



TRABAJO FIN DE GRADO

**EL MÉTODO DE APROXIMACIÓN POR
MEDIA MUESTRAL APLICADO A
OPTIMIZACIÓN ESTOCÁSTICA DISCRETA**

Realizado por: Trinidad Fernández Gómez

Supervisado por: Dr. Eduardo Conde Sánchez

FACULTAD DE MATEMÁTICAS

DEPARTAMENTO DE ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

Índice

1. Introducción.	3
2. Resultados de convergencia.	5
2.1. Convergencia de valores objetivos y soluciones.	5
2.2. Tasa de convergencia.	9
2.3. Propiedades asintóticas de la muestra de valores objetivos.	21
3. Diseño del algoritmo.	24
3.1. Selección del tamaño de la muestra.	24
3.2. Repetición.	25
3.3. Estimación del gap para el diseño de un test de parada.	27
3.4. Algoritmo.	30
4. Aplicaciones.	30
4.1. Problema de asignación de recursos.	31
4.2. El comportamiento empírico de los métodos de muestreo para la programación estocástica.	33
4.3. Optimización de estrategias de detección de virus en redes sociales.	35
4.4. Un método SAA para el problema de balance de línea de desmontaje bajo incertidumbre.	37

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 a random sample is generated and the expected objective function of the stochastic problem is approximated by the corresponding sample average function. The obtained sample average approximating problem is then solved 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 expected value and sample average approximation problems are formally described and some interesting properties are given. In section 2 a statistical inference of the sample average approximation method is discussed; convergence rates of objective values and solutions of such problem are established as well as asymptotic properties of sample objective values. In section 3 we outline an algorithm design for the sample average approximation approach to solve the expected value problem and in particular we discuss various stopping rules. In the last section several applications of the method are presented. First a knapsack problem is described where the capacity of the knapsack is given and the size of each item is modeled as a random variable, for this reason the problem is called the *static stochastic knapsack problem*. In the second application, the quality of the approximate solution obtained from sample average approximations for a two-stage stochastic linear program with recourse is studied. In the next application, the problem of detecting a virus that spreads throughout a contact network with a stochastic behavior is examined. Finally, in the last application, a sample average approximation method for the disassembly line balancing problem under uncertainty is described.

1. Introducción.

La programación lineal comenzó sus estudios a finales de los años 50 y casi simultáneamente lo hizo la programación estocástica. Esta última, como su nombre indica, trata problemas de programación matemática en cuya formulación aparece algún elemento estocástico. Por tanto, mientras que en un problema determinístico de programación matemática todos los datos que aparecen en su formulación son números conocidos, en programación estocástica dichos datos (o al menos algunos de ellos) son desconocidos y en este caso se modelan con el uso de vectores aleatorios.

Estamos interesados en la resolución de problemas de optimización estocástica discreta de la forma

$$\min_{x \in S} \{g(x) := \mathbb{E}[G(x, W)]\}. \quad (1.1)$$

W es un vector aleatorio con distribución de probabilidad P , S es un conjunto finito, $G(x, W)$ es el coste aleatorio en el que incurre el sistema al implementar la solución factible $x \in S$ y $G(x, w)$ es una función real definida para cada par de vectores x y w que es una realización de W , y

$$\mathbb{E}[G(x, W)] = \int G(x, w)P(dw).$$

Suponemos que la función valor esperado $g(x)$ está bien definida, i.e., para cada $x \in S$ la función $G(x, \cdot)$ es una función medible Borel y $\mathbb{E}[|G(x, W)|] < \infty$.

Una de las ventajas que tiene la resolución del problema (1.1) es que la solución obtenida será apropiada para procesos de toma de decisiones repetitivas en las que las condiciones estocásticas permanecen estables. Entre los inconvenientes de la formulación (1.1) se encuentra la posibilidad de que no exista una forma cerrada de la función objetivo $g(x) := \mathbb{E}[G(x, W)]$ o pueda no poseer buenas propiedades desde el punto de vista de la optimización.

En este trabajo vamos a resolver un problema de optimización que determina una aproximación a la solución del problema (1.1) cuya calidad dependerá de la información muestral disponible sobre el vector aleatorio W .

Sea W^1, \dots, W^N una muestra aleatoria de N realizaciones del vector aleatorio W independientes e idénticamente distribuidas (*i.i.d.*). Consideremos la función media muestral

$$\hat{g}_N(x) := \frac{1}{N} \sum_{j=1}^N G(x, W^j)$$

y el problema asociado

$$\min_{x \in S} \hat{g}_N(x). \quad (1.2)$$

Una ventaja del problema (1.2) es que existe una forma cerrada de la función objetivo $\hat{g}_N(x) := \frac{1}{N} \sum_{j=1}^N G(x, W^j)$, que permite dar una solución de compromiso. Su principal inconveniente es que hay que estimar el tamaño muestral adecuado en base al comportamiento aleatorio de la solución respecto a la muestra.

Nos referimos al problema (1.1) como el problema de minimización del coste esperado y al problema (1.2) como el problema SAA (del inglés, *sample average approximation*) ó problema de minimización del coste promedio muestral. Obsérvese que $\mathbb{E}[\hat{g}_N(x)] = g(x)$.

