hicieron después. Llega a la necesidad de probar este resultado al tratar de determinar, hoy decimos estimar, el valor de θ, la probabilidad de éxito en una sola repetición.
El segundo aspecto que interesa destacar de esta parte cuarta, es que en ella introduce lo que llama grado de certeza moral, que es ya una aproximación al concepto de probabilidad subjetiva. Así dice
"Hay certeza moral cuando la probabilidad es casi igual a la certeza total... si una cosa es considerada moralmente cierta cuando tiene 999/1000 de certeza, otra ser moralmente imposible cuando tenga 1/1000 de certeza".
Un estudio bastante detallado de la saga de los Bernoulli puede verse en el libro de Sánchez y Valdés (2001).Una traducción al castellano de la parte cuarta del Arte de la Conjetura puede verse en Rivadulla (1993).
Con la ley débil de los grandes números para la distribución de Bernoulli, el autor ofrece por primera vez una concreción matemática de la medida de la incertidumbre. Bernoulli no sólo prueba cualitativamente que a mayor número de observaciones hay menos incertidumbre en el resultado de la probabilidad de éxito para la distribución que lleva su nombre, él cuantifica además la afirmación, garantizando el número de observaciones que han de realizarse, para que la frecuencia se aproxime a la probabilidad.
Esto nos lleva al tercer aspecto que nos interesa destacar de la última parte del Arte de la Conjetura y es que el trabajo queda incompleto, porque al tratar de determinar el número de observaciones a realizar para obtener la certeza moral de la probabilidad de éxito, con las aproximaciones que utiliza, el número de observaciones al que llega es demasiado grande como para que éste sea de utilidad práctica.
Como dice Stigler (1986) Jacob Bernoulli dió el primer paso de lo que sería un largo y fecundo viaje. Que este aniversario sirva para que cada vez más investigadores se conviertan en compañeros de tan interesante viaje.

Referencias

-Rivadulla, A. (1993) Teoría de probabilidades (Ars Conjectandi, Parte Cuarta, Basilea 1713), Boletín de la Sociedad Española de Historia de las Ciencias, 16, 389-418.
-Sánchez Fernández, C. Y Valdés Castro, C (2001) Los Bernoulli Geómetras y Viajeros, TresCantos (Madrid):Nívola.
-Stigler, S. M. (1986) The Histoty of Statistics. The measurement of uncertainty before 1900, Cambridge: Harvard University Press.

4. ARTÍCULOS DE INVESTIGACIÓN OPERATIVA

EL LEMA DE FARKAS: UNA HERRAMIENTA PARA RESOLVER NUEVAS APLICACIONES

Juan José Salazar González
DEIOC – Universidad de La Laguna (jjsalaza@ull.es)

Resumen

Farkas (1986) dio (entre otros resultados) una condición matemática necesaria y suficiente para determinar cuándo un poliedro es o no vacío, y que posteriormente se ha denominado *Lema de Farkas*. Tradicionalmente este resultado se considera como un "resultado teórico", o al menos así lo clasifican nuestros alumnos. El objetivo de estas notas es mostrar que el Lema de Farkas también tiene gran valor práctico dentro de la Optimización Matemática porque su oportuno uso dentro de algoritmos de "ramificación y corte" permite resolver problemas de tamaño mayor. Para alcanzar este objetivo se presentan cuatro aplicaciones reales, y en cada una se ilustra cómo utilizar el Lema de Farkas.

Introducción

Sea A una matriz con m filas y n columnas, denotadas éstas por $a^1,\ldots,a^n$, sea y un vector de $\Re^n$ y sea b un vector de $\Re^m$. Tradicionalmente el Lema de Farkas se presenta en la siguiente versión:

El sistema $Ay=b,\ y\geq 0$ tiene solución y en $\Re^n$ si, y sólo si, el sistema $A^Tu\geq 0,\ b^Tu<0$ no tiene solución u en $\Re^m$.

Intuitivamente es un resultado trivial: o bien b está en el cono convexo que generan las columnas de A (es decir, existe escalares no negativos $y_1,\dots,y_n$ tales que $b=y_1a^1+\dots+y_na^n$), o bien existe un hiperplano que deja a un lado el vector b y a otro lado las columnas de A (es decir, existe un gradiente u tal que $u^Ta^i\geq 0$ para todo $i=1,\dots,n$, pero $u^Tb<0$). Ahora bien, a la vez es también un resultado fundamental. Así lo hacemos saber a nuestros alumnos, principalmente argumentando que sobre él se sustenta, por ejemplo, la Teoría de la Dualidad en Programación Lineal.

