

TRABAJO FIN DE GRADO

El método de aproximación por media muestral

Presentado por:
Rafael González López

Supervisado por:
DR. RAFAEL BLANQUERO BRAVO
DR. EMILIO CARRIZOSA PRIEGO



FACULTAD DE MATEMÁTICAS
Departamento de Estadística e Investigación Operativa
Sevilla, Junio 2018

Índice general

Abstract	5
Introducción	7
1. Planteamiento teórico del problema	9
1.1. Formulación	9
1.2. Método de aproximación por media muestral	11
1.3. Ejemplos	12
1.3.1. Problema de la cartera restringida	12
1.3.2. Problemas de asignación de recursos	13
1.3.3. Problemas de mezclas probabilístico	14
2. Convergencia	17
2.1. Definiciones y resultados previos	17
2.2. Valor y soluciones óptimas	19
2.2.1. Caso general	19
2.2.2. Restricciones determinísticas	20
2.3. Restricciones de probabilidad	22
3. Tasa de convergencia	25
3.1. Restricciones determinísticas	26
4. Experimentos	29
4.1. Normal multivariante	30
4.2. Mixtura de normales	40
Bibliografía	42

Abstract

The sample average approximation (SAA) method is an approach for solving stochastic optimization problems by using Monte Carlo simulation. The basic idea of such method is that we can approximate the expected objective function by the corresponding sample average function using a random sample. We solve the obtained sample average approximating problem by deterministic optimization techniques, and the process is repeated several times with different samples to obtain candidate solutions along with statistical estimates of their optimality gaps until a stopping criterion is satisfied.

In section 1 we describe the expected value and sample average approximation problems and give a few examples of real cases in which it can be useful. In section 2 we show many results related to convergence of estimators (objective value, optimal solution, etc) under certain assumptions. In section 3 we discuss convergence rates of objective values. In section 4 we implement the method to study two problems (that involve different random variables) to illustrate the power of the method.

Introducción

En ocasiones, cuando tratamos de resolver problemas de programación estocástica que involucran valores esperados, tenemos que lidiar un tamaño del espacio muestral excesivamente grande o con cálculos de esperanzas que pueden ser tremendamente costosos desde un punto de vista computacional. En este trabajo desarrollamos una técnica que nos permite en una gran variedad de casos aproximar estos problemas mediante el método de Monte Carlo. Esta técnica se conoce como Método de aproximación por media muestral.

Tras una presentación inicial del método, proseguiremos dando resultados relacionados con la convergencia, bajo hipótesis adecuadas, de las soluciones, valor objetivo y otros estimadores de los problemas muestrales a los correspondientes del problema real, así como una primera aproximación al orden de convergencia del método.

Finalmente, ejemplificaremos la eficacia del método con experimentos reales. Expondremos dos casos donde utilizaremos una normal multivariante y una mixtura de normales para tener una primera impresión sobre la capacidad de aproximación del método y la utilidad para casos donde, amén de heurística, resulta infactible resolver los problemas por fuerza bruta.

Capítulo 1

Planteamiento teórico del problema

1.1. Formulación

El objetivo de nuestro trabajo consistirá en el estudio del siguiente problema de optimización

$$\begin{aligned} & \min_x f_0(x) \\ & \text{s.a. } f_i(x) \leq 0 \quad i = 1, \dots, q \\ & \quad x \in \chi \end{aligned} \tag{1}$$

La particularidad que caracteriza a este problema frente a otros que hemos estudiado a lo largo del Grado en Matemáticas es que, en nuestro caso, tanto la función objetivo como las restricciones no serán necesariamente determinísticas, sino que ambas pueden estar dadas por esperanzas de variables aleatorias:

$$f_i(x) := \mathbb{E}[F_i(x, \xi)] \quad i = 0, \dots, q$$

En la formulación (1) $\chi \subset \mathbb{R}^n$ es un conjunto determinístico, $\xi \in \mathbb{R}^d$ es un vector que representa unos parámetros indeterminados del problema y F_i son funciones explícitas con imagen real. Es importante notar que en esta formulación asumimos que el vector ξ está modelado por una variable aleatoria con una distribución de probabilidad P y que las esperanzas las calcularemos con respecto a esta distribución. Recordemos que los estas esperanzas están dadas por

$$\mathbb{E}[F_i(x, \xi)] = \int F_i(x, \xi) dP(\xi)$$

Cuando P sea una variable aleatoria discreta, con ξ tomando valores en $\{\xi_1, \dots, \xi_K\}$ y probabilidades (respectivamente) p_1, \dots, p_K , la integral anterior puede escribirse como

$$\mathbb{E}[F_i(x, \xi)] = \sum_{k=1}^K p_k F_i(x, \xi_k)$$

Una referencia básica a problemas de optimización como (1) es Rubinstein y Shapiro [6].

Veamos que esta formulación es lo suficientemente versátil como para formular el problema de restricciones probabilísticas. Restringir el espacio de soluciones factibles con restricciones de la forma $G_i(x, \xi) \leq 0 \forall \xi \in \Xi$ puede introducir un número excesivo de restricciones de manera que no haya solución. En su lugar se pueden tomar unas condiciones menos restrictivas

$$P[G_i(x, \xi) \leq 0] \geq 1 - \alpha$$

con $\alpha \in (0, 1)$. Este tipo de restricciones son denominadas *restricciones probabilísticas* y $1 - \alpha$ es usualmente referido como *nivel de confianza* (confidence level). Tal y como vemos en [5] el problema puede describirse de la siguiente forma. Definimos

$$G(x, \xi) := \max_{1 \leq i \leq q} G_i(x, \xi)$$

Esto nos permite formular las restricciones como una única restricción equivalente

$$P[G(x, \xi) > 0] \leq \alpha$$

Además, como una probabilidad p_A puede verse como una esperanza utilizando una función indicatriz, tomamos $1_{(0, \infty)}(t)$ la función indicatriz de $(0, \infty)$. Como la $\mathbb{E}[c] = c$ y la esperanza es lineal, nuestro problema puede escribirse como

$$\begin{aligned} & \min_x f_0(x) \\ \text{s.a. } & f_1(x) = \mathbb{E}[1_{(0, \infty)}(G(x, \xi)) - \alpha] \leq 0 \\ & x \in \chi \end{aligned} \tag{2}$$

Por tanto, el problema de restricciones de probabilidad puede ser formulado como un problema cuyas restricciones son funciones dadas por esperanzas.

Inconvenientes de la formulación

Un claro inconveniente de la formulación anterior es que las funciones f_i pueden no tener buenas propiedades desde el punto de vista de la optimización. Por ejemplo, en el caso discreto cuando K tome valores grandes (exponencial en el tamaño de la muestra), su cálculo puede no ser directo. Supongamos que las componentes del vector ξ están distribuidas de manera que son independientes entre sí y utilizamos r para la discretización de la distribución marginal de cada componente de ξ . Por tanto, el número total de puntos será r^d , número que crece rápidamente si aumentamos la dimensión d de ξ incluso para valores moderados de r . En el caso de tener distribuciones continuas, las esperanzas estarán dadas por integrales múltiples que rara vez podrán ser evaluadas con una forma cerrada, por lo que tendríamos aproximarlas numéricamente.

1.2. Método de aproximación por media muestral

El método de aproximación por media muestral, SAA, por sus siglas en inglés de Sample Average Approximation, consiste en aproximar el problema de optimización estocástica (1) por un problema determinístico obtenido aplicando el método de Montecarlo para aproximar las funciones f_i , $i = 0, \dots, q$.

Más precisamente, se genera una muestra aleatoria ξ_1, \dots, ξ_N , idénticamente distribuidas a ξ , y se estima por el método de los momentos cada $f_i(x) = E[F_i(x, \xi)]$, de manera que

$$\hat{f}_{iN}(x) = \frac{1}{N} \sum_{k=1}^N F_i(x, \xi_k)$$

y se aproximará el problema (1) mediante el problema de optimización

$$\begin{aligned} & \min_x \hat{f}_{0N}(x) \\ & \text{s.a. } \hat{f}_{iN}(x) \leq 0 \quad i = 1, \dots, q \\ & \quad x \in \chi \end{aligned} \tag{3}$$

En el caso particular en el que las condiciones son de la forma $P[(G_i(x, \xi) \leq 0) \geq 1 - \alpha]$, tal y como vimos en (2), podemos dar una formulación

$$\hat{f}_{1N}(x) = \frac{1}{N} \sum_{k=1}^N 1_{(0, \infty)}(G(x, \xi_k)) - \gamma$$

donde γ es el nivel de confianza asociado al problema muestral. Notemos que el nivel de confianza del problema SAA puede ser diferente del nivel de confianza del problema real. Finalmente, si escribimos

$$P_N := \frac{1}{N} \sum_{k=1}^N \Delta(\xi_k)$$

entonces el problema SAA puede escribirse como

$$\begin{aligned} & \min_x f_0(x) \\ & \text{s.a. } \hat{f}_{1N}(x) = \mathbb{E}_{P_N}[1_{(0, \infty)}(G(x, \xi)) - \gamma] \leq 0 \\ & \quad x \in \chi \end{aligned}$$

1.3. Ejemplos

A continuación se presentarán algunos ejemplos de problemas de optimización estocástica y se describe su SAA asociado.

