



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

EL EFECTO DE LA RETROALIMENTACIÓN DE LOS CONSUMIDORES SOBRE LA
PLANIFICACIÓN DEL SURTIDO

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN GESTIÓN DE OPERACIONES
MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

OSCAR FELIPE CARRASCO HEINE

PROFESOR GUÍA:
DENIS ROLAND SAURÉ VALENZUELA

MIEMBROS DE LA COMISIÓN:
MARCELO OSVALDO OLIVARES ACUÑA
RENÉ ANTONIO CALDENTHEY MORALES

Este trabajo ha sido parcialmente financiado por CONICYT

SANTIAGO DE CHILE
2017

**RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE:** Ingeniero Civil Industrial y grado
de Magíster en Gestión de Operaciones.
POR: Oscar Felipe Carrasco Heine
FECHA: 06/07/2017
PROFESOR GUÍA: Denis Roland Sauré Valenzuela

EL EFECTO DE LA RETROALIMENTACIÓN DE LOS CONSUMIDORES SOBRE LA PLANIFICACIÓN DEL SURTIDO

Los principales antecedentes y la motivación del presente trabajo provienen de la teoría de *Assortment*, los modelos de elección discreta y los procesos de *Social Learning*. Se apunta a reunir parte de la vasta literatura en torno a estos temas, proponiendo un modelo matemático que incorpore el *feedback* de los consumidores en la elección del surtido de productos a ofrecer por parte de un vendedor.

El modelo propuesto incluye un comerciante minorista quien se enfrenta a una secuencia finita de clientes, los que tienen la opción de escoger alguno de los productos ofrecidos por el vendedor. En caso de comprar alguno, descubren su calidad y la reportan de manera pública, siendo esta información utilizada por futuros consumidores en su proceso de decisión mediante actualización bayesiana de sus creencias. Estos reportes a la vez permiten al *retailer* - quien sólo puede ofrecer una cantidad limitada de productos - optimizar su oferta, escogiendo aquel conjunto de alternativas que le entregue un mayor beneficio esperado.

La manera natural para abordar el problema de optimización de assortment enfrentado por el vendedor es mediante Programación Dinámica Estocástica, al tratarse de una situación en tiempo discreto donde período a período se debe tomar una decisión ante un determinado estado del sistema. Sin embargo, el número de estados posibles aumenta explosivamente en función de los parámetros considerados, volviéndose un problema virtualmente imposible de resolver en instancias realistas. Es por ello que se opta por Programación Dinámica Aproximada, proponiéndose 3 heurísticas que permiten obtener resultados en contextos en los cuales no es posible optimizar de manera exacta.

A modo de referencia, se compara el rendimiento de las heurísticas con el de una estrategia bajo la cual el vendedor no considera las evaluaciones hechas por los clientes, ni la incidencia de sus decisiones en períodos futuros. La diferencia en los beneficios llega a ser enorme, observándose en los experimentos realizados aumentos superiores a 300 % al utilizar alguno de los algoritmos sugeridos, en vez de obviar la información disponible. Esto no solo valida la calidad de los métodos heurísticos propuestos, sino que sobre todo ilustra la importancia de considerar el feedback de los consumidores.

*Si de noche lloras por el sol,
las lágrimas te impedirán ver las estrellas.*

Agradecimientos

En primer lugar me gustaría dar las gracias a Denis Sauré. Como profesor guía destaco su dedicación, tomándose siempre el tiempo de recibirme para comentar e involucrarse directamente en mi trabajo. Sin embargo, considero de igual o mayor importancia el hecho que siempre se mostrara abierto a conversar de temas ajenos a mi tesis, como mis dudas vocacionales, fútbol, o simplemente tirar la talla, además de encargarse de quitarme presión mientras atravesaba momentos difíciles en mi vida personal. Creo que esto último no lo haría cualquier profesor, y es una de las razones por las que se ha ganado mi más profunda admiración.

También quiero agradecer los aportes directos a mi tesis hechos por Ignacio Correa y Renzo Fuenzalida, quienes sin esperar nada a cambio hicieron contribuciones fundamentales a este trabajo, simplemente por comentarles mis problemas o pedirles ayuda.

Me gustaría destacar el rol que también tuvieron Felipe Balmaceda, Victor Bucarey, José Correa y Nicolás Inostroza - además de Denis - en mi afán por seguir un camino ligado a la academia y la docencia. Ellos han sido mis principales guías, fuentes de inspiración, motivadores y ejemplos a seguir.

Tampoco puedo dejar de incluir en este listado a mis padres Oscar Carrasco y Cecilia Heine, sin quienes nada de esto habría sido posible y que además me han aguantado todo estos años de estudio.

Además, quisiera destacar el papel fundamental de mis amigos a lo largo de este camino. Por diversos motivos he decidido no hacer una lista, pero creo que cada uno de ellos sabe el rol que ha tenido en mi vida y lo agradezco que siempre estaré por su cariño y compañía.

Finalmente, agradezco a CONICYT por el financiamiento otorgado para realizar este trabajo.

Tabla de contenido

1. Introducción	1
1.1. Motivación	1
1.2. Revisión Bibliográfica	2
2. Modelamiento	6
3. Métodos de Resolución	10
3.1. Oblivious	11
3.2. Greedy	11
3.3. Thompson Sampling	12
3.4. Algoritmo Mixto	12
4. Análisis de Resultados	15
4.1. Programación Dinámica Exacta	16
4.2. Programación Dinámica Aproximada	19
4.2.1. Prior opuesto a la realidad	19
4.2.2. Órdenes de magnitud de α	23
4.2.3. Incremento del número de alternativas	26
4.2.4. Sobre la cantidad de alternativas y los tiempos de resolución	28
5. Conclusiones	30
5.1. Modelo y métodos de resolución	30
5.2. Resultados	31
5.3. Trabajo en progreso y posibles extensiones	32
Bibliografía	35
A. Demostraciones Capítulo 3	39
B. Resultados	40
B.1. Programación Dinámica Exacta	40
B.1.1. Caso a)	40
B.1.2. Caso b)	42
B.1.3. Caso c)	44
B.2. Prior Opuesto a la Realidad	46
B.3. Órdenes de Magnitud de α	47
B.3.1. Caso a)	48
B.3.2. Caso b)	48

B.3.3. Caso c)	49
B.4. Incremento del número de alternativas	50
B.4.1. Órdenes de magnitud de α	50
B.4.2. Prior opuesto a la realidad	52
B.4.3. Parámetros aleatorios	53

Capítulo 1

Introducción

1.1. Motivación

Cualquier agente que ofrezca algún producto o servicio debe tomar una serie de decisiones, como por ejemplo precios, inversión en publicidad o el surtido a presentar, con el fin de influir en la decisión de sus potenciales clientes y así optimizar sus beneficios. Especialmente en el comercio en línea se dispone de información en tiempo real acerca de los consumidores: historial de navegación, compras anteriores, datos demográficos, evolución de las preferencias reflejada a través de evaluaciones cualitativas y cuantitativas hechas por los compradores, entre muchas otras posibilidades. Esto motiva la personalización de las ofertas para cada cliente al momento de su ingreso. Caber mencionar que esta personalización de la oferta no solo se restringe a los productos expuestos, sino que también se puede extender a otras posibilidades, como los anuncios publicitarios presentados por el sitio. En cualquiera de estos casos, se debe escoger el surtido (también llamado *assortment*) óptimo a presentar a cada individuo, con el objetivo de maximizar el ingreso esperado, maximizar la probabilidad de *click* sobre un anuncio, o cualquiera que sea el fin deseado.

Por diversas razones, suele deber considerarse un surtido limitado, ya sea por limitaciones físicas o razones estratégicas como no agobiar al cliente con un exceso de opciones. En muchos casos la decisión también incluye el orden en que se presentan los productos, especialmente en (pero no restringido a) el comercio en línea. Al ingresar por ejemplo una búsqueda en *Amazon.com*, uno se encuentra con un listado de varias páginas. En estos casos, no es raro considerar solo las primeras opciones y obviar las siguientes páginas.

Además de la importancia de las alternativas ofrecidas, y como ya fue mencionado anteriormente, se debe tener en cuenta que muchas personas evalúan su experiencia después de escoger una alternativa, lo cual es información disponible para potenciales futuros clientes. De hecho, para muchas personas estas evaluaciones (también llamadas *ratings*) son información

esencial al momento de tomar una decisión de compra, no solo de productos, sino también al momento de escoger un hotel o un lugar donde comer, por nombrar algunos ejemplos. En el caso particular de productos nuevos, hay dos factores de enorme relevancia antes de y durante su introducción al mercado: la expectativa generada por publicidad y especificaciones técnicas, que seguirán influyendo más allá de esta etapa inicial, así como las primeras evaluaciones registradas. Estas últimas pueden influir altamente la percepción de futuros compradores, haciendo que por ejemplo el producto deje de venderse, incluso tratándose de un artículo de buena calidad. Un caso interesante que ilustra el contraste entre la expectativa y la experiencia real de los consumidores con el producto es el del videojuego *No Man's Sky*, de la compañía de desarrollo de videojuegos *Hello Games*. Este juego de aventura y exploración espacial fue esperado con ansias por muchos, jugando un rol fundamental en ello la campaña publicitaria realizada para promocionarlo. Sin embargo, la percepción cambió rápidamente al momento de su lanzamiento: desde el punto de vista de los primeros compradores, así como de revistas y sitios web especializados, *No Man's Sky* se encontraba muy por debajo de lo prometido, y en consideración de la gran expectativa generada, puede ser calificado como un enorme fracaso¹. El juego original para la consola *Play Station 4* solo cuenta con una evaluación de 2,5 estrellas de 5 posibles en Amazon, habiendo escogido la mayoría de los clientes un rating de 1.

El presente trabajo busca entender el proceso de aprendizaje de los individuos en este tipo de situaciones. ¿Cómo pueden las compañías aprovechar toda esta información? ¿Cómo se benefician los clientes?

1.2. Revisión Bibliográfica

Uno de los fundamentos de esta tesis es la teoría de elección discreta. En este ámbito hay importantes antecedentes, como el trabajo sobre utilidad aleatoria desarrollado en [McFadden, 1973] (basado entre otros en los axiomas de selección introducidos por [Luce, 1959]) que lleva al modelo de elección *Logit*, ampliamente utilizado en disciplinas tan diversas como Investigación Operativa, Marketing y Transporte, entre otras. Luego surgen variadas derivaciones, como el modelo anidado estudiado en [Williams, 1977], y diversas extensiones de éste como las propuestas en [McFadden, 1980]. [McFadden y Train, 2000] trabaja con una mezcla de modelos Logit con diferentes parámetros, conocida como *Mixed Multinomial Logit*. Con una cantidad suficiente de componentes, *Mixed Multinomial Logit* permite aproximar cualquier modelo de elección discreta que se desprenda de una distribución sobre las preferencias

¹El siguiente artículo del sitio *PC Gamer* presenta un interesante análisis de este caso: <http://www.pcgamer.com/the-anatomy-of-hype-how-no-mans-sky-became-the-best-and-worst-game-ever/>

individuales. A pesar de que el modelo de elección utilizado en el presente trabajo presenta algunas diferencias conceptuales con un modelo Logit tradicional, sí aprovecha su estructura y especialmente la fórmula de la probabilidad de selección de cada alternativa.

Otros antecedentes importantes son los relacionados con Assortment. Posiblemente uno de los mayores desafíos en esta disciplina es la complejidad computacional asociada a escoger el surtido óptimo, la cual se encuentra estrechamente ligada al modelo de elección considerado. En [Talluri y Van Ryzin, 2006] por ejemplo se presenta un surtido sencillo para un modelo Logit básico, el cual consiste en ordenar las alternativas de acuerdo al margen que reportan al vendedor. Como se muestra en [Rusmevichientong *et al.*, 2014], optimizar el assortment para un modelo Mixed Multinomial Logit es NP-completo, pero existen maneras eficientes de encontrar resultados aproximados. [Davis *et al.*, 2014] indica que también suele ser difícil optimizar para modelos Logit anidados, aunque como se puede ver en el mismo artículo y en [Li *et al.*, 2015], no siempre es el caso. El problema se complejiza de manera considerable al incorporar inventario, como también es abordado en [van Ryzin y Mahajan, 1999], que presenta uno de los primeros modelos de assortment con *trade-off* entre costos de inventario y variedad de productos. Los mismos autores incorporan quiebres de *stock* y sustitución por parte de los consumidores en [Mahajan y van Ryzin, 2001]. Desde entonces se ha desarrollado una gran variedad de modelos de elección que consideran tanto el problema de Assortment como las limitaciones de inventario. En esta línea se encuentran varios ejemplos, como el modelo de sustitución de la demanda de [Smith y Agrawal, 2000], el *ranking* de preferencias estudiado en [Honhon *et al.*, 2010] y la incorporación de modelos Logit en trabajos como [Liu y van Ryzin, 2008] y [Topaloglu, 2013], entre otros. A diferencia de varios de los casos recién mencionados, en los cuales el inventario inicial es la principal decisión, este trabajo incluye una propuesta de extensión en que el stock inicial es considerado como un dato exógeno con el fin de enfocar el análisis en la optimización del surtido. [Golrezaei *et al.*, 2014] igualmente aborda el problema de Assortment bajo este supuesto, y presenta una función de penalización por escasez de productos que es generalizada y explotada en la presente tesis. [Kallus y Udell, 2016] también presenta un modelo similar al estudiado en este informe: una secuencia finita de clientes escoge una de las alternativas que les son ofrecidas de acuerdo a un modelo Logit, mientras que el vendedor debe escoger un surtido con restricción de capacidad para cada ocasión de compra. Una de las características exógenas de cada producto es el margen que otorga su venta, por lo que no juega un rol la selección óptima de precios.

Por otro lado, la incertidumbre acerca de las preferencias de los clientes da lugar a enfoques estilo *multi-armed bandit* en problemas de assortment: es necesario explorar las alternativas disponibles con el fin de adquirir información suficiente para maximizar los beneficios. [Rusmevichientong *et al.*, 2010] y [Sauré y Zeevi, 2013] son algunos trabajos que estudian esto,

proponiendo algoritmos y entregando cotas superiores para el arrepentimiento, mientras que en [Caro y Gallien, 2007] se estudia el problema de optimización dinámica del surtido bajo aprendizaje en un contexto relacionado con aquel considerado por los dos artículos nombrados previamente, con la diferencia de que la demanda es considerada exógena e independiente de las alternativas ofrecidas.

Esta tesis busca incorporar una nueva dimensión al problema de Assortment: la incertidumbre en cuanto a la calidad de los productos, y cómo los consumidores lidian con ella a través de ratings. Las evaluaciones y críticas *online* de productos y servicios existen hace bastantes años, motivando estudios como el caso de [Chatterjee, 2001]. Evidentemente, la masificación de internet ha incrementado su importancia en los últimos tiempos, lo que ha sido comprobado por ejemplo por [Luca, 2011], donde se muestra una correlación positiva entre los ingresos obtenidos y el rating promedio de los productos. Sin embargo, los primeros antecedentes de esta línea de estudios se remontan a la literatura económica sobre *Social Learning* de comienzos de los años '90. [Banerjee, 1992] y [Bikhchandani *et al.*, 1992] analizan situaciones en que individuos deciden secuencialmente en base a información privada, pero también observando el comportamiento de los demás, lo cual puede llevar a resultados ineficientes. [Smith y Sørensen, 2000] indica condiciones bajo las cuales es posible revertir esta ineficiencia. Los mismos autores implementan en [Smith y Sørensen, 2013] además la idea de no utilizar toda la información disponible, sino realizar un muestreo aleatorio. En un contexto de redes sociales, [Acemoglu *et al.*, 2011] identifica condiciones para que los individuos conozcan el verdadero estado del mundo, dadas señales privadas y la observación de las acciones realizadas por sus pares. En [Acemoglu *et al.*, 2014] se estudia un modelo similar en el cual los agentes además pueden postergar su decisión para obtener más información de los demás, mientras que publicaciones como [Bose *et al.*, 2006] incorporan adicionalmente el problema de *pricing*. Estos conceptos también se han ido incorporando a la literatura de Marketing e Investigación Operativa, estudiándose el efecto de ratings y *reviews*, como por ejemplo intentando responder a dudas como por qué la gente evalúa productos, y la utilidad que estas opiniones tienen para los demás consumidores ([Moe y Schweidel, 2012] y [Pan y Zhang, 2011]), además del ya mencionado artículo de [Luca, 2011] relativo al impacto en los ingresos. [Talwar *et al.*, 2007] estudia la formación de opinión por parte de los clientes, y cómo se determina su evaluación del producto o servicio. Además, los consumidores no solo se quedan con las opiniones de quienes han experimentado los productos, sino que observan las ventas como información adicional relevante, lo cual es estudiado entre otros en [Ye *et al.*, 2013] y [Tucker y Zhang, 2011]. [Cheung *et al.*, 2014] descubre un impacto significativo de la opinión de consumidores anteriores, pero una incidencia aún mayor de la observación de sus decisiones de compra.