Un ejemplo de aplicación del método SAA es el problema del camino más corto, donde los tiempos de viaje son aleatorios y hay penalización cuando se llega tarde a un destino. En este caso, el coste $C(x)$ de un camino x viene dado por

$$C(x) = \sum_{(i,j) \in x} b_{ij} + c \left[\sum_{(i,j) \in x} t_{ij} - q \right]^+,$$

donde b_{ij} es el coste de desplazamiento del arco (i, j) , t_{ij} es el tiempo de desplazamiento del arco (i, j) , q es el tiempo disponible para llegar al destino, y c es la

penalización unitaria por retraso. Entonces, el problema de optimización es

$$\min_{x \in \mathcal{X}} \mathbb{E} [C(x)],$$

donde \mathcal{X} denota el conjunto factible de caminos en la red desde el origen hasta el destino especificados.

2. Resultados de convergencia.

Dado que el conjunto factible S es finito, los problemas (1.1) y (1.2) tienen conjuntos no vacíos de soluciones óptimas, denotados por S^* y \hat{S}_N , respectivamente. Denotemos por v^* y \hat{v}_N a los valores óptimos,

$$v^* := \min_{x \in S} g(x)$$

y

$$\hat{v}_N := \min_{x \in S} \hat{g}_N(x)$$

de los respectivos problemas. También consideraremos conjuntos de soluciones ε -óptimas. Esto es, para cada $\varepsilon \geq 0$ decimos que \bar{x} es una solución ε -óptima de (1.1) si $\bar{x} \in S$ y $g(\bar{x}) \leq v^* + \varepsilon$. Los conjuntos de todas las soluciones ε -óptimas de (1.1) y (1.2) se denotan por S^ε y \hat{S}_N^ε , respectivamente. Claramente, cuando $\varepsilon = 0$ el conjunto S^ε coincide con S^* , y \hat{S}_N^ε coincide con \hat{S}_N .

2.1. Convergencia de valores objetivos y soluciones.

En primer lugar, dado su uso a lo largo del documento ,y así, la importancia de su conocimiento, enunciaremos el concepto de convergencia casi segura.

Definición 2.1.1. (*Rohatgi and Saleh [1]*) Sea $\{X_N\}$ una sucesión de variables

aleatorias definidas sobre el mismo espacio de probabilidad (Ω, A, P) . Se dice que X_N converge casi seguro a X ($X_N \xrightarrow{c.s.} X$) si

$$P(\omega \in \Omega : X_N(\omega) \longrightarrow X(\omega)) = 1.$$

El siguiente resultado se conoce como **Ley Fuerte de los Grandes Números**.

Teorema 2.1.1. (*Rohatgi and Saleh [1]*) Sea $\{X_N\}_{N \geq 1}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con momentos de primer orden finito, entonces

$$\frac{S_N}{N} \xrightarrow{c.s.} \mu$$

con $\mu = \mathbb{E}X_1$ y $S_N = \sum_{i=1}^N X_i$.

Proposición 2.1.1. Sea $\delta_N := \max_{x \in S} |\hat{g}_N(x) - g(x)|$. Se cumple que

$$|\hat{v}_N - v^*| \leq \delta_N.$$

Demostración. Tenemos que demostrar que:

$$|\min_{x \in S} \hat{g}_N(x) - \min_{x \in S} g(x)| \leq \max_{x \in S} |\hat{g}_N(x) - g(x)|.$$

Se cumple,

$$\begin{aligned} \min_{x \in S} \hat{g}_N(x) &= \min_{x \in S} [(\hat{g}_N(x) - g(x)) + g(x)] \geq \\ &\geq \min_{x \in S} (\hat{g}_N(x) - g(x)) + \min_{x \in S} g(x) \geq -\max_{x \in S} |\hat{g}_N(x) - g(x)| + \min_{x \in S} g(x). \end{aligned}$$

La última desigualdad se tiene ya que $-\max_{x \in S} |\hat{g}_N(x) - g(x)| = \min_{x \in S} (-|\hat{g}_N(x) - g(x)|) \leq \min_{x \in S} (\hat{g}_N(x) - g(x))$.

Por tanto:

$$-\max_{x \in S} |\hat{g}_N(x) - g(x)| \leq \min_{x \in S} \hat{g}_N(x) - \min_{x \in S} g(x).$$

Por otro lado,

$$\begin{aligned} \min_{x \in S} g(x) &= \min_{x \in S} [(g(x) - \hat{g}_N(x)) + \hat{g}_N(x)] \geq \\ &\geq \min_{x \in S} (g(x) - \hat{g}_N(x)) + \min_{x \in S} \hat{g}_N(x) \geq -\max_{x \in S} |\hat{g}_N(x) - g(x)| + \min_{x \in S} \hat{g}_N(x). \end{aligned}$$

Por tanto:

$$-\max_{x \in S} |\hat{g}_N(x) - g(x)| \leq \min_{x \in S} g(x) - \min_{x \in S} \hat{g}_N(x),$$

es decir,

$$\min_{x \in S} \hat{g}_N(x) - \min_{x \in S} g(x) \leq \max_{x \in S} |\hat{g}_N(x) - g(x)|.$$

Luego, se tiene la desigualdad deseada.