1.3.1. Problema de la cartera restringida

Consideremos el problema de maximización sujeto a una única restricción

$$\begin{aligned} & \underset{x}{\text{máx}} \mathbb{E}[r^T x] \\ & \text{s.a. } P[r^T x \geq v] \geq 1 - \alpha \\ & \quad x \in X \end{aligned} \tag{4}$$

donde $x \in \mathbb{R}^n$ es un vector de variables de decisión, r es un vector de datos aleatorio con distribución de probabilidad conocida, $v \in \mathbb{R}^n$ y $\alpha \in (0, 1)$ son constantes y

$$X := \{x \in \mathbb{R}^n \mid \sum_{i=1}^n x_i = 1, x \geq 0\}$$

Naturalmente, por propiedades básicas de la esperanza, si denotamos $\bar{r} = E[r]$ entonces $\mathbb{E}[r^T x] = \bar{r}^T x$.

La motivación para estudiar este problema podemos encontrarla en Markowitz [4]. El vector x representa el porcentaje de un total de riqueza de una unidad monetaria (u.m.) invertida en cada una de los n activos disponibles, r es el vector de rendimientos aleatorios de esas inversiones y deseamos maximizar el rendimiento sujeto a tener rendimiento medio mayor o igual a un nivel deseado v con probabilidad al menos $1 - \alpha$. Obviamente, el problema (4) no es realista, pues no tiene en cuenta características fundamentales del mercado como: costes de transacción, ventas en corto, límites inferiores o superiores de almacenamiento, etc. Sin embargo, sirve como un sencillo ejemplo de aplicación SAA.

Aplicación del SAA

Supongamos que r se distribuye según una normal multivariante con media \bar{r} y matriz de covarianzas Σ , es decir, $r \sim N(\bar{r}, \Sigma)$. Por propiedades de la distribución normal, $r^T x \sim N(\bar{r}^T x, x^T \Sigma x)$. En este caso podemos encontrar una forma explícita para la restricción en [2], aunque en general esto no va a ser posible. Entonces el problema SAA puede ser escrito como

$$\begin{aligned} & \underset{x}{\text{máx}} \bar{r}^T x \\ & \text{s.a. } \mathbb{E}_{P_N}[1_{(0, \infty)}(v - r^T x) - \gamma] \leq 0 \\ & \quad x \in \chi \end{aligned}$$

1.3.2. Problemas de asignación de recursos

Un agente tiene que escoger, de entre k proyectos, un subconjunto de ellos. Para este propósito dispone de una cantidad conocida de recursos q para asignar a los distintos proyectos. Cualquier otra cantidad de recursos adicional puede obtenerse con un coste de c , más alto, por unidad de recurso. La cantidad de recurso W_i requerida para emprender cada proyecto no es conocida, pero suponemos que nuestro agente tiene una estimación de cómo se distribuye $W = (W_1, \dots, W_k)$. Denotamos $[x]^+ = \max\{x, 0\}$. Si para cada proyecto i denotamos r_i como la diferencia entre los ingresos y los costes, entonces, como el consumo de recursos extra es $c[W'x - q]^+$, el beneficio neto es la v.a.

$$r'x + c[W'x - q]^+$$

Por tanto, el beneficio neto esperado, a maximizar, es

$$\begin{aligned} \max_x r'x - c\mathbb{E}[W'x - q]^+ \\ \text{s.a. } x \in \{0, 1\}^k \end{aligned} \quad (5)$$

Elegimos presentar este problema en los ejemplos por varias razones. Por ejemplo, muchos problemas interesantes como el problema de la tripulación (una aproximación probabilística a este problema puede verse en [8]) o la versión estocástica del problema del camino más corto en un grafo, que detallaremos más adelante, tienen funciones objetivos parecidas a la de este ejemplo en términos de valores esperados.

Otra razón para su estudio es que en la función objetivo encontramos términos de la forma $\mathbb{E}[\sum_i W_i x_i - q]^+$. En muchos problemas de programación estocástica se pueden obtener soluciones muy buenas si sustituimos las variables aleatorias implicadas en el problema por sus medias y resolviendo el problema determinístico tal y como podemos ver en [1]. Sin embargo, es fácil ver que en general cuando tenemos este tipo de valores esperados no tiene por qué ocurrir así. Dada una solución x , este término puede ser grande y, a pesar de ello, volverse muy pequeño si sustituimos cada W_i por su esperanza.

Finalmente, hemos de tener en cuenta el interés que puede tener este problema en sí mismo. Aplicaciones de este problema pueden ser la toma de decisiones por parte de un contratista que provee a varios usuarios de manera simultánea, como una compañía eléctrica o una empresa de construcción. La cantidad de trabajo que requerirá para cada contrato es desconocida cuando el contrato es firmado, los materiales tienen un bajo costo y si la cantidad de trabajo requerida con un alto coste.

Aplicación del SAA

Denotaremos por W_j^i $j = 1, \dots, N$, la muestra aleatoria de W_i , $i = 1, \dots, k$. Por tanto, podemos formular la aplicación del SAA del problema 5 como

$$\begin{aligned} \max_{x,z} \quad & r^T x - \frac{c}{N} \sum_{i=1}^N z_i \\ \text{s.a.} \quad & z_j \geq \sum_{i=1}^k W_i^j x_i - q \quad j = 1, \dots, N \\ & z_j \geq 0 \quad j = 1, \dots, N \\ & x \in \{0, 1\}^k \end{aligned}$$

Este problema puede ser resuelto por el método de ramificación y acotación, usando el problema lineal relajado para tener cotas superiores.

1.3.3. Problemas de mezclas probabilístico

Vamos a considerar otro ejemplo de problema de restricciones probabilísticas. Supongamos que un granjero desea fertilizar sus cultivos con el fin de aumentar su producción. Él decide utilizar una ciertas cantidades v_1, \dots, v_k de k nutrientes. El granjero tiene a su disposición en el mercado h fertilizantes disponibles, de manera que cada fertilizante i contiene ω_{ij} gramos del nutriente j , $1 \leq i \leq h$, $1 \leq j \leq k$. Precisamente serán ciertos ω_{ij} los que normalmente en este y otros problemas análogos sean parámetros desconocidos distribuidos según alguna variable aleatoria.

Existen muchas formas de modelar el problema de mezclas. Una exposición detallada de las mismas puede encontrarse en [3]. Si denotamos por x_i la cantidad de fertilizante i que compramos a un precio p_i , entonces podemos considerar la formulación

$$\begin{aligned} \min_{x_i} \quad & \sum_{i=1}^h p_i x_i \\ \text{s.a.} \quad & P \left[\sum_{i=1}^h \omega_{ij} x_i \geq v_j \quad \forall j = 1, \dots, k \right] \geq 1 - \alpha \\ & x_i \in \mathbb{R}_{\geq 0} \quad i = 1, \dots, h \end{aligned} \quad (6)$$

Notemos que el problema (6) no está escrito en la forma 2. Sin embargo, si denotamos

$$G(x_1, \dots, x_h) = \min_{1 \leq j \leq k} -v_j + \sum_{i=1}^h \omega_{ij} x_i$$

podemos escribir (6) como

$$\begin{aligned} & \underset{x_i}{\text{mín}} \sum_{i=1}^h p_i x_i \\ & \text{s.a. } P[G(x_1, \dots, x_h, \xi) > 0] \leq \alpha \\ & \quad x_i \in \mathbb{R}_{\geq 0} \qquad \qquad \qquad i = 1, \dots, h \end{aligned}$$

Procediendo de manera análoga al desarrollo teórico podemos encontrar el problema SAA asociado.

Capítulo 2

Convergencia

2.1. Definiciones y resultados previos

Vamos a dar algunas definiciones y resultados de teoría de la probabilidad que serán necesarios para comprender o demostrar los enunciados propios de nuestro trabajo.

Notación 2.1.1. Notaremos por ϑ^* y $\hat{\vartheta}_N$ los valores óptimos del problema real (1) y el problema SAA (2) respectivamente. De la misma forma, S y \hat{S}_N para el conjunto de soluciones óptimas del problema real y el SAA respectivamente y X y \hat{X}_N para referirnos a los conjuntos factibles, es decir

$$\begin{aligned} X &= \chi \cap \{x \mid f_i(x) \leq 0 \forall i = 1, \dots, q\} \\ \hat{X}_N &= \chi \cap \{x \mid \hat{f}_{iN}(x) \leq 0 \forall i = 1, \dots, q\} \end{aligned}$$

Observamos que X y S son subconjuntos de \mathbb{R}^n , mientras que \hat{x}_N y \hat{S}_N , como dependen del espacio probabilístico, son conjuntos aleatorios. Además, dado $\varepsilon \geq 0$ diremos que x es un solución ε -óptima si $f_0(x) \leq \vartheta^* + \varepsilon$. El conjunto de las soluciones ε -óptimas del problema real y el SAA será denotado por S^ε y \hat{S}_N^ε .

