo, pero en cada periodo de tiempo simularemos el cambio de estado de un enlace de acuerdo a su dinámica de estados.

Aproximaciones e Intervalos de Confianza:

Usaremos el mismo intervalo de confianza que en el caso homogéneo

$$\bar{x} \pm t_{(n-1),\frac{\alpha}{2}} \left(\frac{s}{\sqrt{n}} \right).$$

Usando Matlab (Anexo 2), tenemos que:

RecompensaMedia2(P0,P1,R0,R1,beta,D,M,I0,n,sim): Devuelve $\bar{x}$ y los límites del intervalo confianza para la política-índice de Whittle.

RecompensaMediaGreedy2(P0,P1,R0,R1,beta,I0,n,sim): Devuelve $\bar{x}$ y los límites del intervalo confianza para la política-índice Greedy.

Presentamos los resultados en la siguiente tabla:

	Límite Inferior	$\bar{x}$	Límite Superior
Política-índice de Whittle	3737.6	3807.3	3876.9
Política-índice Greedy	2601.8	2656.4	2711

Tabla 4.3: Media e intervalo de confianza para el caso multiproyecto heterogéneo.

La variación porcentual del rendimiento de la política-índice de Whittle con respecto al rendimiento de la política-índice Greedy es

$$\frac{3807.3 - 2656.4}{3807.3} \times 100\,\% = 30.2\,\%.$$

Por lo tanto, tenemos que la política-índice de Whittle mejora a la política-índice Greedy en un 30.2 %, además como los intervalos de confianza son disjuntos se asegura la mejora.

4.6. Cálculo de la Cota de Whittle

Calcularemos la cota de Whittle $W(i^0) = \min_{\lambda \in \mathbb{R}} L(\lambda, i^0)$ siguiendo los pasos descritos en la sección 3.1. La minimización la función $L(\lambda, i^0)$ la haremos usamos el método de la sección áurea tomando λ entre el mayor y el menor índice de Whittle.

Para calcular la cota de Whittle implementamos las siguientes funciones en Matlab (Anexo 3).

CotaWhittleHomo(P0,P1,R0,R1,m,beta,I0,n,M): Para el caso homogéneo.

CotaWhittleHete(P0,P1,R0,R1,M,beta,I0,n,M1): Para el caso heterogéneo.

Tenemos los siguientes resultados:

	Política-Índice de Whittle	Cota de Whittle	Variación Porcentual
Caso Homogéneo	6285.5	6359.5	1.16 %
Caso Heterogéneo	3952.7	4074.3	3.08 %

Tabla 4.4: Cota de Whittle

Podemos observar que en el caso homogéneo (resp. heterogéneo) el rendimiento de la política-índice de Whittle está a 1.16 % (resp. 3.08 %) de la cota de Whittle, esto no dice que en este modelo la política-índice de Whittle nos da valores próximos al valor óptimo real. Además obtuvimos las siguientes gráficas para la función $L(\lambda, i^0)$, las cuales tienen la forma descrita en la Figura 3.1.

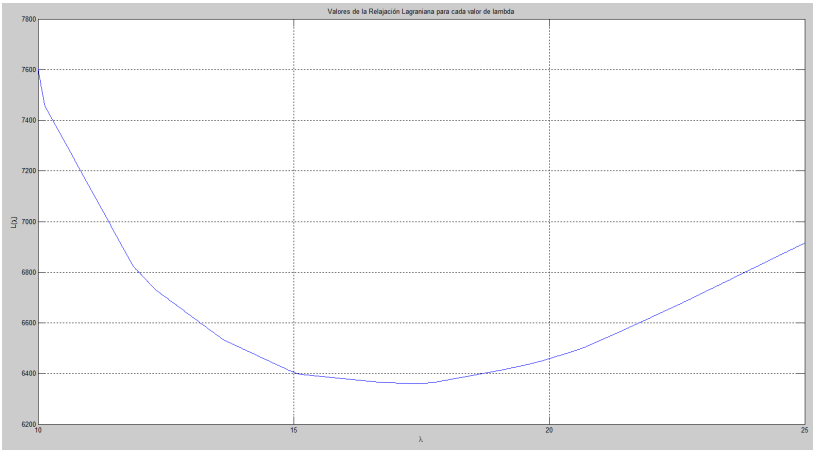

Figura 4.11: Gráfica de la función $L(\lambda, i^0)$ para el caso homogéneo.