□

La siguiente proposición establece convergencia casi segura de los estimadores estadísticos en cuestión.

Proposición 2.1.2. Se cumplen las 2 propiedades siguientes:

(i) $\hat{v}_N \xrightarrow{c.s.} v^*$ cuando $N \rightarrow \infty$.

(ii) Para cualquier $\varepsilon \geq 0$, el suceso $\{\hat{S}_N^\varepsilon \subset S^\varepsilon\}$ ocurre casi seguro (*c.s.*) para N suficientemente grande.

Demostración.

(i) Fijado $x \in S$, $G(x, W^j)$, $j = 1, \dots, N$, son *i.i.d.* a $G(x, W)$. Por la **Ley Fuerte de los Grandes Números** se tiene que $\frac{1}{N} \sum_{j=1}^N G(x, W^j) \xrightarrow{c.s.} \mathbb{E}[G(x, W)]$, es decir, $\hat{g}_N(x) \xrightarrow{c.s.} g(x)$ para cualquier $x \in S$. Entonces, por definición, $\forall x \in S$, existe un conjunto $A(x)$ de medida nula tal que $\hat{g}_N(x) \rightarrow g(x)$ salvo a lo sumo en el conjunto $A = \bigcup_{x \in S} A(x)$ que es de medida nula ya que $P(A) \leq \sum_{x \in S} P(A(x))$ por ser S un conjunto finito. Entonces,

$$\delta_N := \max_{x \in S} |\hat{g}_N(x) - g(x)| \xrightarrow{c.s.} 0 \text{ cuando } N \rightarrow \infty \quad . \quad (2.1)$$

Dado que $|\hat{v}_N - v^*| \leq \delta_N$, entonces, se tiene que $\hat{v}_N \xrightarrow{c.s.} v^*$ cuando $N \rightarrow \infty$.

(ii) Dado $\varepsilon \geq 0$, consideremos el número

$$\rho(\varepsilon) := \min_{x \in S \setminus S^\varepsilon} g(x) - v^* - \varepsilon. \quad (2.2)$$

Se verifican:

(1) Dado que para cualquier $x \in S \setminus S^\varepsilon$ se tiene que $g(x) > v^* + \varepsilon$ y el conjunto S es finito, entonces $\rho(\varepsilon) > 0$.

(2) $\forall x \in S \setminus S^\varepsilon$, $g(x) \geq v^* + \varepsilon + \rho(\varepsilon)$.

Sea N suficientemente grande de forma que $\delta_N < \rho(\varepsilon)/2$, es decir, $\max_{x \in S} |\hat{g}_N(x) - g(x)| < \rho(\varepsilon)/2$. Por tanto, $-\rho(\varepsilon)/2 < \hat{g}_N(x) - g(x) < \rho(\varepsilon)/2$, $\forall x \in S$. Y así, $\hat{g}_N(x) > g(x) - \rho(\varepsilon)/2$, $\forall x \in S$.

En particular, si $x \in S \setminus S^\varepsilon$, usando (2) se tiene

$$g(x) - \rho(\varepsilon)/2 \geq v^* + \varepsilon + \rho(\varepsilon) - \rho(\varepsilon)/2 = v^* + \varepsilon + \rho(\varepsilon)/2.$$

Luego, si $x \in S \setminus S^\varepsilon$, se tiene

$$\hat{g}_N(x) > v^* + \varepsilon + \rho(\varepsilon)/2. \quad (2.3)$$

Como se verifica $|\hat{v}_N - v^*| \leq \delta_N < \rho(\varepsilon)/2$, se tiene que $-\rho(\varepsilon)/2 < \hat{v}_N - v^* < \rho(\varepsilon)/2$, y por tanto $\hat{v}_N < v^* + \rho(\varepsilon)/2$.

Luego, si $x \in S \setminus S^\varepsilon$, con la desigualdad (2.3) se llega a $\hat{g}_N(x) > \hat{v}_N + \varepsilon$ y, por tanto, $x \in S \setminus \hat{S}_N^\varepsilon$. Así se tiene la inclusión $\hat{S}_N^\varepsilon \subset S^\varepsilon$.

□

Tengamos en cuenta que si δ es tal que $0 \leq \delta \leq \varepsilon$, entonces $S^\delta \subset S^\varepsilon$ y $\hat{S}_N^\delta \subset \hat{S}_N^\varepsilon$. En consecuencia, usando la proposición anterior, se tiene que para cualquier $\delta \in [0, \varepsilon]$

el suceso $\{\hat{S}_N^\delta \subset S^\varepsilon\}$ ocurre *c.s.* para N suficientemente grande. También se deduce que si $S^\varepsilon = \{x^*\}$ es unitario, entonces $\hat{S}_N^\varepsilon = \{x^*\}$ *c.s.* para N suficientemente grande. En particular, si el problema (1.1) tiene una única solución óptima x^* , entonces *c.s.* para N suficientemente grande el problema (1.2) tiene una única solución óptima \hat{x}_N que coincide con x^* . También consideremos el conjunto $A := \{g(x) - v^* : x \in S\}$. A es un subconjunto de \mathbb{R}^+ y $|A| \leq |S|$, y por tanto A es finito. Se sigue a partir del análisis anterior que para cualquier $\varepsilon \in \mathbb{R}^+ \setminus A$ el suceso $\{\hat{S}_N^\varepsilon = S^\varepsilon\}$ ocurre casi seguro para N suficientemente grande.