El objetivo de estas notas es acentuar esta importancia, haciendo especial énfasis en sus inmediatas implicaciones prácticas. La clave es observar que el Lema de Farkas da las condiciones que tiene que cumplir A y b para que el poliedro definido por

$P = \{y \in \Re^n : Ay=b, y\geq 0\}$

sea no vacío. Estas condiciones son:

$u^Tb\geq 0$ para toda dirección u del cono poliédrico $C = \{u \in \Re^m : A^Tu\geq 0\}$.

El Teorema de Caratheodory garantiza que este cono está generado por un número finito de vectores $u^1,\dots,u^d$, llamados *direcciones extremas*. Teniendo presente este resultado y la convexidad de C, para asegurar que $P\neq\emptyset$ basta controlar que se cumplan un número finito de desigualdades: $b^Tu^j\geq 0$ para todo $j=1,\dots,d$. En sí mismo éste me parece un resultado destacable porque muchos de nuestros alumnos conocen cómo verificar si un sistema de igualdades lineales es o no factible, pero no tantos responden a cómo se puede determinar si un sistema de desigualdades lineales admite o no solución.

Es cierto que en la práctica el número d de direcciones extremas de un cono poliédrico puede ser muy grande, y por tanto las condiciones anteriores pueden parecer empíricamente inútiles. Sin embargo, es fundamental observar que para verificar si se cumplen *todas* las desigualdades anteriores *no* es necesario controlarlas una a una. Basta resolver el problema de Programación Lineal:

Minimizar	b^Tu
Sujeto a	$u\in C.$

Gracias al Lema de Farkas, $P=\emptyset$ si y sólo si ese programa lineal es no-acotado, y en este caso una dirección de no-acotación u^* (caracterizada por cumplir $A^Tu^*\geq 0$ y $b^Tu^*<0$) da además el certificado que lo demuestra. Lo importante de este resultado no es que la pregunta "*¿$P=\emptyset$?*" equivale a resolver un problema lineal, ya que esto es incluso más evidente resolviendo "*minimizar* 0^Ty *sujeto a* $y\in P$". Lo importante es que para garantizar que P contiene algún punto y sería necesario imponerle a b la condición $b^Tu^*\geq 0$ (llamada *corte*).

Este simple razonamiento es la idea base que permite resolver varias aplicaciones reales, como veremos en las próximas secciones. La característica común a estas (y otras) aplicaciones es que en ellas se debe buscar una decisión x de manera que exista un vector y cumpliendo una familia de desigualdades lineales cuyo término independiente depende de x. En ocasiones estas variables y, junto con las x, pueden suponer demasiadas variables para constituir un modelo matemático resoluble en la práctica. La introducción dinámica de cortes mediante la resolución de problemas lineales da una alternativa que reduce las variables del modelo a sólo las x, haciéndolo más tratable en la práctica. Adicionalmente los cortes que se generan pueden en ocasiones fortalecerse aumentando o disminuyendo (en inglés "lifting") algunos de sus coeficientes o su término independiente, dando así modelos matemáticos con relajaciones lineales más ajustadas.

El esquema global corresponde a una técnica iterativa de hiperplanos de corte. En cada iteración se resuelve un *problema maestro* que es una relajación del problema original y que sólo tiene variables de tipo x. A partir de la solución óptima x^* se resuelve el programa lineal anterior (llamado *subproblema*), y si resulta no acotado entonces se tiene una dirección de no-acotación u^*. Con esta dirección se construye un corte para fortalecer el problema maestro de la siguiente iteración, evitando así que vuelva a aparecer la solución x^*. Dado que el número de direcciones d es finito, este proceso iterativo es un algoritmo que termina con una solución óptima de la aplicación original. Cuando x es un vector de números enteros, puede verse como la técnica de introducida en Benders (1962). Sin embargo, como característica especial, aquí aparece el hecho de que en varias aplicaciones donde x es un vector de enteros, no es necesario mantener que el problema maestro sea un problema de Programación Entera. En efecto, es posible utilizar este esquema iterativo resolviendo sólo problemas maestros y subproblemas lineales si se aplica en cada nodo de un árbol de ramificación sobre el vector de variables x. Es decir, el procedimiento puede verse como una técnica de *ramificación y corte*.

