Departamento de Computación
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Tesis de Licenciatura en Ciencias de la Computación

Subastas estocásticas veraces para búsqueda patrocinada

Pablo Ariel Heiber - pheiber@dc.uba.ar - L.U. 589/02

Director: Esteban Feuerstein

Abril de 2009

A muchos, pero no a todos

Resumen

La venta de publicidad asociada a los resultados de las búsquedas o a los contenidos de una página web, conocida como búsqueda patrocinada, se ha convertido en la mayor fuente de ingresos de las principales empresas del sector, y su importancia relativa continúa creciendo. El esquema más frecuentemente utilizado en ese contexto es el de un conjunto de anunciantes que compiten por la asignación de un conjunto limitado de espacios de publicidad y pagan al editor cuando un usuario hace click en su aviso. Las características particulares de esta actividad han motivado el surgimiento de muchos e interesantes problemas abordables desde distintas disciplinas como la informática, la optimización, la economía y hasta la sociología, en forma mono o multidisciplinaria. Las decisiones acerca de cuántos y cuáles avisos elegir, en qué orden mostrarlos, cómo y cuánto cobrar por el servicio, abren un amplio campo de investigación que vincula las disciplinas mencionadas. Éstas buscan de mecanismos que permitan satisfacer los intereses de los distintos actores involucrados (editores, anunciantes y usuarios), con propiedades de eficiencia y practicidad adecuadas a las características de la aplicación (masividad, velocidad de respuesta requerida, propiedades económicas o de teoría de juegos, etc.).

En esta tesis buscamos analizar propiedades teóricas generales sobre las formas que toman dichas decisiones, así como proponer algunos casos particulares, en forma de algoritmos, con interesantes propiedades. En particular, buscamos una caracterización de la familia de subastas veraces, es decir, mecanismos de selección y pago que incentivan a los avisadores a revelar el valor real que le asignan a un click en su aviso. A partir de dicha caracterización, desarrollaremos distintas herramientas para construir subastas veraces o extender subastas existentes conservando buenas propiedades. Los resultados obtenidos están orientados a proveer nuevas herramientas a los participantes de esta importante y novedosa actividad económica.

Como resultados principales, presentamos la primer clase de subastas veraces para el contexto de las búsquedas patrocinadas sin más restricciones. Dicha clase es una generalización de las subastas *escalonadas* presentadas en [AGM06] y de las subastas estocásticas con método de cobro *condex* presentadas en [MCW05], consolidando ambos contextos en un mismo entorno de trabajo. También probamos que, bajo ciertas condiciones normales para el contexto de aplicación, la clase que damos es la única, lo que nos lleva a un importante corolario de caracterización de todas las subastas veraces para búsquedas patrocinadas en términos puramente aritméticos, que puede tomar el lugar de la definición común, basada en el comportamiento de los agentes involucrados, más difícil de manejar en muchos casos.

Finalmente, introducimos métodos para, dadas subastas simples que cumplen la propiedad de veracidad, generar otras más complejas que la mantienen, con garantías respecto de los cambios en el beneficio del editor al introducir las modificaciones. En el proceso, encontramos cotas superiores a dicho beneficio al pasar de un entorno con un solo aviso publicado a muchos, o al pasar de considerar todos los espacios disponibles como idénticos, a diferenciarlo según su visibilidad. Estas extensiones tienen aplicaciones prácticas directas, que también mencionamos.

Abstract

Selling advertising associated with search results or the content of a web, practice known as sponsored search, has been the major source of income of the main providers of web services, and its relative importance is still growing. The most frequently used framework consist of a set of advertisers competing for a limited set of advertisement slots and pay to the editor when a user clicks on their ad. The particular characteristics of this activity motivated the origin of many interesting problems for many fields, such as informatics, optimization, economy and even sociology, in a mono- or multi-disciplinarian way. The decisions about how many and which ads to show, in which order, how and how much to charge for the service, open a wide field for research for all mentioned disciplines. These search for mechanisms that satisfy the interest of the different involved actors (editors, advertisers and users), with efficiency and practicity properties suitable for the characteristics of the application (massivity, speed of the answer, economic or game theoretic properties, etc.).

In this thesis we analyze general theoretical properties of the mentioned decisions, as well as propose some particular cases, in the form of algorithms, with interesting properties. In particular, we aim for a characterization of the family of truthful auctions, i.e., mechanisms of selection and payment that incentive the advertisers to reveal their true value they assign to a click in their ad. From such a characterization, we develop many tools to build truthful auctions or extend existing auctions while preserving some good properties. The obtained results are oriented to provide new tools to the participants of this important and novel economic activity.

As our main results, we present the first class of truthful auctions for the context of sponsored search without further restrictions. This class generalizes the *laddered* auctions presented in [AGM06] and the stochastic auctions with condex pricing presented in [MCW05], joining both results under the same framework. We also proved that under certain conditions usual in the context of application, the class we provide is unique, which leads us to an important corollary of a purely arithmetic characterization of all truthful auctions for sponsored search, that can replace the regular definition, based on the behavior of the involved agents, which is more difficult to handle in many cases.

Finally, we introduce some methods to, given simple truthful auctions, generate other more complex that maintain truthfulness, with guarantees in respect of the changes on the editor revenue when introducing the modification. In the process, we found upper bounds on the mentioned revenue when passing from a single-slot framework to a multi-slot one, or when going from many identical slots to a context in which slots have different visibility. This extensions have direct practical applications, which we also mention.

Índice general

1. Introducción **3**

1.1. Búsquedas patrocinadas . . . 3

1.2. Motivación . . . 6

1.2.1. ¿Por qué estocásticas? . . . 6

1.2.2. ¿Por qué veraces? . . . 7

1.3. Trabajo relacionado y estado del arte . . . 8

1.4. Trabajo realizado en la tesis . . . 9

1.5. Organización del resto de la presentación . . . 10

2. Preliminares **11**

2.1. Definición del problema y notación . . . 11

2.2. Trabajo previo . . . 15

2.2.1. Subastas determinísticas . . . 15

2.2.2. Subastas con todas las posiciones iguales . . . 15

3. Diseño y caracterización de subastas veraces **17**

3.1. Determinización de precios . . . 17

3.2. Subastas de precio determinístico veraces . . . 19

3.3. Subastas veraces determinísticas . . . 23

3.4. Caracterización completa de subastas veraces . . . 25

3.5. Cantidad de espacios variable . . . 25

4. Extensiones para generar subastas veraces **27**

4.1. Algoritmos auxiliares . . . 27

4.2. De un espacio a múltiples espacios distintos . . . 29

4.3. Extensiones lineales a múltiples espacios . . . 33

5. Conclusiones y trabajo futuro **37**

5.1. Conclusiones . . . 37

5.2. Trabajo Futuro 37

A. Algoritmos adicionales **39**

Bibliografía **41**

Capítulo 1

Introducción

“*Un subastador es aquel que proclama con un martillo que acaba de robarle a alguien con la lengua.*”

Ambrose Bierce

1.1. Búsquedas patrocinadas

La fuente de ingresos más importante en los últimos tiempos para las grandes empresas proveedoras de servicios en la web (como Google o Yahoo!) es la que obtienen al vender espacios de publicidad. Dicha publicidad aparece en diversos lugares: junto a los resultados de una búsqueda (ver Figura 1.1), dentro de los servicios de correo electrónico que esas empresas proveen, o incluso en páginas personales que contratan a la empresa para tercerizar sus ingresos publicitarios (mediante servicios como AdSense, de Google) (Figura 1.2(a)). Últimamente, también se ha comenzado a introducir publicidad en otro tipo de medios no textuales, como videos (Figura 1.2(b)).

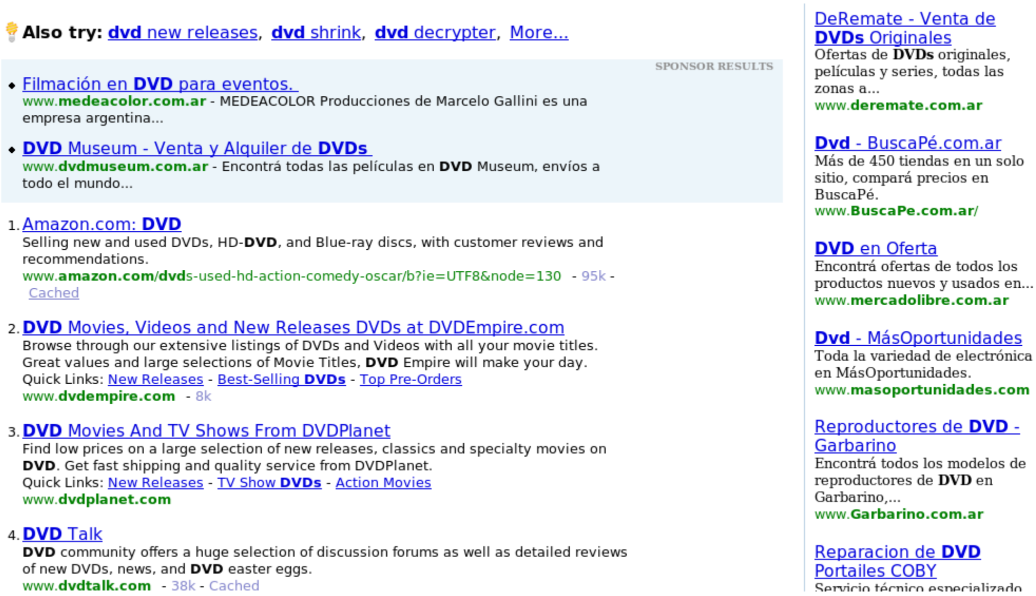
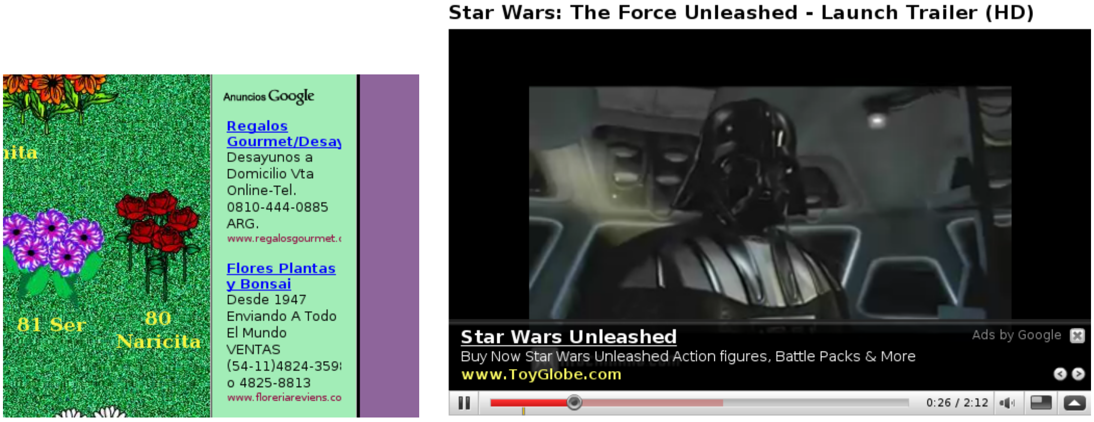


Figura 1.1: Resultados algorítmicos y patrocinados como resultado de la búsqueda “DVD”

La publicidad en internet es cada vez más predominante entre las formas de hacer conocer productos



(a) Avisos publicitarios en una página

(b) Avisos publicitarios en un video de YouTube

Figura 1.2: Avisos publicitarios en otros medios

y servicios, a medida que la inserción de Internet en la sociedad crece con rapidez. Entre otras cosas, esto se debe a que capta lo que se conoce como la "ancha cola de la distribución" en los clientes publicitarios. Al estar los espacios muy precisamente segmentados (comparado con una revista, un programa de televisión o la publicidad callejera en forma de afiches o volantes), permite a empresas pequeñas ubicar sus avisos en lugares muy buenos. Sin embargo en los medios masivos estos no tienen competitividad, debido a que los espacios son copados por empresas grandes con más presupuesto publicitario, quedando los pequeños relegados a la publicidad en medios locales o de baja masividad. Este fenómeno de publicidad en espacios virtuales generados automáticamente se hizo famoso y creció de la mano de los buscadores y los espacios de publicidad en las búsquedas, por lo cual se conoce normalmente a toda esta publicidad con el nombre particular de "búsqueda patrocinada", incluso cuando, como ya dijimos anteriormente, esta forma de monetarización se está esparciendo a otros servicios provistos en la web.

Dada la naturaleza diversa del contenido utilizado como generador de espacios de publicidad y su inmenso volumen, a los que se añaden la diversidad y el volumen de los avisos mismos, los espacios no pueden venderse directamente como se hace en los medios de publicidad tradicionales (televisión, radio, prensa, etc.). Otro impedimento para ello es que parte del contenido (en particular, las búsquedas) y por lo tanto, los espacios de publicidad que lo acompañan, son generados automáticamente, lo que hace imposible asignarlos a priori. Para resolver el mencionado problema, se utiliza un enfoque automático. Los avisadores dan un conjunto de palabras o frases clave que representan el tipo de contenido al cual les interesa que se asocie su aviso. Por ejemplo, un concesionario de automóviles podría querer que su aviso se muestre cerca de una búsqueda que contenga la frase "compra auto", o la palabra "vehículo", o cerca de una página con datos de automotores, o en un video sobre carreras de autos.

Además de tomar la información adicional de los avisadores acerca de sus palabras clave, los proveedores de servicios deben, para cada formato de contenido, generar palabras o frases clave que lo representen. Esto es simple en los casos de búsquedas, un poco más difícil en medios textuales, y aún más en los medios no textuales. Una vez que se tienen las frases o palabras clave tanto de los avisos como del contenido, se puede hacer un apareamiento automático de qué conjunto de avisos son relevantes para un contenido en particular, es decir, qué conjunto de avisos es factible mostrar en los espacios asociados a dicho contenido. Luego, de ese conjunto, la empresa debe decidir, también automáticamente, qué avisos mostrar y, generalmente, cómo mostrarlos, ya que la ubicación espacial de un aviso (i.e., el espacio específico que se le asigna entre todos los posibles asociados al mismo contenido) afecta severamente su visibilidad. Por ejemplo, en los espacios al lado de búsquedas como el ejemplo de la Figura 1.1, los avisos que aparecen sobre los resultados de la búsqueda son los que obtienen más visibilidad y, entre los que aparecen en el costado derecho, los superiores son más visibles que los inferiores.

El modelo de negocio actual para las búsquedas patrocinadas, además de la ya mencionada asignación automática de espacios, incluye un modelo de pago particular. En los inicios de la actividad la forma de pago era similar a la publicidad tradicional, es decir, se pagaba una cantidad de dinero por cada impresión que se hacía del aviso (*pay per impression* en inglés). Esto tenía problemas de negocios, ya que el pago no dependía de si el proceso de decisión llevado a cabo por el proveedor del servicio de poner el aviso en un lugar determinado era bueno o malo. En la publicidad tradicional, es el avisador el que elige el espacio y por lo tanto es natural que cargue con la responsabilidad de pagar independientemente de la eficacia efectiva de poner el aviso en ese lugar particular, pero en las búsquedas patrocinadas el proveedor del servicio también participa de dicha decisión, por lo cual debería dar alguna garantía de efectividad. En concreto, en el modelo de pago por impresión el avisador podría estar pagando impresiones irrelevantes para él, si el proceso de apareo entre avisos y espacios del proveedor del servicio no fuera bueno. Para mitigar esos problemas, se introdujo una forma de garantía de calidad, cambiando el modelo de pago al llamado pago por click (*pay per click*). Los avisos en este modelo incluyen un vínculo al sitio que el avisador desea, y éste paga sólo por las visitas a su sitio producto de clicks en su aviso. En los últimos años se ha estado avanzando hacia un modelo con una garantía aún mayor para el avisador, que es el pago por acción (*pay per action*), o sea, el avisador paga por cada venta que realiza gracias a su aviso. Un modelo como el pago por acción implica, por supuesto, nuevos problemas de distintos campos que están fuera del alcance de este trabajo. Como este último modelo todavía no está ampliamente implementado, durante este trabajo asumiremos el modelo de pago por click.