Ante toda esta evidencia, diversos autores se han dedicado a modelar estos escenarios. Se busca incorporar por ejemplo el hecho de que no todos los clientes evalúan los productos que adquieren, por lo que cabe preguntarse bajo qué condiciones lo hacen. Los modelos de [Dellarocas, 2001] y [Hu *et al.*, 2006] abordan esto de manera similar: los compradores solo reportan cuando su utilidad fue suficientemente alta o bastante baja. Esta evaluación, como es el caso de la mayoría de los modelos en esta línea de investigación, consiste en un registro de la utilidad percibida, de manera que está sujeta a preferencias personales y no es una medida del todo objetiva para futuros clientes. Por lo mismo, quizás la pregunta más fundamental es cómo los clientes utilizan las evaluaciones para tomar su decisión de compra. Para ello, muchos autores optan por dejar de lado la actualización bayesiana de las creencias, dada la complejidad de ésta en muchos casos. Más bien, se propone que los individuos sigan reglas de decisión intuitivas, lo cual ya ha sido considerado por ejemplo en la literatura de Economía, como es el caso de [Ellison y Fudenberg, 1993]. En [Crapis *et al.*, 2015], un monopolista debe escoger el precio de su producto, y los consumidores estiman la calidad basándose directamente en la proporción de evaluaciones positivas y negativas. En el caso de aquel modelo, el hecho que los consumidores aprendan la calidad real del producto genera mayores ganancias para el vendedor. De manera similar, en [Papanastasiou y Savva, 2016] se estudia el problema de pricing de un monopolista, donde la ocurrencia de Social Learning puede incrementar sus ingresos ex-ante. En [Besbes y Scarsini, 2015], el foco está en la actualización de la creencia de los consumidores sobre la calidad. Esto no se realiza de manera bayesiana, y se propone algunas maneras de aproximar la calidad en base a la información disponible. En [Li y Hitt, 2008] se propone un modelo similar, en el cual los consumidores compran en caso de estimar que tendrán utilidad positiva, y estiman la calidad mediante las observaciones hechas hasta el momento. [Papanastasiou *et al.*, 2013] considera el mismo modelo de utilidad que [Li y Hitt, 2008], y estudia diferentes maneras de actualizar las creencias. En [Crapis *et al.*, 2015] y [Besbes y Scarsini, 2015], se tiene en cuenta el sesgo que genera el hecho que el producto sea adquirido por quienes ya tenían una alta valoración a priori, lo cual puede o no ser considerado por los clientes. Al igual que ambos artículos anteriores, en [Ifrach *et al.*, 2014] la utilidad obtenida por el consumidor incluye un término que representa su preferencia idiosincrática por el producto, generándose sesgos en los reportes. En aquel *paper* en particular, la evaluación solo consiste en 3 posibilidades: positiva, negativa u omitida en caso de no comprar. Mediante la presente tesis se desea contribuir incorporando la teoría de Assortment a este tipo de modelos, considerando un mercado con más de un producto, en el cual un vendedor debe escoger el surtido a ofrecer a cada cliente, quienes reportan su experiencia con la alternativa escogida.

Capítulo 2

Modelamiento

Se considera un mercado compuesto por una firma y una cantidad finita de consumidores indexados por su período de llegada $t \in \{1, 2, \dots, \tau\}$. En cada período, la firma ofrece un conjunto de productos X_t dentro de un conjunto I de posibilidades (es decir, $X_t \subset I$), tal que $|X_t| \leq C$. Por lo tanto, C consiste en la cantidad máxima de alternativas que pueden ser ofrecidas simultáneamente. Cada consumidor debe escoger una (o ninguna) de las alternativas a las cuales se enfrenta.

Cada producto $i \in X_t$ se caracteriza por una distribución de probabilidad categórica a partir de la cual se determina su calidad $q_{it} \in \{1, \dots, Q\}$. Ésta varía período a período y solamente es conocida después de la compra, en caso de adquirirse el producto en cuestión. Esta alteración puede deberse a variaciones de un mismo producto, o diferencias en cómo los individuos experimentan o perciben la calidad.

Se definen entonces $p_{ij} := \mathbb{P}(q_{it} = j)$ y $p_i := (p_{ij})_{j=1}^Q$. Este vector de probabilidades se genera aleatoriamente al comienzo del horizonte de evaluación, de manera que tanto los clientes como el vendedor desconocen los valores exactos de sus elementos. Sin embargo, es de conocimiento general que

$$p_i \sim \text{Dir}(\alpha_{i1}, \dots, \alpha_{iQ})$$

¿Por qué una distribución *Dirichlet*? Como fue mencionado anteriormente, recién después de comprar un producto se conoce el valor real de su calidad. Esta información es revelada a la firma y a los futuros clientes, de manera que éstos pueden actualizar su creencia en torno a la distribución de probabilidad de la calidad. La distribución Dirichlet permite una actualización bayesiana sencilla de esta creencia: Sea N_{ij}^t la cantidad de productos i de calidad j que han sido adquiridos hasta el momento en que t debe tomar su decisión de compra (de manera que N^t es la matriz que recopila toda esta información). Entonces, la distribución

condicional observada por t es

$$p_i|N^t \sim \text{Dir}(\alpha_{i1} + N_{i1}^t, \dots, \alpha_{iQ} + N_{iQ}^t)$$

Desde acá en adelante, cuando se escriba α , se estará aludiendo a la matriz de $|I|$ filas y Q columnas cuya i -ésima fila corresponde al vector $(\alpha_{i1}, \dots, \alpha_{iQ})$. Estos parámetros pueden ser interpretados de la siguiente manera: entre mayor sea α_{ij} respecto a los demás valores en la misma fila, más probable es que la distribución p_i resultante tenga un mayor peso para la calidad j . En particular, si $\alpha_{ij} > 0$ y todos los demás elementos de $(\alpha_{i1}, \dots, \alpha_{iQ})$ son iguales a 0, entonces la distribución p_i resultante siempre será $p_{ij} = 1$, y $p_{ik} = 0 \quad \forall k \neq j$. Por otro lado, entre menores sean los elementos de esta matriz, mayor será la varianza.

Al adquirir el producto i , el consumidor t percibe una utilidad

$$U_{it} = f(q_{it}) + \varepsilon_{it}$$

donde $f(q_{it}) = \beta q_{it} + \gamma$. Algunas observaciones:

- ε_{it} representa la valoración intrínseca de la alternativa i por parte de t .
- Si t decide no adquirir ninguno de los productos disponibles, obtiene una utilidad $U_{0t} = \varepsilon_{0t}$.
- El objetivo de aplicar la función lineal f a la calidad es escalar la utilidad en relación a la opción de no comprar.

Como t desconoce el valor exacto de la calidad, estima su utilidad por adquirir el producto i de la siguiente manera:

$$\hat{U}_{it} = f(\hat{q}_{it}) + \varepsilon_{it}$$

donde $\hat{q}_{it} = \mathbb{E}(q_{it}|N^t)$. Gracias a la distribución Dirichlet, este cálculo es bastante directo:

$$\begin{aligned} \mathbb{P}(q_{it} = j|N^t) &= \mathbb{E}(p_{ij}|N^t) = \frac{N_{ij}^t + \alpha_{ij}}{\sum_{k=1}^Q N_{ik}^t + \alpha_{ik}} \\ \Rightarrow \hat{q}_{it} &= \mathbb{E}(q_{it}|N^t) = \sum_{j=1}^Q j \cdot \frac{N_{ij}^t + \alpha_{ij}}{\sum_{k=1}^Q N_{ik}^t + \alpha_{ik}} \end{aligned}$$

Por lo tanto, t escoge aquella alternativa que maximice esta estimación de la utilidad.

Un aspecto fundamental es que no todos los individuos toman la misma decisión enfrentados a una determinada situación. Para incluir esta heterogeneidad en el modelo, se considerará que los consumidores tienen diferentes valoraciones intrínsecas de los productos,

lo cual será representado mediante una distribución continua de los parámetros ε_{it} . En términos generales, se les podría asignar una distribución F_ε . Sin embargo, con el único fin de simplificar los cálculos y obtener un resultado analítico para la probabilidad de elección de una alternativa, se considerará que $\varepsilon_{it} \sim Gumbel(0, 1)$, es decir $F_\varepsilon(x) = e^{-e^{-x}}$. Esto conlleva que la selección de alternativas por parte de los clientes corresponda a un modelo Logit, de manera que

$$\mathbb{P}(t \text{ compra } i \mid N^t) = \frac{e^{f(\hat{q}_{it})}}{1 + \sum_{j \in X_t} e^{f(\hat{q}_{jt})}}$$

Finalmente, se debe considerar que la venta de i genera un ingreso neto r_i para el retailer, de manera que su ingreso total es igual a

$$\sum_{t=1}^{\tau} \sum_{i \in X_t} r_i \cdot \mathbf{1}\{t \text{ escogió } i\}$$

Período a período, el vendedor debe escoger el surtido que maximice su ingreso esperado. Si realiza esta optimización de manera estratégica no solo debe considerar los ingresos del período presente, sino también los potenciales beneficios futuros asociados a una decisión actual. Por lo tanto, con el fin de maximizar su ingreso esperado de manera dinámica, el retailer enfrenta la siguiente ecuación de *Bellman*:

$$V_t(N^t) = \max_{X_t \subset I, |X_t| \leq C} \sum_{i \in X_t} r_i \cdot \mathcal{Q}_i^t + \sum_{i \in X_t} \sum_{j=1}^Q \mathcal{Q}_i^t \cdot \mathcal{P}_{ij}^t \cdot V_{t+1}(N^t + e_{i,j}) + \left(1 - \sum_{i \in X_t} \mathcal{Q}_i^t\right) \cdot V_{t+1}(N^t) \quad (2.1)$$

donde

- $V_{\tau+1}(N^{\tau+1}) = 0 \quad \forall N^{\tau+1}$
- $\mathcal{Q}_i^t = \mathbb{P}(t \text{ escoge } i \mid N^t) = \frac{e^{f(\hat{q}_{it})}}{1 + \sum_{j \in X_t} e^{f(\hat{q}_{jt})}}$
- $\mathcal{P}_{ij}^t = \mathbb{P}(i \text{ posee calidad } j \mid N^t) = \frac{N_{ij}^t + \alpha_{ij}}{\sum_{k=1}^Q N_{ik}^t + \alpha_{ik}}$
- $e_{i,j}$: matriz canónica

$$(e_{i,j})_{k,l} = \begin{cases} 1 & \text{si } k = i, l = j \\ 0 & \text{si no} \end{cases}$$

El primer término del lado derecho de la ecuación 2.1, $\sum_{i \in X_t} r_i \cdot \mathcal{Q}_i^t$, representa el beneficio esperado inmediato. En caso que el retailer enfrentase el problema de manera miope, solo se preocuparía de maximizar este valor. La expresión $\sum_{i \in X_t} \sum_{j=1}^Q \mathcal{Q}_i^t \cdot \mathcal{P}_{ij}^t \cdot V_{t+1}(N^t + e_{i,j})$ por otro lado representa los beneficios futuros esperados en caso que se venda alguno de los

productos del surtido, mientras que el último término indica los beneficios futuros esperados en caso que t opte por no comprar.

Capítulo 3

Métodos de Resolución

A pesar de que en teoría la inducción inversa permite resolver el problema del retailer de manera exacta, la cantidad de escenarios posibles crece rápidamente con el número de productos, períodos y niveles de calidad considerados. Por ejemplo, incluso en un caso básico con 2 productos y $Q = 3$, el número de estados (valores de N^t) posibles al cabo de 18 períodos es igual a 100,947. En una instancia un tanto más realista (pero aún así más bien pequeña), con 12 productos, $Q = 5$ y $\tau = 100$, este número incluso es superior a 10^{44} . No solo se debería optimizar para cada uno de estos casos, sino también para todos los estados de todos los períodos anteriores. Esta dificultad evidentemente no es una característica exclusiva de este modelo, sino que es transversal a una amplia variedad de problemas de Programación Dinámica y es conocida como la maldición de la dimensionalidad: la cantidad de estados del sistema, el número de decisiones posibles, así como el total de valores posibles de las variables aleatorias involucradas pueden ser enormes, tornando virtualmente imposible la optimización exacta. A raíz de ello se ha desarrollado la disciplina de la Programación Dinámica Aproximada, que mediante aproximaciones busca encontrar soluciones factibles para estos casos, en tiempos razonables. Bajo ciertas condiciones es posible asegurar cercanía de estas estimaciones con el óptimo, aunque no siempre es el caso.

Existe una amplia variedad de métodos, así como de herramientas utilizadas en Programación Dinámica Aproximada, como la simulación de los estados futuros, la simplificación de la función objetivo o la aproximación de la función de valor de la ecuación de Bellman. A continuación se propone precisamente algunos métodos heurísticos con el objetivo de encontrar buenas aproximaciones y tomar decisiones en tiempo real, en el contexto del problema de Assortment del retailer.

3.1. Oblivious

Con el objetivo de validar la importancia de considerar las evaluaciones de los consumidores, se implementa una heurística llamada *Oblivious*, bajo la cual el retailer ignora los registros de ratings de los productos. Sin embargo, estos siguen existiendo y los clientes siguen haciendo uso de esta información. Formalmente, bajo una estrategia *Oblivious* el vendedor estima la probabilidad de que el producto i posea calidad j mediante

$$\frac{\alpha_{ij}}{\sum_{k=1}^Q \alpha_{ik}}$$

para cualquier período, obviando las evaluaciones hechas. Por lo tanto, maximizará la siguiente función objetivo:

$$\sum_{i \in X_t} r_i \cdot \frac{e^{f(q_{it}^O)}}{1 + \sum_{j \in X_t} e^{f(q_{jt}^O)}}$$

donde

$$q_{it}^O = \sum_{j=1}^Q j \cdot \frac{\alpha_{ij}}{\sum_{k=1}^Q \alpha_{ik}}$$

Esta situación puede ser interpretada como un contexto dentro del cual el retailer no se preocupa de poner un canal formal para recopilar feedback a disposición de sus clientes, pero donde estos de todas maneras hacen uso de otros medios para compartir sus experiencias, como por ejemplo algún foro en línea.

3.2. Greedy

Para un período t se resuelve

$$\max_{X_t \subset I, |X_t| \leq C} E_G(X_t, N^t)$$

donde

$$E_G(X_t, N^t) = \sum_{i \in X_t} r_i \cdot Q_i^t$$

Como ya había sido mencionado, esto corresponde exactamente al primer término de la ecuación de Bellman. Esto quiere decir que el método *Greedy* solo optimiza los beneficios del período actual, sin considerar las consecuencias que esta decisión tendrá en períodos futuros.