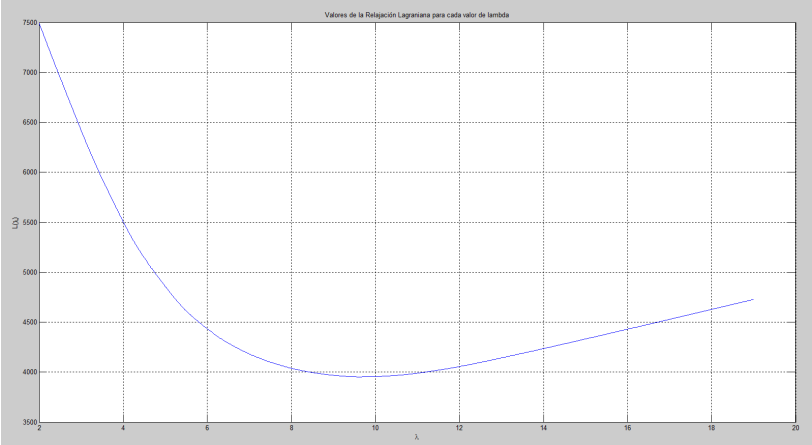

Figura 4.12: Gráfica de la función $L(\lambda, i^0)$ para el caso heterogéneo.

4.7. Conjetura de Whittle

Whittle propuso la siguiente conjetura. Ver Whittle (1988, p. 293).

Conjeruta: *Suppose all proyects indexable. Then* $r_{ind}(\alpha)$ *exists and equals* $r(\alpha)$.

Esto es, si tenemos N proyectos indexables de los cuales activamos M, además si N y M crecen en la misma proporción; entonces el rendimiento producido por la política-índice de Whittle es óptimo en el límite.

En las siguientes tablas valoraremos la conjetura de Whittle para nuestro problema en los casos homogéneo y heterogéneo, también calcularemos y compararemos los rendimientos de la política-índice de Whittle y la política-índice Greedy. Los resultados los obtuvimos haciendo la simulación de Montecarlo mencionada anteriormente y considerando al vector de estados iniciales como copias de $\{i_n^0\}_{n=1}^{12} = \{20, 21, 4, 5, 7, 13, 8, 9, 26, 2, 12, 6\}$ en cada caso.

	Política Índice			Variación Porcentual	
Enlaces Activos	Greedy	Whittle	Cota de Whittle	Greedy\Whittle	Whittle\Cota de Whittle
1 de 12	5174	6285.9	6359.5	17.60 %	1.1 %
2 de 24	10243.0	12489.8	12719.0	17.90 %	1.8 %
3 de 36	15355.0	18666.7	19078.5	17.74 %	2.1 %
4 de 48	20355.5	24780.9	25438.0	17.85 %	2.5 %
5 de 60	25597.2	31218.0	31797.5	18.00 %	1.8 %
6 de 72	30680.1	37566.0	38157.0	18.30 %	1.5 %

Tabla 4.5: Cota de Whittle caso Homogéneo

Podemos observar en la Tabla 4.5 que en el caso homogéneo el rendimiento de la política-índice de Whittle mejora al de la política-índice Greedy hasta en un 18.3 % y que está a menos del 2.5 % de la

cota de Whittle, lo cual habla de un buen rendimiento de la política-índice de Whittle en este modelo. Sin embargo, al aumentar el número de enlaces total y sus respectivos enlaces activos no se observa una marcada reducción a cero de la diferencia entre el rendimiento de la política-índice de Whittle y la cota de Whittle, es decir no se verifica la conjetura de Whittle hasta el caso de tomar como activos 6 enlaces de un total de 72. Se debería intentar aumentar dichas cantidades, sin embargo esto es computacionalmente muy costoso.