Hemos mencionado que la selección de avisos es automática, asignando a cada espacio un aviso de un conjunto generado a partir de claves. También dijimos que el pago que realiza cada avisador es proporcional a la cantidad de clicks que recibe su aviso. Falta, entonces, el paso intermedio entre estos dos eventos. ¿Cómo elegir, del conjunto de avisos relevantes ya generado, la asignación de algunos de éstos a los espacios y un precio por click para cada uno? Ésta es la pregunta principal que da origen a este trabajo. Algunos componentes de la respuesta son dados, por ejemplo, que la forma de selección es una subasta. Esto quiere decir que cada avisador, además del aviso, el vínculo y las palabras clave ya mencionadas, debe indicar al proveedor del servicio una oferta. Dicha oferta es interpretada, en principio, como el precio que estaría dispuesto a pagar por cada click. Quizás contradiciendo la intuición inicial, tomar directamente la oferta como el precio por click trae ciertos problemas que mencionaremos más adelante. Los mecanismos utilizados actualmente y otros todavía no puestos en práctica pero considerados en la investigación, utilizan la oferta como una cota superior al precio por click y generan el precio efectivo a cobrar mediante diversos mecanismos. La subasta, entonces, se da cuando la empresa proveedora de servicio, basándose en las ofertas de los avisadores (y potencialmente en otra información sobre ellos, como veremos más adelante) decide qué aviso mostrar en cada espacio disponible y qué precio cobrarle a cada uno. Profundizaremos sobre esto en el próximo capítulo.

En diversos puntos del proceso descripto hay temas de interés para las Ciencias de la Computación. Como ejemplos podemos poner la extracción automática de palabras clave, tanto de contenido textual como no textual, el apareamiento de avisos, los métodos de selección y de cobro, la detección de fraudes como la generación de clicks falsos, el SPAM de avisos y otros. Para este trabajo nos centraremos en los mecanismos de selección y cobro y propiedades teóricas que éstos pueden o no tener y que estudian el comportamiento estratégico de los avisadores y sus ofertas, que intentan maximizar su propio beneficio, y su relación con el proveedor del servicio, que intenta maximizar el suyo. También abordamos, aunque parcialmente, las dificultades para cuantificar dicho beneficio, en particular el del proveedor del servicio, y presentamos algunas aproximaciones para hacerlo.

1.2. Motivación

Como mencionamos en la sección anterior, en este trabajo lidiaremos con los problemas de selección de avisos a publicar y los mecanismos de cobro a los avisadores. El conjunto de ambos es la subasta de la que hablamos. Las subastas fueron estudiadas por las ciencias económicas desde mucho antes de que existiera el tipo de actividad que aquí tratamos, y en este trabajo intentamos recuperar algunos conceptos de dicho análisis (como el de veracidad, que explicaremos más adelante) sin perder el punto de vista computacional.

El estudio del mecanismo de subastas de avisos en internet es de una gran importancia práctica, ya que cualquier cambio realizado en él afecta a empresas muy grandes como lo son las proveedoras de servicios en la web, y también a muchos negocios menores que la utilizan como principal medio publicitario. La posibilidad de utilizar herramientas teóricas para analizar las consecuencias de dichos cambios, entonces, es prácticamente una necesidad. Éstas herramientas deben ser lo suficientemente generales como para poder analizar en conjunto familias de subastas y para poder elegir la más conveniente, pero también deben ser de fácil aplicación para resultar prácticas, ya que una herramienta demasiado general, pero a su vez demasiado abstracta, no resulta de utilidad a la hora de comparar subastas del mundo real.

Desde un punto de vista teórico, el interés de nuestro trabajo consiste en darle una mirada computacional a conceptos de la economía y entregar resultados en los que, si bien se combina el conocimiento de ambos campos, las técnicas utilizadas provienen del mundo de la computación, y por lo tanto, deberían resultarle familiares a investigadores que provengan de esa disciplina. La actualidad del tema y el hecho de que cada vez más investigadores del mundo de la computación estén adentrándose en el estudio de las búsquedas patrocinadas son motivaciones adicionales para el intento de formular bases teóricas para ese marco desde el punto de vista de las ciencias de la computación.

En este trabajo, como su título lo indica, estudiaremos subastas estocásticas y veraces en el contexto de las búsquedas patrocinadas. A continuación presentaremos las principales razones por las cuales creemos que subastas con esas características son interesantes y merecedoras de un estudio.

1.2.1. ¿Por qué estocásticas?

En este trabajo estudiamos principalmente subastas estocásticas. Esto no quiere decir que dejemos de lado las determinísticas, sino que las vemos sólo como un caso particular de éstas y pretendemos que los resultados puedan aplicarse a cualquier subasta. Es importante notar que las subastas que tratamos aquí son subastas repetitivas. Esto quiere decir que se subastan los mismos espacios para el mismo conjunto de avisos candidatos posiblemente varias veces. Por ejemplo, cada vez que la misma búsqueda, o una similar, se repite. Este hecho implica que el comportamiento de los avisadores es potencialmente realimentado por los resultados previos, generando un ciclo de influencias entre las distintas instancias de la misma subasta y las ofertas de los avisadores.

La motivación para utilizar subastas estocásticas aparece en gran parte de la bibliografía mencionada al respecto en la Sección 1.3, y tiene que ver principalmente con brindar diversidad a los avisos que se publican cuando la búsqueda se repite. Es un escenario normal que una persona realice la misma búsqueda en distintos momentos para reencontrarse con sus resultados, o incluso que realice búsquedas similares iterativamente para refinar lo obtenido, por lo que la repetición de avisos exactamente iguales no sólo desperdicia oportunidades de venta, sino que también puede dar una imagen negativa al usuario. Enumeraremos entonces las principales ventajas de usar subastas estocásticas:

- Las subastas estocásticas son menos vulnerables a los avisadores que intenten ofertar de forma estratégica o incluso maliciosa. En primer lugar, el azar agregado a la forma de producir resultados le hace más difícil a un avisador obtener datos acerca de los avisadores rivales para alimentar su estrategia. En el caso de ofertas maliciosas (por ejemplo, realizadas solamente para subir el precio que pagarán otros avisadores) introducen el riesgo de perder dinero en el corto plazo, ya que el avisador malicioso oferta un importe más alto que el valor real de un click para él, pero no puede asegurarse que su aviso no vaya a ser publicado y clickeado, en cuyo caso pagaría el precio que su oferta falsa induzca, potencialmente más caro que el beneficio que obtiene por el click.

- El hecho de que cualquiera de los avisos (o, al menos, una cantidad mayor que en las subastas determinísticas) tengan la posibilidad de ser mostrados contribuye a tener una mayor base de avisadores, lo que implica más ganancias a mediano plazo (por ejemplo, ver [Kle99], sección 8.2, y [BK96]).

- El incremento en la variedad de los avisos publicados trae ventajas como una mejor experiencia para el usuario, mejor cobertura de sus posibles intenciones y necesidades y por lo tanto mayor posibilidad de que algún aviso produzca ganancias (recordar que esto sólo sucede ante un click).

- Para determinar cuáles son los mejores avisos, se debe estimar la probabilidad de que un aviso en un espacio dado genere un click. Éstas estimaciones se basan en apariciones previas del mismo aviso en el mismo espacio y pueden tener errores. Para evitar dejar afuera avisos con gran esperanza de ganancia por una mala estimación o incluso por falta de información al respecto, se debe alternar entre todos los avisos para mantener dicha información actualizada. Esto se conoce como el equilibrio entre explorar y explotar [PO06] (*explore/exploit trade-off*). Se entiende por explorar a publicar avisos para los cuales no se tiene información fidedigna y explotar a publicar avisos que permitan obtener las mayores ganancias. El equilibrio, entonces, está en poder combinarlas, ya que ambas son necesarias a largo plazo. Una subasta estocástica correctamente balanceada provee una forma implícita de proveer dicho equilibrio [FHMVBY07].

- Los mecanismos estocásticos en general son menos vulnerables al comportamiento fraudulento, tanto de usuarios, como de avisos. Como ejemplo más importante, tenemos el fraude de clicks [Pen04], que consiste en generar clicks sin intención verdadera de visitar la página del avisador, sólo para hacerlo pagar por ellos. Dicho tipo de fraude es más difícil de llevar a cabo cuando no es posible predecir cómo lograr que cierto aviso o ciertos avisos sean publicados, o más específicamente, que sean publicados en cierto espacio particular.

1.2.2. ¿Por qué veraces?

Las subastas veraces inducen a los ofertantes a revelar el verdadero valor que dan a los ítems subastados (en este caso, clicks en sus avisos). Esto quiere decir que cada avisador maximizará su propia ganancia, independientemente de lo que hagan los otros, si oferta exactamente lo que vale un click para él. Las ventajas de una subasta con esta propiedad las podemos resumir en los siguientes puntos:

- Es una forma definitiva de disuadir el comportamiento estratégico o malicioso de los ofertantes, ya que para poder variar su oferta acorde a su estrategia se verían obligados a resignar ganancias.

- La subasta vista como un juego de múltiples jugadores, tiene un equilibrio dominante (aquel donde cada avisador oferta exactamente su valor), lo cual simplifica enormemente los análisis teóricos que se quieran hacer. En particular esto también implica que la subasta tiene un único equilibrio de Nash.

- El hecho de que el punto de maximización de la ganancia de un avisador particular no dependa para nada de los otros avisadores le da estabilidad al sistema como juego repetitivo, lo cual redunda en estabilidad de entradas para el proveedor de servicios y es inmune a problemas de ciclos viciosos descendentes en los precios, o sea, que la estrategia utilizada para maximizar ganancia de los avisadores genere una competencia que resulta en ir bajando las ofertas de todos iterativamente, lo que redunda en una baja de precios.
- El hecho de que el mecanismo mismo asegure las ganancias simplemente revelando el valor del click da más confianza a los potenciales avisadores, ya que por un lado no tienen la necesidad de pensar el juego estratégicamente (lo cual puede requerir costos adicionales de software o personal especializado) y pueden verlo como una inversión publicitaria normal, y por otro lado un modelo de negocio con estas propiedades aparece como más confiable.
- Por su misma definición, las subastas veraces eliminan todo tipo de incentivo a ofertar por debajo o por encima del verdadero valor. En particular, las subastas utilizadas hoy en día (llamadas subastas de segundo precio) no dan incentivo a los ofertantes de sobre-ofertar, pero sí de sub-ofertar, lo que genera problemas ya mencionados sobre las bajas de precios.

1.3. Trabajo relacionado y estado del arte

El trabajo de tesis se basa principalmente en 2 trabajos previos de gran notoriedad en el área de las búsquedas patrocinadas: uno de Aggarwal, Goel y Motwani [AGM06] y otro de Meek, Chickering y Wilson [MCW05]. El primero da una forma de construir subastas determinísticas veraces útil para el contexto de búsquedas auspiciadas y demuestra su unicidad, mientras el segundo da una forma de construir subastas estocásticas veraces, pero sólo cuando todos los ítems a ser subastados son iguales (es decir, para cada ofertante es indistinto obtener cualquiera de los ítems), generalizando el famoso trabajo de Vickrey [Vic61]. El resultado principal de nuestro trabajo generaliza a los dos citados dando una forma de construir subastas potencialmente estocásticas veraces donde los ítems pueden tener un valor distinto para cada avisador, bajo una suposición común en el contexto de búsquedas patrocinadas, demostrando también su unicidad. Sus resultados y los nuestros pueden también verse como consecuencias de trabajos clásicos muy generales de microeconomía [Mye81], que estudian el diseño de mecanismos transaccionales donde la revelación del valor verdadero resulta un equilibrio del juego entre el vendedor y los ofertantes. En forma concisa, los resultados generales en los trabajos nombrados establecen que bajo ciertas condiciones del espacio de las ofertas que pueden hacerse, una forma de selección fija y un pago para el "último" ofertante, hay una única forma de pago que hace la subasta veraz. En nuestro trabajo, por otra parte, derivamos la forma de hacerlo específicamente para el contexto de las búsquedas patrocinadas y definimos técnicas directas para dicho marco, que tiene sus propias particularidades. En el proceso, profundizamos en los detalles técnicos y damos pruebas constructivas de los teoremas, ambos necesarios para su utilización en la práctica. Por otro lado, el trabajo que aquí presentamos está pensado desde un punto de vista computacional/matemático y puede entenderse sin necesidad de estar familiarizado con conceptos generales de economía, como lo son los utilizados por Holmstrom [Hol79] y Myerson [Mye81]. En la próxima sección explicaremos en detalle nuestro propio trabajo. Lo que resta de esta sección está dedicada a otros trabajos relacionados con el contexto de búsquedas patrocinadas y subastas, especialmente estocásticas o veraces, aunque no usados directamente para la tesis.

El uso de subastas estocásticas fue recientemente considerado en varios trabajos: [MCW05], [FHMVBY07] y [BCE$^+$07] para nombrar algunos. Goldberg, Hartline, Karlin, Saks y Wright [GHK$^+$06] estudiaron las formas de diseñar subastas estocásticas competitivas (es decir, con garantías de eficiencia demostrables en términos de ganancia) para ítems idénticos en suministro ilimitado. Abrams y Gosh [AG07] siguieron esa línea, extendiendo sus ideas al contexto de búsquedas patrocinadas, demostrando que es necesario renunciar a la veracidad de la subasta para conseguir competitividad. Una heurística para ofertar en la cual se perturban las ofertas al azar para evitar ciclos es considerada en [BCE+07], con el resultado de converger en subastas de primer y segundo precio a interesantes equilibrios en los cuales los ofertantes comparten los ítems de alguna manera. Finalmente, [FHMVBY07] introduce una familia de formas de selección estocásticas que tienen por objetivo incrementar la diversidad y evitar ciertos tipos de fraude, sin perder réditos de forma significativa.

Hay varios trabajos recientes en el contexto de las búsquedas patrocinadas. Un ejemplo temprano es [BF02], donde se estudia el equilibrio entre los beneficios de los anunciante y el buscador. En [FBP06] se analizan diferentes formas de selección de avisos y el impacto de éste y otros parámetros, como el modelo de ganancias, son analizados teórica y empíricamente. En [Goo05] se estudia un acercamiento al problema de los fraudes de click [AMN+99], usando como idea vender porcentajes de las impresiones en lugar de clicks, lo que podría verse como la determinización de una subasta estocástica. Continuando con los fraudes, en [IJMT05] se presentan formas de estimar las probabilidades de click en un aviso que sean resistentes al fraude de clicks (*click fraud*). [MSVV05] explora el problema con una visión diferente: su objetivo es optimizar la ganancia total para una serie de búsquedas de forma on-line, tratando de consumir el máximo posible del presupuesto de cada avisador. Su resultado es el mejor algoritmo en términos de competitividad para ese contexto. Otros artículos, como [EOS05] se enfocan en otro aspecto fundamental, como lo son las propiedades de diferentes formas de subasta que podrían ser aplicadas a la búsqueda patrocinada. Algunas de esas formas son veraces o tienen propiedades relacionadas con la veracidad, como ciertos equilibrios de Nash (más detalles sobre ésto en el siguiente capítulo).