La primera aplicación que veremos nace en los Institutos de Estadística cuando desean proteger celdas con valores confidenciales en tablas estadísticas. El problema es decidir qué ocultar pero minimizando la pérdida de información en la que se incurre. Otra aplicación nace en el campo del transporte, cuando se necesita diseñar la ruta de un vehículo con capacidad limitada para que recoja y entregue demandas de un producto a través de unos clientes. El objetivo es minimizar los costes de la ruta mientras se mantienen los límites de carga del vehículo y las demandas de los clientes. La tercera aplicación aparece cuando se necesitan depurar encuestas estadísticas que no cumplen un conjunto de reglas de coherencia. En estos procesos el primer objetivo es localizar el menor número de errores potenciales pero garantizando que existirá un valor imputable durante la corrección de dichos errores (segundo objetivo). Por último se presenta una aplicación en el campo de las telecomunicaciones, donde se desea decidir concentradores a instalar de manera que puedan realizarse conexiones entre orígenes y destinos con los menores costes.

Protección de Datos Confidenciales

Los institutos de estadística nacionales o autonómicos recogen información de individuos o empresas bajo el compromiso de garantizar la privacidad. Luego agregan estos datos en tablas donde hay totales (por filas, columnas, etc.). Aplicando ciertas reglas determinan las llamadas *celdas sensibles*, y cuyos valores no pueden hacerse públicos porque revelarían información confidencial. Como consecuencia estos valores se reemplazan por asteriscos, dando lugar a las llamadas *supresiones primarias*. Ahora bien, debido a que las tablas muestran totales marginales, la sola supresión de las celdas sensibles no garantiza que un intruso no calculará su valor original sin más que usar las relaciones lineales que dan los valores totales a partir de los valores internos. Por tanto se deben suprimir también (quizás) otras celdas no sensibles, y que se llaman *supresiones secundarias*. Ahora bien, ¿qué celdas elegir como secundarias? Cada supresión conlleva una pérdida de información, y el objetivo es minimizar la pérdida global como consecuencia de todas las supresiones, pero asegurando en todo momento que un intruso no podrá calcular (ni exacta ni aproximadamente) los valores originales de las celdas sensibles.

La decisión de suprimir o no una celda i puede modelarse como una variable x_i que toma el valor 1 en caso de optar por sustituir su valor por un asterisco, y 0 en otro caso. Una solución x para toda la tabla será válida (aunque quizás no óptima) si garantiza que un intruso no detecta el valor exacto de ninguna celda sensible. Lo deseable, por tanto, es que las celdas i con $x_i=1$ generen suficiente “incertidumbre” para cada celda sensible, entendiendo por incertidumbre el hecho de que existan valores potenciales para los asteriscos que garanticen un nivel de protección mínimo y un nivel de protección máximo para cada celda sensible. Estas garantías se pueden modelizar mediante vectores de variables y, cada uno representando una posible tabla para cada celda y cada para cada nivel. Es decir, se hablaría de dos vectores y^{i+} e y^{i-} para cada celda sensible i en una tabla. Si pensamos en una tabla bidimensional con 100 filas y 100 columnas, y con un 10% de celdas sensibles, esto significaría un modelo con 10,000 variables 0-1 y con 2 millones de variables continuas. En su conjunto esto determina un problema de Programación Matemática demasiado grande para ser manejable en la práctica.

El Lema de Farkas da otra alternativa de resolución. Efectivamente para que una solución x sea factible se requiere que existan los vectores y^{i+} e y^{i-}, es decir, que un conjunto dependiente de x sea no vacío. Ahora bien, es posible reemplazar esta enorme cantidad de variables por unos cortes que se incorporan dinámicamente en un problema maestro a base de resolver una secuencia de subproblemas lineales. El lector puede encontrar detalles técnicos de este procedimiento en Salazar (2004), donde se aplica este esquema iterativo en diferentes metodologías dentro del Control de la Privacidad en Tablas Estadísticas.

Transporte de una mercancía a través de varios clientes

Supongamos que una compañía debe diseñar la ruta de un vehículo que debe mover un producto entre varios clientes. El producto puede ser, por ejemplo, dinero, y los clientes pueden ser sucursales de una misma entidad bancaria. Algunos clientes dan unidades de este producto, mientras otros requieren demandas de ese mismo producto. Se conocen el coste de ir desde un cliente hasta otro y la capacidad máxima del vehículo. ¿Cómo debe servir el vehículo la demanda de los clientes para minimizar el coste global de la ruta?