	Política Índice			Variación Porcentual	
Enlaces Activos	Greedy	Whittle	Cota de Whittle	Greedy\Whittle	Whittle\Cota de Whittle
1 de 12	2883.8	3952.7	4074.3	27.0 %	2.98 %
2 de 24	5192.0	7548.2	7825.5	31.2 %	3.50 %
3 de 36	8358.7	11700	11716	28.5 %	0.14 %
4 de 48	11061	15354	15579	27.9 %	1.40 %

Tabla 4.6: Cota de Whittle caso Heterogéneo

Observamos en la Tabla 4.6 que en el caso heterogéneo el rendimiento de la política-índice de Whittle mejora al de la política-índice Greedy hasta en un 31.2 % y que está a menos del 3.5 % de la cota de Whittle, lo cual también habla de un buen rendimiento de la política-índice de Whittle en este modelo. También podemos observar que al aumentar el número total de enlaces y sus respectivos enlaces activos hay una reducción no muy marcada en la diferencia entre el rendimiento de la política-índice de Whittle y la cota de whittle. Como en el caso anterior deberíamos intentar aumentar el número total de enlaces y sus respectivos enlaces activos, sin embargo esto es computacionalmente muy costoso.

Capítulo 5

Conclusiones y Trabajo Futuro

En la era actual, el internet y los medios de comunicación digitales han cambiado el enfoque del mundo en el que la información era valiosa y escasa, al del mundo en el que la información es abundante y de fácil acceso. En este ultimó contexto nace el concepto de economía de la atención, que básicamente explica que al haber abundancia de información, lo que es escaso es la atención humana. Por lo tanto, lo que los proveedores información buscarán será llamar la atención de los usuarios mostrándoles las alternativas que más los beneficiarán, ya que esto también se traducirá en ganancia para dicho proveedor.

Con estas ideas en mente, planteamos el problema de una página Web que quiere mostrar N enlaces pero solo puede mostrar M ($M < N$) en cada periodo de tiempo, con el objetivo de maximizar el beneficio para el usuario. Modelamos este problema como un problema *multi-armed restless bandit* de doble velocidad y aplicamos la política-índice de Whittle para aproximar su solución.

La conclusión principal que obtenemos es que la política-índice de Whittle en nuestro problema da resultados subóptimos con grado de suboptimalidad pequeño, además mejora sustancialmente al rendimiento de la política-índice Greedy.
A continuación presentamos los resultados obtenidos.

En la sección 4.3, indexamos los estados de un enlace usando el índice de Whittle e hicimos un análisis de sensibilidad de la prioridad de los estados al hacer variar el factor de descuento β, con $0 < \beta < 1$. Obtuvimos los siguientes resultados.

- El análisis de sensibilidad mostrado en la Tabla 4.1 muestra que para valores de β próximos a 0 (en el corto plazo) se priorizan los estados con mayor recompensa actual, es decir se valora más la explotación que la exploración. Sin embargo, para valores de β próximos a 1 (largo plazo) la situación se invierte, es decir, se priorizan los estados con menor recompensa actual, resultado que no es trivial y que nos dice que en el largo plazo se prioriza más la exploración que la explotación.

- Si mantenemos fijo un parámetro del estado (nivel de acceso, número de estrellas) y hacemos aumentar el otro, entonces para $\beta = 0.05, 0.1, 0.2$ y 0.3 podemos observar en la Figura 4.3 que la prioridad aumenta, lo cual es razonable, dado que en corto plazo estamos valorando más la explotación que la exploración. Sin embargo, para $\beta = 0.999$ podemos observar en la Figura 4.10 que la prioridad disminuye, lo cual no es trivial y es debido a que en el largo plazo estamos valorando más la exploración que la explotación.

En las secciones 4.4 y 4.5, comparamos el rendimiento de la política-índice de Whittle con el rendimien-

to de la política-índice Greedy en los casos homogéneo y heterogéneo respectivamente. Obtuvimos el siguiente resultado.

- En el caso homogéneo (resp. heterogéneo), el rendimiento de la política-índice de Whittle mejora al rendimiento de la política-índice Greedy en un 17.6 % (30.2 %).

En la sección 4.6, calculamos la cota de Whittle y la comparamos con el rendimiento de la política-índice de Whittle. Obtuvimos el siguiente resultado.