1.4. Trabajo realizado en la tesis

En esta tesis, como ya mencionamos brevemente en la Sección 1.2, estudiamos los mecanismos para elegir la forma de asignar a cada espacio de publicidad un aviso y a su vez darle a cada aviso asignado un precio que el avisador debe pagar cuando ocurre un click en él. Dicho estudio se puede separar en dos partes: la primera de ellas es una extensión de la teoría existente para el análisis de los mecanismos en el contexto de búsquedas patrocinadas y la segunda es el diseño y análisis de mecanismos específicos, proceso en el cual también damos algunas herramientas útiles para la confección de otros mecanismos además de los nuestros.

En la parte más teórica de nuestro trabajo, tomamos como base los trabajos sobre subastas veraces de Aggarwal, Goel y Motwani [AGM06] y de Meek, Chickering y Wilson [MCW05]. Dichos trabajos proveen bases teóricas para trabajar con subastas veraces, pero ninguno es lo suficientemente general como para considerar las subastas estocásticas en el contexto de búsquedas patrocinadas. Nosotros, entonces, hicimos una generalización de ambos que sí abarcaba dicho contexto, a la vez que unificaba dos trabajos aparentemente no relacionados directamente. Al hacer dicha generalización, además, damos demostraciones más simples y directas a los teoremas, que resultan de utilidad en la práctica. A la vez, dentro del trabajo teórico, obtenemos una caracterización completa de las subastas estocásticas veraces para la búsqueda patrocinada que el trabajo de Meek, Chickering y Wilson, el único que considera subastas no determinísticas, no ofrece. Un objetivo que también cumple el marco teórico introducido es el ser constructivo en sus formas. Esto quiere decir que la teoría introducida es utilizable directamente en la forma en la que se da, lo que no siempre es cierto cuando uno examina resultados teóricos abstractos.

En la parte práctica de diseño y análisis de mecanismos específicos comenzamos proveyendo herramientas prácticas útiles, en la forma de algoritmos auxiliares útiles para armar subastas de búsquedas patrocinadas. Luego, usando tanto las herramientas teóricas de la primera parte como las prácticas mencionadas en la oración anterior, damos mecanismos de simple construcción y utilización que pueden aplicarse a las búsquedas patrocinadas. Algunos de dichos mecanismos no ofrecen resultados positivos para objetivos importantes en el contexto (como ser, buenas garantías de ganancias para la empresa proveedora de servicios), sin embargo, el análisis realizado sobre ellos permite, mediante ejemplos concretos, introducir las formas en las que encarar el diseño ulterior de mecanismos que sí cumplan con dichos objetivos.

Los resultados contenidos en esta tesis fueron publicados en su mayor parte en [FHM08], aunque se hicieron algunos refinamientos, tanto del resultado como de su demostración, para este trabajo. Un resultado particular, el Algoritmo 1 en el Capítulo 4 es parte de otra publicación más reciente: [FHM09]. Las derivaciones de dicho resultado y algunos refinamientos del mismo fueron hechos a posteriori especialmente para este informe.

1.5. Organización del resto de la presentación

El Capítulo 2 está dedicado a introducir los términos y definiciones de manera más formal para poder tratar matemáticamente a los objetos del contexto de las búsquedas patrocinadas, a la vez que presentar el bagaje de conocimiento previo en el cual se basa el trabajo realizado. En el Capítulo 3 se presenta lo que constituye el resultado más relevante de la tesis, que es un conjunto de postulados con sus demostraciones útiles para diseñar subastas veraces, ver si subastas existentes lo son, demostrar que una subasta no es veraz o que una familia de subastas no contiene ninguna subasta veraz, etc. Luego, en el Capítulo 4 utilizamos dichos postulados para presentar métodos para adaptar subastas existentes al contexto de las búsquedas patrocinadas preservando ciertas propiedades. Para finalizar, en el Capítulo 5 se resumen las conclusiones de todo el trabajo y se detallan posibles líneas de investigación para continuarlo.

Capítulo 2

Preliminares

"*Quizás nadie haya sido aún suficientemente veraz acerca de lo que la veracidad es.*"

Friedrich Nietzsche

En este capítulo introducimos formalmente el problema a estudiar, dando un conjunto de definiciones e introduciendo la notación necesaria. También mencionamos aquí los resultados previos principales en los que está basado el trabajo, unificando su notación. Dado que uno de los resultados principales que aquí presentamos es una generalización de dichos teoremas, y que las demostraciones por lo tanto incluyen muchas de las mismas ideas, en lugar de incluir demostraciones a los teoremas previos daremos una demostración directa (es decir, sin citar a estos teoremas) de nuestros propios resultados para luego utilizarlo fácilmente para dar una demostración alternativa (y entendemos que más simple) de los mencionados teoremas existentes.

2.1. Definición del problema y notación

Como mencionamos en la introducción, el trabajo de esta tesis se centra en la subasta de espacios de publicidad. Esto requiere que exista un trabajo previo de recopilación de avisos candidatos y de búsqueda de los espacios, pero para el problema que estudiaremos supondremos que ambos conjuntos (el de espacios disponibles y el de avisos candidatos a ser publicados) ya fueron calculados.

Llamaremos n a la cantidad de avisos candidatos a ser publicados y k a la cantidad de espacios. Si bien para todo el trabajo asumiremos que k es un número fijo, notaremos al final que pueden extenderse los mismos resultados a casos en los que la determinación de ese número es también parte del problema. Dado que ésto último requiere de notación y explicaciones innecesariamente más complicadas, no lo describiremos formalmente en este informe.

Un mecanismo de selección de avisos es tal que dado un conjunto de espacios y un conjunto de avisos, decide qué aviso es mostrado en qué espacio. Para dicha decisión, la única restricción inherente al problema es que ningún aviso puede aparecer más de una vez y qué deben aparecer k avisos, por lo cual asumiremos que siempre $k \leq n$. En casos particulares, puede haber también restricciones adicionales referidas a los espacios o a los avisos, por ejemplo, un ente externo que requiera que cierto aviso sea publicado para mejorar las estadísticas sobre el mismo (ver equilibrio entre explorar y explotar en el Capítulo 1).

Un mecanismo de asignación de precios es tal que dado un conjunto de avisos y uno de espacios decide el precio por click que paga cada aviso. En este trabajo asumiremos que el beneficio de un avisador por obtener un click en su aviso no depende de dónde fue mostrado el aviso, y por lo tanto también nos restringiremos a mecanismos que eligen cobrar lo mismo por un click en cualquier espacio.

Una subasta entonces se define por sus mecanismos de selección y de asignación de precios, que, por supuesto, no son necesariamente independientes.

Para cada aviso y cada espacio posible, existe una probabilidad de que haya un click en el aviso correspondiente, si es publicado en ese espacio. Asumiremos que las probabilidades de click en cada uno de los avisos mostrados son independientes y notaremos por $z_{i,j}$ la probabilidad de que haya un click en el aviso i si es mostrado en el lugar j. Es importante notar que esto implica asumir que la probabilidad de click en un aviso sólo depende de su ubicación, y no de, por ejemplo, qué otros avisos son mostrados. Hay casos en la literatura donde esto no es así (por ejemplo, el modelo introducido en el problema [Hei07] y estudiado luego en [KM08]). Para nuestro trabajo asumiremos también la *separabilidad* [AGM06] de las probabilidades de click. Esto significa que dichas probabilidades pueden ser separadas en dos componentes, una dependiente del espacio y otra del aviso. Llamaremos entonces a_i a la componente provista por el aviso i (CTR del aviso, por las siglas de *click through rate*) y w_j a la componente provista por el espacio j (CTR del espacio o CTR posicional) resultando $z_{i,j} = a_i w_j$. Los valores a_i y w_j los tomaremos como dados, ignorando también otro problema asociado a las búsquedas patrocinadas: la estimación de las probabilidades de click [IJMT05]. Asumiremos que los espacios son numerados de tal manera que $w_1 \geq w_2 \geq ... \geq w_k$ y que, sin pérdida de generalidad, $w_1 = 1$, es decir, los w_j representan la visibilidad de cada espacio relativa al número 1 en términos de probabilidad de click. Dado esto, podemos ver a los a_i como la probabilidad intrínseca del aviso de ser clickeado cuando su visibilidad es máxima y luego los w_j resultan un ajuste sobre dicha probabilidad reflejando la probabilidad de click perdida por haber sido asignado a un espacio menos visible. Asumiremos también que los CTR de avisos son siempre estrictamente positivos, ya que un aviso con $a_i = 0$ no tiene posibilidad de click, y por lo tanto no hay posibilidad de ganancias ni para el avisador ni para el editor, entonces nunca es conveniente mostrarlo.

Como ya hemos mencionado, cada avisador, además de su aviso y las palabras clave para realizar la asociación a espacios posibles, presenta una oferta. Denotaremos como b_i a la oferta que realiza el avisador i (asociado al aviso i). Esta oferta se interpreta como un máximo para el precio que puede pagar por cada click el avisador i (recordemos que decidir dicho precio también es parte del problema que estudiamos).

Como mencionamos en la introducción, las subastas que consideraremos son estocásticas. Lo azaroso en ellas está dado exclusivamente por la forma en la que se asignan los espacios y los precios, que puede contener un sorteo. Por ejemplo, si hay un solo espacio y 2 avisos, y tomamos probabilidades de ser asignado al espacio proporcionales a las ofertas, el aviso 1 será asignado con probabilidad $b_1/(b_1+b_2)$ y el aviso 2 con probabilidad $b_2/(b_1+b_2)$. De acuerdo con esto, toda la notación que introdujimos hasta aquí (cantidad de avisos y espacios, CTRs de avisos, CTRs de espacios y ofertas) son consideradas fijas en el sentido de que analizamos el caso para cualquier combinación de dichos parámetros (con las restricciones de dominio ya mencionadas por supuesto). A continuación introduciremos la notación para representar la parte aleatoria de la subasta, para lo cual hay que tener en cuenta que dicha aleatoriedad solo está considerada sobre el sorteo que forma parte de los mecanismos de asignación de espacios y de precios y no sobre incertidumbre potencialmente existente sobre los valores de los parámetros introducidos hasta aquí.

Es importante notar los mecanismos son públicos. Si bien los parámetros n, k, a_i, w_j y b_i no lo son necesariamente (en la práctica no lo son), no podemos saber si los avisadores lo aprenden o no, por lo cual consideramos el "peor caso" para el editor , es decir, que los avisadores conocen todos los parámetros involucrados (inclusive las ofertas de los otros avisadores).

Dados $b_1, ..., b_{i-1}, b_{i+1}, ..., b_n$ fijos notaremos por $W_i(x)$ a la variable aleatoria que representa el CTR posicional obtenido por el avisador i si oferta x. Llamaremos $p_{i,j}(x)$ a la probabilidad de que el aviso i sea asignado al espacio j si su oferta es x. Cuando escribimos simplemente $p_{i,j}$ refiriéndonos a la constante que representa la probabilidad del aviso i de ser mostrado en el lugar j, esto es simplemente $p_{i,j}(b_i)$. También definimos por comodidad la función de probabilidad de click posicional agregada de i, denotada como q_i y dada por la ecuación:

$$q_i(x) = \sum_{j=1}^{k} w_j p_{i,j}(x).$$

Puede verse que $q_i(x) = E[W_i(x)]$. Es importante notar que las funciones W_i, $p_{i,j}$ y q_i dependen de las constantes implícitas $b_1, ..., b_{i-1}, b_{i+1}, ..., b_n$. Dada esta notación, cada mecanismo de selección puede ser visto a través de las funciones W_i o de las $p_{i,j}$ y por lo tanto las q_i que determina.

Por su definición, las $p_{i,j}(b_i)$ cumplen las siguientes propiedades:

$$\begin{aligned}
\sum_{i=1}^{n} p_{i,j}(b_i) &= 1 \text{ (siempre se asigna un y sólo un aviso a un espacio)} \\
\sum_{j=1}^{k} p_{i,j}(b_i) &\leq 1 \text{ (cada avisador puede ser asignado a lo sumo a un espacio).}
\end{aligned}$$

En particular, si $k = n$ la segunda ecuación es una igualdad (por conveniencia a veces se asume que hay n espacios, asignando $w_j = 0$ para los espacios que no son reales).

Finalmente, notaremos como $M_i(x)$ al precio por click que pagará el avisador i si su oferta es x. Este precio puede ser una variable aleatoria.

Definición 2.1 (Precio determinístico).
Una subasta es de *precio determinístico* si y sólo si el precio cobrado a un avisador es siempre el mismo cuando todas las ofertas están fijas.

En las subastas de precio determinístico, denotaremos por $\mu_i(x)$ al precio por click que pagará el avisador i si su oferta es x (en ambos casos, μ_i y M_i, se toman como constantes implícitas las otras ofertas $b_1, ..., b_{i-1}, b_{i+1}, ..., b_n$). Un mecanismo de asignación de precios, entonces, puede ser visto a través de las funciones M_i o μ_i que define.

Cualquier mecanismo de precios debe ser tal que $P[M_i(x) > x] = 0$, ya que no se puede cobrar más que la oferta. Cuando dicho precio es determinístico, esto quiere decir que $\mu_i(x) \leq x$, siempre para cualesquiera otras ofertas y para todo x. Nuevamente, cuando hablemos simplemente del precio como una constante y no interesa la forma de la función escribiremos μ_i para representar $\mu_i(b_i)$.

Asumiremos también que todos los avisadores son neutrales al riesgo. Esto quiere decir que su comportamiento está determinado únicamente por la intención de maximizar su propia ganancia esperada medida en dinero. Para cuantificar ese beneficio, debemos introducir cuánto vale un click para el avisador i. Notaremos ese valor como v_i. Por supuesto, v_i solo es conocido en principio por el avisador i y no por los otros avisadores ni por el proveedor del servicio. Ahora podemos sí escribir la ganancia esperada del avisador como la probabilidad de que reciba un click multiplicada por lo que gana con un click, es decir, su valor menos lo que debe pagar. Llamando $\beta_i(x)$ al beneficio esperado del avisador i cuando oferta x (y asumiendo separabilidad), esto resulta:

$$\beta_i(x) = E[a_i W_i(x)(v_i - M_i(x))].$$

En particular, si la subasta es de precio determinístico, $M_i(x)$ es constantemente $\mu_i(x)$, quedando

$$\begin{aligned}\beta_i(x) &= E[a_i W_i(x)(v_i - \mu_i(x))] \\ &= a_i E[W_i(x)](v_i - \mu_i(x)) \\ &= a_i \sum_{j=1}^{n} p_{i,j}(x) w_j (v_i - \mu_i(x)) \\ &= a_i q_i(x)(v_i - \mu_i(x)).\end{aligned}$$

Definición 2.2 (Veracidad).
Una subasta es *veraz* si y sólo si para todo avisador ofertar su propio valor maximiza su ganancia esperada, independientemente de cuáles sean las ofertas de los otros avisadores.

Dada la notación presentada, si vemos la subasta a través de las funciones W_i y M_i o q_i y μ_i en el caso de precio determinístico, y el beneficio de un avisador según la definida β_i, esto implica que una subasta es veraz si y sólo si todo $\beta_i(x)$ posible (para cualesquiera $b_1, ..., b_{i-1}, b_{i+1}, ..., b_n$) tiene un máximo global en $x = v_i$.