Definición 2.1.1. Una sucesión de variables aleatorias $\{X_N\}$ definida sobre un espacio probabilístico (Ω, A, P) converge de manera casi segura a X o con probabilidad 1, lo cual notaremos $X_N \xrightarrow{CS} X$, si se verifica que:

$$P(\{\omega \in \Omega \mid X_N(\omega) \longrightarrow X(\omega)\}) = 1$$

Definición 2.1.2. Una sucesión de variables aleatorias $\{X_N\}$ definida sobre un espacio probabilístico (Ω, A, P) converge en probabilidad a X , que notaremos $X_N \xrightarrow{P} X$ si se verifica:

$$\lim_{N \rightarrow \infty} P(\{\omega \in \Omega \mid |X_N(\omega) - X(\omega)| \leq \varepsilon\}) = 1$$

Definición 2.1.3. Una sucesión de variables aleatorias $\{X_N\}$ definida sobre un espacio probabilístico (Ω, A, P) con funciones de distribución F_{X_N} converge en Ley a X , lo cual notaremos $X_N \xrightarrow{L} X$ si se verifica que:

$$F_{X_N}(x) \longrightarrow F_X(x) \forall x \text{ punto de continuidad de } F_X$$

Proposición 2.1.1. La convergencia casi segura implica convergencia en probabilidad, y esta implica la convergencia en Ley.

Definición 2.1.4. Una sucesión de funciones $f_N: \chi \times \Omega \longrightarrow \mathbb{R}$ con (Ω, A, P) espacio probabilístico, converge uniformemente de manera casi segura a la función $f: \chi \times \Omega \longrightarrow \mathbb{R}$ si

$$P(\{\omega \in \Omega \mid \sup_{x \in \chi} |f_N(x, \omega) - f(x, \omega)| \rightarrow 0\}) = 1$$

Teorema 2.1.1 (Ley fuerte de los grandes números). Sea $\{X_N\}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas (i.i.d) con esperanza finita. Sea $\mu = E[X_N]$ entonces:

$$\frac{1}{N} \sum_{i=1}^N X_i \xrightarrow{CS} \mu$$

Teorema 2.1.2 (Teorema central del límite). Sea $\{X_N\}$ una sucesión de variables aleatorias i.i.d con varianza finita y no nula. Sean $\mu = E[X_N]$ y $\sigma^2 = V[X_N]$ entonces:

$$\sqrt{N} \frac{\frac{1}{N} \sum_{i=1}^N X_i - \mu}{\sigma} \xrightarrow{L} N(0, 1)$$

donde $N(0, 1)$ denota la distribución normal de media 0 y varianza 1.

Definición 2.1.5. Sea $f: \mathbb{R}^n \times \Xi \longrightarrow \mathbb{R}$. Diremos que $f(x, y)$ es una función de Carathéodory si $f(x, \cdot)$ es medible para cada $x \in \mathbb{R}^n$ y $f(\cdot, y)$ es continua para cada $y \in \Xi$.

Definición 2.1.6. Sea una sucesión de funciones $f_N: \chi \subset \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$. Diremos que $\{f_N(x)\}$ epiconverge a $f(x)$ y denotaremos $f_N \xrightarrow{e} f$ si se verifica

- Para cada sucesión $\{x_N\}$ tal que $x_N \rightarrow x$

$$\liminf_{N \rightarrow \infty} f_N(x_N) \geq f(x)$$

- Existe $\{x_N\}$ tal que $x_N \rightarrow x$ y tal que

$$\limsup_{N \rightarrow \infty} f_N(x_N) \leq f(x)$$

Definición 2.1.7. Sea $f: \chi \rightarrow \overline{\mathbb{R}}$. Diremos que f es semicontinua inferiormente en x_0 si

$$\liminf_{x \rightarrow x_0} f(x) \geq f(x_0)$$

Análogamente f se dice semicontinua superiormente en x_0 si

$$\limsup_{x \rightarrow x_0} f(x) \leq f(x_0)$$

Definición 2.1.8. Sea $f: \chi \rightarrow \overline{\mathbb{R}}$. Para cada $\alpha \in \mathbb{R}$ definimos los conjuntos de nivel inferiores como $\{x \in \chi \mid f(x) \leq \alpha\}$.

Proposición 2.1.2. Una función es semicontinua inferiormente si y solo si los conjuntos de nivel inferiores son cerrados.

Definición 2.1.9. Sean $A, B \subset \mathbb{R}^n$. Definimos la distancia de $x \in \mathbb{R}^n$ a A como

$$\text{dist}(x, A) := \inf_{y \in A} \|x - y\|$$

y la desviación del conjunto A del conjunto B como

$$\mathbb{D}(A, B) := \sup_{x \in A} \text{dist}(x, B)$$

Por convenio, diremos $\text{dist}(x, \emptyset) = \infty$.

2.2. Valor y soluciones óptimas

2.2.1. Caso general

Primeramente veremos un teorema que nos asegure convergencia del valor óptimo y del conjunto de soluciones factibles en la formulación general del problema.

Teorema 2.2.1. Supongamos que X , conjunto factible de (1) es cerrado y que existe un compacto de $C \subset \mathbb{R}^n$ tal que

1. El conjunto S de soluciones óptimas de (1) es no vacío y está contenido en C .
2. La función $f_0(x)$ tiene imagen finita y es continua en C .
3. $\hat{f}_{0N} \xrightarrow{CS} f_0$ uniformemente en C cuando $N \rightarrow \infty$.
4. Con probabilidad 1 para N suficientemente grande \hat{S}_N es no vacío y $\hat{S}_N \subset C$
5. Para toda sucesión $\{x_N\}$ con $x_N \in \hat{X}_N$ y $x_N \xrightarrow{CS} x$, entonces $x \in X$.
6. $\exists x \in S$ tal que $\exists \{x_N\}$ con $x_N \in \hat{X}_N$ tal que $x_N \xrightarrow{CS} x$.

Entonces $\hat{\vartheta}_N \rightarrow \vartheta^*$ y $\mathbb{D}(\hat{S}_N, S) \rightarrow 0$ con probabilidad 1 cuando $N \rightarrow \infty$.

Una prueba de este teorema puede encontrarse en [7].

2.2.2. Restricciones determinísticas

Veamos algunos resultados para el caso en el que $\hat{X}_N = X$, es decir, cuando las restricciones son determinísticas. Vamos a enunciar algunos resultados con características sobre X , que bien podrían ser enunciados sobre χ , pero sería claramente innecesario.

Nota 2.2.1. *Si las restricciones del problema son determinísticas, entonces las dos últimas condiciones del Teorema 2.2.1 son innecesarias y se tiene el mismo resultado.*

Proposición 2.2.1. *Supongamos que $\hat{f}_{0N}(x)$ converge a $f_0(x)$ con probabilidad 1 cuando $N \rightarrow \infty$ uniformemente en X . Entonces $\hat{\vartheta}_N$ converge con probabilidad 1 a ϑ^* cuando $N \rightarrow \infty$.*

Demostración. *Aunque lo omitamos, la función \hat{f}_{0N} es aleatoria, pues depende de la muestra, es decir $\hat{f}_{0N}(x) = \hat{f}_{0N}(x, \omega)$. La convergencia uniforme con probabilidad 1 nos dice que $\forall \varepsilon > 0$ y para casi todo $\omega \in \Xi \exists N^* = N^*(\varepsilon, \omega)$ tal que, $\forall N \geq N^*$*

$$\sup_{x \in X} \left| \hat{f}_{0N}(x, \omega) - f_0(x) \right| \leq \varepsilon$$

Se sigue inmediatamente que $\left| \hat{\vartheta}_N - \vartheta^ \right| \leq \varepsilon$ para todo $N \geq N^*$. □*

Proposición 2.2.2. *Si X es compacto y las funciones involucradas son continuas, entonces*

$$\left| \hat{\vartheta}_N - \vartheta^* \right| \leq \delta_N := \max_{x \in X} \left| \hat{f}_{0N}(x) - f_0(x) \right|$$

Nota 2.2.2. *El resultado se seguiría manteniendo de manera trivial si prescindimos de las hipótesis sobre las funciones pero imponemos que el conjunto X sea finito.*

Demostración. *La hipótesis de compacidad se usa para garantizar la existencia de máximos y mínimos*

$$-\max_{x \in X} |f(x)| = \min_{x \in X} -|f(x)| \leq \min_{x \in X} f(x)$$

Se tiene

$$\begin{aligned} \min_{x \in X} \hat{f}_{0N}(x) &= \min_{x \in X} (\hat{f}_{0N}(x) + f_0(x) - f_0(x)) \\ &\geq \min_{x \in X} (\hat{f}_{0N}(x) - f_0(x)) + \min_{x \in X} f_0(x) \\ &\geq -\max_{x \in X} \left| \hat{f}_{0N}(x) - f_0(x) \right| + \min_{x \in X} f_0(x) \end{aligned}$$

Por lo que

$$\delta_N \geq \hat{\vartheta}_N - \vartheta^*$$

La otra desigualdad se deduce del siguiente razonamiento

$$\begin{aligned} \min_{x \in X} f_0(x) &= \min_{x \in X} (f_0(x) - \hat{f}_{0N}(x) + \hat{f}_{0N}(x)) \\ &\geq \min_{x \in X} (f_0(x) - \hat{f}_{0N}(x)) + \min_{x \in X} \hat{f}_{0N}(x) \\ &\geq -\max_{x \in X} |\hat{f}_{0N}(x) - f_0(x)| + \min_{x \in X} \hat{f}_{0N}(x) \end{aligned}$$

Por tanto

$$\delta_N \geq \vartheta^* - \hat{\vartheta}_N$$

□

Conjunto factible finito

Si mantenemos que $X = \hat{X}_N$ y además imponemos que X sea conjunto finito, podemos mejorar algunos resultados.