Claramente hay que diseñar una ruta x, y en este sentido el problema se parece al clásico problema del Viajante de Comercio. El vector x tiene una variable decisoria 0-1 para cada arco de la red. La gran diferencia es que ahora hay demandas (positivas y negativas) en los clientes y capacidad en el vehículo, lo que hace que no todo circuito Hamiltoniano sea válido. Para garantizar la validez de una ruta x es necesario que exista una carga y_a asociada a cada arco a con $x_a=1$. Y es aquí donde vuelve a aparecer la utilidad del Lema de Farkas. Por un lado se puede construir un modelo con las x y con las y, pero las características del problema hacen mucho más conveniente usar el Lema de Farkas para reemplazar las variables y por cortes en las variables x. Es más, el subproblema que aparece en esta aplicación tiene una estructura combinatoria especial, y es posible caracterizar todas las direcciones extremas. En consecuencia la familia de cortes, aunque siempre de gran tamaño, admite ser fortalecida, originando así problemas maestros con mejores características. Estas consideraciones han permitido resolver problemas con dimensiones antes no resueltas. Los detalles de estas investigaciones pueden consultarse en el artículo Hernández y Salazar (2004).

Depuración de Encuestas

Muchos organismos públicos recogen información de individuos o empresas como base para análisis estadísticos sobre los que tomar de decisiones. Haciendo uso de las llamadas *reglas de coherencia*, estos organismos detectan que algunas encuestas contienen errores y deben ser depuradas antes de entrar con las otras en el proceso de análisis. Por ejemplo, en base a una de estas reglas se puede considerar que hay un error en la encuesta de un individuo cuya edad dice “12 años” y cuyo estado civil dice “divorciado”. En base a discusiones previas, tales como la original en Fellegi y Holt (1979), está ampliamente aceptado que la forma de corregir una encuesta que no satisfaga todas las reglas es alterando el menor número de sus valores. Por tanto, el primer objetivo a afrontar durante la depuración de una encuesta es el problema de optimización que busca el menor número de campos a modificar de manera que existan luego valores imputables. Sea x_i la variable decisoria que toma el valor 1 si el campo i debe modificarse, y 0 en otro caso, y sea y_i el valor imputable al campo i cuando $x_i=1$. Una vez más sucede que se buscan soluciones x asociadas a un conjunto que debe contener alguna solución y. Cuando las respuestas son números reales y las reglas son relaciones lineales, estos conjuntos son poliedros, y el Lema de Farkas vuelve a dar una alternativa eficaz para resolver problemas de grandes dimensiones.

En esta aplicación la vinculación entre las variables x_i y las variables y_i es a través de los extremos del intervalo de valores para la respuesta i. Es decir, si el campo i puede tomar valores entre f_i y g_i, y si el valor original en la encuesta es h_i, entonces esta vinculación viene dada por las desigualdades $h_i + (f_i-h_i)x_i \leq y_i \leq h_i + (g_i-h_i)\, x_i$. Cuando f_i y g_i son valores muy alejados, el modelo matemático con sólo las variables x, más los cortes que vienen de aplicar el Lema de Farkas, ofrece diversas mejoras frente a un modelo matemático sin estos cortes y con las variables y. El lector puede encontrar los detalles técnicos en Riera y Salazar (2004).

Diseño de una red de comunicaciones

En el diseño de una red de comunicaciones se deben localizar concentradores y establecer las conexiones con el menor costo posible de manera que se garantice la existencia de caminos para enviar paquetes de información desde unos terminales a otros. La existencia de límites en las capacidades de los concentradores y de las conexiones obliga a fraccionar los paquetes. Por tanto las comunicaciones pueden ser flujos que circulan por una red con capacidades. Lo fundamental es que hay un costo asociado a todos los concentradores y conexiones por los que pase alguna cantidad de flujo. La decisión de si se abre o no un concentrador en un nodo i puede modelizarse mediante una variable x_i, y la decisión de si se establece una conexión o no a través de un arco a de la red se puede modelizar mediante una variable x_a. Para que una decisión x sea válida debe garantizarse que exista un flujo y que permita todos las comunicaciones deseadas, y es aquí donde vuelve a intervenir el Lema de Farkas. En efecto, sin el Lema de Farkas sería necesario trabajar con un modelo con variables (x,y) mientras que gracias al Lema de Farkas es posible eliminar las variables y a cambio de insertar dinámicamente algunos cortes. En esta aplicación concreta el subproblema que genera tales cortes tiene una especial estructura que permite una segunda descomposición reduciendo su resolución a una serie de problemas de caminos mínimos en un grafo. Se pueden consultar los detalles en Rodríguez y Salazar (2004).