3.3. Thompson Sampling

Mientras Greedy es una heurística enfocada en explotar la información disponible, el objetivo de *Thompson Sampling* es la exploración. Se trata de uno de los primeros algoritmos desarrollados para enfrentar el problema del *multi-armed bandit*, utilizado precisamente para ejemplificar el clásico dilema de explotación contra exploración. Propuesto originalmente en [Thompson, 1933], sugiere enfrentar mediante simulación un problema de optimización bajo incertidumbre por etapas en el cual se desconoce la distribución real de los beneficios. Para ello, se cuenta con un *prior* y la información observada a lo largo de las etapas anteriores.

En el caso particular del modelo estudiado en esta tesis, para un determinado período t , se simula un vector de probabilidades π_i^t para cada producto según una distribución $Dir(\alpha_{i1} + N_{i1}^t, \dots, \alpha_{iQ} + N_{iQ}^t)$. Esto permite hacer una estimación alternativa $q_{it}^T = \sum_{j=1}^Q j \cdot \pi_{ij}^t$ de la calidad. Luego, se busca resolver

$$\max_{X_t \subset I, |X_t| \leq C} E_T(X_t, \pi^t)$$

donde

$$E_T(X_t, \pi^t) = \sum_{i \in X_t} r_i \cdot \frac{e^{f(q_{it}^T)}}{1 + \sum_{j \in X_t} e^{f(q_{jt}^T)}}$$

La simulación de la distribución de la calidad permite por ejemplo exhibir productos que a pesar de poseer un mal prior, podrían tener una alta calidad real. De esta manera, se sacrifican beneficios en el presente con el fin de obtener más información y eventualmente obtener mayores ganancias en el largo plazo.

3.4. Algoritmo Mixto

Como fue mencionado en la revisión bibliográfica, en [Golrezaei *et al.*, 2014] se estudia el problema de Assortment con inventario. Se busca algoritmos mediante los cuales sea posible actualizar el surtido en tiempo real, considerando que los intervalos temporales entre clientes sucesivos pueden consistir en fracciones de segundo. En particular, se propone una heurística llamada *Inventory Balancing with a Penalty Function*. A grandes rasgos ésta pareciera ser muy similar a una política Greedy, en la cual simplemente se busca maximizar la suma de los beneficios entregados por la venta de cada producto, ponderados por sus respectivas probabilidades de compra. Sin embargo, también se multiplica cada uno de estos beneficios por un factor de descuento que considera una reducción en las ganancias en la medida que cae el stock disponible del producto correspondiente. Para ello, se considera una función Ψ

con las siguientes características:

- (i) $\Psi : [0, 1] \rightarrow [0, 1]$
- (ii) Ψ es creciente.
- (iii) $\Psi(0) = 0$
- (iv) $\Psi(1) = 1$

Al evaluar esta función en la razón entre el inventario disponible y el inventario inicial de cada producto, se obtiene el factor de descuento deseado. De esta manera, en la medida que quedan menos unidades disponibles de cierto artículo, se considera cada vez menos beneficiosa su venta, previniendo así quiebres de stock.

Un ejemplo propuesto en el mismo artículo es la función $\Psi(x) = \frac{1-e^{-x}}{1-e^{-1}}$. En esta tesis se propone la siguiente generalización:

$$\Psi_\lambda(x) = \begin{cases} \frac{1-e^{\lambda x}}{1-e^\lambda} & \text{si } \lambda \neq 0 \\ x & \text{si } \lambda = 0 \end{cases}$$

Esta función sigue cumpliendo las condiciones (i), (ii), (iii) y (iv). Además, si $\lambda = -1$ se obtiene la función original de [Golrezaei *et al.*, 2014].

Se propone enfrentar el problema de Assortment mediante la siguiente función objetivo:

$$\max_{X_t \subset I, |X_t| \leq C} \Psi_\lambda(t/\tau) \cdot E_G(X_t, N^t) + (1 - \Psi_\lambda(t/\tau)) \cdot E_T(X_t, \pi^t)$$

Este algoritmo mixto pretende rescatar las ventajas de las dos heurísticas anteriores. Para $x \in [0, 1]$, $\Psi_\lambda(x)$ es una función estrictamente creciente tal que $\Psi_\lambda(0) = 0$ y $\Psi_\lambda(1) = 1$. Por lo tanto, la función objetivo del algoritmo mixto es una combinación convexa de las funciones objetivo de los métodos anteriores, otorgándole un mayor peso a *Thompson Sampling* al comienzo del horizonte de evaluación, para luego ir asignándole cada vez mayor importancia a *Greedy*. Tal como fue mencionado anteriormente, *Thompson Sampling* es una heurística exploratoria, por lo que se desea aprovecharla durante el comienzo del horizonte de evaluación con el fin de conocer mejor la distribución de la calidad de los productos. Por otro lado, el objetivo del método *Greedy* es explotar la información ya disponible, lo cual resulta cada vez más conveniente en la medida que se recopila esta información. En particular, dado que $V_{\tau+1}(N^{\tau+1}) = 0 \quad \forall N^{\tau+1}$, *Greedy* es la estrategia óptima en el último período.

Como se puede ver en la figura 3.1, cuando $\lambda < 0$, $\Psi_\lambda(x)$ es una función cóncava, mientras que si $\lambda > 0$ entonces $\Psi_\lambda(x)$ es convexa. Este parámetro permite ajustar a qué ritmo ocurre la

transición de un método al otro. Si por ejemplo $\lambda > 0$, entonces Thompson Sampling tendrá una mayor ponderación durante la mayoría del horizonte temporal, aunque hacia los últimos períodos se tenderá rápidamente a privilegiar Greedy como función objetivo.

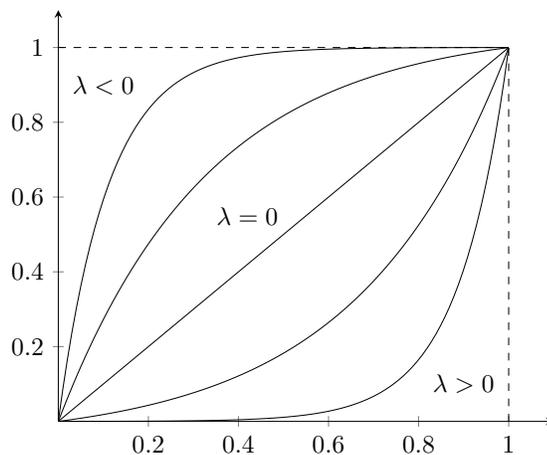


Figura 3.1: Efecto de λ sobre la función Ψ_λ

Todas las propiedades de Ψ_λ mencionadas son demostradas en la sección A de Apéndices.

Capítulo 4

Análisis de Resultados

El objetivo del presente capítulo es exponer y analizar los resultados obtenidos al resolver el problema de Assortment del vendedor mediante los métodos explicados en la sección anterior. A modo de referencia, se comienza resolviendo el problema de manera exacta para instancias de menor tamaño, considerando secuencias de hasta 18 clientes. A continuación se compara el rendimiento de las heurísticas en el caso de haber varios cientos de clientes, para así tener una idea de qué tan bien funcionan estas técnicas en el mediano y largo plazo. A lo largo de gran parte de este análisis solo se considerará 2 productos - de los cuales solamente se puede mostrar uno - y 3 niveles de calidad, tanto para efectos didácticos como de complejidad computacional. En la sección final sin embargo se experimentará en un ambiente menos controlado, con una mayor cantidad de productos y generación aleatoria de parámetros, entre otras variaciones.

También se debe escoger valores adecuados para los parámetros β y γ que definen la función f . Cabe recordar que el objetivo de esta función es escalar la utilidad con respecto a la opción de no comprar. Se escoge estos valores de manera de crear un intervalo simétrico centrado en 0. De esta manera - obviando las preferencias idiosincráticas de los individuos - un producto de calidad 1 sería peor que la alternativa externa, un producto de calidad esperada igual a 3 sería comprado, y un producto de calidad 2 generaría indiferencia.

Es deseable que la heterogeneidad de los clientes tenga un impacto en sus decisiones de compra, pero que el foco siga siendo la estimación de la calidad de los productos en base a las evaluaciones registradas. Sea $[-a, a]$ el intervalo simétrico en cuestión, de manera que

- $f(1) = -a$
- $f(2) = 0$
- $f(3) = a$

Como medida de la relevancia de ε_{it} en la toma de decisión, se calcula $P(-a \leq \varepsilon_{it} \leq a)$. Si

el intervalo es demasiado grande, esta probabilidad es cercana a 1, teniendo las preferencias idiosincráticas un impacto demasiado bajo en la toma de decisión. Por otro lado, si el intervalo es muy pequeño y consecuentemente esta probabilidad es muy baja, las evaluaciones se volverán irrelevantes y las acciones de los consumidores serán determinadas exclusivamente por ε_{it} . Se opta por $\beta = 3$ y $\gamma = -6$ de manera que $a = 3$ y $\mathbb{P}(-3 \leq \varepsilon_{it} \leq 3) = 0,95$.

Finalmente se debe asignar un valor a λ para definir la función Ψ_λ de la estrategia Mixta. Como fue explicado en 3.4, este parámetro determina de qué manera evoluciona la ponderación entre Thompson Sampling y Greedy. En este caso se optará por privilegiar ampliamente el carácter exploratorio de Thompson Sampling, y recién aprovechar repentinamente las virtudes de Greedy en la medida que se acerca el último período. Si $\lambda = 1$, entonces Ψ_λ es convexa, obteniéndose el efecto deseado.

4.1. Programación Dinámica Exacta

Como primer paso, se procede a resolver el problema de manera exacta bajo diferentes escenarios, para luego comparar el rendimiento de las heurísticas, incluyendo *Oblivious* para destacar la importancia de considerar las evaluaciones de los consumidores. Se considera los siguientes 9 valores de τ : 2, 4, 6, \dots , 16, y 18. Para cada uno de ellos, se considera 3 escenarios donde la única diferencia es el orden de magnitud de los valores de la matriz α . ¿En qué incide esto? Cabe recordar que

$$\mathbb{P}(q_{it} = j | N^t) = \mathbb{E}(p_{ij} | N^t) = \frac{N_{ij}^t + \alpha_{ij}}{\sum_{k=1}^Q N_{ik}^t + \alpha_{ik}}$$

En caso que los valores de α sean extremadamente grandes, las evaluaciones no tendrán mayor incidencia en la estimación de la calidad, o al menos no mientras no se acumule una cantidad importante de observaciones. Por el contrario, pequeños valores de los elementos de α implican que incluso las primeras evaluaciones ya tendrán una gran importancia en la actualización de la creencia de los futuros consumidores. De manera arbitraria, se escogen los siguientes casos:

$$\begin{aligned} \text{a) } \alpha &= \begin{pmatrix} 3 & 2 & 4 \\ 2 & 3 & 1 \end{pmatrix} \\ \text{b) } \alpha &= \begin{pmatrix} 0,3 & 0,2 & 0,4 \\ 0,2 & 0,3 & 0,1 \end{pmatrix} \end{aligned}$$

$$c) \alpha = \begin{pmatrix} 0,03 & 0,02 & 0,04 \\ 0,02 & 0,03 & 0,01 \end{pmatrix}$$

Para cada una de estas situaciones, se considera ingresos $r_1 = 0,48$ y $r_2 = 0,51$. Al ser valores similares, no juegan un rol demasiado gravitante en la elección por parte del vendedor.

Cada uno de los métodos propuestos es implementado mediante *Matlab*, obteniéndose los resultados detallados en la sección B.1 de Apéndices.

- **Caso a):** Debido a la escasa influencia de los primeros ratings, la estrategia Oblivious arroja resultados bastante cercanos al óptimo, entregando beneficios similares a los obtenidos mediante Greedy. Incluso la diferencia con el resultado obtenido a través de optimización exacta no es significativa en varios de los escenarios considerados. En cambio, las estrategias Mixta y especialmente Thompson Sampling presentan un rendimiento considerablemente inferior. En la figura 4.1 se puede apreciar un caso particular de estos resultados, considerando 4 períodos.

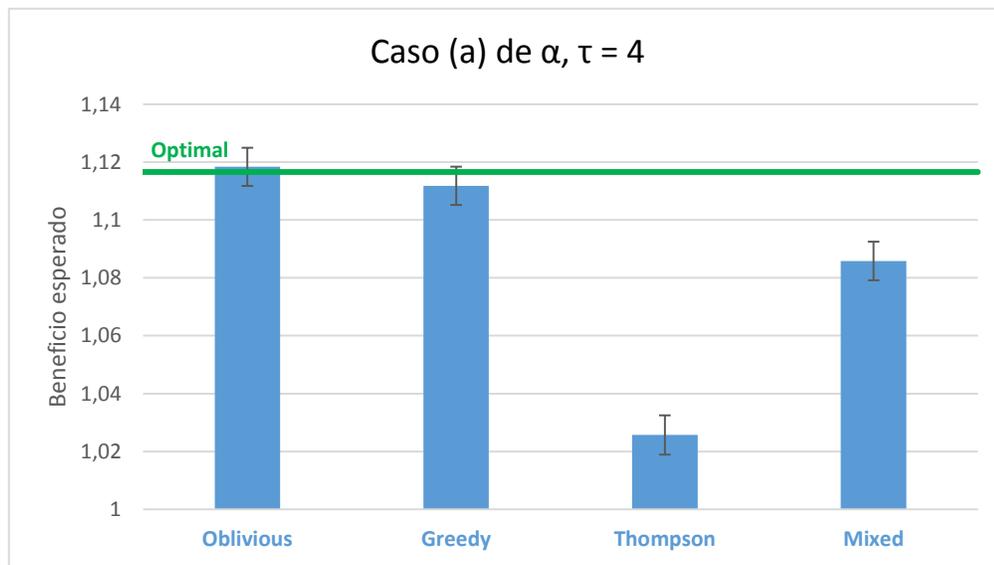


Figura 4.1: Comparación de rendimiento de heurísticas respecto al óptimo, caso (a)

- **Caso b):** Bajo estos parámetros comienza a ser más notoria la relevancia de las evaluaciones, incluso para horizontes temporales cortos. Como se puede ver por ejemplo en la figura 4.2, Oblivious deja de entregar resultados cercanos al óptimo, y en la medida que se considera una mayor cantidad de compradores, peor es su rendimiento, tratándose de la peor entre las 4 heurísticas. Por el contrario, Greedy se presenta como la estrategia de mejores resultados para este escenario, siendo la única cuyo beneficio en algunos casos no es significativamente inferior a la utilidad óptima.
- **Caso c):** La alta influencia desde un comienzo de los ratings obliga al vendedor a des-

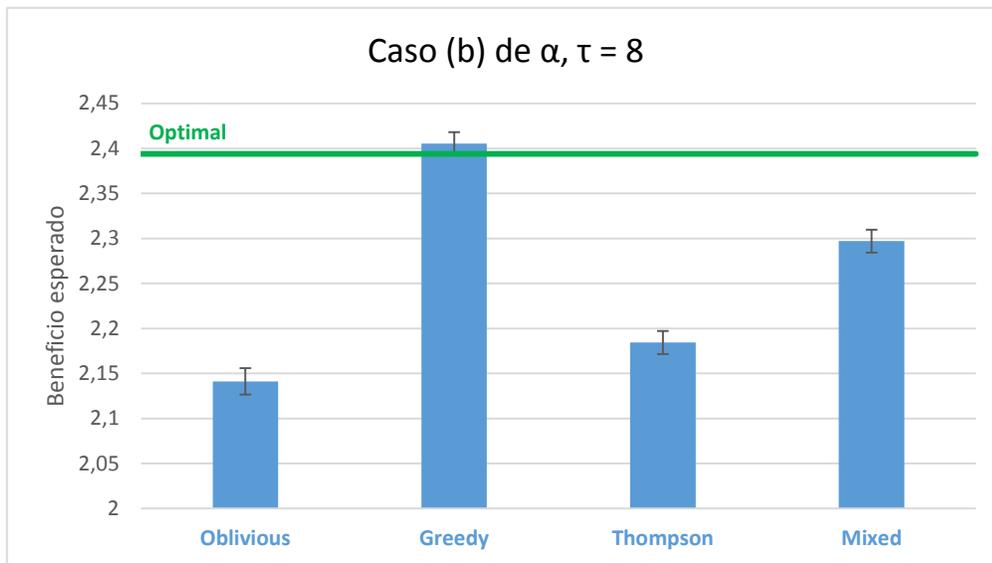


Figura 4.2: Comparación de rendimiento de heurísticas respecto al óptimo, caso (b)

cartar definitivamente la opción de obviar el feedback de sus consumidores. Con cualquiera de las demás estrategias conseguirá incrementar considerablemente sus ingresos. Por otro lado también cabe destacar que, a pesar de que Greedy sigue arrojando los mejores resultados entre las heurísticas aplicadas, Thompson Sampling y especialmente la estrategia Mixta arrojan resultados cada vez mejores en la medida que aumenta el número de períodos considerados, pudiéndose prever que su rendimiento incluso pueda superar a Greedy ante una mayor cantidad de compradores potenciales. A modo de ejemplo la figura 4.3 ilustra esto en el caso de considerarse un horizonte de 18 consumidores.