Proposición 2.2.3. *Se tienen las siguientes propiedades*

- $\hat{\vartheta}_N \xrightarrow{CS} \vartheta^*$ cuando $N \rightarrow \infty$.
- Para todo $\varepsilon \geq 0$ el suceso $\{\hat{S}_N^\varepsilon \subset S^\varepsilon\}$ ocurre con probabilidad 1 para N suficientemente grande.

Demostración.

- Se sigue de manera inmediata a partir de la ley fuerte de los grandes números (Teorema 2.1.1) que para cada $x \in S$, $\hat{f}_{0N}(x)$ converge de manera casi segura a $f_0(x)$. Dado que X es finito y la unión finita de conjuntos de medida nula es de medida nula, tenemos que $\hat{f}_{0N}(x)$ converge de manera uniforme $f_0(x)$, es decir,

$$\max_{x \in X} |\hat{f}_{0N}(x) - f_0(x)| \xrightarrow{CS} 0 \quad N \rightarrow \infty$$

Junto a la Proposición 2.2.2 se tiene el primer punto.

- Dado un $\varepsilon \geq 0$ podemos considerar

$$\rho(\varepsilon) := \min_{x \in X \setminus S^\varepsilon} f_0(x) - \vartheta^* - \varepsilon$$

Dado cualquier $x \in X \setminus S^\varepsilon$, por definición $f_0(x) > \vartheta^* + \varepsilon$ y siendo X un conjunto finito, se sigue que $\rho(\varepsilon) > 0$. Luego, $\forall x \in X$ $f_0(x) \geq \vartheta^* + \varepsilon + \rho(\varepsilon)$.

Sea N lo suficientemente grande como para asegurar que $\delta_N < \rho(\varepsilon)/2$. En particular, tenemos que $\hat{f}_{0N}(x) > f_0(x) - \rho(\varepsilon)/2 \forall x \in X$. Si $x \in \bar{X} \setminus S^\varepsilon$, tenemos además

$$\hat{f}_{0N}(x) > f_0(x) - \rho(\varepsilon)/2 \geq \vartheta^* + \varepsilon + \rho(\varepsilon) - \rho(\varepsilon)/2 = \vartheta^* + \varepsilon + \rho(\varepsilon)/2$$

Usando la Proposición 2.2.2 tenemos que $\hat{\vartheta}_N - \vartheta^* \leq \delta_N < \rho(\varepsilon)/2$, Por tanto, si $x \in X \setminus S^\varepsilon$ entonces tenemos

$$\hat{f}_{0N}(x) > \vartheta^* + \varepsilon + \rho(\varepsilon)/2 > \hat{\vartheta}_N + \varepsilon$$

Por lo que $x \in X \setminus \hat{S}_N^\varepsilon$, de donde se deduce el resultado. \square

2.3. Restricciones de probabilidad

Veamos algunos resultados referentes al problema (2). En esta sección supondremos que χ es cerrado.

Proposición 2.3.1. *Sea $G(x, \xi)$ una función de Carathéodory. Entonces $\hat{f}_{1N}(x)$ y $f_1(x)$ son semicontinuas inferiormente y $\hat{f}_{1N}(x) \xrightarrow{e} f_1(x)$ con probabilidad 1. Es más, si $\forall x \in \chi G(x, \xi) \neq 0$ con probabilidad 1, entonces $\hat{f}_{1N}(x)$ es continua en χ y converge a $f_1(x)$ uniformemente en cada compacto $C \subset \chi$.*

Una demostración de este resultado puede encontrarse en [5].

Proposición 2.3.2. *Supongamos que el nivel de confianza de SAA y el problema real es el mismo ($\alpha = \gamma$), χ es compacto y la función $f(x)$ es continua, $G(x, \xi)$ es una función de Carathéodory y se verifica la siguiente condición*

- (A) *Hay una solución x^* del problema real tal que para todo $\varepsilon > 0$ existe $x \in \chi$ tal que $\|x^* - x\| \leq \varepsilon$ y $f_1(x) < \alpha$.*

Entonces $\hat{\vartheta}_N \rightarrow \vartheta^*$ y $\mathbb{D}(\hat{S}_N, S) \rightarrow \infty$ con probabilidad 1 cuando $N \rightarrow \infty$.

Demostración. *Por la condición (A) el conjunto S es no vacío y existe un x' tal que $f_1(x') < \alpha$. Que $\hat{f}_{1N}(x') \xrightarrow{CS} f_1(x')$ se tiene por la ley fuerte de los grandes números. Por tanto, con probabilidad 1 y para N suficientemente grande $\hat{f}_{1N}(x') < \alpha$ y el problema SAA tiene solución factible. De hecho, $x' \in X_N$ para N suficientemente grande con probabilidad 1. Además, usando la semicontinuidad inferior de \hat{f}_{1N} , el conjunto factible del problema SAA \hat{X}_N es compacto (es la intersección de un compacto y un cerrado [Proposición 2.1.2]) y, por tanto, por continuidad de f_0 , \hat{S}_N es no vacío con probabilidad 1 para N suficientemente grande. Si $x_N \in \hat{X}_N$, entonces $f_0(x_N) \geq \hat{\vartheta}_N$. Utilizando (A) y el razonamiento anterior podemos considerar una sucesión x_N con $x_N \in \hat{X}_N$ con probabilidad 1 para N suficientemente grande, tal que converge a x^* . Usando que f es continua, tenemos que con probabilidad 1*

$$\limsup_{N \rightarrow \infty} \hat{\vartheta}_N \leq \limsup_{N \rightarrow \infty} f(x_N) = f(x^*) = \vartheta^*$$

Sea ahora $x_N \in \hat{S}_N$. Dado que χ es compacto, renombrando una subsucesión si fuera necesario, tenemos una sucesión x_N converge a un punto $\bar{x} \in \chi$ con probabilidad

1. Junto con el hecho de que $\hat{f}_{1N} \xrightarrow{e} f_1(x)$ con probabilidad 1, tenemos que con probabilidad 1

$$\liminf_{N \rightarrow \infty} \hat{f}_{1N}(x_N) \geq f_1(\bar{x})$$

Se sigue que $f_1(\bar{x}) \leq \alpha$ y, por tanto, \bar{x} es punto factible del problema real, por lo que $f_0(\bar{x}) \geq \vartheta^*$. Como también se tiene que $f(x_N) \xrightarrow{CS} f(\bar{x})$, tenemos con probabilidad 1 que

$$\liminf_{N \rightarrow \infty} \hat{\vartheta}_N \geq \vartheta^*$$

Por tanto, $\hat{\vartheta}_N \rightarrow \vartheta^*$ con probabilidad 1. Además, \bar{x} es solución óptima del problema real, luego $\mathbb{D}(\hat{S}_N, S) \rightarrow \infty$ con probabilidad 1. \square

Nota 2.3.1. Notemos que la condición (A) es esencial para probar la convergencia casi segura de $\hat{\vartheta}_N$ y \hat{S}_N . Pensemos, por ejemplo, en una situación donde las restricciones del problema dejen un único punto factible \bar{x} tal que $f_0(\bar{x}) = \alpha$. Cualquier perturbación arbitrariamente pequeña en la restricción $\hat{f}_{1N}(x) \leq \alpha$ puede provocar que el conjunto factible del problema SAA sea vacío.

Nota 2.3.2. Observemos además que no hemos usado la condición (A) para probar

$$\liminf_{N \rightarrow \infty} \hat{\vartheta}_N \geq \vartheta^*$$

Capítulo 3

Tasa de convergencia

En el capítulo anterior hemos probado la consistencia de los estimadores del problema SAA bajo ciertas hipótesis. El hecho que estos converjan al menos en probabilidad asegura que el error en la estimación de los parámetros se aproxima a cero cuando el tamaño de la muestra crece. A pesar de la importancia desde un punto de vista teórico, dada una muestra no tenemos ninguna indicación de la magnitud del error que podemos estar cometiendo.