2.2. Tasa de convergencia.

Los resultados anteriores no dicen nada acerca de la tasa de convergencia de \hat{v}_N y de \hat{S}_N^δ a sus contrapartes del problema original. Haciendo uso de la Teoría de las Grandes Desviaciones (LD, del inglés, *large deviations*), se demuestra que, bajo condiciones débiles de regularidad y $\delta \in [0, \varepsilon]$, la probabilidad del suceso $\{\hat{S}_N^\delta \subset S^\varepsilon\}$ se aproxima a 1 exponencialmente rápido cuando $N \rightarrow \infty$.

Consideremos una variable aleatoria X que toma valores reales, con media $\mu := \mathbb{E}[X]$. Suponemos que su función generatriz de momentos cumple que $M(t) > 0$ para todo $t \in \mathbb{R}$, $M(0) = 1$, y el dominio $\{t : M(t) < +\infty\}$ es un intervalo que contiene al cero en su interior. La función conjugada

$$I(z) := \sup_{t \in \mathbb{R}} \{tz - \Lambda(t)\} \quad (2.4)$$

de la función generatriz de momentos logarítmica $\Lambda(t) := \log M(t)$, se denomina función tasa LD de X .

Proposición 2.2.1. La función generatriz de momentos $M(t)$ es *logconvex*, es decir, $\Lambda(t) := \log M(t) = \log \mathbb{E}[e^{tX}]$ es convexa.

Demostración. Sea $\lambda \in [0, 1]$ cualquiera.

$$\Lambda(\lambda t_1 + (1 - \lambda)t_2) = \log \mathbb{E} [e^{(\lambda t_1 + (1 - \lambda)t_2)X}] = \log \mathbb{E} [e^{\lambda t_1 X} e^{(1 - \lambda)t_2 X}].$$

Haciendo uso de la desigualdad de Hölder se tiene

$$\log E [e^{\lambda t_1 X} e^{(1 - \lambda)t_2 X}] \leq \log [E(e^{t_1 X})^\lambda E(e^{t_2 X})^{(1 - \lambda)}] = \lambda \log E [e^{t_1 X}] + (1 - \lambda) \log E [e^{t_2 X}].$$

Por tanto,

$$\Lambda(\lambda t_1 + (1 - \lambda)t_2) \leq \lambda \Lambda(t_1) + (1 - \lambda)\Lambda(t_2).$$

□

Sean X_1, \dots, X_N *i.i.d.* a X y sea $\bar{X}_N := N^{-1} \sum_{i=1}^N X_i$ su correspondiente media muestral. Entonces, para cualesquiera números reales a y $t \geq 0$ se cumple que $P(\bar{X}_N \geq a) = P(e^{t\bar{X}_N} \geq e^{ta})$, y entonces, haciendo uso de la desigualdad de Markov se tiene

$$P(e^{t\bar{X}_N} \geq e^{ta}) \leq e^{-ta} E [e^{t\bar{X}_N}] = e^{-ta} [M(t/N)]^N$$

y, por tanto,

$$P(\bar{X}_N \geq a) \leq e^{-ta} [M(t/N)]^N.$$

Haciendo el cambio de variable $t' := t/N$,

$$P(\bar{X}_N \geq a) \leq e^{-t'Na} [M(t')]^N,$$

tomando logaritmo en ambos lados de la desigualdad, se tiene

$$\log P(\bar{X}_N \geq a) \leq -t'Na + N \log M(t'),$$

y tomando ínfimo cuando $t' \geq 0$, para $a \geq \mu$ se tiene que

$$\frac{1}{N} \log [P(\bar{X}_N \geq a)] \leq -I(a). \quad (2.5)$$

Sabemos que $\Lambda(t)$ es convexa, y si $M(t)$ existe en un entorno del cero es de clase C^∞ (**Teorema de Convergencia Dominada**), además, como $\Lambda'(t) = \frac{M'(t)}{M(t)}$, $\Lambda'(0) = \mu$, por lo que $\frac{d}{dt} [tz - \log M(t)]|_{t=0} = z - \mu \geq 0, \forall z \geq \mu$. Como $tz - \log M(t)$ es cóncava (pues $\Lambda(t)$ es convexa), su derivada es decreciente y en el cero es mayor o igual que cero, luego el máximo o el supremo se alcanza en $t \geq 0$. Es decir, si $z \geq \mu$ se tiene:

$$I(z) := \sup_{t \in \mathbb{R}} \{tz - \Lambda(t)\} = \sup_{t \in \mathbb{R}^+} \{tz - \Lambda(t)\}.$$

Luego, para $a \geq \mu$ es suficiente tomar el supremo en la definición (2.4) de $I(a)$ para $t \geq 0$, por tanto esta restricción se omite. La desigualdad (2.5) corresponde con el límite superior del Teorema LD de Cramér.