Un ejemplo muy simple es cuando hay un solo espacio y la subasta la gana quien oferta más. Supongamos que decidiéramos cobrar la oferta del avisador que ganó, es decir $\mu_i(x) = x$. Para simplificar aún más, supongamos solo 2 avisadores, con $a_1 = a_2 = 1$ y el avisador 2 ofertando $b_2 = 1$. Supongamos también que el valor de un click para el avisador 1 es $v_1 = 10$. Si el avisador 1 ofertara su valor $b_1 = v_1$, el espacio le sería asignado con un precio de 10, obteniendo como ganancia

$$\beta_1(b_1) = (v_1 - b_1)a_1 w_1 = 0.$$

Por supuesto, al avisador 1 le conviene ofertar una cantidad mucho más baja, aunque más alta que b_2 (por ejemplo, $b_1 = 2$), de manera de obtener siempre la asignación del mismo espacio, pero a un precio más barato. En el caso de $b_1 = 2$ la ganancia del avisador 1 sería

$$\beta_1(b_1) = (v_1 - b_1)a_1 w_1 = 10 - 2 = 8.$$

Como el avisador 1 puede obtener mayor ganancia ofertando algo distinto de su valor, la subasta hasta aquí descripta no es veraz.

Por otro lado, si modificamos la subasta de modo que, en lugar de cobrar la oferta de quien ganó, cobrar la segunda oferta más grande, la subasta resulta veraz. En nuestro ejemplo, no importa cuanto oferte el avisador 1, siempre pagará lo mismo si su oferta está por encima de b_2, teniendo como ganancia

$$\beta_1(b_1) = (v_1 - b_2)a_1 w_1 = v_1 - b_2 > 0.$$

Si oferta por debajo de b_2, por otro lado, su ganancia será 0 porque no será asignado, con lo cuál es menos conveniente para él. Vemos entonces que para el caso particular de nuestro ejemplo, desde el punto de vista del avisador 1, la subasta es veraz, ya que no hay forma de obtener más ganancia que ofertando su valor (aunque hay otras formas de obtener la misma ganancia).

Las dos subastas que presentamos son conocidas como subasta de primer precio y de segundo precio, respectivamente, y son conocidos ejemplos de subastas. El hecho de que la subasta de segundo precio es veraz cuando se subasta un solo ítem es también ampliamente conocido, y puede verse como un caso particular de la subasta escalonada presentada al final de este capítulo.

2.2. Trabajo previo

2.2.1. Subastas determinísticas

Aquí describiremos, adaptada a nuestro contexto, una subasta con mecanismo de selección determinístico. Los mecanismos de selección usados en la práctica, tanto histórica como actualmente, entran dentro de los representados por esta subasta que llamaremos subasta escalonada (*laddered* en su nombre original en inglés). Esta subasta fue presentada por Aggarwal, Goel y Motwani en [AGM06].

El mecanismo de selección de avisos de la subasta escalonada implica ordenarlos por oferta b_i ó por ganancia esperada $b_i a_i$. Sea entonces o_i el coeficiente que se multiplica por b_i, pudiendo ser $o_i = 1$ para el primer caso y $o_i = a_i$ para el segundo. Supongamos sin pérdida de generalidad que los avisos están numerados de forma que $b_1 o_1 \geq b_2 o_2 \geq ... \geq b_n o_n$, es decir, el aviso 1 se publica en el espacio 1, el aviso 2 en el espacio 2, ..., el aviso k en el espacio k y el resto de los avisos no se publican. El pago por click del avisador i, para $i \leq k$ (el pago por click de avisos no publicados es irrelevante) está dado por:

$$\mu_i = \sum_{j=i}^{k} \frac{w_j - w_{j+1}}{o_i w_i} o_{j+1} b_{j+1}.$$

La subasta escalonada le cobra a cada avisador lo mínimo necesario para retener el CTR posicional que tiene. Es decir, por los primeros w_k cobra lo necesario para ganarle al $(k+1)$-ésimo avisador, por los siguientes $w_{k-1} - w_k$ lo necesario para ganarle al k-ésimo, etc.

Teorema 2.1 ([AGM06]).
La subasta escalonada es veraz.

Teorema 2.2 ([AGM06]).
Si una subasta determinística es veraz y su mecanismo de selección consiste en ordenar los avisos por un ranking basado en la oferta multiplicada por un coeficiente independiente de las ofertas, entonces la subasta es una subasta escalonada.

Este segundo teorema indica que la subasta escalonada es la única veraz con mecanismo de selección determinístico y basado en ranking de la forma $o_i b_i$.

2.2.2. Subastas con todas las posiciones iguales

En este caso describiremos una familia de subastas estocásticas, también adaptada al contexto descrito en la sección anterior, que utilizamos durante todo el trabajo. Esta familia de subastas sólo está pensada para casos donde $w_1 = w_2 = ... = w_k$. Asumiendo eso, la separación en $p_{i,j}$ pierde un poco de sentido ya que no interesa a qué espacio es asignado un aviso, sino simplemente si es asignado o no. Dado esto, miraremos directamente la q_i implicada por el mecanismo de selección, que en este caso indica la probabilidad de ser asignado algún espacio (ya que no importa cuál). El mecanismo de asignación de precios llamado *condex* es el utilizado en esta familia de subastas, que dice que al avisador i se le cobra por cada click:

$$\mu_i(b_i) = \begin{cases} 0 & \text{si } q_i(b_i) = 0 \\ \dfrac{\int_0^{b_i} x dq_i(x)}{q_i(b_i)} & \text{en otro caso} \end{cases}.$$

Este precio surge de cobrar la esperanza de ganar algún ítem apostando b_i dado que se ganó, o sea, una esperanza condicional de donde surge el nombre condex por *conditional expectation* en inglés.

Llamaremos subasta condex a una subasta cuyo mecanismo de asignación de precios sigue la regla condex. Otra forma de ver dicha regla, que no precisa de integrales de Riemann-Stieltjes, es la siguiente (integrando por partes):

$$\begin{aligned}
\mu_i(b_i) &= \frac{\displaystyle\int_0^{b_i} x dq_i(x)}{q_i(b_i)} \\
&= \frac{q_i(b_i)b_i - q_i(0)0 - \displaystyle\int_0^{b_i} q_i(x)dx}{q_i(b_i)} \\
&= b_i - \frac{\displaystyle\int_0^{b_i} q_i(x)dx}{q_i(b_i)}.
\end{aligned}$$

Teorema 2.3 ([MCW05]).
Si en una subasta condex la probabilidad de ser mostrado es siempre creciente en la oferta, entonces dicha subasta es veraz.

Si bien no está demostrado en el trabajo previo (más específicamente, en [MCW05]) es cierto que condex es única en el mismo sentido que la subasta escalonada de la subsección anterior, esto es, si una subasta es veraz en el contexto para el cual condex fue definida, entonces la subasta es condex (veremos la demostración en el próximo capítulo durante la presentación de nuestro trabajo).

Capítulo 3

Diseño y caracterización de subastas veraces

"La estúpida creencia de que ser veraz es fácil; solo el artista, el gran artista, sabe cuan difícil es."

Willa Sibert Cather

En este capítulo daremos los principales resultados teóricos del trabajo. Éstos forman una base teórica para trabajar con subastas veraces que incluye métodos para diseñarlas ajustando los precios a cobrar, relación entre subastas de precio determinístico y subastas en general y una caracterización de todas las subastas veraces en el contexto de búsquedas patrocinadas.

3.1. Determinización de precios

Las subastas en general, como fueron presentadas, permiten tener precios aleatorios. A continuación veremos cómo la aleatoriedad de los precios no agrega beneficios a los usuarios ni al proveedor del sistema en términos de ganancias, ya que siempre hay una subasta equivalente de precio determinístico.

Supongamos una subasta en general A con funciones de selección de avisos que determinan ciertas funciones W_i y asignación de precios que determinan ciertas funciones M_i. Para poder tener un precio determinístico sin afectar las ganancias, precisamos que el pago esperado ante un click no cambie. Si llamamos $\mu_i(x)$ al precio determinístico por click que pagará el avisador i cuando oferta x dentro de una subasta que tiene el mismo mecanismo de selección de avisos que A (o sea, éste está representado también por W_i) precisamos que:

$$\begin{aligned} E[W_i(x)a_iM_i(x)] &= E[W_i(x)a_i\mu_i(x)] \\ E[W_i(x)M_i(x)] &= E[W_i(x)]\mu_i(x) \end{aligned}$$

Sea $D(A)$ la determinización de A que resulta en hacer la misma selección de avisos, pero cobrarle al avisador i cuando oferta x el siguiente precio determinístico, que cumple la igualdad necesaria mencionada anteriormente:

$$\mu_i(x) = \begin{cases} f_i(x) & \text{si } E[W_i(x)] = 0 \\ \frac{E[W_i(x)M_i(x)]}{E[W_i(x)]} & \text{en otro caso,} \end{cases}$$

donde $f_i(x)$ es una función cualquiera. Notemos que igualmente dicha función es irrelevante ya que $E[W_i(x)] = 0$ implica $W_i(x)$ constantemente 0 (ya que nunca es negativo) y por lo tanto, el aviso en cuestión nunca será asignado, con lo cual nunca obtendrá clicks y entonces nunca pagará ningún precio que se le asigne. De aquí en adelante utilizaremos $f_i(x) = 0$ por comodidad. El siguiente lema ilustra como A y $D(A)$ son equivalentes en términos de ganancias esperadas.

Lema 3.1 (Determinización de subastas). *La ganancia esperada de cada avisador y del proveedor de servicios en una subasta A es la misma que en la subasta $D(A)$.*

Dem:

Consideremos primero a los avisadores y fijemos un avisador i. Su ganancia esperada en A si oferta x está dada por la función

$$\beta_i(x) = E[a_i W_i(x)(v_i - M_i(x))]$$

y en $D(A)$ por la función

$$D(\beta_i)(x) = E[a_i W_i(x)(v_i - \mu_i(x))].$$

Veamos entonces que ambas son iguales:

$$\begin{aligned}
E[a_i W_i(x)(v_i - M_i(x))] &= a_i v_i E[W_i(x)] - a_i E[W_i(x) M_i(x)] \\
&= a_i v_i E[W_i(x)] - a_i \mu_i(x) E[W_i(x)] \\
&= a_i E[W_i(x)](v_i - \mu_i(x)) \\
&= E[a_i W_i(x)(v_i - \mu_i(x))].
\end{aligned}$$

Ahora tomemos la ganancia esperada del proveedor de servicios. Dicha ganancia en una subasta está dada por la suma de los pagos de todos los clicks, es decir, en A

$$R = E\left[\sum_{i=1}^{n} a_i W_i(b_i) M_i(b_i)\right]$$

y en $D(A)$ dicha ganancia es

$$D(R) = E\left[\sum_{i=1}^{n} a_i W_i(b_i) \mu_i(b_i)\right].$$

Veamos que $R = D(R)$:

$$\begin{aligned}
E\left[\sum_{i=1}^{n} a_i W_i(b_i) M_i(b_i)\right] &= \sum_{i=1}^{n} a_i E[W_i(b_i) M_i(b_i)] \\
&= \sum_{i=1}^{n} a_i \mu_i(b_i) E[W_i(b_i)] \\
&= E\left[\sum_{i=1}^{n} a_i \mu_i(b_i) W_i(b_i)\right].
\end{aligned}$$

□

El Lema 3.1 implica que el operador D retiene también muchas propiedades deseables. En particular retiene la veracidad, ya que ésta está dada sólo en términos de ganancias esperadas (Definición 2.2). Dado esto, para el estudio de las subastas veraces empezaremos con subastas de precio determinístico, ya que son más simples de observar (a través de las funciones p, q y μ, todas las cuales tienen como co-dominio números reales y no variables aleatorias).

Una vez definida una subasta de precio determinístico B se puede hacer algún camino inverso al propuesto por el operador D para definir A de manera que $B = D(A)$ y de esa manera A resultará equivalente a B y por lo tanto, retendrá la propiedad de veracidad. Un ejemplo simple es propuesto en [MCW05] y consiste en construir la asignación de precios de B independiente de la selección de avisos y de manera que $E[M_i(x)] = \mu_i(x)$, ya que esto implica que la igualdad $\mu_i(x) = E[M_i(x)W_i(x)]/E[W_i(x)]$ del operador D se cumple.

3.2. Subastas de precio determinístico veraces

En esta sección trabajaremos exclusivamente con subastas de precio determinístico, por lo cual omitiremos decirlo explícitamente y hablaremos simplemente de subastas. En particular, el único resultado que tiene sentido en subastas en general, el Lema 3.2, es verdadero en general, pero daremos la demostración en el caso de subastas de precio determinístico para facilitar la comprensión (de todas maneras, la demostración para el caso general es análoga).

La intuición indica que en una subasta el subir la oferta no puede derivar en un pago final menor ni tampoco en un menor CTR posicional "comprado", ya que eso da lugar a especulación. Con los primeros dos lemas, mostraremos que una subasta veraz necesariamente tiene estas 2 propiedades.

Lema 3.2. *Si una subasta es veraz, entonces la función de probabilidad de click posicional agregada de cualquier ofertante es no decreciente.*

Dem:

Dado un avisador i, como la subasta es veraz, la ganancia $\beta_i(x)$ del avisador i se maximiza ofertando su valor $x = v_i$. Supongamos v y w ofertas cualesquiera tal que $v < w$. Tomando $v_i = v$ se cumple:

$$\begin{aligned}
(v - \mu_i(v))q_i(v)a_i &\geq (v - \mu_i(w))q_i(w)a_i \\
(v - \mu_i(v))q_i(v) &\geq (v - \mu_i(w))q_i(w),
\end{aligned}$$

y haciendo $v_i = w$ se cumple:

$$\begin{aligned}
(w - \mu_i(w))q_i(w)a_i &\geq (w - \mu_i(v))q_i(v)a_i \\
(w - \mu_i(w))q_i(w) &\geq (w - \mu_i(v))q_i(v).
\end{aligned}$$

Sumando ambas ecuaciones obtenemos:

$$\begin{aligned}
(v - \mu_i(v))q_i(v) + (w - \mu_i(w))q_i(w) &\geq (v - \mu_i(w))q_i(w) + (w - \mu_i(v))q_i(v) \\
vq_i(v) - \mu_i(v)q_i(v) + wq_i(w) - \mu_i(w)q_i(w) &\geq vq_i(w) - \mu_i(w)q_i(w) + wq_i(v) - \mu_i(v)q_i(v) \\
vq_i(v) - wq_i(w) &\geq vq_i(w) + wq_i(v) \\
v(q_i(v) - q_i(w)) &\geq w(q_i(v) - q_i(w))
\end{aligned}$$

y como $v < w$ por hipótesis, $q_i(v) - q_i(w)$ debe ser no positivo, por lo tanto, $q_i(w) \geq q_i(v)$. Hemos probado que para cualesquiera dos valores $v < w$, $q_i(v) \leq q_i(w)$, o sea, q_i es no decreciente, que es lo que queríamos demostrar.