Agradecimientos

El trabajo de investigación sobre recogidas y entregas de una mercancía a través de clientes ha sido realizado conjuntamente con Hipólito Hernández Pérez (hhperez@ull.es), el trabajo sobre depuración de encuestas estadísticas con Jorge Riera Ledesma (jriera@ull.es), y el trabajo sobre la red de comunicaciones con Inmaculada Rodríguez Martín (irguez@ull.es). Todos somos miembros del *Departamento de Estadística, Investigación Operativa y Computación* de la Universidad de La Laguna, financiados en parte por los proyectos nacionales TIC2000-1750-C06-02 y TIC2003-05982-C05-02. Estas notas nacieron de una charla impartida en las *I Jornadas SEIO-RSME sobre Programación Matemática*, Elche, 6-7 de mayo de 2004.

Bibliografía

J.F. Benders, "Partitioning procedures for solving mixed-variables programming problems", *Numerische Mathematik* 4 (1962) 238-252.

G. Farkas, "A Fourier-féle mechanikai elv alkalmazásainak algebrai alapjáról", *Mathematikaiés Physikai Lapok* 5 (1986) 49-54. [en húngaro]

I.P. Fellegi, D. Holt, "A systematic approach to automatic edit and imputation", *Journal of the American Statistical Association* 71 (1976) 17-35.

H. Hernández-Pérez, J.J. Salazar-González, "A Branch-and-Cut Algorithm for the Traveling Salesman Problem with Pickups and Deliveries", *Discrete Applied Mathematics* 145 (2004) 126-139.

J. Riera-Ledesma, J.J. Salazar-González, "Algorithms for automatic data editing", *Statistical Journal of the United Nations ECE* 20 (2003) 255-264.

I. Rodríguez-Martín, J.J. Salazar-González, "Decomposition Approaches for a Capacitated Hub Problem", *Lecture Notes in Artificial Intelligence* 3315 (2004) 154-163.

J.J. Salazar-González, "A Unified Mathematical Programming Framework for different Statistical Disclosure Limitation Methods", en prensa en *Operations Research*, 2005.

4. ARTÍCULOS DE APLICACIÓN

ANÁLISIS ESTADÍSTICO DE LA ESTABILIDAD EN ASIENTOS DE USO DOMÉSTICO Y PÚBLICO

Mónica Alacreu1 y Carmen Armero
Área d'Epidemiologia. Direcció General de Salut Pública.
Departament d'Estadística i Investigació Operativa. Universitat de València.

1. Introducción

La certificación de calidad de un producto tiene como objetivo garantizar al consumidor la resistencia y seguridad del artículo que va a adquirir. Además, desde el punto de vista comercial, le proporciona un signo de distinción que permite diferenciarlo de la competencia de otros en el mercado. En el sector del mueble, esta garantía se obtiene sometiendo al mobiliario a unos ensayos de acuerdo a su naturaleza y función según las especificaciones que determina la norma prevista para su uso. Estas pruebas se realizan en laboratorios especializados, que disponen de la acreditación y de la maquinaria adecuada para simular, en poco tiempo, las acciones que el mueble tendría que soportar durante varios años en su régimen normal de uso.

La Asociación de Investigación y Desarrollo en la Industria del Mueble y Afines (AIDIMA) es una entidad privada de ámbito nacional y sin ánimo de lucro, cuyo objetivo básico es lograr el progreso y desarrollo tecnológico del sector español del mueble e industrias afines, el incremento en la calidad de producción y el fortalecimiento de las exportaciones. Entre los diferentes servicios que oferta se encuentra el Laboratorio para el Mueble Acabado (LMA), que es un centro técnico donde se realizan los ensayos y certificaciones de las diversas familias de muebles. Una de las más comunes y en la que centramos este trabajo, es la sillería de uso doméstico y público, que abarca tanto las sillas (asientos sin reposabrazos) como los sillones (asientos con reposabrazos).

El análisis estadístico que presentamos en este trabajo es parte del realizado durante el disfrute de una beca de formación profesional práctica que la primera autora realizó en AIDIMA sobre la recopilación y tratamiento estadístico de resultados del LMA. Durante ese periodo analizamos los resultados obtenidos con la norma UNE EN 1022: 1998, que es una publicación que regula la metodología y las especificaciones a seguir en todos los ensayos que evalúan la estabilidad de los asientos de uso doméstico y público en España.

Este trabajo está estructurado en cuatro apartados. Tras esta introducción, explicaremos en el siguiente apartado las dificultades que tuvimos en el diseño y recopilación del banco de datos. Después, presentaremos una descripción básica de las variables que caracterizan la estabilidad de los asientos y finalizaremos aplicando técnicas básicas de supervivencia para analizar el comportamiento de la fuerza