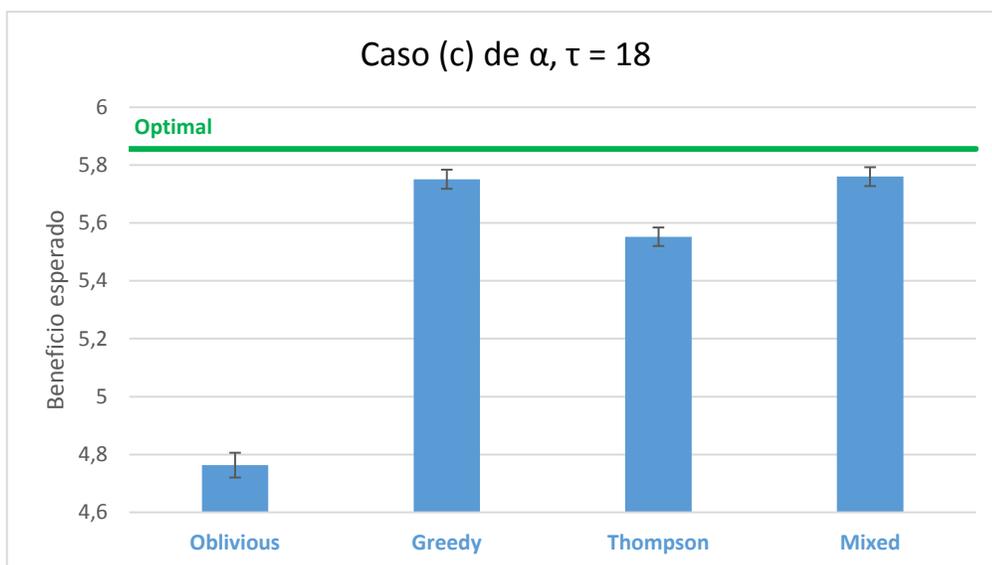


Figura 4.3: Comparación de rendimiento de heurísticas respecto al óptimo, caso (c)

4.2. Programación Dinámica Aproximada

Luego de resolver el problema de manera exacta y haber visto el rendimiento de las heurísticas en aquellas pequeñas instancias, se evalúa sus resultados en mayores horizontes temporales, de 100, 500 y hasta 1200 clientes. A modo de referencia, se sigue incluyendo la estrategia Oblivious, especialmente considerando que ya no se contará con el beneficio óptimo a modo de comparación.

Para evaluar el rendimiento de estas estrategias, se realiza una gran cantidad de simulaciones bajo diferentes escenarios, observándose el beneficio promedio obtenido en cada uno de éstos. En primer lugar, se considerará una situación en la cual la distribución real de la calidad de los productos es bastante diferente a lo que se esperaría a partir de los parámetros de la distribución Dirichlet. Luego, se retomará la idea de alterar los órdenes de magnitud de α , aunque esta vez considerando mayores cantidades de clientes. Finalmente, se aumentará el número de alternativas ofrecidas e incluso se incorporará cierto grado de aleatoriedad en algunos parámetros.

4.2.1. Prior opuesto a la realidad

Se busca destacar el potencial de las heurísticas exploratorias, Thompson Sampling y la estrategia Mixta, a través del siguiente experimento: Los parámetros de la distribución Dirichlet indican claramente la tendencia de la calidad de los 2 productos considerados, habiendo un producto que a priori se perfila como significativamente mejor que el otro. Sin embargo, se impondrá un vector de probabilidades resultante (en vez de generarlo aleatoriamente mediante la distribución Dirichlet) cuya tendencia sea exactamente la opuesta: el producto que a priori pareciera ser de calidad inferior tendrá una alta calidad esperada, y viceversa. De esta manera, tanto el vendedor como los consumidores comenzarán estimando las calidades erróneamente, vendiéndose mayoritariamente el peor producto. Se espera que una política exploratoria permita descubrir y corregir este error, mientras que posiblemente el proceso de compra se detenga completamente bajo una heurística más básica: los consumidores descubren que el producto que están comprando en verdad es de mala calidad, pero no están dispuestos a probar el otro debido a su mal prior.

En la sección B.2 de Apéndices se puede observar el detalle de los resultados obtenidos. Para ello se consideró que

$$\alpha = \begin{pmatrix} 0,7 & 0,3 & 0,1 \\ 0,1 & 0,2 & 0,8 \end{pmatrix}$$

Esto indica que, a priori, la distribución de probabilidad de la calidad del producto 1 debie-

ra presentar una tendencia hacia valores bajos, mientras que el producto 2 por el contrario debiera presentar con mayor frecuencia una alta calidad. Para reforzar aún más la aparente ventaja del segundo producto, se le asigna un beneficio neto bastante superior en proporción al del producto 1: $r_1 = 0,4$ y $r_2 = 0,7$. Con esta información, la decisión parece obvia: $i = 2$ generará mayores ingresos y con una alta probabilidad presentará mejor calidad que $i = 1$. Sin embargo, se incorpora un elemento adicional al experimento: se impone en cada simulación que $p_1 = (0,05 \ 0,15 \ 0,8)$ y $p_2 = (0,8 \ 0,1 \ 0,1)$, en vez de generar estas distribuciones aleatoriamente. Se trata de realizaciones factibles de la distribución Dirichlet considerada, pero con valores opuestos a lo esperado. De esta manera, será el producto 1 el que presentará frecuentemente una alta calidad, mientras que el artículo 2 será de baja calidad con probabilidad 0,8. Tanto los consumidores como el vendedor desconocen estos dos vectores, de manera que inicialmente solo contarán con el prior, para luego ir actualizando su creencia a través de los ratings registrados.

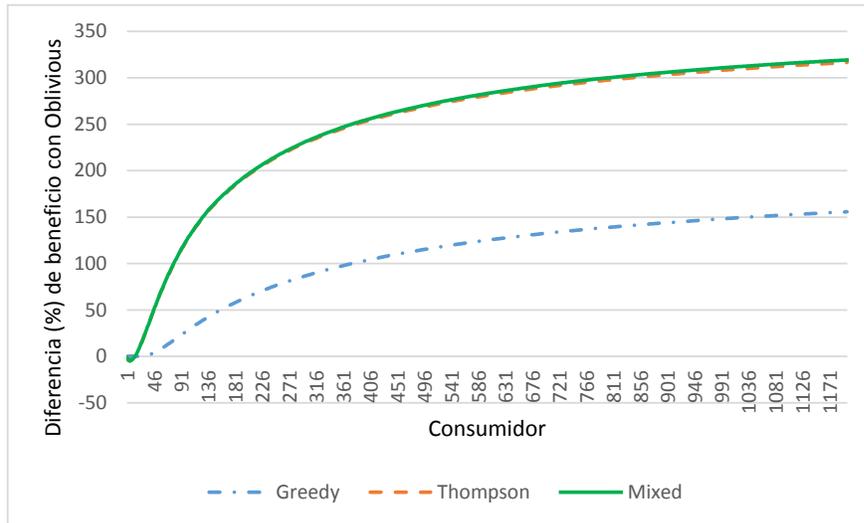


Figura 4.4: Beneficio acumulado: prior opuesto a la realidad, 1200 consumidores

Como se puede observar en las tablas B.28, B.29 y B.30, la diferencia entre no considerar las evaluaciones hechas por los clientes, e incorporarlas de alguna manera en el proceso de decisión del assortment se vuelve abismal. Incluso en un contexto con solo 100 clientes, Greedy entrega aproximadamente un 28 % de beneficio superior a Oblivious. Sin embargo, son los métodos exploratorios los que realmente logran sacar ventajas en este contexto. Thompson Sampling llega a superar a Oblivious hasta en un 318 % (!) al considerarse 1200 consumidores. Finalmente, la estrategia mixta logra perfeccionar este resultado algo más, logrando para el mismo caso una diferencia de aproximadamente 321 %. Cabe destacar que esta estrategia no solo arroja resultados considerablemente mejores que el caso base (es decir, la heurística Oblivious), sino que también supera a Greedy en un 75 %, 73 % y 65 % para los casos de

100, 500 y 1200 clientes, respectivamente. Estos resultados no son sorprendentes: Oblivious no considera de ninguna manera las observaciones de los clientes, siendo imposible que el vendedor se entere en algún momento de la importante diferencia entre su creencia previa sobre la distribución de la calidad de los productos, y la distribución real, y por lo tanto no llegando a modificar nunca su oferta. Por otro lado, la estrategia Greedy sí considera las evaluaciones de los consumidores, pero falla al postergar darle una oportunidad al producto de mejor calidad. Con el objetivo de ilustrar esta situación, la figura 4.4 refleja la evolución de esta diferencia entre el beneficio acumulado obtenido a través de Oblivious y cada una de las demás heurísticas, a lo largo de un horizonte de 1200 clientes. En todo momento, las estrategias exploratorias superan a Greedy, y con mayor razón a Oblivious.

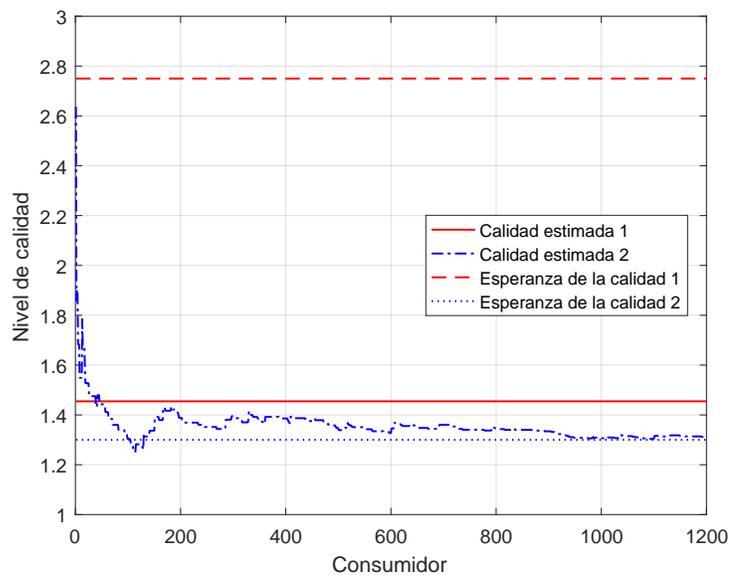


Figura 4.5: Posible evolución de la estimación de la calidad, estrategia Oblivious

Las consecuencias en el aprendizaje de los consumidores y el vendedor pueden ser apreciadas al graficar la evolución de la estimación de la calidad. Cada una de las figuras 4.5, 4.6, 4.7, 4.8 y 4.9 representa una simulación en particular. En la figura 4.5 por ejemplo se observa cómo los clientes estiman que el producto 1 tiene calidad más bien baja (línea continua roja), en base al *prior* conocido. Sin embargo, la distribución de probabilidad de la calidad de esta alternativa en verdad posee la esperanza representada por la línea punteada roja. Análogamente, la línea punteada azul representa la esperanza real de la calidad del producto 2, mientras que la línea alternada azul muestra la creencia de los clientes acerca de este valor. Como bajo la estrategia Oblivious el retailer jamás considera el feedback de los compradores, sino que toma todas sus decisiones en base al prior, siempre ofrece el segundo producto. La estimación de la calidad por parte de los consumidores en cambio evoluciona constantemente, dejando eventualmente atrás su creencia inicial y descubriendo la real calidad esperada del

peor producto, aunque sin poder nunca experimentar la otra alternativa. Como es posible apreciar en la figura 4.6, el mismo resultado puede ocurrir bajo una política Greedy. Sin embargo, gracias a que en tal caso sí se considera las evaluaciones del público, esto no ocurre siempre y existe la posibilidad de revertir la situación, lo cual se ve reflejado en la figura 4.7. En este último caso, tanto los clientes como el vendedor descubren que la calidad de la alternativa 2 es bastante inferior a lo considerado inicialmente, y al darle una oportunidad al producto 1 se revela su alta calidad.

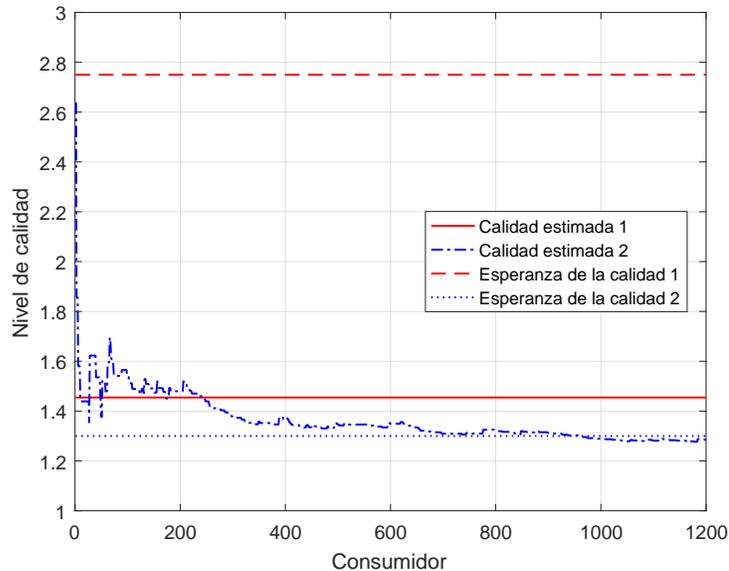


Figura 4.6: Posible evolución de la estimación de la calidad, estrategia Greedy (1)

A diferencia de los dos métodos anteriores, tanto Thompson Sampling como la estrategia Mixta suelen permitir descubrir rápidamente la diferencia entre el prior y la distribución real. En la figura 4.8 se observa como el producto 1 es escogido bastante rápidamente por el vendedor como la opción a ser ofrecida. Por lo mismo, el público no demora mucho en tener una buena estimación de su calidad. Sin embargo, la desventaja que conlleva Thompson Sampling en este caso es que se sigue experimentando con el producto de baja calidad prácticamente hasta los últimos períodos, perdiéndose oportunidades de ofrecer la mejor alternativa. La estrategia Mixta en cambio abandona paulatinamente los esfuerzos en explorar, y en la medida que se acerca el final del horizonte de evaluación, opta por explotar la información disponible. En la situación ilustrada en la figura 4.9 por ejemplo se observa cómo no se sigue experimentando con la alternativa 2 mucho más allá del período 600.