$\square$

Lema 3.3. *Si una subasta es veraz, entonces la función de asignación de precios de cualquier ofertante es no decreciente.*

Dem:

Sea i un ofertante fijo. Sean $v < w$ dos ofertas posibles cualquiera. Si $v_i = v$ como la subasta es veraz sabemos que:

$$\begin{aligned}(v - \mu_i(v))q_i(v)a_i &\geq (v - \mu_i(w))q_i(w)a_i \\ (v - \mu_i(v))q_i(v) &\geq (v - \mu_i(w))q_i(w) \\ (v - \mu_i(v))q_i(v)/q_i(w) &\geq (v - \mu_i(w)),\end{aligned}$$

y como $q_i(v)/q_i(w) \leq 1$ porque $v < w$ y q_i es no decreciente por el Lema 3.2, tenemos

$$\begin{aligned}(v - \mu_i(v)) &\geq (v - \mu_i(v))q_i(v)/q_i(w) \geq (v - \mu_i(w)) \\ v - \mu_i(v) &\geq v - \mu_i(w) \\ \mu_i(w) &\geq \mu_i(v).\end{aligned}$$

Hemos probado que para cualesquiera dos valores $v < w$, $\mu_i(v) \leq \mu_i(w)$, o sea, μ_i es no decreciente.

$\square$

A continuación veremos cómo diseñar subastas veraces. Empezaremos tomando como dado un mecanismo de selección de avisos, sólo requiriendo que las probabilidades de click posicional agregadas sean crecientes (que, como ya vimos, es necesario en toda subasta veraz) y diseñaremos un mecanismo de asignación precios al que llamamos *condex* por su analogía con el trabajo publicado en [MCW05], que resulta en una subasta veraz.

El precio *condex* para una subasta de búsquedas patrocinadas se define como:

$$\mu_i(b_i) = \begin{cases} 0 & \text{si } q_i(b_i) = 0 \\ \frac{\int_0^{b_i} x dq_i(x)}{q_i(b_i)} & \text{en otro caso} \end{cases}$$

o escrito de otra forma (integrando por partes)

$$\mu_i(b_i) = \begin{cases} 0 & \text{si } q_i(b_i) = 0 \\ b_i - \frac{\int_0^{b_i} q_i(x) dx}{q_i(b_i)} & \text{en otro caso.} \end{cases}$$

Llamaremos subasta condex a una subasta tal que se cumple la igualdad arriba mencionada.

Teorema 3.1 (Veracidad de condex para búsquedas patrocinadas).
Si la probabilidad de click posicional agregada es no decreciente en una subasta condex, entonces dicha subasta es veraz.

Dem:

Fijemos un ofertante cualquiera i de la subasta, y sea q_i su función de probabilidad de click agregada y v_i el valor de un click para él. Sabemos por hipótesis que q_i es no decreciente. El avisador ofertará de manera de maximizar su ganancia dada por la función:

$$\beta_i(x) = q_i(x) a_i \left(v_i - \mu_i(x)\right).$$

a_i es positivo, y por lo tanto el máximo de β_i es el mismo que el máximo de

$$\beta'_i(x) = q_i(x)(v_i - \mu_i(x)).$$

Dado que la subasta es condex, sabemos exactamente cómo es $\mu_i(x)$ y podemos reemplazarlo. Utilizando la segunda expresión dada para el precio condex, esto resulta en que el avisador intentará maximizar [1].

$$\begin{aligned}
\beta'_i(x) &= q_i(x)\left(v_i - b_i + \frac{\int_0^x q_i(t)dt}{q_i(x)}\right) \\
&= q_i(x)(v_i - b_i) + \int_0^x q_i(t)dt.
\end{aligned}$$

Si $x = v_i$, la función resulta

$$\begin{aligned}
\beta'_i(v_i) &= q_i(v_i)(v_i - v_i) + \int_0^{v_i} q_i(t)dt \\
&= \int_0^{v_i} q_i(t)dt
\end{aligned}$$

Si $x > v_i$ entonces podemos escribirlo como $x = v_i + \delta$ con δ positivo, resultando

$$\begin{aligned}
\beta'_i(v_i + \delta) &= q_i(v_i + \delta)(v_i - v_i - \delta) + \int_0^{v_i+\delta} q_i(t)dt \\
&= -\delta q_i(v_i + \delta) + \int_0^{v_i} q_i(t)dt + \int_{v_i}^{v_i+\delta} q_i(t)dt
\end{aligned}$$

como q_i es no decreciente por hipótesis, podemos acotar superiormente $q_i(t)$ con $t \in (v_i, v_i + \delta)$ por $q_i(v_i + \delta)$

$$\begin{aligned}
\beta'_i(v_i + \delta) &\leq -\delta q_i(v_i + \delta) + \int_0^{v_i} q_i(t)dt + \int_{v_i}^{v_i+\delta} q_i(v_i + \delta)dt \\
&\leq -\delta q_i(v_i + \delta) + \int_0^{v_i} q_i(t)dt + \delta q_i(v_i + \delta) \\
&\leq \int_0^{v_i} q_i(t)dt.
\end{aligned}$$

Si, por otro lado, $x < v_i$ podemos escribirlo como $x = v_i - \delta$ para algún δ positivo, resultando

$$\begin{aligned}
\beta'_i(v_i - \delta) &= q_i(v_i - \delta)(v_i - v_i + \delta) + \int_0^{v_i-\delta} q_i(t)dt \\
&= \delta q_i(v_i - \delta) + \int_0^{v_i} q_i(t)dt - \int_{v_i-\delta}^{v_i} q_i(t)dt
\end{aligned}$$

[1] Analizamos solamente qué pasa con la función en el caso de $q_i(x) > 0$. La expresión a la que se llega al final efectivamente da una ganancia de 0 cuando $q_i(x) = 0$, por lo que la simplificación en realidad es solo notacional.

como q_i es no decreciente por hipótesis, podemos acotar inferiormente $q_i(t)$ con $t \in (v_i - \delta, v_i)$ por $q_i(v_i - \delta)$

$$\begin{aligned} \beta'_i(v_i - \delta) &\leq \delta q_i(v_i - \delta) + \int_0^{v_i} q_i(t)dt - \int_{v_i-\delta}^{v_i} q_i(v_i - \delta)dt \\ &\leq \delta q_i(v_i + \delta) + \int_0^{v_i} q_i(t)dt - \delta q_i(v_i + \delta) \\ &\leq \int_0^{v_i} q_i(t)dt. \end{aligned}$$

Por lo tanto, si $x \neq v_i$ entonces $\beta'_i(x) \leq \beta'_i(v_i)$ lo que implica $\beta_i(x) \leq \beta_i(v_i)$, por lo cual cada avisador maximiza su ganancia ofertando su verdadero valor, es decir, por la Definición 2.2, la subasta es veraz. □

Notemos en particular que el precio condex definido es siempre no decreciente, lo que también puede deducirse como corolario de que la subasta definida es veraz y que en una subasta veraz el precio es efectivamente no decreciente (Lema 3.3)

Se puede notar fácilmente cómo este teorema es una generalización del Teorema 2.3 del trabajo previo, como ya habíamos mencionado, tomando $w_1 = w_2 = ... = w_k$ con lo cual q_i resulta ser efectivamente la probabilidad de ganar algún ítem. Más adelante mostraremos que también es una generalización del Teorema 2.1, también citado anteriormente.

Para finalizar la presente sección, demostraremos que la estrategia para armar subastas veraces dado un mecanismo de selección factible para hacerlo es la única posible. Ilustramos este hecho formalmente con el siguiente teorema.

Teorema 3.2 (Unicidad de precio veraz).
Dado un mecanismo de selección de avisos con probabilidades de click posicional agregadas no decrecientes, existe un único mecanismo de asignación de precios que resulta en una subasta de precio determinístico y veraz.

Dem:

La existencia está demostrada en el Teorema 3.1, así que sólo resta probar la unicidad. Por el absurdo, supongamos que se pueden hacer dos subastas distintas pero con idéntico mecanismo de selección de avisos. Sean A y B subastas veraces con idéntico mecanismo de selección de avisos y sea q_i la función de probabilidad de click posicional agregada en ambos para el avisador i. Sea $\mu_i(x)$ el precio asignado al aviso i cuando oferta x en A y $\nu_i(x)$ lo propio en B. Dado que cuando $q_i(x) = 0$ el precio es irrelevante, asumimos que en esos casos $\mu_i(x) = \nu_i(x) = 0$ (es importante notar que sí es posible $\mu_i(x) = 0$ o $\nu_i(x) = 0$ pero $q_i(x) > 0$).

Cómo μ y ν son distintas por hipótesis, existen valores de n, i, $b_1, ..., b_{i-1}, b_{i+1}, ..., b_n$ tal que hay un t positivo para el cual $\mu_i(t) \neq \nu_i(t)$. Fijemos dichos valores y sin pérdida de generalidad, supongamos $\mu_i(t) > \nu_i(t)$.

Como A y B son veraces, sabemos que para cualquier valor v del avisador i:

$$v \in \underset{x}{\operatorname{argmax}}(v - \mu_i(x))q_i(x)$$

y

$$v \in \underset{x}{\operatorname{argmax}}(v - \nu_i(x))q_i(x).$$

Sea c tal que $\mu_i(t) - \nu_i(t) > c > 0$ y $t/c > \lfloor t/c \rfloor$. Es decir, del intervalo $(0, \mu_i(t) - \nu_i(t))$ tomemos un valor c tal que t/c no sea entero. Cómo el intervalo es abierto y no es vacío, existe tal valor. Sea

$$k = \text{máx}\{h \in \{0, ..., \lfloor t/c \rfloor\} : \mu_i(t - hc) - \nu_i(t - hc) > c\}.$$

$k = \lfloor t/c \rfloor$ no es posible porque $t - \lfloor t/c \rfloor c < c$ y la imagen de ν_i es no negativa, así que no puede ser porque quedaría $\mu_i(t - \lfloor t/c \rfloor c) = \mu_i(c) > c$. Entonces $k < \lfloor t/c \rfloor$, por lo que:

$$\begin{aligned} \mu_i(t - kc) - \nu_i(t - kc) &> c \\ \mu_i(t - (k+1)c) - \nu_i(t - (k+1)c) &\leq c. \end{aligned}$$

Sea $a = t - kc$. Notemos que $\mu_i(a) = \mu_i(t - kc) > c > 0$ por lo cual $q_i(a) > 0$. La ecuación anterior nos da

$$(a - \mu_i(a))q_i(a) < (a - c - \nu_i(a))q_i(a)$$

y, utilizando la veracidad, tenemos

$$(a - \mu_i(a - c))q_i(a - c) \leq (a - \mu_i(a))q_i(a)$$

y

$$(a - c - \nu_i(a))q_i(a) \leq (a - c - \nu_i(a - c))q_i(a - c).$$

Por transitividad de las últimas desigualdades llegamos a

$$(a - \mu_i(a - c))q_i(a - c) < (a - c - \nu_i(a - c))q_i(a - c).$$

Como la desigualdad es estricta, $q_i(a - c)$ no puede ser cero, y tampoco puede ser negativo porque es suma de probabilidades multiplicadas por pesos no negativos, por lo cual $q_i(a - c)$ es positivo y podemos cancelarlo manteniendo la desigualdad, quedando

$$\begin{aligned} a - \mu_i(a - c) &< a - c - \nu_i(a - c) \\ c &< \mu_i(a - c) - \nu_i(a - c) = \mu_i(t - (k+1)c) - \nu_i(t - (k+1)c) \end{aligned}$$

Que se contradice con las ecuaciones dadas para $\mu_i(t - (k+1)c)$ y $\nu_i(t - (k+1)c)$ por la definición de k. Por lo tanto, la premisa de que μ y ν son distintas debe ser falsa, completando la demostración.

$\square$

Finalmente, combinando el Teorema 3.1 y el Teorema 3.2 obtenemos el siguiente resultado.

Corolario 3.1 (Unicidad de condex)**.** *Si una subasta es de precio determinístico, entonces es veraz si y sólo si es una subasta condex.*

3.3. Subastas veraces determinísticas

Una subasta determinística es un caso particular de una subasta de precio determinístico en la cual $p_{i,j}(b_i)$ siempre da 0 o 1 y por lo tanto $q_i(b_i)$ es siempre igual a algún w_j o a 0. Como son un caso particular, podemos definir qué da condex en una subasta determinística. Consideraremos como ejemplo (aunque el trabajo se puede repetir análogamente para otras familias) las subastas determinísticas de ranking. Dichas subastas tienen n funciones de ranking $f_1, ..., f_n$, cada una de las cuales puede depender de parámetros del aviso como el CTR del aviso a_i u otros y la asignación se realiza de la siguiente manera:

Para cada aviso se calcula su puntaje $f_i(b_i)$ y luego se asignan los espacios por orden decreciente de puntaje (es decir, el aviso de mayor puntaje al espacio 1, el segundo de mayor puntaje al espacio 2, etc). Ejemplos de funciones utilizadas son $f_i(x) = x$ (ranking de ofertas), $f_i(x) = xa_i$ (ranking de ganancia esperada del aviso) y combinaciones similares mencionadas en [FBP06]. Las dos primeras fueron consideradas en el Capítulo 2 para las subastas escalonadas. Por simplicidad asumiremos que las funciones están diseñadas para tener imágenes disjuntas, o sea, nunca hay empate en los puntajes (se puede sumar un ϵ_i distinto en cada caso al resultado suficientemente chico como para sólo discernir en los empates).

Para que una subasta de ranking como las descriptas sea veraz, sus respectivos q_i deben ser no decrecientes (Lema 3.2). Como q_i toma el valor del espacio en el que se ubica el avisador i, ésto quiere decir que las funciones de puntaje f_i deben ser no decrecientes. Por simplicidad asumiremos que son estrictamente crecientes, y por lo tanto, inversibles. Ésta suposición no trae impedimentos ya que las funciones de ranking utilizadas en la práctica y mencionadas en la literatura la cumplen (ver los ejemplos dados en el párrafo anterior).

Supongamos que enumeramos a los avisadores de tal manera que $f_1(b_1) \geq f_2(b_2) \geq ... \geq f_n(b_n)$, o sea, $q_i(b_i) = w_i$. Asumiremos para esta parte que $k = n$ siendo potencialmente $w_j = 0$ para los espacios que no existen en realidad. Veamos la forma de $q_i(x)$ en este caso particular (asumiendo que no hay empates):

$$q_i(x) = \begin{cases} w_1 & x > f_i^{-1}(f_1(b_1)) \\ w_j & j < i \wedge f_i^{-1}(f_{j-1}(b_{j-1})) \geq x > f_i^{-1}(f_j(b_j)) \\ w_{j-1} & j \geq i \wedge f_i^{-1}(f_{j-1}(b_{j-1})) \geq x > f_i^{-1}(f_j(b_j)) \end{cases}$$

Por claridad, sea $d_j = f_i^{-1}(f_j(b_j))$, con lo cual q_i queda definida de la siguiente manera:

$$q_i(x) = \begin{cases} w_1 & x > d_1 \\ w_j & j < i \wedge d_{j-1} \geq x > d_j \\ w_{j-1} & j \geq i \wedge d_{j-1} \geq x > d_j \end{cases}$$

Notamos que $d_i = f_i^{-1}(f_i(b_i)) = b_i$, que la función q_i es constante por tramos y que como la integral de condex va hasta b_i, solo se utiliza el último caso de la definición de q_i. Por claridad, definimos $d_{n+1} = 0$. De esta manera, podemos reescribir el precio condex para el caso de subastas de ranking.

$$\begin{aligned} \mu_i(b_i) &= b_i - \frac{\int_0^{b_i} q_i(x)dx}{q_i(b_i)} = b_i - \frac{\sum_{j=i}^{n}(d_j - d_{j+1})w_j}{w_i} = b_i - \frac{\sum_{j=i}^{n} d_j w_j - \sum_{j=i}^{n} d_{j+1} w_j}{w_i} \\ &= b_i - \frac{b_i w_i + \sum_{j=i+1}^{n} d_j w_j - \sum_{j=i+1}^{n} d_j w_{j-1}}{w_i} = b_i - b_i - \frac{\sum_{j=i+1}^{n} d_j(w_j - w_{j-1})}{w_i} \\ &= \frac{\sum_{j=i+1}^{n} d_j(w_{j-1} - w_j)}{w_i} = \frac{\sum_{j=i+1}^{n} f_i^{-1}(f_j(b_j))(w_{j-1} - w_j)}{w_i} \end{aligned}$$

Aquí vemos como las subastas condex generalizan las subastas escalonadas presentadas en el Capítulo 2 usando $f_i(x) = o_i x$. Los teoremas allí presentados como Teorema 2.1 y Teorema 2.2 son entonces corolarios del Teorema 3.1 y del Corolario 3.1, respectivamente.