- En el caso homogéneo (resp. heterogéneo), el rendimiento de la política-índice de Whittle está a 1.16 % (3.08 %) de la cota de Whittle, esto habla de un buen rendimiento de política-índice de Whittle en este modelo.

En la sección 4.7, estudiamos la conjetura de Whittle y obtuvimos los siguientes resultados.

- A través de las tablas 4.5 y 4.6, podemos observar que en el caso homogéneo (resp. heterogéneo) el rendimiento de la política-índice de Whittle está a menos del $2.5\,\%$($3.5\,\%$) de la cota de Whittle y a más del $17.6\,\%$($27.9\,\%$) del rendimiento de la política-índice Greedy, lo cual nos habla de un buen rendimiento de la política-índice de Whittle en este modelo.
- En los experimentos realizados no se verifica la conjetura de Whittle. Se debería intentar aumentar el número total de enlaces y sus respectivos enlaces activos, sin embargo, esto es computacionalmente muy costoso.

Como trabajo futuro, se propone calcular los parámetros de un enlace (dados en la sección 4.2) usando datos reales y aplicar la metodología de la presente memoria.

Anexos

Anexo 1

Mostraremos la simulación de Montecarlo para aproximar el rendimiento de la política-índice de Whittle en el caso homogéneo.

```
function [I,Recompensa]=RecompensaMedia(P0,P1,R0,R1,beta,d,m,I0,n,sim)
% DATOS DE ENTRADA
% P0  : Matriz con la dinamica de estados bajo la accion pasiva.
% P1  : Matriz con la dinamica de estados bajo la accion activa.
% R0  : Vector de recompensas bajo la accion pasiva.
% R1  : Vector de recompensas bajo la accion activa.
% beta: Factor de descuento.
% d   : Vector con los estados del Output del algoritmo AG.
% m   : vector con los IPM del Ouput del algoritmo AG.
% I0  : Vector conteniendo los estados iniciales de la  cartera de proyectos.
% n   : Numero de enlaces a los cuales se aplica la accion activa.
% sim : Numero de simulaciones.
% DATOS DE SALIDA
% Recompensa : Aproximacion de la esperanza de la ganancia total descontada .
% I          : Intervalo de confianza.
V=zeros(1,sim);
s=0;
for i=1:sim
    V(i)=RecomMonte(P0,P1,R0,R1,beta,d,m,I0,n);
end
Recompensa=mean(V)
s=std(V);
I=Recompensa+tinv([0.01 0.99],(sim-1))*(s/sqrt(sim));
```

```
function V=RecomMonte(P0,P1,R0,R1,beta,d,m,I0,n)
% DATOS DE SALIDA
% V   : Aproximacion de esperanza de la recompensa total descontada de una
%       Simulacion.
V=0;
a=zeros(1,n);
b=zeros(1,length(I0)-n);
for t=0:74000 % Contador de periodos, estamos considerando 74000 periodos.
    % Formamos el vector M con los IPM de los estados de I0.
    M=zeros(1,length(I0));
    for k=1:length(I0)
          for j=1:26
            if (I0(k)==d(j))
                % Vector con los IPM de los estados de I0.
               M(k)=m(j);
            end
          end
    end
    % Ordenamos los estados de mayor a menor IPM y a los n primeros
    % correspondientes estados le tomamos la accion activa.
    p=zeros(1,length(I0));
    I=zeros(1,length(I0));
    for i=1:length(I0)
       p(i)=find(M==max(M), 1, 'last' );
       M(p(i))=min(M)-1;
        % I es el vector con los estados iniciales ordenados de
        % mayor a menor IPM.
       I(i)=I0(p(i));
```

```
    end
    % Calculamos las recompensas descontadas.
    for j=1:n
        a(j)=R1(I(j));
    end
    for j=(n+1):length(I0)
        b(j)=R0(I(j));
    end
    V=V+(sum(a)+sum(b))*(beta^t);
    % Generamos los estados a los que transitan los estados de I0.
    Q=zeros(1,26);
    r=0;
    s=0;
    for k=1:n
        Q(1)=P1(I(k),1);
        for i=2:26
        % Q vector con las probabilidades acumuladas.
        Q(i)=Q(i-1)+P1(I(k),i);
        end
        r=rand;
           if (0<= r)&&(r < Q(1))
              I(k)=1;
           else
               for i=2:26
                   if(Q(i-1)<= r)&&(r < Q(i))
                   I(k)=i;
                   end
               end
           end
    end
    for k=(n+1):length(I0)
        Q(1)=P0(I(k),1);
        for i=2:26
        Q(i)=Q(i-1)+P0(I(k),i);
        end
        s=rand;
           if (0<= s)&&(s < Q(1))
              I(k)=1;
           else
               for i=2:26
                   if(Q(i-1)<= s)&&(s < Q(i))
                   I(k)=i;
                   end
               end
           end
    end
    I0=I;
end
```