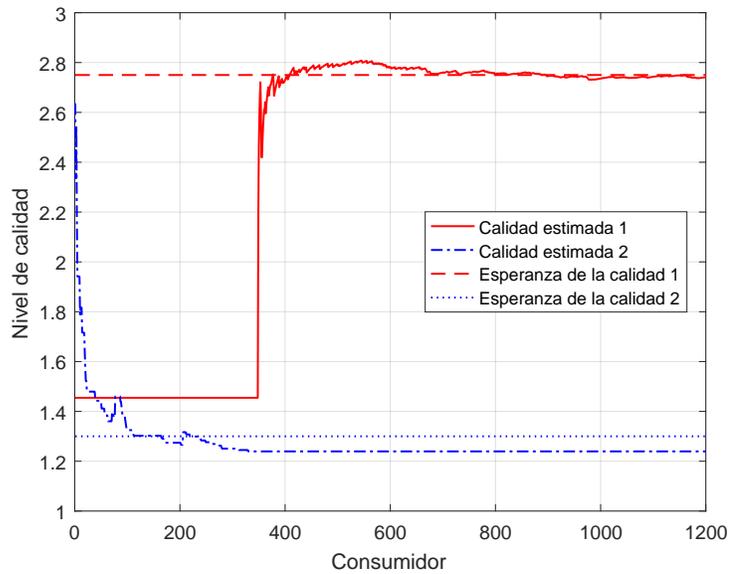


Figura 4.7: Posible evolución de la estimación de la calidad, estrategia Greedy (2)

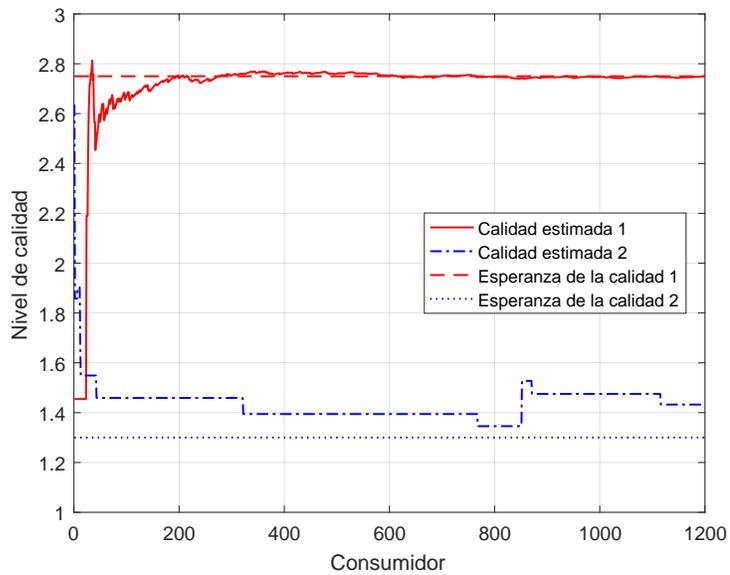


Figura 4.8: Posible evolución de la estimación de la calidad, Thompson Sampling

4.2.2. Órdenes de magnitud de α

En esta sección se busca llevar a cabo el mismo experimento que fue realizado al resolver el problema de manera exacta: dependiendo de los órdenes de magnitud de los parámetros de la distribución Dirichlet, las evaluaciones tienen una diferente incidencia. Se verá entonces el impacto que tiene esto en el rendimiento de las diferentes heurísticas. Los casos a considerar serán exactamente los mismos que en la sección 4.1:

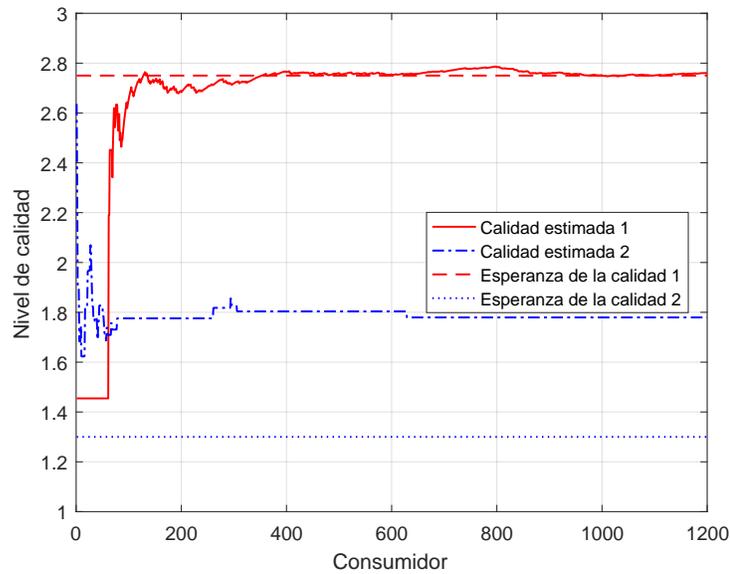


Figura 4.9: Posible evolución de la estimación de la calidad, estrategia Mixta

a) $\alpha = \begin{pmatrix} 3 & 2 & 4 \\ 2 & 3 & 1 \end{pmatrix}$

b) $\alpha = \begin{pmatrix} 0,3 & 0,2 & 0,4 \\ 0,2 & 0,3 & 0,1 \end{pmatrix}$

c) $\alpha = \begin{pmatrix} 0,03 & 0,02 & 0,04 \\ 0,02 & 0,03 & 0,01 \end{pmatrix}$

Para cada una de estas situaciones, se considera nuevamente ingresos $r_1 = 0,48$ y $r_2 = 0,51$. Sin embargo, se debe recordar que en esta ocasión se trabajará con $\tau = 100, 500, \text{ y } 1200$.

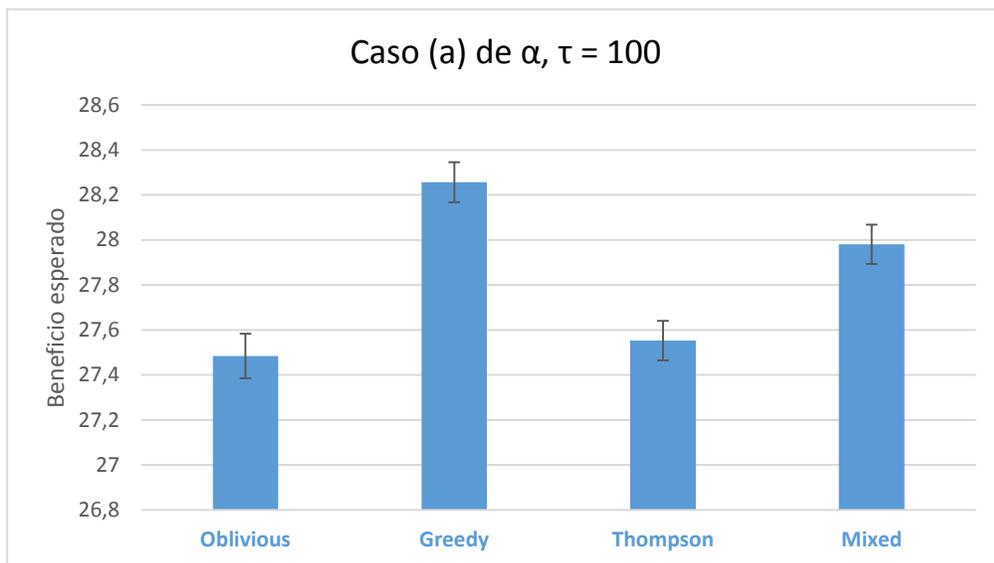


Figura 4.10: Comparación de rendimiento de heurísticas, caso (a) con 100 consumidores

Como se puede ver en la sección B.3 de Apéndices, a excepción del caso a) con $\tau = 100$ - en el cual Greedy presenta el mejor desempeño, lo cual se encuentra ilustrado en la figura 4.10 -, los mejores resultados siempre son obtenidos a través de Thompson Sampling y la estrategia Mixta. Esta última en particular a veces también entrega un beneficio significativamente superior al de Thompson Sampling, como lo ejemplifica la figura 4.11, perfilándose como la mejor de las 3 heurísticas en la mayoría de los casos estudiados.

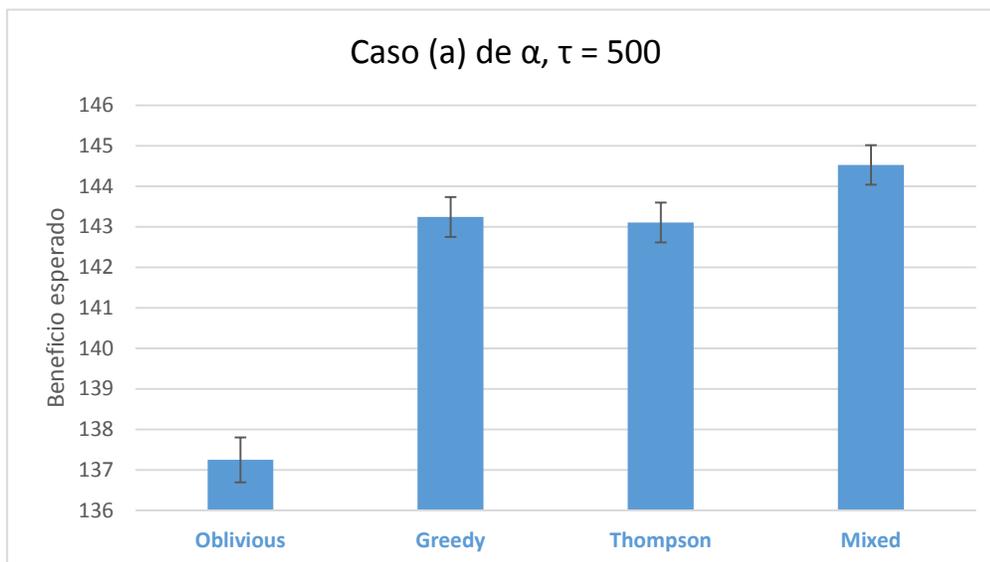


Figura 4.11: Comparación de rendimiento de heurísticas, caso (a) con 500 consumidores

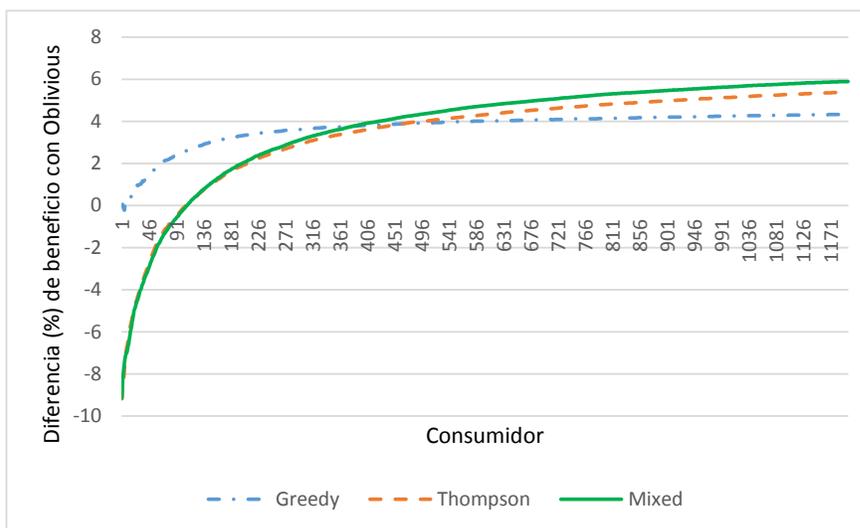


Figura 4.12: Beneficio acumulado: caso (a) de α , 1200 consumidores

Cabe destacar cómo los resultados indican que la mayor influencia de las evaluaciones conlleva un mayor beneficio esperado. En la sección B.3 se observa cómo para cada valor de τ estudiado, es el caso (c) en el cual se obtiene la mayor utilidad posible. Como ha sido

mencionado anteriormente, el valor inferior de los elementos de la matriz α conlleva un mayor peso relativo de los ratings respecto al prior, siendo más valiosa la información entregada por los consumidores y facilitando de esta manera el proceso de aprendizaje, tanto del público como del retailer.

Finalmente, tal como fue presentado en la sección anterior, las figuras 4.12, 4.13 y 4.14 reflejan la evolución de la diferencia entre el beneficio obtenido a través de Oblivious y cada una de las demás heurísticas, a lo largo de un horizonte de 1200 consumidores. En cada una de ellas, especialmente para el caso (a), se puede apreciar cómo Greedy comienza aventajando al resto de las estrategias. Thompson Sampling y la estrategia Mixta incluso tardan algunos períodos en superar a Oblivious, pero eventualmente se perfilan como las mejores alternativas de largo plazo, siendo el algoritmo Mixto levemente superior.

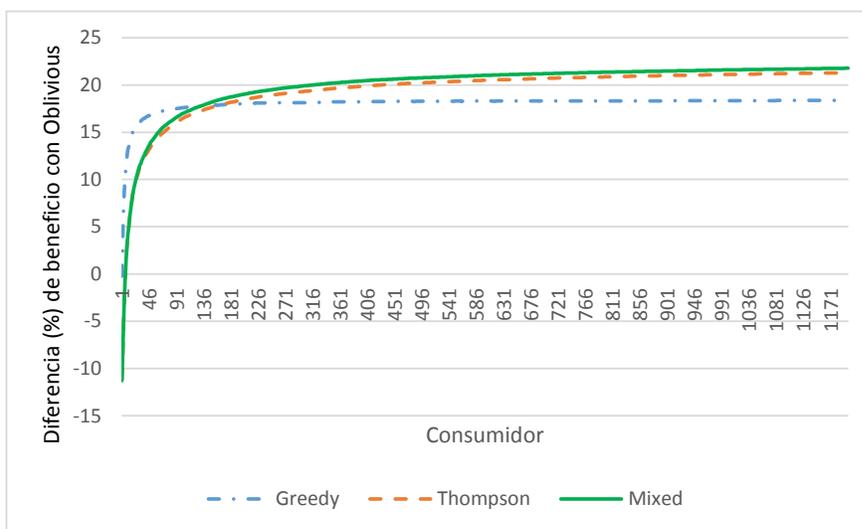


Figura 4.13: Beneficio acumulado: caso (b) de α , 1200 consumidores

4.2.3. Incremento del número de alternativas

Con el objetivo de complementar los resultados anteriores, se considera una mayor cantidad de productos, y posteriormente se explora aleatoriamente valores de algunos parámetros para comparar las heurísticas en estos casos. Para poder obtener estos resultados en tiempos acotados y razonables, se impone que todos los clientes observados compren alguno de los productos disponibles, acelerando así el proceso de aprendizaje. De esta manera, la probabilidad de elección del artículo i , condicional en que no se opta por la alternativa externa, es igual a

$$\frac{e^{f(\hat{q}_{it})}}{\sum_{j \in X_t} e^{f(\hat{q}_{jt})}}$$

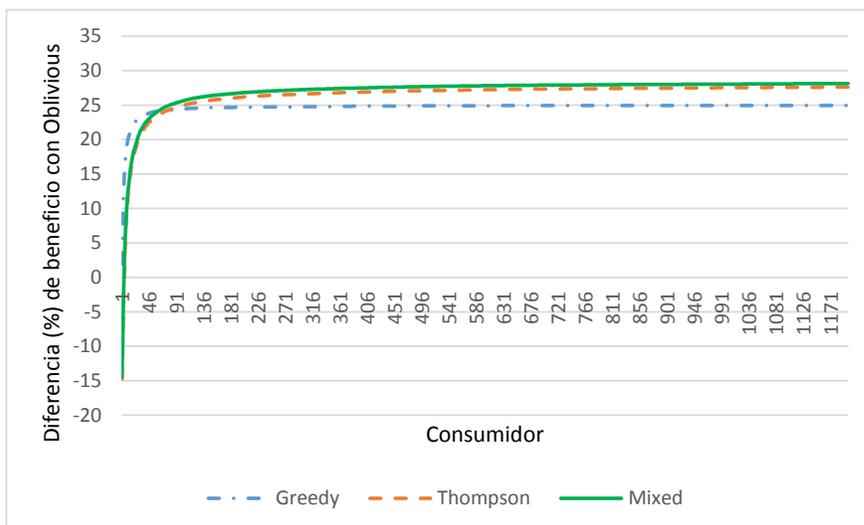


Figura 4.14: Beneficio acumulado: caso (c) de α , 1200 consumidores

Por otro lado, la probabilidad de que t compre alguno de los productos ofrecidos es

$$b_t := \frac{\sum_{j \in X_t} e^{f(\hat{q}_{jt})}}{1 + \sum_{j \in X_t} e^{f(\hat{q}_{jt})}}$$

Se puede interpretar la cantidad de clientes que transcurre hasta que uno de ellos compre como una variable aleatoria de distribución geométrica con parámetro b_t , de manera que la cantidad esperada de clientes entre las compras de los compradores $t - 1$ y t viene dada por $\frac{1}{b_t}$. Este resultado será utilizado para estimar la cantidad de individuos que pasan por el sistema hasta completar un total de $\tau = 1000$ ventas. Dado esto, cada caso estudiado tendrá un horizonte temporal aproximado, y diferente de los demás, por lo cual el indicador de rendimiento a considerar no es el beneficio total obtenido tras 1000 ventas, sino el beneficio promedio por cliente, incluyendo aquellos que no compraron.