La constante $I(a)$ en (2.5) da la mejor tasa exponencial con la cual la probabilidad $P(\bar{X}_N \geq a)$ converge a cero cuando $a > \mu$. Esto se desprende la cota inferior

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \log [P(\bar{X}_N \geq a)] \geq -I(a) \quad (2.6)$$

del Teorema LD de Cramér. Una condición suficiente para que se tenga (2.6) es que la función generatriz de momentos $M(t)$ tome un valor finito para todo $t \in \mathbb{R}$.

Proposición 2.2.2. La función $I(z)$ tiene las siguientes propiedades:

- (i) Es una función convexa.
- (ii) Alcanza su mínimo en $z = \mu$ y $I(\mu) = 0$.
- (iii) El desarrollo de Taylor de orden 2 de $I(z)$ es:

$$I(z) = \frac{(z - \mu)^2}{2\sigma^2} + o(|z - \mu|^2).$$

Demostración. Supongamos que la función generatriz de momentos $M(t)$ toma valores finitos para todo t en un entorno de $t = 0$. Entonces, X tiene momentos finitos, y por el **Teorema de Convergencia Dominada** se tiene que $M(t)$, y entonces la función $\Lambda(t)$, es infinitamente derivable en $t = 0$.

(i) Sea $\lambda \in [0, 1]$ cualquiera.

$$\begin{aligned} I(\lambda z_1 + (1 - \lambda) z_2) &= \sup_{t \in \mathbb{R}} \{\lambda t z_1 + (1 - \lambda) t z_2 - \lambda \Lambda(t) - (1 - \lambda) \Lambda(t)\} \leq \\ &\leq \lambda \sup_{t \in \mathbb{R}} \{t z_1 - \Lambda(t)\} + (1 - \lambda) \sup_{t \in \mathbb{R}} \{t z_2 - \Lambda(t)\} = \lambda I(z_1) + (1 - \lambda) I(z_2). \end{aligned}$$

(ii) En primer lugar vamos a demostrar que

$$I(\mu) = \sup_{t \in \mathbb{R}} \{t\mu - \log M(t)\} = 0.$$

Llamemos $\varphi(t) = t\mu - \log M(t)$. Dado que $\varphi(t)$ es cóncava y $\varphi'(t) = \mu - \frac{M'(t)}{M(t)}$, y por tanto, $\varphi'(0) = \mu - \frac{M'(0)}{M(0)} = \mu - \frac{\mu}{1} = 0$, se tiene que el supremo se alcanza para $t = 0$. Además, como $\varphi(0) = 0$, $I(\mu) = 0$.

Veamos ahora que $I(z)$ alcanza su mínimo en $z = \mu$. Llamemos $\phi(t) = tz - \log M(t)$, se tiene que $\phi(0) = 0$ y $\phi'(t) = z - \frac{M'(t)}{M(t)}$, y por tanto, $\phi'(0) = z - \frac{M'(0)}{M(0)} = z - \mu$.

Si $z > \mu$, entonces $\phi'(0) > 0$ y además tenemos que $\phi(0) = 0$, por tanto, existe $\varepsilon_1 > 0$ tal que $\phi(t) > 0, \forall t \in (0, \varepsilon_1)$. Luego, si $z > \mu$, $I(z) = \sup_{t \in \mathbb{R}} \{\phi(t)\} > 0$.

Análogamente, si $z < \mu$, entonces $\phi'(0) < 0$ y además tenemos que $\phi(0) = 0$, por tanto, existe $\varepsilon_2 > 0$ tal que $\phi(t) > 0, \forall t \in (-\varepsilon_2, 0)$. Luego, si $z < \mu$, $I(z) = \sup_{t \in \mathbb{R}} \{\phi(t)\} > 0$.

En definitiva, $I(z) \geq 0 = I(\mu), \forall z \in \mathbb{R}$.

(iii) El desarrollo de Taylor de orden 2 de la función $I(z)$ en μ es:

$$I(z) = \frac{1}{0!} I(\mu) + \frac{1}{1!} I'(\mu)(z - \mu) + \frac{1}{2!} I''(\mu)(z - \mu)^2 + o(|z - \mu|^2).$$

Ya sabemos por el apartado (ii) que $I(\mu) = 0$.

Dado que $\phi(t) = tz - \log M(t)$ es cóncava y derivable, alcanza su máximo en $t(z)$ tal que $\phi'(t(z)) = 0$, es decir, $t(z)$ es tal que cumple:

$$z - \frac{M'(t(z))}{M(t(z))} = 0. \quad (2.7)$$

Por tanto, $I(z) = zt(z) - \log M(t(z))$.