3.4. Caracterización completa de subastas veraces

Combinando los resultados previos de este capítulo en la Sección 3.1, que relacionan a cualquier subasta con una subasta de precio determinístico y en la Sección 3.2, que analizan la forma de las subastas veraces de precio determinístico, podemos lograr una caracterización puramente aritmética de las subastas veraces en general, es decir, una forma de determinar, dada una subasta, si es o no veraz, sin necesidad de hacer análisis de comportamiento comunes en teoría de juegos, como indica su definición.

Para esto definimos la esperanza del precio por click de un avisador i. Al contrario de lo que pueda parecer, dicha esperanza no está dada por $E[M_i(b_i)]$ sino por $E[W_i(b_i)M_i(b_i)]/E[W_i(b_i)]$ ya que hay que considerar que no todo precio posible se paga la misma cantidad de veces y que las variables aleatorias $W_i(b_i)$ y $M_i(b_i)$ no son necesariamente independientes (si lo fueran, las dos expresiones mencionadas coincidirían). Por completitud, definimos como 0 la esperanza del precio por click de un avisador cuando no tiene posibilidad de click (es decir, $W_i(b_i)$ es constantemente 0).

Teorema 3.3 (Caracterización de subastas veraces).
Una subasta es veraz si y sólo si la esperanza del precio por click de todo avisador coincide con el precio condex para ese avisador.

Dem:

Sea A una subasta. Por el Lema 3.1 A es veraz si y sólo si $D(A)$ lo es. Por definición de $D(A)$, su precio es la esperanza del precio por click de cada avisador en A. Por el Corolario 3.1 $D(A)$ es veraz si y sólo si es una subasta condex, y por lo tanto, tiene precio condex. Como la selección de avisos de A y $D(A)$ coinciden, también coincide su precio condex, completando la demostración.

$\square$

3.5. Cantidad de espacios variable

En la introducción mencionamos que consideraríamos k, la cantidad de espacios disponibles, como un número fijo, pero que los resultados eran aplicables también a casos en los que k es variable. No consideramos dichos casos formalmente en este capítulo porque, como ya fue mencionado, requiere de un modelo previo de cómo se comportan los CTR posicionales cuando k es variable e incluso teniendo un modelo validado para ese caso, requeriría introducir la notación adicional para éste. Puesto que los análisis realizados son en función de W_i ó q_i y dichas funciones son definibles también en un contexto con k variable (aunque dependiendo del modelo es posible que la definición sea complicada), podemos aplicar los resultados de este capítulo a dicho contexto.

Capítulo 4

Extensiones para generar subastas veraces

"*Las palabras veraces no son bellas. Las palabras bellas no son veraces.*"

Lao Tsé

En este capítulo presentaremos un aporte para el diseño de subastas veraces. Este aporte consta de algoritmos auxiliares útiles a la hora de hacer subastas y formas de extender subastas veraces ya existentes para otros contextos al marco de las búsquedas patrocinadas, incluyendo un análisis de la relación entre la subasta original y la extendida.

4.1. Algoritmos auxiliares

En esta sección presentaremos algunos algoritmos útiles para diseñar subastas. Al hacer análisis de ganancias (sea del publicador o de los avisadores) o más en particular, de veracidad, lo importante acerca del mecanismo de selección de avisos es la forma que toma cada función de probabilidad de click posicional agregada q_i. Esto está ilustrado en el Capítulo 3, ya que todos los análisis allí realizados se basan en q_i. El problema, sin embargo, es que al diseñar un mecanismo de selección de avisos estocástico uno diseña un sorteo, es decir, un algoritmo potencialmente azaroso que entrega una subconjunto ordenado del conjunto de avisos posibles.

La relación entre el sorteo y la forma que pueda tomar la q_i no siempre es clara, ya que el algoritmo podría tener $\frac{n!}{(n-k)!}$ salidas distintas. Calcular la probabilidad de cada una de esas salidas individualmente dudosamente sea algo factible, por su explosión exponencial, y el sorteo puede estar diseñado de forma suficientemente complicada como para que no se pueda extraer una generalización que permita examinarlas de forma implícita. Como ejemplo, en la extensión que introducimos más adelante el Algoritmo 3 produce que la función de probabilidad $p_{i,j}$ dependa de $p_{h,j-1}$ para todo h, con lo cual para poder calcular la q_i, que precisa por su definición las $p_{i,j}$, se deben considerar una cantidad exponencial de funciones de probabilidad condicional. Recalcaremos esto nuevamente más adelante, al introducir dicha extensión. Dado este problema, querríamos poder diseñar simplemente una q_i (o unas $p_{i,j}$), eventualmente con algunas restricciones necesarias, y luego construir el sorteo para ellas. Los algoritmos de esta sección sirven para hacer dicha conexión.

En primer lugar daremos un algoritmo para, dadas probabilidades $p_{i,j}$, diseñar un sorteo tal que la probabilidad del aviso i de ser asignado al lugar j es justamente $p_{i,j}$. Supongamos por ahora n avisos y

n espacios. Como ya mencionamos en el Capítulo 2, las $p_{i,j}$ deben cumplir

$$\begin{aligned}\sum_{i=1}^{n} p_{i,j} &= 1 \\ \sum_{j=1}^{n} p_{i,j} &= 1.\end{aligned}$$

Si pensamos en una matriz $M \in [0,1]^{n\times n}$ tal que $M[i,j] = p_{i,j}$, esto significa que la matriz es *doble estocástica* [Bir46].

Definición 4.1 (Doble estocástica)**.**
Una matriz $M \in [0,1]^{n\times n}$ es *doble estocástica* si y sólo si todas sus filas y columnas suman 1.

Las matrices doble estocásticas ya han sido analizadas en el pasado, y un resultado importante que rescataremos es el siguiente:

Teorema 4.1 ([Bir46])**.**
Una matriz doble estocástica es una combinación convexa de matrices de permutación.

Esto quiere decir que podemos hayar matrices de permutación $P_1, ..., P_m$ y coeficientes $\lambda_1, ..., \lambda_m$ con $\lambda_i \in [0,1]$ tal que

$$M = \sum_{i=1}^{m} \lambda_i P_i.$$

En el Apéndice A, el Algoritmo 4 encuentra una combinación posible con $m \leq n^2$ en $\mathcal{O}(n^4)$.

Algoritmo 1 sortearSegún

---

sortearSegún(M)
Requiere: $M \in [0,1]^{n\times n}$ doble estocástica
Sean $P_1, ..., P_m$ y $\lambda_1, ..., \lambda_m$ tal que $\lambda_i \in [0,1]$ y $\sum_{i=1}^{m} \lambda_i P_i = M$.
Elegir el orden dado por la permutación $i \in \{1, ..., m\}$ con probabilidad λ_i.

Es claro por la construcción el hecho de que la probabilidad del aviso i de ser asignado al lugar j es $M[i,j]$.

Si bien el sorteo es cuadrático en el tamaño de la entrada, y esto puede resultar prohibitivo en el contexto de uso de las búsquedas patrocinadas porque se realizan muchas por segundo y deben ser muy rápidas, es posible perfeccionarlo. Si bien el sorteo se realiza muy frecuentemente, las modificaciones a la matriz de entrada M no lo son. Podemos entonces memorizar los P_i y λ_i para una matriz dada y con ello memorizado (en $\mathcal{O}(n^3)$ memoria, n por cada permutación habiendo como máximo n^2 de ellas) hacer el sorteo en $\mathcal{O}(\log n)$. Los detalles técnicos para esto pueden encontrarse en el Apéndice A.

Ahora daremos un algoritmo similar, pero que no precisa la suposición de tener necesariamente n espacios. Asumiendo entonces que tenemos $p_{i,j} \in [0,1]$ con $i \in \{1, ..., n\}$ y $j \in \{1, ..., k\}$ cumpliendo las propiedades mencionadas en el Capítulo 1:

$$\begin{aligned}\sum_{i=1}^{n} p_{i,j} &= 1 \\ \sum_{j=1}^{n} p_{i,j} &\leq 1,\end{aligned}$$

podemos crear una matriz $M \in [0,1]^{n\times n}$ tal que $M[i,j] = p_{i,j}$ para los i, j que $p_{i,j}$ existe y M es doble estocástica, con lo cual podemos usar el Algoritmo 1 para hacer el sorteo y luego quedarnos con los primeros k. Formalmente, definimos M como:

$$M[i,j] = \begin{cases} p_{i,j} & j \leq k \\ \frac{1 - \sum_{j'=1}^{k} M[i,j']}{n-k} & j > k \end{cases}$$

Para ver que M es doble estocástica, notamos que las primeras k columnas suman 1 por requerimiento de $p_{i,j}$. Si la columna j es una de las últimas $n-k$ columnas:

$$\sum_{i=1}^{n} M[i,j] = \frac{\sum_{i=1}^{n} \left(1 - \sum_{j'=1}^{k} M[i,j']\right)}{n-k} = \frac{n - \sum_{j'=1}^{k}\sum_{i=1}^{n} M[i,j']}{n-k} = \frac{n-k}{n-k} = 1.$$

Para i fijo:

$$\begin{aligned} \sum_{j=1}^{n} M[i,j] &= \sum_{j=1}^{k} M[i,j] + \sum_{j=k+1}^{n} M[i,j] = \sum_{j=1}^{k} M[i,j] + \sum_{j=k+1}^{n} \frac{1 - \sum_{j'=1}^{k} M[i,j']}{n-k} \\ &= \sum_{j=1}^{k} M[i,j] + 1 - \frac{n-k}{n-k}\sum_{j'=1}^{k} M[i,j'] = 1. \end{aligned}$$

Notamos que de tener $k << n$ espacios, lo que se requiere guardar para hacer el sorteo muchas veces es $\mathcal{O}(n^2 k)$ y no $\mathcal{O}(n^3)$. El algoritmo entonces dada una matriz $M \in [0,1]^{n\times k}$ representando los $p_{i,j}$ para k espacios es

Algoritmo 2 sortearSegúnK

sortearSegúnK(M)
Requiere: $\sum_{i=1}^{n} M[i,j] = 1$, $\sum_{j=1}^{n} M[i,j] \leq 1$
Sea $M' \in [0,1]^{n\times n}$ tal que:

$$M'[i,j] = \begin{cases} M[i,j] & j \leq k \\ \frac{1 - \sum_{j'=1}^{k} M[i,j']}{n-k} & j > k \end{cases}$$

$P :=$ **sortearSegún**(M')
devolver Los primeros k elementos de P

4.2. De un espacio a múltiples espacios distintos

Un mecanismo de selección de avisos refleja una intención de cómo las oportunidades de click deben ser distribuidas entre los avisadores. Dado que las oportunidades de click no están “todas juntas” para ser repartidas, sino en varios espacios y que hay restricciones de cómo pueden repartirse (i.e., ningún avisador puede ser mostrado en más de un espacio al mismo tiempo) no cualquiera de estas intenciones puede ser implementada directamente. Por ejemplo, si dos espacios idénticos son subastados (es decir, $k = 2$, $w_1 = w_2 = 1$) ningún aviso puede ser asignado a más del 50 % del total de bienes a repartir, sin importar cuánto oferte el avisador. Por otro lado, si hubiera un solo espacio, cualquier distribución del total de bienes es posible. Es útil, entonces, tener una forma de emular cualquier filosofía de distribución de un solo espacio en el caso más realista de que exista más de uno. Notamos como ejemplo de la dificultad de dar subastas estocásticas para múltiples espacios que si bien [MCW05] estudia el caso de más de un espacio (aunque todos idénticos), todos los ejemplos de mecanismos de selección de avisos que provee son para el caso de un único espacio.

Presentaremos en esta sección una familia de mecanismos de selección para varios espacios potencialmente distintos, que satisface las premisas del Teorema 3.1 y por lo tanto puede ser utilizado junto con un precio condex para definir subastas veraces en ese escenario.

El mecanismo que proponemos asigna el espacio más importante (el primero, que tiene el CTR de espacio mayor) según la filosofía básica para un espacio tomada como parámetro y luego cada espacio subsecuente de acuerdo con las mismas probabilidades escaladas para cubrir los lugares que faltan. Formalmente, dadas probabilidades para un espacio $p_1, ..., p_n$ el siguiente algoritmo produce el sorteo mencionado: El algoritmo **sortearPermutación** funciona para una gran familia de mecanismos para

Algoritmo 3 sortearPermutación

sortearPermutación$(p_1, ..., p_n, k)$
Requiere: $n \geq k \geq 1$, $p_i \in (0, 1]$, $\sum_{i=1}^{n} p_i = 1$
$S := \{1, ..., n\}$
para $j = 1$ **hasta** k **hacer**
Elegir $i \in S$ con probabilidad $p_i / \sum_{i' \in S} p_{i'}$
Asignar el aviso i al espacio j
$S := S \setminus \{i\}$.
fin para

un solo espacio, suficientemente grande para abarcar todos los utilizados en la práctica y otros presentes en la literatura. Si bien el requerimiento es un poco más fuerte que el tener la función de probabilidad de click no decreciente, no es algo que intuitivamente no sea razonable pedir. La siguiente definición formaliza el mencionado requerimiento.

Definición 4.2 (Consistentemente monótono).
Un mecanismo de selección de avisos para un espacio es *consistentemente monótono* si y sólo si siempre que un avisador aumenta su oferta y el resto quedan fijos, su probabilidad de ser asignado no decrece y la probabilidad de cualquiera de los otros avisos de ser asignados no crecen.

Es importante notar que los mecanismos de selección de avisos mencionados en este trabajo y en la bibliografía citada, así como los usados en la práctica por Google y Yahoo!, tanto históricamente como los actuales, son todos consistentemente monótonos (al aplicarlos a un sólo espacio).

La compatibilidad de este algoritmo con el uso de precios condex (y por lo tanto, de veracidad) se formaliza con el siguiente resultado.

Teorema 4.2.
*Si un mecanismo de selección de avisos para un espacio es consistentemente monótono, entonces su extensión a múltiples espacios vía el algoritmo **sortearPermutación** resulta en un mecanismo de selección de avisos que tiene una función de probabilidad de click posicional agregada no decreciente.*

Dem:

Sean $p_1, ..., p_n$ las funciones de probabilidad de que cada aviso sea asignado al único espacio en el mecanismo de selección original. Consideremos para toda la demostración un avisador fijo i. Dado un subconjunto C de $\{1, ..., n\}$, llamaremos $R(C, x)$ a la probabilidad de que los elementos de C sean asignados a los primeros $|C|$ espacios por el algoritmo **sortearPermutación** si el avisador i oferta x (todos las otras ofertas las consideramos fijas). Por definición, para cada entero no negativo m,

$$\sum_{C:|C|=m} R(C, x) = 1 \tag{4.1}$$

(la suma de las probabilidades de todos los subconjuntos de un tamaño dado es 1, porque exactamente uno de ellos es siempre elegido y son todas elecciones disjuntas).