Anexo 2

Mostraremos la simulación de Montecarlo para aproximar el rendimiento de la política-índice de Whittle en el caso heterogéneo.

```
function [I,Recompensa]=RecompensaMedia2(P0,P1,R0,R1,beta,D,M,I0,n,sim)
% DATOS DE ENTRADA
% P0  : Matriz con la dinamica de estados bajo la accion pasiva
%       (ordenados por bloques fila de orden 26*26 para cada proyecto).
% P1  : Matriz con la dinamica de estados bajo la accion activa
%       (ordenados por bloques fila de orden 26*26 para cada proyecto).
% R0  : Vector de recompensas bajo la accion pasiva.
% R1  : Vector de recompensas bajo la accion activa.
% beta: Factor de descuento.
% D   : Vector que contiene los estados generados por el algoritmo AG
%      (ordenados de 26 en 26 para cada proyecto).
% M   : Vector que contiene los IPM generados por el algoritmo AG
%     : (ordenados de 26 en 26 para cada proyecto).
% I0  : Vector conteniendo los estados iniciales de la cartera
%       de proyectos I0=[20,21,4,5,7,13,8,9,26,2,12,6], si el numero total
%       de enlaces es mayor de 12, entonces tomamos copias de I0
% n   : Numero de enlaces a los cuales se aplica la accion activa.
% sim : Numero de simulaciones.
% DATOS DE SALIDA
% I          , Intervalo al 98% de confianza.
% Recompensa , Aproximacion de la recompensa total descontada media usando IPM.
V=zeros(sim,1);
s=0;
for i=1:sim
    V(i)=RecomMonte2(P0,P1,R0,R1,beta,D,M,I0,n);
end
Recompensa=mean(V)
s=std(V);
I=Recompensa+tinv([0.01 0.99],(sim-1))*(s/sqrt(sim));
```

```
function V=RecomMonte2(P0,P1,R0,R1,beta,D,M,I0,n)
% DATOS DE SALIDA
% V   , Aproximacion de la recompensa total descontada media usando IPM
%       en el caso Heterogeneo
%
V=0;
a=zeros(1,n);
b=zeros(1,length(I0)-n);

for t=0:74000  % contador de periodos, estamos considerando 74000 periodos.
    % Formamos el vector M con los IPM de los estados de I0.
    M1=zeros(1,length(I0));
    d=zeros(1,26);
    m=zeros(1,26);
    for k=1:length(I0)
        d=D((26*k-25):(26*k));
        m=M((26*k-25):(26*k));
        for j=1:26
            if (I0(k)==d(j))
                M1(k)=m(j);  % Vector con los IPM de los estados de I0.
            end
        end
```

```
    end
    % Ordenamos los estados de mayor a menor IPM y a los 2 primeros
    % correspondientes estados le tomamos la acción activa.
    p=zeros(1,length(I0));
    I=zeros(1,length(I0));
    for i=1:length(I0)
         p(i)=find(M1==max(M1), 1, 'last' );
         M1(p(i))=min(M1)-1;
        % I es el vector con los estados iniciales ordenados de mayor a menor IPM.
         I(i)=I0(p(i));
    end
    % Calculamos las recompensas descontadas.
    for j=1:n
        a(j)=R1(I(j));
    end
    for j=(n+1):length(I0)
        b(j)=R0(I(j));
    end
    V=V+(sum(a)+sum(b))*(beta^t);
    % Generamos los estados a los que transitan los estados de I0.
    Q=zeros(1,26);
    r=0;
    s=0;
    P=zeros(26,26);
    for k=1:n
        P=P1(:,[(26*k-25):(26*k)]);
        Q(1)=P(I(k),1);
        for i=2:26
        Q(i)=Q(i-1)+P(I(k),i);  % Q vector con las probabilidades acumuladas.
        end
        r=rand;
           if (0<= r)&&(r < Q(1))
              I(k)=1;
           else
               for i=2:26
                   if(Q(i-1)<= r)&&(r < Q(i))
                   I(k)=i;
                   end
               end
           end
    end
    R=zeros(26,26);
    for k=(n+1):length(I0)
        R=P0(:,[(26*k-25):(26*k)]);
        Q(1)=R(I(k),1);
        for i=2:26
        Q(i)=Q(i-1)+R(I(k),i);  % Q vector con las probabilidades acumuladas.
        end
        s=rand;
           if (0<= s)&&(s < Q(1))
              I(k)=1;
           else
               for i=2:26
                   if(Q(i-1)<= s)&&(s < Q(i))
                   I(k)=i;
                   end
               end
           end
    end
    I0=I;
end
```