Si asumimos que existe una solución $t(z)$ de la ecuación funcional (2.7) que es derivable, $I(z)$ también es derivable y

$$I'(z) = t(z) + zt'(z) - \frac{M'(t(z))}{M(t(z))}t'(z) = t(z) + \left(z - \frac{M'(t(z))}{M(t(z))} \right) t'(z),$$

$$I''(z) = t'(z),$$

luego, usando (2.7), se tiene que $I'(z) = t(z)$.

Por otro lado, derivando en ambos términos de (2.7):

$$1 - \frac{M''(t(z))M(t(z))t'(z) - (M'(t(z)))^2t'(z)}{(M(t(z)))^2} = 0 \quad (2.8)$$

Sustituyendo $z = \mu$ en (2.8) y sabiendo por el apartado (ii) que $t(\mu) = 0$, se tiene:

$$1 - \frac{M''(0)M(0) - (M'(0))^2}{(M(0))^2}t'(\mu) = 0.$$

Como $M(0) = 1$, $M'(0) = \mathbb{E}[X]$ y $M''(0) = \mathbb{E}[X^2]$, entonces:

$$1 - (E[X^2] - E[X]^2) t'(\mu) = 0,$$

es decir, $t'(\mu) = \frac{1}{\sigma^2}$. Con esto se justifica que el desarrollo de Taylor de orden 2 de

$I(z)$ es:

$$I(z) = \frac{(z - \mu)^2}{2\sigma^2} + o(|z - \mu|^2).$$

□

Para $a > \mu$ la derivada de $\psi(t) := ta - \Lambda(t)$ en $t = 0$ es mayor que cero, y por tanto $\psi(t) > 0$ para $t > 0$ suficientemente pequeño. En este caso, se tiene que $I(a) > 0$. Por el desarrollo de Taylor:

$$I(a) = \frac{(a - \mu)^2}{2\sigma^2} + o(|a - \mu|^2). \quad (2.9)$$

Como consecuencia, para a cercano a μ se puede aproximar $I(a)$ por $\frac{(a-\mu)^2}{2\sigma^2}$. Además, para cualquier $\tilde{\epsilon} > 0$ hay un entorno \mathcal{N} de μ tal que

$$I(a) \geq \frac{(a - \mu)^2}{(2 + \tilde{\epsilon})\sigma^2}, \quad \forall a \in \mathcal{N}. \quad (2.10)$$

En particular, se puede tomar $\tilde{\epsilon} = 1$.

Ahora volvemos a los problemas (1.1) y (1.2). Consideremos los números $\varepsilon \geq 0$, $\delta \in [0, \varepsilon]$, y el suceso $\{\hat{S}_N^\delta \subset S^\varepsilon\}$. Se cumple que

$$\{\hat{S}_N^\delta \not\subset S^\varepsilon\} = \bigcup_{x \in S \setminus S^\varepsilon} \bigcap_{y \in S} \{\hat{g}_N(x) \leq \hat{g}_N(y) + \delta\} \quad (2.11)$$

y por tanto

$$P(\hat{S}_N^\delta \not\subset S^\varepsilon) \leq \sum_{x \in S \setminus S^\varepsilon} P\left(\bigcap_{y \in S} \{\hat{g}_N(x) \leq \hat{g}_N(y) + \delta\}\right). \quad (2.12)$$

Consideremos una función arbitraria $u : S \setminus S^\varepsilon \mapsto S$. De (2.12) se tiene que

$$P(\hat{S}_N^\delta \not\subset S^\varepsilon) \leq \sum_{x \in S \setminus S^\varepsilon} P(\hat{g}_N(x) - \hat{g}_N(u(x)) \leq \delta). \quad (2.13)$$

Supongamos que la función $u(x)$ es elegida de forma que para algún $\varepsilon^* > \varepsilon$

$$g(u(x)) \leq g(x) - \varepsilon^*, \quad \forall x \in S \setminus S^\varepsilon. \quad (2.14)$$

Tengamos en cuenta que si $u(x) \in S^*$ para todo $x \in S \setminus S^\varepsilon$, entonces (2.14) se cumple con

$$\varepsilon^* := \min_{x \in S \setminus S^\varepsilon} g(x) - v^* \quad (2.15)$$

y $\varepsilon^* > \varepsilon$ ya que el conjunto S es finito. Por tanto, existe siempre una función $u(\cdot)$ que satisfaga (2.14).

Para cada $x \in S \setminus S^\varepsilon$, sea

$$H(x, w) := G(u(x), w) - G(x, w).$$

Hay que tener en cuenta que $\mathbb{E}[H(x, W)] = g(u(x)) - g(x)$, y entonces $\mathbb{E}[H(x, W)] \leq -\varepsilon^*$. Sea W^1, \dots, W^N una muestra aleatoria *i.i.d.* a W , y consideremos la función media muestral

$$\hat{h}_N(x) := \frac{1}{N} \sum_{j=1}^N H(x, W^j) = \hat{g}_N(u(x)) - \hat{g}_N(x).$$

De (2.13) se tiene que

$$P\left(\hat{S}_N^\delta \not\subset S^\varepsilon\right) \leq \sum_{x \in S \setminus S^\varepsilon} P\left(\hat{h}_N(x) \geq -\delta\right). \quad (2.16)$$

Denotemos por $I_x(\cdot)$ a la función tasa LD de $H(x, W)$. La desigualdad (2.16) junto con (2.5) implica que

$$P\left(\hat{S}_N^\delta \not\subset S^\varepsilon\right) \leq \sum_{x \in S \setminus S^\varepsilon} e^{-NI_x(-\delta)}. \quad (2.17)$$

Es importante señalar que la desigualdad anterior (2.17) no es asintótica y es válida para cualquier muestra aleatoria de tamaño N .