Primero, mostraremos que para cualquier subconjunto C de $\{1, ..., n\} \setminus \{i\}$, cuando C es fijo, $R(C, x)$ es no creciente. Haremos la prueba por inducción en $|C|$.

Para $|C| = 0$ solo hay un conjunto $C = \emptyset$ y por definición $R(\emptyset, x) = 1$ es constante y por lo tanto no creciente. Para el paso inductivo, asumiendo que $R(C, x)$ es no creciente para conjuntos de tamaño $m - 1$, mostraremos que lo mismo se cumple para conjuntos de tamaño m. Sea $r_h(x)$ la probabilidad básica en el caso de un espacio del avisador h cuando el avisador i oferta x. Por hipótesis, r_h es no creciente si $h \neq i$. También, por definición de R:

$$R(C, x) = \sum_{h \in C} R(C \setminus \{h\}, x) \frac{r_h(x)}{1 - \sum_{c \in C \setminus \{h\}} r_c(x)} \tag{4.2}$$

donde la fracción representa la probabilidad de h de ser elegido en la m-ésima iteración, dado que los avisadores en $C \setminus \{h\}$ fueron elegidos en las primeras $m - 1$, en algún orden.

$R(C \setminus \{h\}, x)$ es no creciente por hipótesis inductiva. $r_h(x)$ es no creciente porque $h \in C$ y por hipótesis esto quiere decir $h \neq i$. $1/(1 - \sum_{c \in C \setminus \{h\}} r_c(x))$ es no creciente porque como cada r_c es no creciente, la sumatoria también lo es. Dado todo esto, la suma en (4.2) es no creciente, terminando la demostración.

Ahora, sea $S_j(x)$ la probabilidad de que el avisador i reciba alguno de los primeros j espacios cuando oferta x. Por (4.1) se sigue que:

$$S_j(x) = \sum_{C:|C|=j \wedge i \in C} R(C, x) = 1 - \sum_{C:|C|=j \wedge i \notin C} R(C, x)$$

Dado que todos los términos en la suma son de la forma $R(C, x)$ para algún C que no contiene a i, todos ellos son no crecientes y $S_j(x)$ es no decreciente.

También por definición $S_j(x) = S_{j-1}(x) + p_{i,j}(x)$ donde $p_{i,j}$ es, en el mecanismo resultante, la probabilidad de que el aviso i sea asignado al lugar j dado que el avisador i oferta x. Entonces ya podemos escribir la función de probabilidad de click posicional agregada de i:

$$\begin{aligned} q_i(x) &= \sum_{j=1}^{k} w_j p_{i,j}(x) = \sum_{j=1}^{k} w_j (S_j(x) - S_{j-1}(x)) \\ &= \sum_{j=1}^{k} w_j S_j(x) - \sum_{j=1}^{k} w_j S_{j-1}(x) \\ &= \sum_{j=1}^{k-1} S_j(x)(w_j - w_{j+1}) + w_k S_k(x) - w_1 S_0(x) \end{aligned}$$

y todos los términos en la última expresión son no decrecientes (dado que $S_0(x) = 0$, el último es constante), por lo que queda probado que $q_i(x)$ es no decreciente.

oferta			E. CTR 2 esp.			prob. 1 esp.			precio 2 esp.			precio 1 esp.			gan. 2 esp.	gan. 1 esp.
b_1	b_2	b_3	q_1	q_2	q_3	p_1	p_2	p_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3		
1,00	1,00	1,00	0,50	0,50	0,50	0,33	0,33	0,33	0,39	0,39	0,39	0,43	0,43	0,43	0,58	0,43
1,00	1,00	2,00	0,42	0,42	0,67	0,25	0,25	0,50	0,40	0,40	0,65	0,45	0,45	0,77	0,77	0,61
1,00	2,00	2,00	0,33	0,58	0,58	0,20	0,40	0,40	0,43	0,70	0,70	0,46	0,83	0,83	0,96	0,76
1,00	10,00	10,00	0,09	0,70	0,70	0,05	0,48	0,48	0,48	2,43	2,43	0,49	3,94	3,94	3,46	3,77
0,10	0,10	1,00	0,30	0,30	0,91	0,08	0,08	0,83	0,04	0,04	0,16	0,05	0,05	0,23	0,17	0,20
0,10	0,10	10,00	0,25	0,25	0,99	0,01	0,01	0,98	0,04	0,04	0,37	0,05	0,05	0,60	0,38	0,59

Cuadro 4.1: Ejemplo para diferentes combinaciones de ofertas con $n = 3$ avisadores, $k = 1$ vs. 2 espacios con $w_1 = 1$ y $w_2 = 2$

□

Finalmente, para relacionar directamente el Algoritmo 3 con la veracidad, damos el siguiente resultado.

Corolario 4.1. *Si un mecanismo de selección de avisos para un espacio es consistentemente monótono, entonces su extensión a múltiples espacios vía el algoritmo* ***sortearPermutación*** *combinado con el uso de un precio condex da una subasta veraz.*

Un ejemplo

A continuación daremos un ejemplo de aplicación del Algoritmo 3 y el Corolario 4.1. Supongamos que las probabilidades en una subasta para un espacio son proporcionales a las ofertas de los avisadores. Esto es $p_i(b_i) = \frac{b_i}{\sum_{h=1}^{n} b_h}$. Extenderemos dichas probabilidades a un escenario con $k = 2$ espacios con $w_1 = 1$ y $w_2 = 1/2$ y $n = 3$ avisadores. Sólo haremos el desarrollo de las formulas para el avisador $i = 1$, ya que para el resto son simétricas. El precio condex para la subasta original para un espacio es ilustrado en [MCW05] y está dado por:

$$\mu_1(b_1) = (b_2 + b_3)\left[\left(1 + \frac{b_2 + b_3}{b_1}\right)\ln\left(1 + \frac{b_1}{b_2 + b_3}\right) - 1\right].$$

Si aplicamos el Algoritmo 3 **sortearPermutación** obtenemos la función de probabilidad de click posicional agregada:

$$q_1(b_1) = \frac{b_1}{b_1 + b_2 + b_3}\left(1 + \frac{b_2}{2(b_1 + b_3)} + \frac{b_3}{2(b_1 + b_2)}\right).$$

Invitamos al lector a comprobar que dicha función es creciente en b_1, tal como indica el Teorema 4.2, lo que nos habilita a calcular el precio correspondiente usando la regla condex, obteniendo

$$\mu_1(b_1) = -\frac{\frac{{b_2}^2}{b_2+b_1} + \frac{{b_3}^2}{b_3+b_1} + b_2\ln(b_2 + b_1) + b_3\ln(b_3 + b_1)}{2q_1(b_1)}$$

En el cuadro 4.1 presentamos las probabilidades, precios y ganancia esperada para el proveedor del servicio, para diferentes combinaciones de ofertas en ambos escenarios (subasta original de un espacio y la subasta extendida a dos espacios). Observamos que dicha ganancia esperada con dos espacios es más grande cuando las ofertas son similares, y decrece a medida que se separan. Esto es porque la "pendiente" de q_3 es menor cuanto mas grande es b_2, por lo cual la integral que aparece restando en la segunda expresión para el precio condex decrece con b_2, o sea que el precio del avisador 3 crece, y al ser éste el avisador mas importante, es el que mas influye en la ganancia obtenida.

4.3. Extensiones lineales a múltiples espacios

Los resultados de las secciones anteriores proveen formas muy generales de definir subastas veraces para el entorno de múltiples espacios dada una para un solo espacio, pero su mecanismo de asignación de precios es usualmente complicado y dificultoso de calcular, ya que requiere buscar muchas funciones de probabilidad que dependen de otras funciones condicionales y luego sobre ésta calcular la integral siguiendo la definición de condex. Por otro lado, la relación entre los precios resultantes y los precios de la subasta original, y por lo tanto la relación entre las ganancias en uno y otro caso, tanto para los avisadores como para el proveedor del servicio, no es nada clara. En esta sección exploramos otras formas de extender un mecanismo de selección de avisos para un solo espacio de modo que las relaciones mencionadas sean manejables. Si bien los resultados no son prometedores en términos de practicidad, proveen un primer acercamiento hacia otras formas de extensión y generación de subastas para el contexto de las búsquedas patrocinadas.

Primero veremos cómo usar un mecanismo de selección de avisos para un espacio para generar mecanismos de selección para múltiples espacios suponiéndolos idénticos. Luego, veremos cómo pasar de los espacios idénticos a espacios no idénticos de una forma simple. Entre ambas conforman una herramienta muy simple para extender mecanismos de un espacio a múltiples espacios distintos. Como veremos, el beneficio de dichas extensiones en términos de la ganancia percibida por el proveedor del servicio depende de ciertas propiedades del mecanismo original para un espacio utilizado.

Sean $\overline{p}_i$ las funciones representantes de un mecanismo de selección de avisos para un solo espacio, o sea, $\overline{p}_i(x)$ es la probabilidad de que el aviso i sea asignado al espacio subastado cuando el avisador i oferta x. Por supuesto, $\sum_{i=1}^{n} \overline{p}_i(b_i) = 1$.

El objetivo, entonces, es definir funciones de probabilidad p_i para el caso de k espacios iguales (por lo cual simplemente representamos la probabilidad de ser asignado a cualquiera de ellos), de tal modo que

$$\sum_{i=1}^{n} p_i(b_i) = k \text{ y } 0 \leq p_i(b_i) \leq 1 \ \forall i \tag{4.3}$$

Recordemos que el mecanismo de asignación de precios condex puede ser difícil de calcular para algunas funciones, por las integrales que contiene. Por lo tanto, quisiéramos tener la posibilidad de usar funciones de probabilidad que "se lleven bien" con dichas integrales. En el comienzo de este capítulo presentamos algoritmos para seleccionar avisos dadas las probabilidades deseadas que satisfagan (4.3). Consecuentemente, nos enfocaremos aquí en diseñar directamente las funciones que cumplan con dicho requerimiento, ya que luego podremos diseñar un sorteo específico que tenga exactamente dichas funciones utilizando los algoritmos ya presentados.

Consideraremos extensiones lineales, o sea, funciones de la forma $p_i(x) = a\overline{p}_i(x) + c$, donde a y c son constantes a ser definidas. De (4.3) se derivan condiciones necesarias y suficientes para a y c para la factibilidad de p_i. Primero, se sigue que

$$k = \sum_{i=1}^{n} p_i(b_i) = \sum_{i=1}^{n} (a\overline{p}_i(b_i) + c) = a + nc,$$

así que:

$$(\textit{cota del coeficiente constante}) \ c = (k - a)/n. \tag{4.4}$$

El hecho de que $p_i(b_i)$ tiene que ser una probabilidad nos permite derivar una cota superior para a. Es importante tener en mente para las próximas ecuaciones que el supremo de la función de probabilidad $\sup(\overline{p}_i) \geq 1/n$ y su ínfimo $\inf(\overline{p}_i) \leq 1/n$, dado que en el caso extremo todas las probabilidades son constantemente iguales y son todas $1/n$. Dicho caso extremo lo obviaremos, ya que una subasta que asigna siempre $1/n$ de probabilidad no es interesante (y además, para ser veraz, debe cobrar 0 a todos sus participantes, ya que q_i es constante y por lo tanto la integral de condex es 0), por lo cual tomaremos $\sup(\overline{p}_i) > 1/n$ y $\text{ínf}(\overline{p}_i) < 1/n$.

$$\begin{aligned}
1 &\geq p_i(b_i) = a\overline{p}_i(b_i) + c = a\overline{p}_i(b_i) + (k-a)/n \\
1 - k/n &\geq a(\overline{p}_i(b_i) - 1/n) \\
1 - k/n &\geq a(\sup(\overline{p}_i) - 1/n) \\
& \quad \text{notando que } \sup(\overline{p}_i) - 1/n \text{ debe ser positivo} \\
\frac{1-k/n}{\sup(\overline{p}_i) - 1/n} &\geq a \\
\frac{n-k}{n\sup(\overline{p}_i) - 1} &\geq a \\
0 &\leq p_i(b_i) = a\overline{p}_i(b_i) + c = a\overline{p}_i(b_i) + (k-a)/n \\
-k/n &\leq a(\overline{p}_i(b_i) - 1/n) \\
-k/n &\leq a(\text{ínf}(\overline{p}_i) - 1/n) \\
& \quad \text{notando que } \text{ínf}(\overline{p}_i) - 1/n \text{ debe ser negativo} \\
\frac{k}{1 - n\,\text{ínf}(\overline{p}_i)} &\geq a
\end{aligned}$$

Poniendo ambos resultados juntos, obtenemos:

$$(\textit{cota del coeficiente lineal})\ a \leq \text{mín}\left(\frac{n-k}{n\sup(\overline{p}_i) - 1}, \frac{k}{1 - n\,\text{ínf}(\overline{p}_i)}\right). \tag{4.5}$$

Dado que $a < 0$ da funciones de probabilidad no crecientes (opuesto a la necesidad de que sea no decreciente), asumiremos $a \geq 0$.

Ahora, si tomamos $\overline{\mu}$ como el precio de la subasta original para un espacio y $\overline{E}$ como la ganancia esperada en dicha subasta para el proveedor del servicio, podemos ver que el precio condex para la subasta resultante es

$$\begin{aligned}
\mu_i(b_i) &= \frac{\int_0^{b_i} x\ dp_i(x)}{p_i(b_i)} = \frac{\int_0^{b_i} x\ d(a\overline{p}_i(x) + c)}{a\overline{p}_i(b_i) + c} \\
&= \frac{a\int_0^{b_i} x\ d\overline{p}_i(x)}{a\overline{p}_i(b_i) + c} = \frac{a\overline{\mu}_i(b_i)\overline{p}_i(b_i)}{a\overline{p}_i(b_i) + c} \\
&= \overline{\mu}(b_i)\frac{a\overline{p}_i(x)}{a\overline{p}_i(b_i) + c} = \overline{\mu}(b_i)\left(1 - \frac{c}{a\overline{p}_i(b_i) + c}\right),
\end{aligned}$$

y la ganancia esperada para el proveedor del servicio es

$$\begin{aligned}
E &= \sum_{i=0}^{n} a_i p_i(b_i)\mu_i(b_i) \\
&= \sum_{i=0}^{n} a_i(a\overline{p}_i(b_i) + c)\frac{a\overline{\mu}_i(b_i)\overline{p}_i(b_i)}{a\overline{p}_i(b_i) + c} \\
&= a\sum_{i=0}^{n} a_i\overline{p}_i(b_i)\overline{\mu}_i(b_i) = a\overline{E},
\end{aligned} \tag{4.6}$$

Notamos que para (4.6) asumimos que el CTR posicional de cada uno de los espacios en el contexto donde hay múltiples es el mismo que el del único espacio en el contexto original. Si esta suposición no fuera realista, podemos ajustar el resultado dado que, bajo la hipótesis de separabilidad, la diferencia es sólo una constante apropiada.

De (4.4), (4.5) y (4.6) se derivan inmediatamente condiciones necesarias y suficientes para saber en qué casos $\overline{p}_i$ puede ser extendida a un contexto de múltiples espacios generando un mayor ingreso al proveedor de servicios, manteniendo veracidad y sin necesidad de cálculos adicionales complejos. Esto se resume formalmente en el siguiente resultado.