Definición 3.0.1. Diremos que un estimador X_n está estocásticamente acotado por la sucesión $\{a_n\}$ y notaremos $X_n = O_p(a_n)$ si $\forall \varepsilon > 0 \exists M, N > 0$ finitos tales que

$$P(|X_n/a_n| > M) < \varepsilon \quad \forall n > N$$

Proposición 3.0.1. Consideremos el problema (1). Si la varianza $V[F_0(x, \xi)]$ es finita y N es el tamaño de la muestra, entonces $f_{0N}(x) = O_p(N^{-1/2})$.

Demostración. Por el Teorema Central del Límite sabemos que $\forall x \in \chi$

$$\sqrt{N} \frac{\hat{f}_{0N}(x) - f_0(x)}{\sigma} \xrightarrow{L} N(0, 1)$$

donde $\sigma^2 = V[F_0(x, \xi)]$. Esto significa que $\hat{f}_{0N}(x)$ está asintóticamente distribuida como una normal $N(f_0(x), \sigma^2/N)$. Por tanto, tenemos un intervalo de confianza asintótico al $100(1 - \alpha)\%$ para $f_0(x)$ dado por

$$\left[\hat{f}_{0N}(x) - \frac{z_{\alpha/2} \hat{\sigma}(x)}{\sqrt{N}}, \hat{f}_{0N}(x) + \frac{z_{\alpha/2} \hat{\sigma}(x)}{\sqrt{N}} \right]$$

donde $z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$, $\Phi(\cdot)$ es la función de distribución de la normal estándar y $\hat{\sigma}^2(x)$ es la cuasivarianza muestral dada por

$$\hat{\sigma}^2(x) = \frac{1}{N-1} \sum_{j=1}^N \left(F(x, \xi) - \hat{f}_{0N}(x) \right)^2$$

De aquí se sigue el resultado. □

3.1. Restricciones determinísticas

Volvamos a considerar el caso donde $X_N = X$. Sabemos que en estos casos podemos hablar indistintamente de χ o X , renombrando $\chi := X$ si hiciera falta.

Proposición 3.1.1. *Se verifica*

$$\vartheta^* \geq \mathbb{E}[\hat{\vartheta}_N]$$

Demostración. *Claramente, $\forall x' \in X$ tenemos que $\hat{f}_{0N}(x') \geq \inf_{x \in X} \hat{f}_{0N}(x)$. Ahora podemos tomar esperanza a ambos lados y tomar ínfimo en el izquierdo, obteniendo así*

$$\inf_{x \in X} \mathbb{E}[\hat{f}_{0N}(x)] \geq \mathbb{E}\left[\inf_{x \in X} \hat{f}_{0N}(x)\right]$$

Dado que $\mathbb{E}[\hat{f}_{0N}(x)] = f_0(x)$ y además $\inf_{x \in X} \hat{f}_{0N}(x) = \hat{\vartheta}_N$ obtenemos

$$\hat{f}_{0N}(x) \geq \vartheta^* \geq \mathbb{E}[\hat{\vartheta}_N]$$

□

Proposición 3.1.2. *Para todo $n \in \mathbb{N}$ se verifica*

$$\mathbb{E}[\hat{\vartheta}_N] \leq \mathbb{E}[\hat{\vartheta}_{N+1}] \leq \vartheta^*$$

Demostración. *Es inmediato por la proposición anterior que $\mathbb{E}[\hat{\vartheta}_{N+1}] \leq \vartheta^*$. Podemos escribir*

$$\hat{f}_{0N+1}(x) = \frac{1}{N+1} \sum_{i=1}^{N+1} \left[\frac{1}{N} \sum_{j \neq i} F_0(x, \xi^j) \right]$$

Se sigue

$$\begin{aligned} \mathbb{E}[\hat{\vartheta}_{N+1}] &= \mathbb{E}\left[\inf_{x \in X} \hat{f}_{0N+1}(x)\right] \\ &= \mathbb{E}\left[\inf_{x \in X} \frac{1}{N+1} \sum_{i=1}^{N+1} \left(\frac{1}{N} \sum_{j \neq i} F_0(x, \xi^j) \right)\right] \\ &\geq \mathbb{E}\left[\frac{1}{N+1} \sum_{i=1}^{N+1} \left(\inf_{x \in X} \frac{1}{N} \sum_{j \neq i} F_0(x, \xi^j) \right)\right] \\ &= \frac{1}{N+1} \sum_{i=1}^{N+1} \mathbb{E}\left[\inf_{x \in X} \frac{1}{N} \sum_{j \neq i} F_0(x, \xi^j)\right] \\ &= \frac{1}{N+1} \sum_{i=1}^{N+1} \mathbb{E}[\hat{\vartheta}_N] \\ &= \mathbb{E}[\hat{\vartheta}_N] \end{aligned}$$

□

Proposición 3.1.3. *Supongamos que X es compacto y se verifican las condiciones*

(A1) *Existe algún $\bar{x} \in X$ tal que $\mathbb{E}[F(\bar{x}, \xi)^2]$ es finito.*

(A2) *Existe una función medible $C: \Xi \rightarrow \mathbb{R}_+$ tal que $\mathbb{E}[C(\xi)^2]$ es finito y*

$$|F_0(x, \xi) - F_0(x', \xi)| \leq C(\xi) \|x - x'\|$$

para todo $x, x' \in X$ y casi todo $\xi \in \Xi$.

entonces se tiene:

$$\hat{\vartheta}_N = \inf_{x \in S} \hat{f}_N(x) + O_p(N^{-1/2})$$

Si $S = \{\bar{x}\}$ es un conjunto unitario, entonces

$$N^{1/2} \left(\hat{\vartheta}_N - \vartheta^* \right) \xrightarrow{L} N(0, \sigma^2(\bar{x}))$$

Una demostración de esta proposición puede encontrarse en [7]

Capítulo 4

Experimentos

Con el fin de ilustrar el potencial del método, vamos a aplicarlo a dos experimentos y presentaremos los resultados. Para modelar y resolver los problemas utilizaremos el software Gurobi (v. 8.0.0) con una interfaz de Python (v. 3.6.2).

Nuestro objetivo será minimizar los costes al elegir dentro de una determinada gama de n proyectos, imponiendo que la probabilidad de que el rendimiento sea mayor o igual que una cierta constante v sea mayor o igual que una cierta probabilidad $1 - \alpha$. Aceptar el proyecto i implica asumir un coste c_i fijo y proporciona un rendimiento aleatorio w_i . Denotemos $W = (w_1, \dots, w_n)$, vector aleatorio de marginales independientes. De esta forma, el problema de optimización considerado es

$$\begin{aligned} \min_x \quad & \sum_{i=1}^n c_i x_i \\ \text{s.a.} \quad & P \left[\sum_{i=1}^n w_i x_i \geq v \right] \geq 1 - \alpha \\ & x \in \{0, 1\}^n \end{aligned}$$

Para la distribución de los rendimientos plantearemos los casos. En primer lugar que W es un vector normal multivariante, y en el segundo una mixtura de normales independientes.

En cualquier caso, el problema real induce un problema SAA que tenemos que resolver. Sea W_1, \dots, W_N una muestra de tamaño N del vector aleatorio W . Dadas las limitaciones que tenemos para resolver problemas de programación entera con restricciones cuadráticas es necesario que hagamos uso de una constante auxiliar $K > 0$ que

nos permita formular el problema, que puede formularse de la siguiente forma

$$\begin{aligned} \min_{x,z} \quad & \sum_{i=1}^n c_i x_i \\ \text{s.a.} \quad & W_i' x - v \geq -K(1 - z_i) \quad \forall i = 1, \dots, N \\ & \frac{1}{N} \sum_{i=1}^N z_i \geq 1 - \alpha \\ & x \in \{0, 1\}^n \\ & z \in \{0, 1\}^N \end{aligned}$$

Explicamos esta formulación. Cuando tratamos de traducir la restricción probabilística a una que pueda implementarse en nuestro software aparece de manera natural unas ciertas variables z_i que deben contar el número de veces que la muestra verifica la desigualdad deseada. Si $z_i = 1$ entonces la restricción $W_i' x - v$ se verifica mientras que si $z_i = 0$ entonces no se impone nada sobre su cumplimiento, pues tomamos K tal que $\forall x P[W' x - v \leq -K]$ sea despreciable.

4.1. Normal multivariante

Para el caso práctico supondremos que $n = 17$, $1 - \alpha = 0,95$, $v = 18$ y vector de costes

$$\begin{aligned} c_{1:9} &= (0,55 \ 0,57 \ 0,46 \ 0,02 \ 0,25 \ 0,34 \ 0,89 \ 0,49 \ 0,50) \\ c_{10:18} &= (0,91 \ 0,52 \ 0,4 \ 0,6 \ 0,2 \ 0,225 \ 0,344 \ 0,9 \ 0,45) \end{aligned}$$

Los parámetros de las variables los generaremos de manera aleatoria pero con una semilla fija.