Hipótesis (H). Para cada $x \in S$ la función generatriz de momentos de la variable aleatoria $H(x, W)$ es finita en un entorno que contiene al origen en su interior.

La hipótesis (H) se tiene, por ejemplo, si $H(x, W)$ es una variable aleatoria acotada, o si $H(x, \cdot)$ crece a lo sumo linealmente y W tiene una distribución de la familia exponencial.

Proposición 2.2.3 Sean ε y δ números no negativos tales que $\delta \leq \varepsilon$. Entonces

$$P(\hat{S}_N^\delta \not\subset S^\varepsilon) \leq |S \setminus S^\varepsilon| e^{-N\Upsilon(\delta, \varepsilon)}, \quad (2.18)$$

donde

$$\Upsilon(\delta, \varepsilon) := \min_{x \in S \setminus S^\varepsilon} I_x(-\delta). \quad (2.19)$$

Además, si se cumple la hipótesis (H), entonces $\Upsilon(\delta, \varepsilon) > 0$.

Demostración. La desigualdad (2.18) es consecuencia inmediata de la desigualdad (2.17). Se tiene que $-\delta > -\varepsilon^* \geq \mathbb{E}[H(x, W)]$, y entonces por la hipótesis (H) se tiene que $I_x(-\delta) > 0$ para cada $x \in S \setminus S^\varepsilon$. Esto implica que $\Upsilon(\delta, \varepsilon) > 0$.

□

El siguiente resultado asintótico es una consecuencia inmediata de la desigualdad (2.18),

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \left[1 - P(\hat{S}_N^\delta \subset S^\varepsilon) \right] \leq -\Upsilon(\delta, \varepsilon). \quad (2.20)$$

La desigualdad (2.20) nos dice que la probabilidad del suceso $\{\hat{S}_N^\delta \subset S^\varepsilon\}$ se aproxima a 1 exponencialmente rápido cuando $N \rightarrow \infty$. Esto sugiere que el muestreo de tipo Monte Carlo, combinado con un método eficaz para resolver el problema SAA, puede resolver de forma eficiente los problemas que estudiamos, siempre que la constante $\Upsilon(\delta, \varepsilon)$ no sea demasiado pequeña.

Se deduce a partir de (2.9) que

$$I_x(-\delta) \approx \frac{(-\delta - \mathbb{E}[H(x, W)])^2}{2\sigma_x^2} \geq \frac{(\varepsilon^* - \delta)^2}{2\sigma_x^2}, \quad (2.21)$$

donde ε^* está definido en (2.15) y

$$\sigma_x^2 := \text{Var}[H(x, W)] = \text{Var}[G(u(x), W) - G(x, W)].$$

Por tanto, la constante $\Upsilon(\delta, \varepsilon)$, dada en (2.19), se puede aproximar por

$$\Upsilon(\delta, \varepsilon) \approx \min_{x \in S \setminus S^\varepsilon} \frac{(-\delta - \mathbb{E}[H(x, W)])^2}{2\sigma_x^2} \geq \frac{(\varepsilon^* - \delta)^2}{2\sigma_{max}^2} > \frac{(\varepsilon - \delta)^2}{2\sigma_{max}^2}, \quad (2.22)$$

donde

$$\sigma_{max}^2 := \max_{x \in S \setminus S^\varepsilon} \text{Var}[G(u(x), W) - G(x, W)]. \quad (2.23)$$

Para ilustrar algunas de las implicaciones de la desigualdad (2.18) para la complejidad de resolución de problemas estocásticos, vamos a fijar un nivel de significación $\alpha \in (0, 1)$ y a estimar el tamaño de la muestra N que es necesario para que la probabilidad $P(\hat{S}_N^\delta \subset S^\varepsilon)$ sea al menos $1 - \alpha$. Al exigir que el lado derecho de (2.18) sea menor o igual que α , tenemos que

$$N \geq \frac{1}{\Upsilon(\delta, \varepsilon)} \log \left(\frac{|S \setminus S^\varepsilon|}{\alpha} \right). \quad (2.24)$$

Por otro lado, se deduce de (2.10) y de (2.19) que $\Upsilon(\delta, \varepsilon) \geq (\varepsilon - \delta)^2 / (3\sigma_{max}^2)$ para todo $\varepsilon \geq 0$ suficientemente pequeño. Por tanto, se tiene que para todo $\varepsilon > 0$ suficientemente pequeño y $\delta \in [0, \varepsilon)$, una condición suficiente para que se cumpla (2.24) es que

$$N \geq \frac{3\sigma_{max}^2}{(\varepsilon - \delta)^2} \log \left(\frac{|S|}{\alpha} \right). \quad (2.25)$$