Teorema 4.3.
Si una subasta condex para un solo espacio es extendida linealmente a una subasta para múltiples espacios idénticos, respetando las cotas de coeficiente lineal y de coeficiente constante, vale lo siguiente:

- *El mecanismo de asignación de precios condex de la extensión es* $1 - \frac{c}{a\overline{p}_i(b_i)+c}$ *veces el precio en la subasta original de un solo espacio, donde* a *y* c *son los coeficientes lineal y constante, respectivamente.*
- *El beneficio esperado para el proveedor del servicio es* a *veces el beneficio en la subasta original de un solo espacio.*
- *La subasta extendida es veraz.*

De los resultados previos también surge una cota en las ganancias del proveedor del servicio.

Corolario 4.2. *Si una subasta para un contexto de un solo espacio es llevada con una extensión lineal a una para varios espacios, entonces la ganancia para el proveedor del servicio de la subasta resultante no es más que* $\frac{n-k}{n\sup(\overline{p}_i)-1}$ *veces el de la subasta original.*

Este último corolario implica un resultado negativo si la subasta original no tenía cotas para la probabilidad que podía alcanzar un aviso de ser publicado.

Corolario 4.3. *Si una subasta para un contexto de un espacio sin cota superior para la probabilidad que un avisador puede obtener es llevada mediante una extensión lineal a una subasta para múltiples espacios, la ganancia del proveedor de servicios decrecerá respecto de la que tenía en la subasta original de un espacio.*

Si los espacios no fueran todos idénticos (como es el caso en la realidad), una forma simple de continuar la extensión es elegir un un conjunto de k avisos sin orden siguiendo un mecanismo previamente existente (por ejemplo, usando la extensión anteriormente presentada) y permutarlos al azar para ubicarlos en los espacios. Es claro que el comportamiento en cuanto a ganancias esperadas en dicho caso, tanto de los avisadores como del proveedor de servicios, se modifica únicamente dependiendo de la razón entre el las probabilidades de click posicional totales de uno y otro caso, ya que la permutación perfectamente al azar reparte de maneras iguales entre los avisos elegidos dichas probabilidades (es decir, cada aviso recibirá en promedio ω/k probabilidad de click posicional agregada, donde ω es el click posicional sumado de todos los espacios disponibles).

Capítulo 5

Conclusiones y trabajo futuro

“*Una conclusión es simplemente un lugar donde alguien se cansó de pensar*”

Arthur Block

5.1. Conclusiones

Hemos mostrado cómo derivar subastas veraces para amplias clases de mecanismos de selección de avisos. También mostramos que dichas derivaciones son únicas para el análisis de ganancias. Nuestro trabajo generaliza resultados previos, generando un único marco de trabajo para subastas veraces y permitiendo abarcar todo el espectro necesario por las búsquedas patrocinadas, espectro que dichos resultados previos no cubrían totalmente. Este trabajo derivó en una caracterización directa de las subastas veraces para el contexto de búsquedas patrocinadas, solamente en términos de la forma de selección de avisos y de asignación de precios de la subasta, evitando la necesidad de complicados análisis de comportamiento.

Presentamos también algunas formas de diseño de subastas basados en subastas previas para contextos más simples, preservando propiedades importantes como la veracidad y encontrando relaciones para el análisis de ganancias entre la subasta original y la generada. Parte de este trabajo introduce algoritmos generales que pueden resultar útiles en la generación de otras familias de subastas, no solamente extendiendo subastas previas como en nuestros ejemplos, sino también en la creación de subastas nuevas.

5.2. Trabajo Futuro

Sería interesante hacer un análisis más profundo en términos de ganancias desde el punto de vista del proveedor de servicios, tanto desde el punto de vista teórico como experimental. También es de interés un análisis de ganancias desde el punto de vista de los avisadores, práctica aún más reciente [ZCL08].

Algunas preguntas de interés aún no resueltas son: ¿Cual es la diferencia real en términos de ganancia para los varios mecanismos propuestos? Más aún, ¿Se pueden encontrar garantías teóricas en la línea propuesta por [AG07] o [GHK$^+$06]? ¿Hay algún límite a priori a dichas garantías impuesto por la veracidad?

Aunque en este trabajo asumimos un número fijo de espacios, nuestros resultados son extensibles fácilmente a un contexto donde este número puede variar. Hemos trabajado, aunque no es parte de esta tesis, en algunos modelos para dicha flexibilidad, pero no fueron validados por datos experimentales. Tanto los modelos como el análisis teórico de los mismos para esta variante son preguntas aún abiertas en la investigación de las búsquedas patrocinadas.

Finalmente, hay mucho campo de investigación no abarcado por esta tesis pero relacionado al campo de las búsquedas patrocinadas, como ser la estimación de las probabilidades de click, tanto de avisos como de espacios y la validación de la separabilidad, el agregado de otros límites a las posibilidades de publicación como un presupuesto máximo por cada avisador, validación de la premisa de neutralidad al riesgo de los avisadores y análisis teóricos para casos donde no sea válida, y muchos otros.

Apéndice A

Algoritmos adicionales

Algoritmo 4 encontrarCombinaciónConvexa

encontrarCombinaciónConvexa(M)
Requiere: $M \in [0,1]^{n \times n}$ doble estocástica
Asegura: $m \leq n^2$, $\lambda_i \in (0,1]$ y $\sum_{i=1}^{m} \lambda_i P_i = M$.
Sea $G = (V, E)$ un grafo tal que $V = \{f_1, ..., f_n, c_1, ..., c_n\}$ tiene un nodo por cada fila y columna de M y $E = \{(f_i, c_j) | M[i, j] \neq 0\}$. Es claro que G es bipartito.
Sea A un apareamiento de los nodos de G inicialmente vacío.
$m := 0$
mientras M tenga elementos distintos de 0 **hacer**
  $m := m + 1$
  Poner en A un apareamiento perfecto de los nodos de G utilizando caminos de aumento.
  Tomar P_m la permutación inducida por los ejes (f_i, c_j) de A
  Tomar λ_m como el mínimo entre los elementos de M que aparecen como ejes en A.
  $M := M - \lambda_m P_m$
  Eliminar de G y de A los ejes asociados a elementos de M que se hayan vuelto 0
fin mientras

Teorema A.1.
El algoritmo ***encontrarCombinaciónConvexa*** *es correcto, siempre termina y su complejidad es* $\mathcal{O}(n^4)$.

Dem:

Empecemos viendo la correctitud. Puede verse fácilmente que $M \in [0,1]^{n \times n}$ en todo momento, ya que siempre se resta una permutación multiplicada por el mínimo de los elementos a restar. Esto también indica que $\lambda_i \leq 1$. $\lambda_i > 0$ puede verse por el hecho de que todos los elementos de A son ejes de G que son elementos distintos de 0 de M. Dado que en cada paso M tiene al menos un elemento igual a 0 más que en el paso anterior, el algoritmo termina e itera a lo sumo n^2 veces, con lo cual $m \leq n^2$. Asumiendo que A siempre se puede extender a un flujo que representa un apareamiento perfecto de los nodos de G, siempre se encuentran una matriz de permutación y un coeficiente adecuados, y la condición $\sum_{i=1}^{m} \lambda_i P_i = M$ se satisface trivialmente por construcción.

Veamos que si M tiene elementos distintos de 0, entonces el grafo G tiene un apareamiento perfecto. Como invariante tomaremos que M siempre tiene todas sus filas y columnas de igual suma, y como los elementos son todos positivos y alguno es distinto de 0, mientras está en la iteración, dicha suma no es 0.

El hecho de que el invariante se preserva es trivial, dado que siempre se actualiza M restando un múltiplo de una matriz de permutación. Siendo así, sea s la suma de cada fila y columna de M y consideremos $M' = 1/sM$. Como M' es doble estocástica, por el Teorema 4.1 del Capítulo 4 es combinación convexa de matrices de permutación y por lo tanto, tomando alguna de las matrices de permutación asociadas a coeficientes positivos, podemos ver que hay al menos un apareamiento perfecto en el grafo bipartito inducido por los elementos distintos de 0 de M', que es el mismo que el de M, o sea G.

Para ver la complejidad, notamos que la parte más cara del algoritmo es buscar los caminos de aumento para llevar A a un apareamiento perfecto. Cada uno de los caminos de aumento tarda $\mathcal{O}(n^2)$ porque el grafo tiene $\mathcal{O}(n^2)$ ejes a lo sumo. Ahora, cada camino de aumento necesario proviene o bien de la primer iteración, cuando A está vacío, o bien de un eje que se retiró de A por haber producido un 0 en M. Como hay n caminos de aumento iniciales y $\mathcal{O}(n^2)$ caminos de aumento necesarios por ceros producidos en M, hay $\mathcal{O}(n^2)$ caminos de aumento en total, resultando en una complejidad temporal total de $\mathcal{O}(n^4)$.

□

Para el algoritmo de sorteo, además de memorizar los P_i y λ_i memorizaremos la función acumulada de λ es decir un vector Λ tal que $\Lambda[i] = \sum_{j=1}^{i} \lambda_i$.

Algoritmo 5 sortear

---

sortear$(P_1, ..., P_m, \lambda_1, ..., \lambda_m, \Lambda)$
Requiere: $\Lambda[i] = \sum_{j=1}^{i} \lambda_i$, $\lambda_i \in (0, 1]$
$x :=$ un número real uniformemente distribuido en $[0, 1)$
$i :=$ primer posición de Λ tal que $\Lambda[i] > x$
devolver P_i

Teorema A.2.
El algoritmo ***sortear*** *toma tiempo* $\mathcal{O}(\log n)$ *y elige* P_i *con probabilidad* λ_i.

Dem:

Para toda la demostración se debe tener en cuenta que Λ está ordenado de forma creciente.

La cota de complejidad temporal es simple. El primer y el último paso son $\mathcal{O}(1)$. El paso de la búsqueda en Λ implementado con una búsqueda binaria toma $\mathcal{O}(\log n)$.

Veamos la correctitud: P_1 es elegido si y sólo si $\Lambda[1] > x$, por lo cual el intervalo de valores de x para el cual se lo elige es $[0, \Lambda[1])$, de tamaño $\Lambda[1] = \lambda_1$. Como x se elige uniformemente de un intervalo de longitud 1, la probabilidad de que se elija en un intervalo de longitud λ_1 es, por definición de uniformidad, λ_1. Si $i \neq 1$, P_i es elegida si y sólo si $\Lambda[i] > x$ y $\Lambda[i-1] \leq x$. Esto quiere decir que el intervalo de valores de x para los cuales P_i es elegido es $[\Lambda[i-1], \Lambda[i])$ que tiene tamaño $\Lambda[i] - \Lambda[i-1] = \lambda_i$. Análogamente al caso anterior, la probabilidad de que esto suceda es λ_i.

□

Bibliografía

[AG07] Z. Abrams and A. Ghosh. Auctions with revenue guarantees for sponsored search. In *16th International World Wide Web Conference (WWW2007)*, 2007.

[AGM06] Gagan Aggarwal, Ashish Goel, and Rajeev Motwani. Truthful auctions for pricing search keywords. In *ACM Conference on Electronic Commerce*, pages 1–7, 2006.

[AMN+99] Vinod Anupam, Alain J. Mayer, Kobbi Nissim, Benny Pinkas, and Michael K. Reiter. On the security of pay-per-click and other web advertising schemes. *Computer Networks*, 31(11-16):1091–1100, 1999.

[BCE+07] C. Borgs, J. Chayes, O. Etesami, N. Immorlica, K. Jain, and M. Mahdian. Dynamics of bid optimization in online advertisement auctions. In *16th International World Wide Web Conference (WWW2007)*, 2007.

[BF02] H.K. Bhargava and J. Feng. Paid placement strategies for internet search engines. In *11th International World Wide Web Conference (WWW2002)*, 2002.

[Bir46] G. Birkhoff. Tres observaciones sobre el algebra lineal. In *Univ. Nac. Tucumán. Revista*, pages 147–151, 1946.

[BK96] J.I. Bullow and Paul Klemperer. Auctions vs. negotiations. *American Economic Review*, 86:180–94, 1996.

[EOS05] B.G. Edelman, M. Ostrovsky, and M. Schwarz. Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords. Stanford Graduate School of Business Research Paper No. 1917 Available at SSRN: http://ssrn.com/abstract=861164, 2005.

[FBP06] J. Feng, H.K. Bhargava, and D.M. Pennock. Implementing sponsored search in web search engines: Computational evaluation of alternative mechanisms. *Informs Journal on Computing*, 2006.

[FHM08] E. Feuerstein, P. Heiber, and M. Mydlarz. Truthful stochastic and deterministic auctions for sponsored search. In *6th Latin American Web Congress (LA-WEB 2008)*, 2008.

[FHM09] E. Feuerstein, P. Heiber, and M. Mydlarz. Optimal auctions capturing constraints in sponsored search. In *5th International Conference on Algorithmic Aspects in Information and Management (AAIM2009)*, 2009.

[FHMVBY07] E. Feuerstein, P. Heiber, J. Martínez-Viademonte, and R. Baeza-Yates. New stochastic algorithms for placing ads in sponsored search. In *5th Latin American Web Congress (LA-WEB 2007)*, 2007.

[GHK+06] Andrew Goldberg, Jason Hartline, Anna Karlin, Mike Saks, and Andrew Wright. Competitive auctions. *Games and Economic Behavior*, 2006.

[Goo05] J. Goodman. Pay-per-percentage of impressions: an advertising method that is highly robust to fraud. In *ACM E-Commerce Workshop on Sponsored Search Auctions*, 2005.

[Hei07] P. Heiber. Adplacement. Problem statement for TopCoder Marathon Match 14, 2007.

[Hol79] Bengt Holmstrom. Groves' scheme on restricted domains. *Econometrica*, 47(5):1137–1144, 1979.

[IJMT05] N. Immorlica, K. Jain, M. Mahdian, and K. Talwar. Click fraud resistant methods for learning click-through rates. In *Workshop on Internet and Network Economics (WINE)*, 2005.

[Kle99] Paul Klemperer. Auction theory: A guide to the literature. *Journal of Economic Surveys*, 13(3):227–286, 1999.

[KM08] David Kempe and Mohammad Mahdian. A cascade model for externalities in sponsored search. In *Proceedings of the 4th International Workshop on Internet and Network Economics (WINE08)*, pages 585–596, Berlin, Heidelberg, 2008. Springer-Verlag.

[MCW05] C. Meek, D.M. Chickering, and D.B. Wilson. Stochastic and contingent-payment auctions. In *Workshop on Sponsored Search Auctions - ACM Conference on Electronic Commerce (EC'05)*, 2005.

[MSVV05] A. Mehta, A. Saberi, U. Vazirani, and V. Vazirani. Adwords and generalized on-line matching. In *Symposium on Foundations of Computer Science*, 2005.

[Mye81] R. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6:58–73, 1981.

[Pen04] A. Penemberg. Click fraud threatens web. *Wired news*, October 13, 2004.

[PO06] S. Pandey and C. Olston. Handling advertisements of unknown quality in search advertising. In *Twentieth Annual Conference on Neural Information Processing Systems (NIPS), Vancouver, Canada*, 2006.

[Vic61] W. Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance, Vol. 16, No. 1*, pages 8–37, 1961.

[ZCL08] Yunhong Zhou, Deeparnab Chakrabarty, and Rajan Lukose. Budget constrained bidding in keyword auctions and online knapsack problems. In *17th International World Wide Web Conference (WWW2008)*, pages 1243–1244, New York, NY, USA, 2008. ACM.