Comencemos generando los datos del problema. En este caso, utilizamos una semilla fija para asegurar la reproducibilidad del experimento.

```

1 set.seed(481516)
2 mu = abs(round(rnorm(17, 1.5, 1), 2))
3 # Imponemos que mu sea un vector de números positivos
4 # pues suponemos que son rendimientos económicos de los
5 # distintos proyectos.
6 sigma = abs(round(rnorm(17, 0, 0.3), 2))

```

Ahora que ya tenemos todos los datos del problema, vamos a resolverlo en R por enumeración completa. Esto nos facilitará posteriormente una primera impresión sobre la capacidad del método.

```
1 #####
2 #
3 # Scripts para solucionar el problema #
4 #   real por enumeración completa   #
5 #
6 #####
7
8 # Unos comandos para limpiar el entorno
9 closeAllConnections()
10 rm(list=ls())
11
12 # Dimensión del problema
13 n = 17
14
15 # Con el siguiente comando obtenemos una enumeración
16 # completa de {0,1}^n
17
18 A <-expand.grid(rep(list(0:1),n))
19 colnames(A)<-NULL
20 B = as.matrix(A)
21
22 # Escribimos el vector de costes
23 c = c(0.55, 0.57, 0.46, 0.02,0.25, 0.34, 0.89, 0.49 ,0.50,
24       0.91, 0.52, 0.4, 0.6, 0.2,0.225, 0.344, 0.9)
25
26 # Fijamos la constante del rendimiento
27 v = 18
28
29 # Y finalmente el parámetro de probabilidad 1-alfa
30 alfa = 0.05
31 beta = 1-alfa
32
33 # Distribución normal
34 #####
35
36 # Como suponemos independencia, calculamos la probabilidad
37 # de una normal univariante de media c'mu y varianza
38 # (c^2)'sigma^2, donde mu y sigma son los vectores de
39 # las respectivas medias y desviaciones.
40
41 set.seed(481516)
42 mu = abs(round(rnorm(17,1.5,1),2))
43 sigma = abs(round(rnorm(17,0,0.3),2))
44
```

```

45 # Con la siguiente función calculamos el valor de la restricci
    ón
46 normi <- function(q){
47   mus = t(q) %* %mu;
48   sigmas = sqrt((t(q) %* %sigma^2));
49   # La desviación típica de la suma es la raíz de la suma de
    las varianzas.
50   a = 1-pnorm(v,mean = mus, sd = sigmas)
51   return(a) }
52
53 # Este método nos devuelve las posibilidades que verifican la
    restricción
54 m = c();
55 for (i in 1:2^n){
56   if (normi(B[i,])<beta){
57     m = append(m,i) }
58 }
59 if (length(m)==0){
60   D = B
61 }else
62 {
63 D = B[-m,]}
64 head(D)
65
66 # A continuación minimizamos
67
68 obje <- function(q){
69   t(c) %* %q
70 }
71 H = apply(D,1,obje)
72 head(H)
73
74 # Obtenemos la/s solución/es
75 S = D[which(H==min(H)),]
76 S
77
78 # Y el valor objetivo
79 nu = min(H)
80 nu

```

```

1 > S
2 [1] 0 1 0 1 1 1 0 0 0 0 1 1 1 1 1 0 1
3 > nu
4 [1] 4.369

```

Por tanto, tenemos solución única con valor objetivo 4,369. A continuación vamos a presentar un script, también en R, para generar un conjunto de muestras de una normal 10-variante con los parámetros anteriores (las componentes se suponen independientes). En nuestro experimento vamos a resolver, para cada tamaño muestral considerado, 20 problemas y considerar la solución media como solución asociada a dicho tamaño muestral.

```

1 #####
2 #                                     #
3 #   Scripts para generar muestras   #
4 #                                     #
5 #####
6
7 # Unos comandos para limpiar el entorno
8 closeAllConnections()
9 rm(list=ls())
10
11 # Número de problemas que vamos a resolver para cada tamaño
    muestral
12 M = 20;
13
14 library(mvtnorm)
15
16 # Fijamos la semilla
17 set.seed(481516)
18 medias = abs(round(rnorm(17, 1.5, 1), 2))
19 desvi = abs(round(rnorm(17, 0, 0.3), 2))
20 matcov<-diag(desvi^2)
21
22 # Generamos las muestras de normal multivariante N(mu,sigma)
23 # de tamaño 7*i cuyas componentes son independientes entre sí.
24 set.seed(123456)
25 dir = paste(getwd(), "/muestreo", sep="")
26 for (i in 1:40) {
27   direc = paste(dir, "/datos", i, ".txt", sep = "")
28   direc
29   data = rmvnorm(M*7*i, mean=medias, sigma=matcov)
30   write.table(data, direc, row.names=F, sep="\t")
31 }

```

Finalmente, resolvemos usando Gurobi (con interfaz de Python 3.6) los problemas muestrales. Notemos que el código que hemos utilizado para la resolución en Python, así como el planteamiento del problema muestral, naturalmente es independiente de la distribución que siga W , por lo que no necesitamos volver a citarlo en un caso posterior.

```
1 # Utilizamos como interfaz para el software Gurobi.
2
3 from gurobipy import *
4 import numpy as np
5
6 # Leemos todas las muestras.
7 datos = [np.loadtxt("muestreo/datos" + str(i) + ".txt",
8                 skiprows=1) for i in range(1,41)]
9
10 # Aunque nuestro tamaño muestra varía, vamos a resolver
11     siempre "20" problemas.
12 # Es decir, cada elemento de datos es una lista con "20*(7*i)"
13     listas de 10 elementos.
14
15 # Con la siguiente función resolvemos un único problema
16     asociado a una muestra
17 # W1,...,Wn.
18 def problema(w):
19
20     # Dimensión
21     n = 17
22     N = len(w)
23     # Creamos un modelo
24     m = Model("escenarios");
25
26     # Vector de costes
27     c = [0.55, 0.57, 0.46, 0.02,0.25, 0.34, 0.89, 0.49 ,0.50,
28         0.91,
29         0.52, 0.4, 0.6, 0.2,0.225, 0.344, 0.9]
30
31     # Parámetro de rendimiento
32     v = 18
33
34     # Probabilidad
35     alfa = 0.05
36     beta = 1 - alfa
37
38     # Creamos las variables. El tipo puede ser
39     # 'C' para continuas, 'B' para binarias, 'I' para enteras,
40     # 'S' para semicontinuas, or 'N' for semienteras.
41     x = m.addVars(range(n), vtype='B', name="x");
42     z = m.addVars(range(N), vtype='B', name="z")
```

```
40 # Definimos un producto escalar
41 def productx(v):
42     k = 0
43     for i in range(len(v)):
44         k = k + v[i]*x[i]
45     return(k)
46
47 # Definimos la función objetivo y queremos maximizar/
48 # minimizar.
49 m.setObjective(productx(c), GRB.MINIMIZE);
50
51 # Añadimos las restricciones
52 def productz(v):
53     k = 0
54     for i in range(len(v)):
55         k = k + v[i]*z[i]
56     return(k)
57
58 unos = np.repeat(1,N)
59 m.addConstr(productz(unos) >= N*beta, "c0");
60
61 for i in range(N):
62     m.addConstr(productx(w[i])-v >= -100*(1-z[i]), "cad");
63
64 m.optimize();
65
66 for v in m.getVars():
67     print(v.varName, v.x)
68 return(m.objVal)
69
70 # Hemos exportado las 20 muestras, para cada tamaño muestral,
71 # todas
72 # juntas, por lo que es necesario dividir las en paquetes de
73 # tamaño adecuado.
74 def divide(v,n):
75     return([v[x:x+n] for x in range(0, len(v), n)])
76
77 # Utilizamos este comando para ver la solución más repetida en
78 # cada caso y tomarla como solución asociada al tamaño
79 # muestral.
80 def modaux(L):
81     SL = sorted((x, i) for i, x in enumerate(L))
82     groups = itertools.groupby(SL, key=operator.itemgetter(0))
83     def _auxfun(g):
84         item, iterable = g
```

```
80     count = 0
81     min_index = len(L)
82     for _, where in iterable:
83         count += 1
84         min_index = min(min_index, where)
85     return count, -min_index
86 return max(groups, key=_auxfun)[0]
87
88 # Finalmente, utilizamos este método para calcular la media y
89 # la mediana
90 # truncadas con 3 decimales, así como la solución modal
91 # asociada.
92 def resuelve(w):
93     N = len(w)//15
94     a = divide(w,N)
95     s = []
96     p = []
97     for i in range(len(a)):
98         try:
99             kr= problema(a[i])[0]
100             kp = problema(a[i])[1]
101             s.append(kr)
102             p.append(kp)
103         except GurobiError as e:
104             print('Error code ' + str(e.errno) + ": " + str(e)
105                 )
106         except AttributeError:
107             print('Encountered an attribute error ')
108     mean = np.mean(s)
109     med = np.median(s)
110     k = np.round(mean, 3)
111     l = np.round(med, 3)
112     lp = modaux(p)
113     return ([[k,l],lp])
```

Resultados numéricos

Comencemos por el estudio de la solución óptima. Con la enumeración completa observamos que el problema tiene solución única asociada al vector x^*