El presente análisis considera los parámetros expuestos en la sección B.4 del apéndice. Entre otras definiciones importantes, se consideró 10 productos, de los cuales se puede exponer 5. Con el objetivo de reducir los tiempos requeridos para realizar todos los cálculos necesarios en el caso de tantos productos, se reduce a 100 el número de simulaciones. Esto conlleva que en varias ocasiones los intervalos de confianza calculados no permitan afirmar con certeza que todos los resultados sean significativos, pero esto no impide observar ciertas tendencias que responden a la intuición.

- 1) **Órdenes de magnitud de α :** Nuevamente se estudia el efecto de alterar el orden de magnitud de los elementos de la matriz α . Los 3 casos considerados así como los resultados pueden ser observados en la sección B.4.1 del apéndice. La tendencia en el

beneficio esperado por cliente sigue lo observado en secciones anteriores: la heurística de mejor rendimiento promedio en cada caso es la estrategia Mixta, mientras que Oblivious entrega las utilidades más bajas. Además, se confirma que menores valores de los elementos de α conllevan mayores ganancias.

- 2) **Prior opuesto a la realidad:** Como se puede ver en los parámetros expuestos en B.4.2, una vez más se impuso una distribución fija para la calidad de cada producto, la cual en la mayoría de los casos se opone a lo que los clientes y el vendedor esperarían a partir del prior. Esto conlleva que las heurísticas exploratorias presenten un mejor rendimiento (como se puede ver en la tabla B.43), al permitir corregir este error, siendo en particular la estrategia Mixta la que entrega el mayor beneficio por cliente.
- 3) **Parámetros aleatorios:** Como es explicado en la sección B.4.3 del apéndice, se genera aleatoriamente los valores de α y r de acuerdo a distribuciones uniformes entre 0 y 1. En la tabla B.44 se puede ver cómo Greedy presenta la mejor razón de beneficio por cliente, seguida de cerca por la estrategia Mixta. Oblivious presenta el peor rendimiento, como era de esperar. Sin embargo, la gran varianza resultante de la generación aleatoria implica que los intervalos de confianza obtenidos sean bastante más extensos que aquellos obtenidos en los dos experimentos anteriores. Por lo tanto, no es posible ordenar las utilidades de cada heurística con certeza.

4.2.4. Sobre la cantidad de alternativas y los tiempos de resolución

En escenarios reales, un comerciante usualmente no debe elegir solamente entre unas pocas alternativas, sino entre varias decenas o incluso cientos de productos. A su vez, en caso que para la elección del surtido se deba considerar información que se actualiza cada vez que se realiza una compra (como es el caso de los ratings estudiados en esta tesis), es necesario ser capaz de determinar el assortment en fracciones de segundo, al menos cuando se trata de retail online con un alto flujo de consumidores. Esto justifica la importancia de la implementación de heurísticas eficientes, capaces de entregar buenos resultados en lapsos temporales acotados. Particularmente en los experimentos realizados a lo largo de este estudio, se observó tiempos de resolución de algunas horas, pero se debe tener en cuenta que se buscó estimar el valor esperado del beneficio total a lo largo de todo el horizonte de evaluación. Por lo tanto, por cada simulación realizada, cada heurística fue aplicada τ veces. Además, dependiendo del caso, se realizaron 100 o 20000 simulaciones. Un cálculo simple e informal en base a los resultados observados en la tabla B.9 por ejemplo sería dividir el tiempo transcurrido para obtener el valor esperado del beneficio obtenido a través de la política Greedy por $\tau = 18$ y 20000, obteniéndose una estimación preliminar del tiempo de resolución de $1,039 \cdot 10^{-4}$ segundos.

Aunque este caso solo contempla 2 productos, cabe destacar que el tiempo total a su vez incluye otros procesos que no es necesario realizar cada vez que se actualizan la información disponible y el surtido a ofrecer. Además, el problema de Assortment en tiempo real con modelos de elección discreta ya ha sido enfrentado anteriormente con resultados satisfactorios. Como ya ha sido mencionado, en [Golrezaei *et al.*, 2014] se estudia esta situación, justificando que el problema de optimización subyacente a enfrentar en cada ocasión puede ser resuelto en tiempo polinomial, y también evaluando casos numéricos con varias decenas de productos.

Capítulo 5

Conclusiones

En este último capítulo se presenta las principales conclusiones del trabajo realizado, además de proponerse posibles extensiones y mejoras al modelo estudiado.

5.1. Modelo y métodos de resolución

En esta tesis se propone un modelo matemático con el fin de modelar la selección del conjunto de productos a ofrecer por parte de un vendedor así como el proceso de decisión de compra y la evaluación de su experiencia por parte de los consumidores, considerándose elementos como un límite al tamaño del surtido y las preferencias inherentes de los clientes, entre otros. Para ello, se considera tanto un modelo de elección discreta así como el proceso de aprendizaje de los compradores y el retailer.

El modelo desarrollado da lugar al problema de optimización enfrentado por el vendedor, quien debe escoger qué productos ofrecer. Este problema resulta ser bastante difícil de resolver en tiempos razonables: en la medida que se considera una mayor cantidad de clientes, posibles niveles de calidad y productos, la cantidad de estados del sistema aumenta de manera exponencial. En efecto, el tamaño del conjunto de estados posibles en el último período, al considerarse $|I|$ alternativas, Q valores posibles de la calidad y τ consumidores, viene dado por

$$\sum_{t=1}^{\tau} \binom{Q \cdot |I| + t - 2}{Q \cdot |I| - 1}$$

Basta reemplazar con algunos valores para dimensionar la complejidad de esto: mientras que una instancia más bien pequeña, con 2 productos, 3 niveles de calidad y 10 consumidores ya implica 5005 estados, un escenario en que $|I| = 20$, $Q = 5$ y $\tau = 30$ conlleva $6,0285 \cdot 10^{28}$ posibilidades al término del horizonte de evaluación. Una solución exacta a este problema, enfrentado mediante programación dinámica, consiste en especificar el assortment óptimo

para todo estado posible de todo período, lo cual es virtualmente imposible bajo escenarios como el recién mencionado. Por lo tanto, se debe descartar la opción de resolver instancias grandes de manera exacta, motivando el desarrollo de métodos heurísticos. De todas maneras es posible encontrar los surtidos óptimos para escenarios de menor tamaño, sirviendo esto como referencia para el rendimiento de los 3 algoritmos propuestos:

- **Greedy:** Optimización del período actual, sin considerar futuras consecuencias
- **Thompson Sampling:** Algoritmo exploratorio que simula la distribución de la calidad en base a una distribución que depende de las observaciones hechas hasta el momento.
- **Algoritmo Mixto:** Método que pondera las funciones objetivo de los dos métodos anteriores, priorizando la exploración en un comienzo pero otorgándole mayor importancia a la explotación en la medida que se acerca el último período.

De esta manera, el presente trabajo busca aportar incluyendo la teoría económica de Social Learning en los problemas de Assortment, además de desarrollar heurísticas capaces de entregar el surtido ideal bajo diferentes escenarios en tiempos de resolución razonables.

5.2. Resultados

Como fue mencionado anteriormente, el problema rápidamente se vuelve difícil (prácticamente imposible) de resolver de manera exacta en la medida que aumenta el número de consumidores y la cantidad de alternativas. Sin embargo, un retailer debe ser capaz de obtener resultados en tiempo real en la medida que vende sus productos y por lo tanto se actualiza el estado del sistema. Las heurísticas propuestas a lo largo de este informe presentan la rapidez necesaria para ello: basta considerar los tiempos de resolución registrados y darse cuenta que estos fueron obtenidos al realizar miles de simulaciones y a lo largo de cientos de períodos. En la práctica, solo es necesario resolver para un período -dado un determinado estado- y una sola vez, con el fin de obtener el assortment ideal para el siguiente cliente, lo cual puede ser llevado a cabo en tan solo una fracción de segundo.

Cada heurística tiene características particulares que la vuelven más o menos apta ante determinados escenarios. Sin embargo, es la estrategia Mixta la que presenta una mayor flexibilidad y entrega el mayor beneficio en la mayoría de los escenarios considerados. Oblivious solo presenta un rendimiento razonable en la medida que se considera un bajo número de consumidores, contexto en el cual aún no cobran gran relevancia las evaluaciones de productos registradas. Sin embargo, en el mediano y largo plazo rápidamente se transforma en una alternativa enormemente inferior a las demás. Esto reafirma que es fundamental considerar el feedback de los clientes: Oblivious refleja precisamente lo que sucedería en caso que un

retailer no considere la revelación de la calidad realizada por los compradores, y la diferencia puede ser abismal. En el peor caso registrado en este estudio, la utilidad obtenida a través de la estrategia Mixta supera a Oblivious en un 321 %. Por otro lado, Greedy no se presenta como una mala opción, al ser incluso óptima desde un punto de vista miope bajo el cual solo se considera el período actual. Sin embargo, su debilidad es no considerar las consecuencias futuras de las decisiones presentes. Por el contrario, Thompson Sampling estima la distribución de la calidad de los productos mediante simulación, permitiendo una exploración más extensa de las posibilidades y por lo tanto teniendo un buen desempeño en situaciones en que se cuenta con poca información, o incluso información engañosa inicialmente. El problema de esta heurística es que no deja nunca de lado su carácter exploratorio, sin aprovechar de hacer un uso más eficiente de los datos obtenidos, especialmente en la medida que se acerca el fin del horizonte de evaluación y se vuelve menos necesario investigar las posibilidades futuras. La estrategia Mixta aprovecha las virtudes de estos dos algoritmos, lo cual explica sus mejores resultados en general. Sin embargo, en algunos casos las diferencias no son significativas, sugiriendo la necesidad de realizar una mayor cantidad de simulaciones y considerar mayores horizontes temporales.

5.3. Trabajo en progreso y posibles extensiones

Lo desarrollado hasta el momento presenta oportunidades para seguir trabajando con el modelo, ya sea incorporando nuevos aspectos o perfeccionando aquellos que ya han sido incluidos. A continuación se presenta algunas ideas a considerar a futuro:

- 1) **Distribución de los beneficios de cada heurística:** Un análisis interesante que por el momento ha sido realizado de manera más bien superficial es el de la distribución de los beneficios obtenidos a través de cada heurística. Esto permitiría hacer observaciones más detalladas que tan solo comparar el beneficio esperado, y analizar otros aspectos como la varianza o la frecuencia con que cada heurística supera a las demás bajo escenarios idénticos. En efecto, al estudiar diversas situaciones cuyo factor común fue la existencia de 3 productos y un horizonte temporal de 1000 consumidores, se obtuvo que la estrategia Mixta presentaba generalmente el mayor beneficio esperado, pero Greedy resultó ser el algoritmo que con mayor frecuencia superaba a todos los demás. ¿A qué se debe esto? ¿Bajo qué criterio debiera escogerse la heurística a utilizar?
- 2) **¿Cuándo reportan los consumidores?** Los sitios de retail no suelen exigir a sus clientes que evalúen los productos que adquieren. Se trata de un acto voluntario, y por

lo mismo no todo el mundo lo realiza. Entonces, ¿por qué lo hacen? Como fue mencionado en la Revisión Bibliográfica, varios investigadores se han hecho esta pregunta y modelado la situación en un esfuerzo por entender y analizar este comportamiento. Una solución sencilla es considerar una cierta probabilidad de que cada persone evalúe un producto. De esta manera, se incluye en el modelo el hecho que no todos reportan, aunque no se explica el motivo ni se hace diferencia alguna entre un individuo u otro. Es más: bajo este supuesto, se esperaría observar los mismos resultados que cuando todos los compradores evalúan su experiencia, aunque a lo largo de un horizonte temporal más extenso. Una solución que se hace cargo de las diferencias entre las preferencias de los consumidores, así como del motivo por el cual deciden compartir su experiencia con los demás, es la propuesta - con leves diferencias - en [Dellarocas, 2001] y [Hu *et al.*, 2006]. En ambos artículos solo se reporta cuando la utilidad obtenida es suficientemente alta o baja, sosteniendo entonces que las personas suelen comunicar su impresión solamente cuando se encuentran altamente satisfechas, o profundamente decepcionadas. Bajo ciertas consideraciones, esto podría ser incorporado en el modelo estudiado en el presente informe.

- 3) **Reportes sesgados:** Estrechamente ligado al punto anterior, la subjetividad a la cual suele estar sujeta este tipo de evaluaciones es un punto importante a considerar. A diferencia de lo supuesto a lo largo de este trabajo, los reportes suelen estar profundamente sesgados por la preferencia personal de los individuos. A pesar de que en el modelo propuesto acá se considera la heterogeneidad de los consumidores como un elemento fundamental en la selección de una alternativa, cabe recordar que los clientes finalmente reportan la calidad real del producto comprado. Se propone entonces buscar una manera alternativa de modelar esta situación, debiendo seguramente modificarse el formato de los reportes, la forma en que los compradores deciden qué calificación asignar a un producto, y la manera en que utilizan la información disponible para escoger una u otra alternativa.
- 4) **Inventario:** Otro elemento de interés es incorporar inventario limitado de cada producto, considerado como una variable exógena. De esta manera, el vendedor no podría ofrecer un producto cuyo stock ya se haya agotado, ni podría reabastecerse. En cada período, se contaría con S_{it} unidades del producto i . S_{i1} sería el inventario inicial, y se le debería descontar una unidad cada vez que un consumidor opte por la alternativa correspondiente. Luego,

$$S_{it} = S_{i1} - \sum_{j=1}^Q N_{ij}^t$$

Esta nueva restricción no solo limitaría la venta de productos en la medida que se reduce

el inventario disponible, sino que significaría también una reducción de la cantidad de estados factibles del sistema, pudiendo aprovecharse esto al momento de optimizar el assortment mediante programación dinámica.

Bibliografía

- [Acemoglu *et al.*, 2014] Acemoglu, D., Bimpikis, K., y Ozdaglar, A. (2014). Dynamics of information exchange in endogenous social networks. *Theoretical Economics*, 9(1):41–97.
- [Acemoglu *et al.*, 2011] Acemoglu, D., Dahleh, M. A., Lobel, I., y Ozdaglar, A. (2011). Bayesian learning in social networks. *The Review of Economic Studies*, 78(4):1201–1236.
- [Banerjee, 1992] Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, pp. 797–817.
- [Besbes y Scarsini, 2015] Besbes, O. y Scarsini, M. (2015). On information distortions in online ratings. *Columbia Business School Research Paper*, (13-36).
- [Bikhchandani *et al.*, 1992] Bikhchandani, S., Hirshleifer, D., y Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy*, pp. 992–1026.
- [Bose *et al.*, 2006] Bose, S., Orosel, G., Ottaviani, M., y Vesterlund, L. (2006). Dynamic monopoly pricing and herding. *The RAND Journal of Economics*, 37(4):910–928.
- [Caro y Gallien, 2007] Caro, F. y Gallien, J. (2007). Dynamic assortment with demand learning for seasonal consumer goods. *Management Science*, 53(2):276–292.
- [Chatterjee, 2001] Chatterjee, P. (2001). Online reviews: do consumers use them?
- [Cheung *et al.*, 2014] Cheung, C. M., Xiao, B. S., y Liu, I. L. (2014). Do actions speak louder than voices? the signaling role of social information cues in influencing consumer purchase decisions. *Decision Support Systems*, 65:50–58.
- [Crapis *et al.*, 2015] Crapis, D., Ifrach, B., Maglaras, C., y Scarsini, M. (2015). Monopoly pricing in the presence of social learning. *Available at SSRN 1957924*.
- [Davis *et al.*, 2014] Davis, J. M., Gallego, G., y Topaloglu, H. (2014). Assortment optimization under variants of the nested logit model. *Operations Research*, 62(2):250–273.
- [Dellarocas, 2001] Dellarocas, C. (2001). Analyzing the economic efficiency of ebay-like online reputation reporting mechanisms. En *Proceedings of the 3rd ACM Conference on Electronic Commerce*, pp. 171–179. ACM.