Anexo 3

Mostraremos el cálculo de la cota de Whittle en el caso homogéneo.

```
function W=CotaWhittleHomo(P0,P1,R0,R1,m,beta,I0,n,M)
% DATOS DE ENTRADA
% P0  : Matriz con la dinamica de estados bajo la accion pasiva.
% P1  : Matriz con la dinamica de estados bajo la accion activa.
% R0  : Vector de recompensas bajo la accion pasiva.
% R1  : Vector de recompensas bajo la accion activa.
% m   : Vector con los IPM del Ouput del algoritmo AG.
% beta: Factor de descuento.
% I0  : Vector conteniendo los estados iniciales de la  cartera de proyectos
%       I0=[20,21,4,5,7,13,8,9,26,2,12,6] .
% n   : Numero de enlaces a los cuales se aplica la acción activa.
% M   : Numero total de enlaces.
% DATOS DE SALIDA
% W   : Cota de Whittle.
ak=min(m);
bk=max(m);
r=0.618;
% Puntos generados con el numero aureo.
xk=bk-r*(bk-ak);
yk=ak+r*(bk-ak);
% Evaluamos xk e yk
fx=FuncionLambdaHomo(P0,P1,R0,R1,beta,I0,n,M,xk);
fy=FuncionLambdaHomo(P0,P1,R0,R1,beta,I0,n,M,yk);

while abs(bk-ak)>= 0.01
      if (fx < fy)
         bk=yk;
         yk=xk;
         fy=fx;
         xk=bk-r*(bk-ak);
         fx=FuncionLambdaHomo(P0,P1,R0,R1,beta,I0,n,M,xk);
      else if (fx > fy)
         ak=xk;
         xk=yk;
         fx=fy;
         yk=ak+r*(bk-ak);
         fy=FuncionLambdaHomo(P0,P1,R0,R1,beta,I0,n,M,yk);
           end
      end
end
for k=1:length(m)
    if (xk <= m(k))&&(m(k) <= yk)
       W=FuncionLambdaHomo(P0,P1,R0,R1,beta,I0,n,M,m(k));
    end
end
```

```
function L=FuncionLambdaHomo(P0,P1,R0,R1,beta,I0,n,M,lambda)
% DATOS DE SALIDA
% L : valor de la funcion L(Lambda,i°).
s=0;
A1=zeros(26,26);
A2=zeros(26,26);
b1=zeros(26,1);
b2=zeros(26,1);
```

```
for k=1:n
    f=linspace(1/26,1/26,26);
    A1=beta*P1-eye(26);
    A2=beta*P0-eye(26);
    A=[A1;A2];
    b1=(linspace(lambda,lambda,26))'-(R1)';
    b2=-R0';
    b=[b1;b2];
    x=linprog(f,A,b);
    y=x(I0(k));
    s=s+y;
end
L=((lambda*M)/(1-beta))+s
```

Referencias

Bellman, R. (1957). E. 1957. dynamic programming. *Princeton UniversityPress. BellmanDynamic programming1957*, 151.