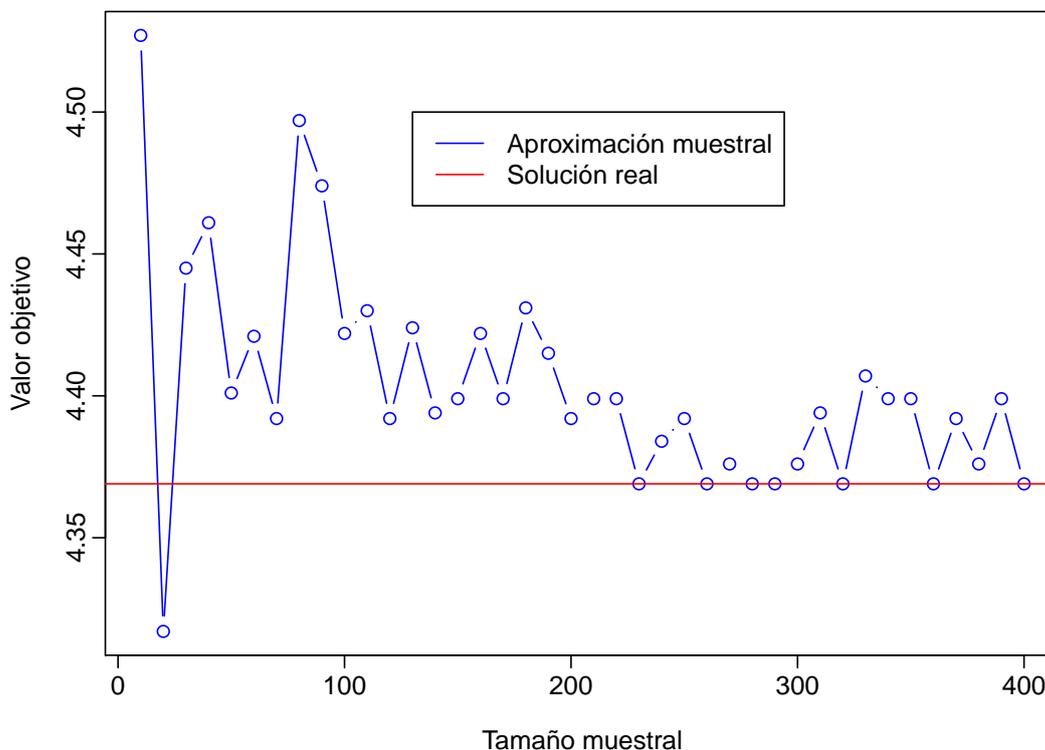
```

1 > x^*
2 [1] 0 1 0 1 1 1 0 0 0 0 1 1 1 1 1 0 1
    
```

Aunque en principio podríamos asignar una interpretación a la solución media de cada uno de los veinte problemas asociados a cada tamaño muestral, vamos a asociar como solución óptima a cada tamaño muestral la solución modal de los problemas asociados. No necesitamos hacer ningún otro tipo de análisis en este caso, pues, obtenemos para todos los tamaños muestrales la misma solución modal, que coincide con x^* .

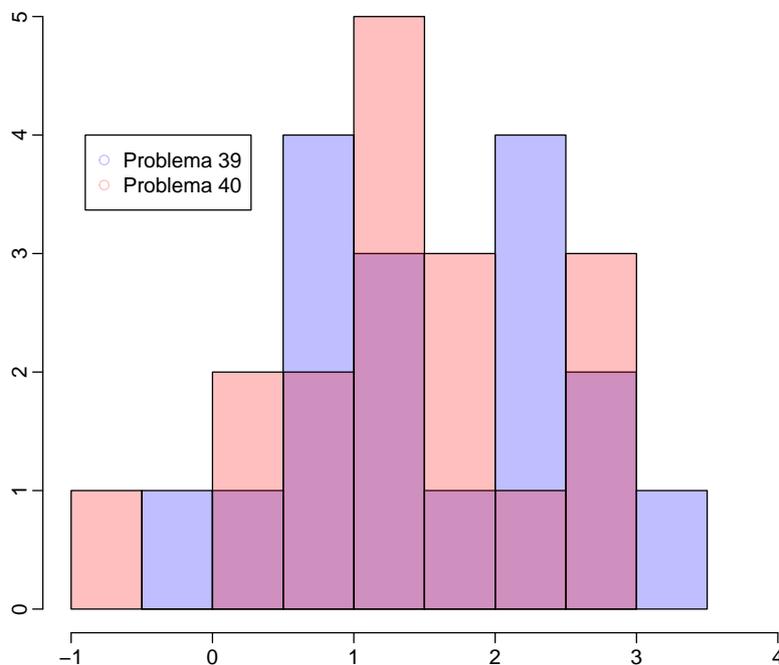
Continuemos con el análisis del valor objetivo. Sabemos, por la resolución por enumeración completa, que el valor objetivo del problema real es $\vartheta^* = 4,369$. En la siguiente gráfica veremos representados los valores asociados a cada tamaño muestral, cada uno de los cuáles ha sido obtenido resolviendo veinte problemas y tomando la media de las soluciones.

Resolución del Problema SAA



Observamos que, como cabía esperar, según aumenta el tamaño muestral nuestra solución media se aproxima a la solución real. Experimentalmente observamos que a medida que el conjunto factible del problema real disminuye (aumentando el valor de v), mayor tamaño muestral es necesario para obtener una buena aproximación.

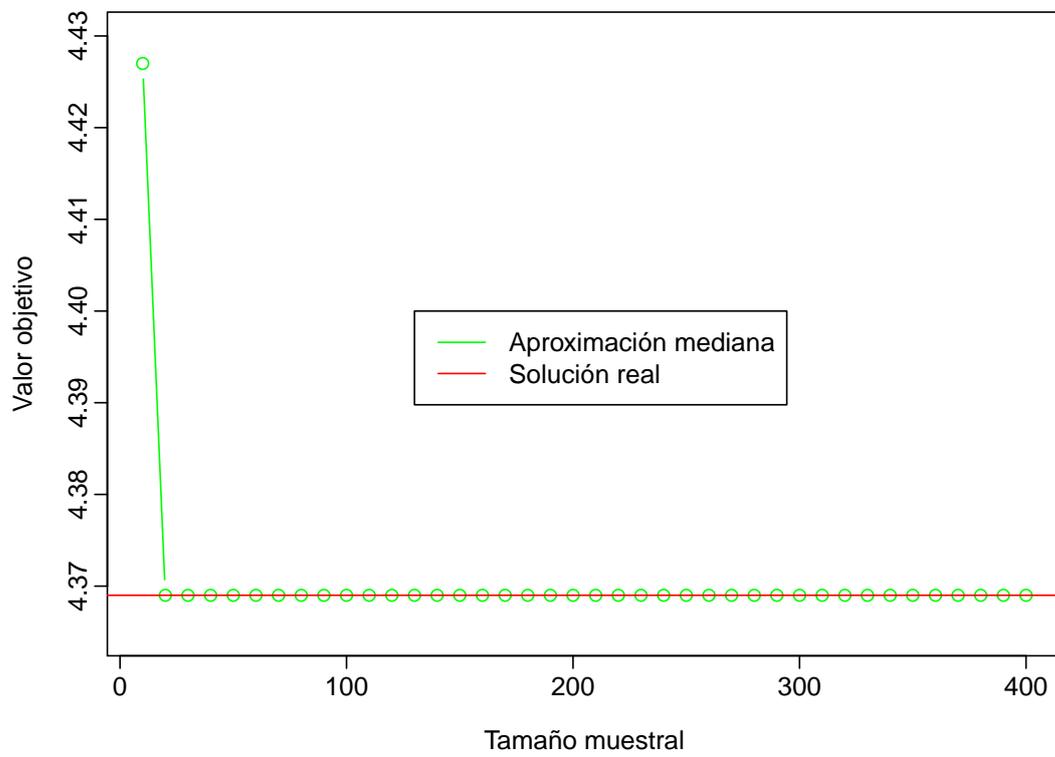
Sin embargo, si la cantidad de problemas que resolvemos no es demasiado grande, una muestra atípica puede afectar a la media. Por ejemplo, estudiemos los Problemas 39 y 40. El valor objetivo del Problema 39 es 4,39, mientras que el del Problema 40 coincide con el valor objetivo del problema real. Si hacemos un análisis de las correlaciones entre las i -ésimas componentes de los datos, observamos que hay algunas donde las mismas son menores que 0,9. Veamos a continuación un histograma comparativo de una de ellas.



Histograma de la componente 14

Notemos que los datos del Problema 39 se encuentran más alejados de los valores centrales que los del Problema 40, lo que puede explicar el valor atípico asociado a estas muestras. En general, dado que el muestreo puede proporcionar de manera aleatoria datos que pudieran ser engañosos, es interesante comprobar además la solución mediana asociada a cada tamaño muestral.

Desde la segunda muestra (tamaño 20) obtenemos el resultado exacto. Gracias al uso de la mediana podemos corregir en parte el problema de tener muestras que se desvíen razonablemente de las respectivas medias.