- [Ellison y Fudenberg, 1993] Ellison, G. y Fudenberg, D. (1993). Rules of thumb for social learning. *Journal of political Economy*, pp. 612–643.
- [Golrezaei *et al.*, 2014] Golrezaei, N., Nazerzadeh, H., y Rusmevichientong, P. (2014). Real-time optimization of personalized assortments. *Management Science*, 60(6):1532–1551.
- [Honhon *et al.*, 2010] Honhon, D., Gaur, V., y Seshadri, S. (2010). Assortment planning and inventory decisions under stockout-based substitution. *Operations research*, 58(5):1364–1379.
- [Hu *et al.*, 2006] Hu, N., Pavlou, P. A., y Zhang, J. (2006). Can online reviews reveal a product’s true quality?: empirical findings and analytical modeling of online word-of-mouth communication. En *Proceedings of the 7th ACM conference on Electronic commerce*, pp. 324–330. ACM.
- [Ifrach *et al.*, 2014] Ifrach, B., Maglaras, C., y Scarsini, M. (2014). Bayesian social learning with consumer reviews. *ACM SIGMETRICS Performance Evaluation Review*, 41(4):28–28.
- [Kallus y Udell, 2016] Kallus, N. y Udell, M. (2016). Dynamic assortment personalization in high dimensions. *arXiv preprint arXiv:1610.05604*.
- [Li *et al.*, 2015] Li, G., Rusmevichientong, P., y Topaloglu, H. (2015). The d-level nested logit model: Assortment and price optimization problems. *Operations Research*, 63(2):325–342.
- [Li y Hitt, 2008] Li, X. y Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research*, 19(4):456–474.
- [Liu y van Ryzin, 2008] Liu, Q. y van Ryzin, G. (2008). On the choice-based linear programming model for network revenue management. *Manufacturing & Service Operations Management*, 10(2):288–310.
- [Luca, 2011] Luca, M. (2011). Reviews, reputation, and revenue: The case of yelp. com. *Com (September 16, 2011). Harvard Business School NOM Unit Working Paper*, (12-016).
- [Luce, 1959] Luce, R. D. (1959). *Individual Choice Behavior a Theoretical Analysis*. John Wiley and sons.
- [Mahajan y van Ryzin, 2001] Mahajan, S. y van Ryzin, G. (2001). Stocking retail assortments under dynamic consumer substitution. *Operations Research*, 49(3):334–351.
- [McFadden, 1973] McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.
- [McFadden, 1980] McFadden, D. (1980). Econometric models for probabilistic choice among products. *Journal of Business*, pp. S13–S29.

- [McFadden y Train, 2000] McFadden, D. y Train, K. (2000). Mixed mnl models for discrete response. *Journal of applied Econometrics*, pp. 447–470.
- [Moe y Schweidel, 2012] Moe, W. W. y Schweidel, D. A. (2012). Online product opinions: Incidence, evaluation, and evolution. *Marketing Science*, 31(3):372–386.
- [Pan y Zhang, 2011] Pan, Y. y Zhang, J. Q. (2011). Born unequal: a study of the helpfulness of user-generated product reviews. *Journal of Retailing*, 87(4):598–612.
- [Papanastasiou *et al.*, 2013] Papanastasiou, Y., Bakshi, N., y Savva, N. (2013). Social learning from early buyer reviews: Implications for new product launch. *History*.
- [Papanastasiou y Savva, 2016] Papanastasiou, Y. y Savva, N. (2016). Dynamic pricing in the presence of social learning and strategic consumers. *Management Science*.
- [Rusmevichientong *et al.*, 2010] Rusmevichientong, P., Shen, Z.-J. M., y Shmoys, D. B. (2010). Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Operations research*, 58(6):1666–1680.
- [Rusmevichientong *et al.*, 2014] Rusmevichientong, P., Shmoys, D., Tong, C., y Topaloglu, H. (2014). Assortment optimization under the multinomial logit model with random choice parameters. *Production and Operations Management*, 23(11):2023–2039.
- [Sauré y Zeevi, 2013] Sauré, D. y Zeevi, A. (2013). Optimal dynamic assortment planning with demand learning. *Manufacturing & Service Operations Management*, 15(3):387–404.
- [Smith y Sørensen, 2000] Smith, L. y Sørensen, P. (2000). Pathological outcomes of observational learning. *Econometrica*, 68(2):371–398.
- [Smith y Sørensen, 2013] Smith, L. y Sørensen, P. N. (2013). Rational social learning by random sampling. *Available at SSRN 1138095*.
- [Smith y Agrawal, 2000] Smith, S. A. y Agrawal, N. (2000). Management of multi-item retail inventory systems with demand substitution. *Operations Research*, 48(1):50–64.
- [Talluri y Van Ryzin, 2006] Talluri, K. T. y Van Ryzin, G. J. (2006). *The theory and practice of revenue management*, volumen 68. Springer Science & Business Media.
- [Talwar *et al.*, 2007] Talwar, A., Jurca, R., y Faltings, B. (2007). Understanding user behavior in online feedback reporting. En *Proceedings of the 8th ACM conference on Electronic commerce*, pp. 134–142. ACM.
- [Thompson, 1933] Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294.
- [Topaloglu, 2013] Topaloglu, H. (2013). Joint stocking and product offer decisions under the multinomial logit model. *Production and Operations Management*, 22(5):1182–1199.

- [Tucker y Zhang, 2011] Tucker, C. y Zhang, J. (2011). How does popularity information affect choices? a field experiment. *Management Science*, 57(5):828–842.
- [van Ryzin y Mahajan, 1999] van Ryzin, G. y Mahajan, S. (1999). On the relationship between inventory costs and variety benefits in retail assortments. *Management Science*, 45(11):1496–1509.
- [Williams, 1977] Williams, H. C. (1977). On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and planning A*, 9(3):285–344.
- [Ye *et al.*, 2013] Ye, Qiang and Fang, Bin and others (2013). Learning from other buyers: The effect of purchase history records in online marketplaces. *Decision Support Systems*, 56:502–512.

Apéndice A

Demostraciones Capítulo 3

Sea $\Psi_\lambda : [0, 1] \rightarrow [0, 1]$ la función definida por

$$\Psi_\lambda(x) = \begin{cases} \frac{1-e^{\lambda x}}{1-e^\lambda} & \text{si } \lambda \neq 0 \\ x & \text{si } \lambda = 0 \end{cases}$$

Propiedad 1: Ψ_λ es creciente.

Demostración. Cuando $\lambda \neq 0$, $\frac{d}{dx}\Psi_\lambda(x) = \frac{\lambda e^{\lambda x}}{e^\lambda - 1} > 0$. Cuando $\lambda = 0$, $\frac{d}{dx}\Psi_0(x) = 1 > 0$. \square

Propiedad 2: $\Psi_\lambda(0) = 0$, $\Psi_\lambda(1) = 1$.

Demostración. Si $\lambda \neq 0$, $\Psi_\lambda(0) = \frac{1-e^0}{1-e^\lambda} = 0$ y $\Psi_\lambda(1) = \frac{1-e^\lambda}{1-e^\lambda} = 1$. Si $\lambda = 0$, $\Psi_0(0) = 0$ y $\Psi_0(1) = 1$. \square

Propiedad 3: Si $\lambda < 0$, $\Psi_\lambda(x)$ es una función cóncava. Si $\lambda > 0$, entonces $\Psi_\lambda(x)$ es convexa.

Demostración. $\frac{d^2}{dx^2}\Psi_\lambda(x) = \frac{\lambda^2 e^{\lambda x}}{e^\lambda - 1}$. Si $\lambda < 0$, entonces $\frac{\lambda^2 e^{\lambda x}}{e^\lambda - 1} < 0$. Si $\lambda > 0$, $\frac{\lambda^2 e^{\lambda x}}{e^\lambda - 1} > 0$. \square

Apéndice B

Resultados

Todos los intervalos de confianza incluidos en esta sección fueron calculados con una significancia del 5%.

B.1. Programación Dinámica Exacta

Para los resultados expuestos a continuación, se consideró los siguientes parámetros:

- $|I| = 2$
- $C = 1$
- $Q = 3$
- $\beta = 3$
- $\gamma = -6$
- $r = (0,48 \quad 0,51)$

El beneficio de cada heurística fue obtenido simulando el proceso 20000 veces y calculando el ingreso promedio.

B.1.1. Caso a)

$$\alpha = \begin{pmatrix} 3 & 2 & 4 \\ 2 & 3 & 1 \end{pmatrix}$$

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	0,55890631	0,55890631	0,55890631	0,05562889
Oblivious	0,558216	0,55357774	0,56285426	5,4546718
Greedy	0,558528	0,55389338	0,56316262	4,99518822
Thompson	0,51675	0,51197389	0,52152611	6,17177754
Mixed	0,5459505	0,5412897	0,5506113	7,66525235

Figura B.1: Resultados para 2 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	1,11659427	1,11659427	1,11659427	0,08610829
Oblivious	1,118376	1,11178156	1,12497044	9,93256796
Greedy	1,111818	1,10519831	1,11843769	9,10889642
Thompson	1,025691	1,0188953	1,0324867	11,3952856
Mixed	1,085841	1,07917177	1,09251023	14,4183315

Figura B.2: Resultados para 4 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	1,67389157	1,67389157	1,67389157	0,47038075
Oblivious	1,677936	1,66974278	1,68612922	14,8814792
Greedy	1,6769985	1,66874763	1,68524937	13,2760363
Thompson	1,5568005	1,54826177	1,56533923	17,0314973
Mixed	1,61544	1,60705099	1,62382901	21,5338694

Figura B.3: Resultados para 6 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	2,23141516	2,23141516	2,23141516	3,64887216
Oblivious	2,235144	2,22537892	2,24490908	18,8193661
Greedy	2,233308	2,22357461	2,24304139	17,1603328
Thompson	2,078772	2,06877945	2,08876455	21,7699729
Mixed	2,163063	2,15312139	2,17300461	27,7210154

Figura B.4: Resultados para 8 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	2,7894849	2,7894849	2,7894849	29,1828362
Oblivious	2,781528	2,77020629	2,79284971	23,8367452
Greedy	2,7871635	2,77589981	2,79842719	22,6101654
Thompson	2,599317	2,5879101	2,6107239	29,4208982
Mixed	2,7008955	2,68956674	2,71222426	37,5583855

Figura B.5: Resultados para 10 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	3,34826135	3,34826135	3,34826135	181,877099
Oblivious	3,331824	3,31903943	3,34460857	28,1037933
Greedy	3,359688	3,34691887	3,37245713	25,5434837
Thompson	3,1429395	3,13012091	3,15575809	33,3682041
Mixed	3,240426	3,22776004	3,25309196	43,5743566

Figura B.6: Resultados para 12 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	3,90781728	3,90781728	3,90781728	1289,84686
Oblivious	3,8874	3,87300863	3,90179137	32,5358206
Greedy	3,9067455	3,89244709	3,92104391	29,552457
Thompson	3,684639	3,67054959	3,69872841	37,5887868
Mixed	3,78603	3,77173763	3,80032237	48,4114244

Figura B.7: Resultados para 14 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	4,46817389	4,46817389	4,46817389	6300,81767
Oblivious	4,460616	4,44488211	4,47634989	36,5383839
Greedy	4,4704575	4,4547301	4,4861849	33,5350933
Thompson	4,1928	4,17721416	4,20838584	42,3994927
Mixed	4,3469925	4,33120338	4,36278162	54,6782938

Figura B.8: Resultados para 16 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	5,02932202	5,02932202	5,02932202	25549,0232
Oblivious	4,983072	4,96572658	5,00041742	41,2976683
Greedy	5,013507	4,99627884	5,03073516	37,4143397
Thompson	4,7459505	4,72894821	4,76295279	48,609147
Mixed	4,8828555	4,86557052	4,90014048	62,1284922

Figura B.9: Resultados para 18 períodos

B.1.2. Caso b)

$$\alpha = \begin{pmatrix} 0,3 & 0,2 & 0,4 \\ 0,2 & 0,3 & 0,1 \end{pmatrix}$$

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	0,57294474	0,57294474	0,57294474	0,04078573
Oblivious	0,555144	0,55055927	0,55972873	5,62907514
Greedy	0,575478	0,57076542	0,58019058	5,15753573
Thompson	0,5029575	0,49804514	0,50786986	6,29043859
Mixed	0,544623	0,53980662	0,54943938	7,77122909

Figura B.10: Resultados para 2 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	1,17313124	1,17313124	1,17313124	0,08697946
Oblivious	1,09356	1,08611106	1,10100894	10,2692453
Greedy	1,173903	1,16664611	1,18115989	9,39349641
Thompson	1,0512075	1,04370944	1,05870556	11,839318
Mixed	1,1102205	1,10280995	1,11763105	14,81756

Figura B.11: Resultados para 4 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	1,78126132	1,78126132	1,78126132	0,46020755
Oblivious	1,626216	1,61532001	1,63711199	14,6502671
Greedy	1,781121	1,77126245	1,79097955	13,4228981
Thompson	1,616418	1,60642443	1,62641157	17,1649889
Mixed	1,69746	1,68747791	1,70744209	21,5215381

Figura B.12: Resultados para 6 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	2,39393747	2,39393747	2,39393747	3,83477607
Oblivious	2,141208	2,1265392	2,1558768	20,4549334
Greedy	2,405382	2,39265849	2,41810551	18,5566309
Thompson	2,184357	2,17154672	2,19716728	24,0726038
Mixed	2,296989	2,2842702	2,3097078	29,9480957

Figura B.13: Resultados para 8 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	3,00978581	3,00978581	3,00978581	29,3585019
Oblivious	2,67276	2,65433602	2,69118398	25,048937
Greedy	3,0011985	2,98545943	3,01693757	22,8553081
Thompson	2,770332	2,75475662	2,78590738	29,606724
Mixed	2,898111	2,8824495	2,9137725	37,469538

Figura B.14: Resultados para 10 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	3,62883872	3,62883872	3,62883872	189,020071
Oblivious	3,225576	3,20312812	3,24802388	29,9322267
Greedy	3,618699	3,59989525	3,63750275	27,3125529
Thompson	3,402249	3,383828	3,42067	35,479948
Mixed	3,5102055	3,49148576	3,52892524	44,4485135

Figura B.15: Resultados para 12 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	4,25155585	4,25155585	4,25155585	1353,49212
Oblivious	3,73188	3,70555189	3,75820811	34,8690962
Greedy	4,258716	4,23675008	4,28068192	31,5463824
Thompson	3,9703005	3,94848465	3,99211635	41,0494383
Mixed	4,1111895	4,08950791	4,13287109	52,2626858

Figura B.16: Resultados para 14 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	4,87723503	4,87723503	4,87723503	6236,80261
Oblivious	4,232664	4,20220046	4,26312754	37,2763975
Greedy	4,874361	4,84908312	4,89963888	35,1761759
Thompson	4,5723405	4,54759926	4,59708174	45,763267
Mixed	4,7248305	4,69975861	4,74990239	59,6101239