Bertsimas, D., y Niño-Mora, J. (1996). Conservation laws, extended polymatroids and multiarmed bandit problems; a polyhedral approach to indexable systems. *Mathematics of Operations Research*, *21*(2), 257–306.

Bertsimas, D., y Niño-Mora, J. (2000). Restless bandits, linear programming relaxations, and a primal-dual index heuristic. *Operations Research*, *48*(1), 80–90.

Cho, J., y Roy, S. (2004). Impact of search engines on page popularity. En *Proceedings of the 13th international conference on world wide web* (pp. 20–29).

Clark, J. B. (1902). *The distribution of wealth: a theory of wages, interest and profit*. Macmillan.

Gittins, J. (1989). Multi-armed bandit allocation indices. wiley-interscience series in systems and optimization.

Gittins, J., y Jones, D. (1974). *A dynamic allocation index for the sequential allocation of experiments, in (j. gani, et al, eds.) progress in statistics.* North Holland, Amsterdam.

Gittins, J. C. (1979). Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society. Series B (Methodological)*, 148–177.

Glazebrook, K., Niño-Mora, J., y Ansell, P. (2002). Index policies for a class of discounted restless bandits. *Advances in Applied Probability*, 754–774.

Huberman, B. A., y Wu, F. (2006, septiembre 13). *Determining most valuable ordering of items for presentation.* Google Patents. (US Patent App. 11/531,652)

Huberman, B. A., y Wu, F. (2008). The economics of attention: maximizing user value in information-rich environments. *Advances in Complex Systems*, *11*(04), 487–496.

Klimov, G. (1975). Time-sharing service systems. i. *Theory of Probability & Its Applications*, *19*(3), 532–551.

Niño-Mora, J. (2001). Restless bandits, partial conservation laws and indexability. *Advances in Applied Probability*, *33*(1), 76–98.

Niño-Mora, J. (2002). Dynamic allocation indices for restless projects and queueing admission control: a polyhedral approach. *Mathematical programming*, *93*(3), 361–413.

Niño-Mora, J. (2006a). Marginal productivity index policies for scheduling a multiclass delay-/loss-sensitive queue. *Queueing Systems*, *54*(4), 281–312.

Niño-Mora, J. (2006b). Restless bandit marginal productivity indices, diminishing returns, and optimal control of make-to-order/make-to-stock m/g/1 queues. *Mathematics of Operations Research*, *31*(1), 50–84.

Niño-Mora, J. (2007). Dynamic priority allocation via restless bandit marginal productivity indices. *Top*, *15*(2), 161–198.

Pandey, S., Roy, S., Olston, C., Cho, J., y Chakrabarti, S. (2005). Shuffling a stacked deck: the case for partially randomized ranking of search engine results. En *Proceedings of the 31st international conference on very large data bases* (pp. 781–792).

Papadimitriou, C. H., y Tsitsiklis, J. N. (1999). The complexity of optimal queuing network control. *Mathematics of Operations Research*, *24*(2), 293–305.

Puterman, M. L. (1994). Markov decision processes: Discrete dynamic stochastic programming. *New York, NY: John Wiley. doi*, *10*, 9780470316887.

Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, *58*(5), 527–535.

Servia-Rodríguez, S., Huberman, B. A., y Asur, S. (2014). Deciding what to display: maximizing the information value of social media. *arXiv preprint arXiv:1411.3214*.

Simon, H. A. (1971). Designing organizations for an information-rich world.

Stidham Jr, S. (2002). Analysis, design, and control of queueing systems. *Operations Research*, *50*(1), 197–216.

Varaiya, P., Walrand, J., y Buyukkoc, C. (1985). Extensions of the multiarmed bandit problem: the discounted case. *IEEE transactions on automatic control*, *30*(5), 426–439.

Weber, R. (1992). On the gittins index for multiarmed bandits. *The Annals of Applied Probability*, *2*(4), 1024–1033.

Weber, R., y Weiss, G. (1990). On an index policy for restless bandits. *Journal of Applied Probability*, 637–648.

Whittle, P. (1988). Restless bandits: Activity allocation in a changing world. *Journal of applied probability*, 287–298.

Whittle, P. (2002). Applied probability in great britain. *Operations Research*, *50*(1), 227–239.