Resolución del Problema SAA

4.2. Mixtura de normales

Vamos a estudiar un caso donde W se distribuye según una mixtura de normales. En este caso no podemos resolver el problema por enumeración completa, pues tomaremos una n relativamente alta, lo suficiente para que con nuestros medios podamos computar el SAA en un tiempo razonable pero no la enumeración. Comencemos con adjuntando el código con el que hemos generado todos los parámetros involucrados en R.

```

1 #####
2 #                               #
3 #   Scripts para generar muestras #
4 #                               #
5 #####
6
7 # Unos comandos para limpiar el entorno
8 closeAllConnections()
9 rm(list=ls())
10
11 # Número de problemas que vamos a resolver para cada tamaño
    muestral
12 M = 8
13
14 # Dimensión del problema
15 n = 25
16
17 # Construimos el vector de costes
18 c = abs(round(rnorm(n,1,0.5),2))
19
20 # Escojamos los parámetros de la convinación convexa.
21 # Elejimos 2*15 parámetros y el restante será 1 - la
22 # suma de dos de ellos.
23 set.seed(2342)
24
25 # Como tenemos una mixtura de 3 normales, necesitamos 3*n
26 # medias y varianzas y realizar la suma convexa.
27 mediaux = abs(round(rnorm(3*n,1.5,1),2))
28 desviaux = abs(round(rnorm(3*n,0,2),2))
29 medias = matrix(mediaux,nrow=n)
30 desvi = matrix(desviaux, nrow=n)
31
32 # Generemos la probabilidad de tomar cada normal en los
33 # distintos casos
34 lista = 1:50/100
35 paraux = sample(lista,n*2,replace=TRUE)

```

```

36
37 # Generamos la matriz de parámetros de la mixtura
38 comp = c()
39 for (j in 1:n) {
40   probaux = paraux[(2*j-1):(2*j)]
41   probi = append(probaux, 1-sum(probaux))
42   comp = rbind(comp, probi)
43 }
44 row.names(comp) <- NULL
45
46 # Por tanto, la variable i-ésima es una mixtura de 3 normales
47 # con medias: medias[i,] y sd: desvi[i,], dadas por la suma
48 # convexa con componentes comp[i,]. Vamos a construir una
49 # función que para cada i nos dé una muestra de tamaño i
50 muestra <- function(N) {
51   s = c()
52   for (i in 1:n){
53     compo <- sample(1:3, prob=comp[i,], size=N, replace=TRUE)
54     samples <- rnorm(n=N, mean=medias[i,][compo],
55                    sd=desvi[i,][compo])
56     s = cbind(s, samples)   }
57   colnames(s) <- NULL
58   return(s) }
59
60 # Con el siguiente script generamos y exportamos las muestras
61 dir = paste(getwd(), "/muestreoM", sep="")
62 for (i in 1:50) {
63   tam = 8*i*M
64   direc = paste(dir, "/datos", i, ".txt", sep = "")
65   write.table(muestra(tam), direc, row.names=F, sep="\t")
66 }

```

Resultados numéricos

Comencemos por el estudio de la solución óptima. En este caso las soluciones modales no coinciden y se hace necesario un estudio algo más elaborado. Primeramente, la moda de las 50 soluciones modales, que denotaremos x_0 , es

```

1 x0 = [0.0, 1.0, 0.0, 1.0, 0.0, 0.0, 1.0, 1.0, 1.0, 0.0, 1.0,
      0.0, 0.0, 1.0, 1.0, 0.0, 0.0]

```

Otra posibilidad es considerar para cada i , la moda x_i de las últimas $50 - i$ soluciones modales. Para valores no demasiado grandes de i solo tendremos en cuenta las soluciones que provienen de tamaños muestrales grandes. Mostremos algunos valores de x_i

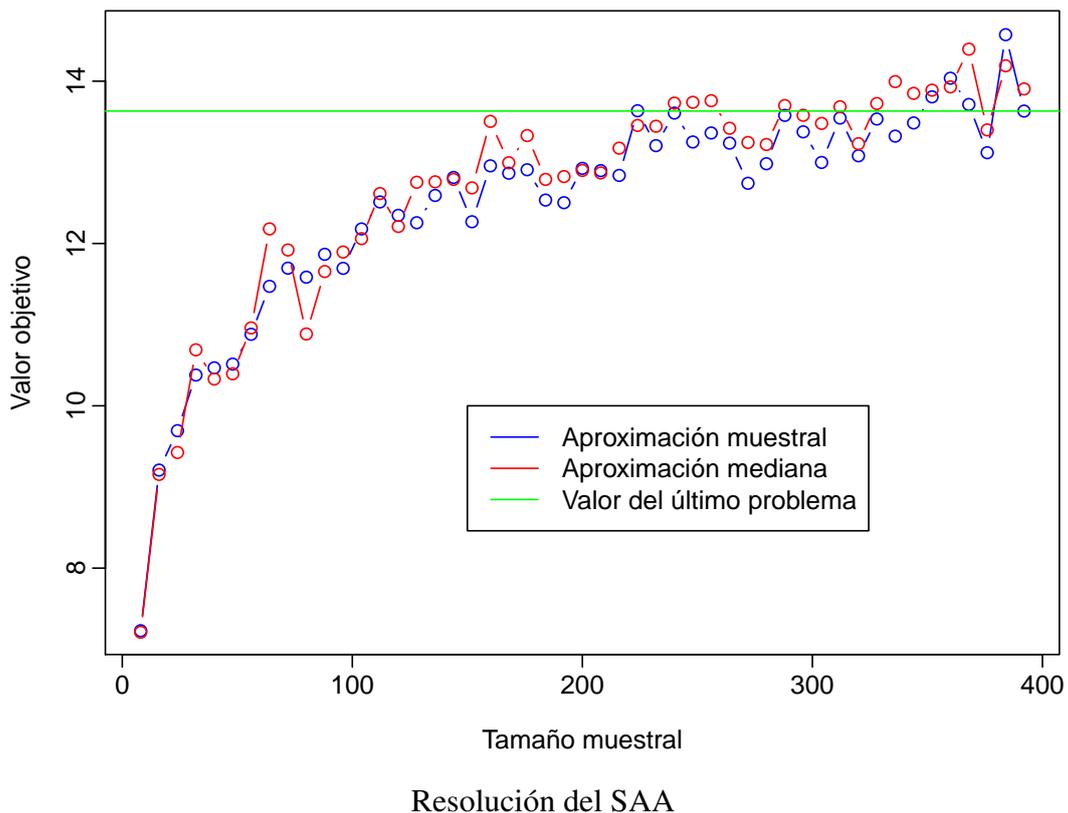
```

1 x5 = [0.0, 1.0, 0.0, 1.0, 1.0, 0.0, 1.0, 1.0, 1.0, 0.0, 1.0,
2       0.0, 0.0, 1.0, 1.0, 0.0, 0.0]
3 x10 = [0.0, 1.0, 0.0, 1.0, 0.0, 0.0, 1.0, 1.0, 1.0, 0.0, 1.0,
4        0.0, 0.0, 1.0, 1.0, 0.0, 0.0]
5 x15 = [0.0, 1.0, 0.0, 1.0, 0.0, 0.0, 1.0, 1.0, 1.0, 0.0, 1.0,
        0.0, 0.0, 1.0, 1.0, 0.0, 0.0]

```

Notemos que $x_0 = x_{10} = x_{15}$ y solo difieren en un valor de x_5 .

Pasemos al análisis del valor objetivo. Pasar de $n = 17$ a $n = 24$ supone un coste computacional importante, por lo que solo resolveremos 8 problemas cada vez. Obtenemos la siguiente gráfica, donde vemos representados la media y la mediana de dichos problemas para cada tamaño muestral considerado. La solución del último problema tiene valor objetivo 13,632.



Bibliografía

- [1] John R. Birge and François Louveaux. *Introduction to Stochastic Programming*. Springer Publishing Company, Incorporated, 2nd edition, 2011.
- [2] Stephen Boyd and Lieven Vandenbergh. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.
- [3] WK Klein Haneveld and MH van der Vlerk. Stochastic programming (lecture notes). *Available on request, email mh van. der. vlerk@ eco. rug. nl*, 2000.
- [4] Harry Markowitz. Portfolio selection*. *The Journal of Finance*, 7(1):77–91, 1952.
- [5] B. K. Pagnoncelli, S. Ahmed, and A. Shapiro. Sample average approximation method for chance constrained programming: Theory and applications. *Journal of Optimization Theory and Applications*, 142(2):399–416, Aug 2009.
- [6] Reuven Y. Rubinstein and Alexander Shapiro. Optimization of static simulation models by the score function method. *Mathematics and Computers in Simulation*, 32(4):373 – 392, 1990.
- [7] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming*. Society for Industrial and Applied Mathematics, 2009.
- [8] Joyce W. Yen and John R. Birge. A stochastic programming approach to the airline crew scheduling problem. *Transportation Science*, 40(1):3–14, 2006.