Figura B.17: Resultados para 16 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	5,50510311	5,50510311	5,50510311	25460,2061
Oblivious	4,764216	4,7296924	4,7987396	41,7176398
Greedy	5,5178805	5,48946678	5,54629422	37,7599958
Thompson	5,1600735	5,13200913	5,18813787	49,9845537
Mixed	5,3585205	5,33034786	5,38669314	63,7841542

Figura B.18: Resultados para 18 períodos

B.1.3. Caso c)

$$\alpha = \begin{pmatrix} 0,03 & 0,02 & 0,04 \\ 0,02 & 0,03 & 0,01 \end{pmatrix}$$

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	0,5822701	0,5822701	0,5822701	0,0446178
Oblivious	0,549888	0,54531459	0,55446141	5,61318929
Greedy	0,581799	0,57699584	0,58660216	5,19208887
Thompson	0,495573	0,49051308	0,50063292	6,41979584
Mixed	0,5602545	0,55541516	0,56509384	8,05571234

Figura B.19: Resultados para 2 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	1,21359408	1,21359408	1,21359408	0,08639168
Oblivious	1,083888	1,07564523	1,09213077	10,2771353
Greedy	1,220103	1,21248652	1,22771948	9,496875
Thompson	1,0699365	1,06194575	1,07792725	12,5642254
Mixed	1,1449395	1,13711572	1,15276328	15,1913288

Figura B.20: Resultados para 4 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	1,85889163	1,85889163	1,85889163	0,47117666
Oblivious	1,606536	1,59380308	1,61926892	14,8498931
Greedy	1,8447975	1,83408005	1,85551495	13,4155736
Thompson	1,6699815	1,65905941	1,68090359	17,6505899
Mixed	1,7813175	1,77054931	1,79208569	22,5188772

Figura B.21: Resultados para 6 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	2,51550191	2,51550191	2,51550191	3,86013956
Oblivious	2,130312	2,11284334	2,14778066	19,1344824
Greedy	2,502675	2,48849415	2,51685585	17,6164497
Thompson	2,3122755	2,29831858	2,32623242	22,7557485
Mixed	2,4279285	2,41375666	2,44210034	29,0741569

Figura B.22: Resultados para 8 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	3,17926838	3,17926838	3,17926838	29,2794513
Oblivious	2,662464	2,63988746	2,68504054	24,4429465
Greedy	3,147255	3,12949888	3,16501112	22,0220713
Thompson	2,95242	2,93502344	2,96981656	28,427128
Mixed	3,0651225	3,04753163	3,08271337	37,4819597

Figura B.23: Resultados para 10 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	3,84662885	3,84662885	3,84662885	178,212798
Oblivious	3,189192	3,16161085	3,21677315	28,1677156
Greedy	3,805296	3,78382449	3,82676751	25,3646628
Thompson	3,6057435	3,58498557	3,62650143	33,4877094
Mixed	3,7453305	3,72403049	3,76663051	42,765786

Figura B.24: Resultados para 12 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	4,51571449	4,51571449	4,51571449	1327,60312
Oblivious	3,683856	3,65116656	3,71654544	32,651927
Greedy	4,4640795	4,43880764	4,48935136	29,5977993
Thompson	4,260915	4,23659692	4,28523308	41,6069055
Mixed	4,3828245	4,35782714	4,40782186	50,9681425

Figura B.25: Resultados para 14 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	5,1856627	5,1856627	5,1856627	6233,79324
Oblivious	4,237176	4,19946571	4,27488629	37,1749581
Greedy	5,076711	5,04731961	5,10610239	33,3947754
Thompson	4,9117995	4,88391541	4,93968359	44,0624175
Mixed	5,051445	5,02254171	5,08034829	56,1582827

Figura B.26: Resultados para 16 períodos

Método	Beneficio	Intervalo de confianza		Tiempo [s]
Optimal	5,85607938	5,85607938	5,85607938	25413,5803
Oblivious	4,763136	4,72027232	4,80599968	41,9829087
Greedy	5,751306	5,71828737	5,78432463	37,4035631
Thompson	5,5525395	5,52053087	5,58454813	47,5814538
Mixed	5,7602925	5,72764754	5,79293746	60,8610192

Figura B.27: Resultados para 18 períodos

B.2. Prior Opuesto a la Realidad

Para los resultados expuestos a continuación, se consideró los siguientes parámetros:

- $|I| = 2$
- $C = 1$
- $Q = 3$
- $\beta = 3$
- $\gamma = -6$
- $r = (0,4 \quad 0,7)$

Se consideró que

$$\alpha = \begin{pmatrix} 0,7 & 0,3 & 0,1 \\ 0,1 & 0,2 & 0,8 \end{pmatrix}$$

Sin embargo, se impuso la siguiente distribución de probabilidades de la calidad para todas las simulaciones: $p_1 = (0,05 \ 0,15 \ 0,8)$, y $p_2 = (0,8 \ 0,1 \ 0,1)$. El beneficio de cada heurística fue obtenido simulando el proceso 20000 veces y calculando el ingreso promedio.

Método	Beneficio	Intervalo de confianza	
Oblivious	12,495595	12,4395386	12,5516514
Greedy	16,03882	15,9566094	16,1210306
Thompson	28,08046	27,9974283	28,1634917
Mixed	28,10391	28,0094215	28,1983985

Figura B.28: Resultados para 100 períodos

Método	Beneficio	Intervalo de confianza	
Oblivious	45,852625	45,7037276	46,0015224
Greedy	98,5326	97,8060006	99,2591994
Thompson	169,51457	169,381483	169,647657
Mixed	170,899825	170,736911	171,062739

Figura B.29: Resultados para 500 períodos

Método	Beneficio	Intervalo de confianza	
Oblivious	100,523535	100,2743	100,77277
Greedy	257,166935	255,173802	259,160068
Thompson	420,36826	420,212819	420,523701
Mixed	423,15539	422,989547	423,321233

Figura B.30: Resultados para 1200 períodos

B.3. Órdenes de Magnitud de α

Para los resultados expuestos a continuación, se consideró los siguientes parámetros:

- $|I| = 2$
- $C = 1$
- $Q = 3$
- $\beta = 3$
- $\gamma = -6$
- $r = (0,48 \ 0,51)$

El beneficio de cada heurística fue obtenido simulando el proceso 20000 veces y calculando el ingreso promedio.

B.3.1. Caso a)

$$\alpha = \begin{pmatrix} 3 & 2 & 4 \\ 2 & 3 & 1 \end{pmatrix}$$

Método	Beneficio	Intervalo de confianza	
Oblivious	27,484272	27,3854344	27,5831096
Greedy	28,2564735	28,1674539	28,3454931
Thompson	27,5524965	27,4646225	27,6403705
Mixed	27,980403	27,8927986	28,0680074

Figura B.31: Resultados para 100 períodos

Método	Beneficio	Intervalo de confianza	
Oblivious	137,246448	136,691643	137,801253
Greedy	143,243517	142,750962	143,736072
Thompson	143,107883	142,616265	143,5995
Mixed	144,529186	144,041257	145,017116

Figura B.32: Resultados para 500 períodos

Método	Beneficio	Intervalo de confianza	
Oblivious	330,831816	329,4414	332,222232
Greedy	343,88949	342,672881	345,106099
Thompson	348,511656	347,300487	349,722825
Mixed	349,864931	348,653947	351,075914

Figura B.33: Resultados para 1200 períodos

B.3.2. Caso b)

$$\alpha = \begin{pmatrix} 0,3 & 0,2 & 0,4 \\ 0,2 & 0,3 & 0,1 \end{pmatrix}$$

Método	Beneficio	Intervalo de confianza	
Oblivious	26,58996	26,3808031	26,7991169
Greedy	31,1652585	30,9967119	31,3338051
Thompson	30,973944	30,8032578	31,1446302
Mixed	31,4752395	31,3055273	31,6449517

Figura B.34: Resultados para 100 períodos

Método	Beneficio	Intervalo de confianza	
Oblivious	133,603176	132,525198	134,681154
Greedy	157,841306	156,973968	158,708643
Thompson	161,130629	160,250021	162,011236
Mixed	161,852565	160,980634	162,724496

Figura B.35: Resultados para 500 períodos

Método	Beneficio	Intervalo de confianza	
Oblivious	320,15976	317,553685	322,765835
Greedy	378,501399	376,39973	380,603068
Thompson	389,853005	387,73954	391,966469
Mixed	390,565916	388,451334	392,680497

Figura B.36: Resultados para 1200 períodos

B.3.3. Caso c)

$$\alpha = \begin{pmatrix} 0,03 & 0,02 & 0,04 \\ 0,02 & 0,03 & 0,01 \end{pmatrix}$$

Método	Beneficio	Intervalo de confianza	
Oblivious	26,132136	25,8778947	26,3863773
Greedy	32,8189335	32,6247772	33,0130898
Thompson	33,11334	32,9159324	33,3107476
Mixed	33,299004	33,1025189	33,4954891

Figura B.37: Resultados para 100 períodos

Método	Beneficio	Intervalo de confianza	
Oblivious	133,224048	131,948966	134,49913
Greedy	163,584545	162,591036	164,578053
Thompson	167,938539	166,921844	168,955234
Mixed	168,427236	167,414985	169,439487

Figura B.38: Resultados para 500 períodos

Método	Beneficio	Intervalo de confianza	
Oblivious	315,028968	311,948135	318,109801
Greedy	394,671726	392,270679	397,072773
Thompson	404,743958	402,301431	407,186484
Mixed	406,880304	404,453224	409,307384

Figura B.39: Resultados para 1200 períodos

B.4. Incremento del número de alternativas

Para los resultados expuestos a continuación, se consideró los siguientes parámetros:

- $|I| = 10$
- $C = 5$
- $Q = 3$
- $\beta = 3$
- $\gamma = -6$

El beneficio de cada heurística fue obtenido simulando el proceso 100 veces y calculando el ingreso promedio.

B.4.1. Órdenes de magnitud de α

Se consideró que

$$r = (0,48 \ 0,46 \ 0,51 \ 0,33 \ 0,63 \ 0,41 \ 0,58 \ 0,69 \ 0,52 \ 0,55)$$

a)

$$\alpha = \begin{pmatrix} 5 & 3 & 6 \\ 7 & 8 & 7 \\ 9 & 2 & 7 \\ 2 & 3 & 4 \\ 2 & 1 & 3 \\ 4 & 5 & 6 \\ 3 & 3 & 2 \\ 1 & 7 & 4 \\ 3 & 2 & 4 \\ 7 & 6 & 6 \end{pmatrix}$$

Método	Beneficio	Cientes totales	Beneficio/Ciente	Intervalo de confianza
Oblivious	631,3478	1176,094934	0,538915178	0,532052966 0,54577739
Greedy	632,5034	1166,964654	0,542578453	0,537105389 0,548051518
Thompson	632,5503	1164,883037	0,543897273	0,537723085 0,550071461
Mixed	628,8576	1153,231929	0,545941191	0,539949443 0,551932939

Figura B.40: Resultados caso a)

b)

$$\alpha = \begin{pmatrix} 0,5 & 0,3 & 0,6 \\ 0,7 & 0,8 & 0,7 \\ 0,9 & 0,2 & 0,7 \\ 0,2 & 0,3 & 0,4 \\ 0,2 & 0,1 & 0,3 \\ 0,4 & 0,5 & 0,6 \\ 0,3 & 0,3 & 0,2 \\ 0,1 & 0,7 & 0,4 \\ 0,3 & 0,2 & 0,4 \\ 0,7 & 0,6 & 0,6 \end{pmatrix}$$

Método	Beneficio	Cientes totales	Beneficio/Ciente	Intervalo de confianza
Oblivious	622,5918	1149,569057	0,558781161	0,543887062 0,573675259
Greedy	616,7636	1102,485559	0,560736588	0,551049525 0,570423651
Thompson	624,1192	1105,891753	0,567749082	0,555458107 0,580040058
Mixed	627,0528	1100,190648	0,57149155	0,562110272 0,580872828

Figura B.41: Resultados caso b)

c)

$$\alpha = \begin{pmatrix} 0,05 & 0,03 & 0,06 \\ 0,07 & 0,08 & 0,07 \\ 0,09 & 0,02 & 0,07 \\ 0,02 & 0,03 & 0,04 \\ 0,02 & 0,01 & 0,03 \\ 0,04 & 0,05 & 0,06 \\ 0,03 & 0,03 & 0,02 \\ 0,01 & 0,07 & 0,04 \\ 0,03 & 0,02 & 0,04 \\ 0,07 & 0,06 & 0,06 \end{pmatrix}$$

Método	Beneficio	Cientes totales	Beneficio/Cliente	Intervalo de confianza
Oblivious	620,5395	1170,366545	0,559621478	0,5416973 0,577545656
Greedy	610,4145	1062,362272	0,575501223	0,56345828 0,587544167
Thompson	619,516	1062,252081	0,585007026	0,572568328 0,597445723
Mixed	622,0469	1055,76929	0,58946664	0,577673125 0,601260155

Figura B.42: Resultados caso c)

B.4.2. Prior opuesto a la realidad

Se consideró que

$$r = (0,1 \ 0,15 \ 0,2 \ 0,25 \ 0,3 \ 0,35 \ 0,4 \ 0,45 \ 0,5 \ 0,55)$$

Además,

$$\alpha = \begin{pmatrix} 0,9 & 0,2 & 0,1 \\ 0,9 & 0,5 & 0,1 \\ 0,8 & 0,5 & 0,2 \\ 0,7 & 0,5 & 0,3 \\ 0,6 & 0,5 & 0,4 \\ 0,4 & 0,5 & 0,6 \\ 0,3 & 0,5 & 0,7 \\ 0,2 & 0,5 & 0,8 \\ 0,1 & 0,5 & 0,9 \\ 0,1 & 0,2 & 0,9 \end{pmatrix}$$

Sin embargo, se impuso la siguiente distribución de probabilidades de la calidad para todas las simulaciones:

- $p_1 = (0,05 \ 0,05 \ 0,9)$
- $p_2 = (0,05 \ 0,15 \ 0,8)$
- $p_3 = (0,15 \ 0,15 \ 0,7)$

- $p_4 = (0,2 \quad 0,2 \quad 0,6)$
- $p_5 = (0,2 \quad 0,3 \quad 0,5)$
- $p_6 = (0,5 \quad 0,3 \quad 0,2)$
- $p_7 = (0,6 \quad 0,2 \quad 0,2)$
- $p_8 = (0,7 \quad 0,15 \quad 0,15)$
- $p_9 = (0,8 \quad 0,15 \quad 0,05)$
- $p_{10} = (0,9 \quad 0,05 \quad 0,05)$

Método	Beneficio	Cientes totales	Beneficio/Cliente	Intervalo de confianza
Oblivious	521,321	6159,343949	0,084849327	0,084019267 0,085679387
Greedy	349,549	1394,495051	0,252567498	0,249571142 0,255563854
Thompson	338,432	1317,786586	0,256941459	0,255523373 0,258359545
Mixed	339,096	1318,868486	0,257200876	0,255843213 0,258558539

Figura B.43: Resultados prior opuesto a la realidad

B.4.3. Parámetros aleatorios

α y r son generados aleatoriamente en cada simulación, siguiendo cada elemento de la matriz y del vector una distribución uniforme entre 0 y 1.

Método	Beneficio	Cientes totales	Beneficio/Cliente	Intervalo de confianza
Oblivious	807,368477	1267,912332	0,679477627	0,37595781 0,982997445
Greedy	806,855065	1142,455111	0,713926809	0,454027768 0,973825849
Thompson	791,655771	1163,311129	0,687752487	0,4530506 0,922454374
Mixed	818,995094	1169,434269	0,709334445	0,477593206 0,941075685

Figura B.44: Resultados parámetros aleatorios