Parece ser que la cota (2.25) pueda llegar a ser demasiado conservativa para estimaciones prácticas de los tamaños muestrales requeridos. Sin embargo, (2.25) tiene consecuencias importantes para problemas complejos. Una característica clave de (2.25) es que N depende logarítmicamente tanto del cardinal del conjunto factible S como de α . Una consecuencia importante de esto es la siguiente. Supongamos que (i) el tamaño del conjunto factible S crece a lo sumo exponencialmente en el tamaño de la entrada del problema, (ii) la varianza σ_{max}^2 crece polinomialmente en el tamaño de la entrada del problema, y (iii) la complejidad de búsqueda de una solución δ -óptima de (1.2) crece polinomialmente en el tamaño de la entrada del problema y el tamaño de la muestra N . Entonces, una solución puede ser generada en tiempo que crece polinomialmente en el tamaño de la entrada del problema de tal manera que, con probabilidad $1 - \alpha$ al menos, la solución es ε -óptima para (1.1).

Ahora, supongamos que el problema (1.1) tiene una única solución óptima x^* , i.e., $S^* = \{x^*\}$ es unitario, y consideremos el caso en el que el problema SAA tiene una única solución óptima \hat{x}_N . Consideremos el suceso aleatorio $\{x^* = \hat{x}_N\}$. Además, consideremos la función $u : S \setminus S^\varepsilon \mapsto \{x^*\}$, i.e., $u(x) \equiv x^*$, y la constante correspondiente $\Upsilon^* := \Upsilon(0, 0)$.

$$\Upsilon^* = \min_{x \in S \setminus \{x^*\}} I_x(0), \quad (2.26)$$

con $I_x(\cdot)$ la función tasa LD de $G(x^*, W) - G(x, W)$. Tengamos en cuenta que $\mathbb{E}[G(x^*, W) - G(x, W)] = g(x^*) - g(x)$, y entonces $\mathbb{E}[G(x^*, W) - G(x, W)] < 0$ para cada $x \in S \setminus \{x^*\}$. Por lo tanto, si se cumple la hipótesis (H), i.e., la función generatriz de momentos de $G(x^*, W) - G(x, W)$ es finita en un entorno abierto del origen, entonces $\Upsilon^* > 0$.

Proposición 2.2.4. Supongamos que el problema (1.1) tiene una única solución óptima x^* y la función generatriz de momentos de cada variable aleatoria $G(x^*, W)$ —

$G(x, W)$, $x \in S \setminus \{x^*\}$, es finita en \mathbb{R} . Entonces,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log [1 - P(\hat{x}_N = x^*)] = -\Upsilon^*. \quad (2.27)$$

Demostración. Usando (2.20) tenemos que:

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log [1 - P(\hat{x}_N = x^*)] \leq -\Upsilon^*. \quad (2.28)$$

Consideremos el complementario del suceso $\{\hat{x}_N = x^*\}$, el cual denotamos por $\{\hat{x}_N \neq x^*\}$ y es igual a la unión de los sucesos $\{\hat{g}_N(x) \leq \hat{g}_N(x^*)\}$, $x \in S \setminus \{x^*\}$.

Por lo tanto, para cualquier $x \in S \setminus \{x^*\}$,

$$P(\hat{x}_N \neq x^*) \geq P(\hat{g}_N(x) \leq \hat{g}_N(x^*)).$$

Usando la cota inferior (2.6) del Teorema LD de Crámer, se sigue que se cumple la desigualdad

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \log [1 - P(\hat{x}_N = x^*)] \geq -I_x(0) \quad (2.29)$$

para cada $x \in S \setminus \{x^*\}$. Las desigualdades (2.28) y (2.29) implican (2.27).

□

Supongamos que $S^* = \{x^*\}$ y consideremos el número

$$\kappa := \max_{x \in S \setminus \{x^*\}} \frac{\text{Var} [G(x, W) - G(x^*, W)]}{[g(x) - g(x^*)]^2}. \quad (2.30)$$

Se tiene por (2.9) y (2.26) que $\kappa \approx 1/(2\Upsilon^*)$. Podemos ver κ como un número de condición para el problema (1.1). Es decir, el tamaño muestral requerido para que el suceso $\{\hat{x}_N = x^*\}$ ocurra con probabilidad dada es aproximadamente proporcional a κ .

Para un problema cuyo conjunto factible S es grande, si $S^* = \{x^*\}$, el número $\min_{x \in S \setminus \{x^*\}} g(x) - g(x^*)$, aunque positivo, tiende a ser pequeño. Por tanto, el tamaño muestral requerido para calcular la solución óptima exacta x^* con probabilidad alta puede ser grande, incluso si la solución óptima x^* es única. Para problemas mal condicionados tiene sentido buscar aproximaciones ε -óptimas del problema (1.1). En este caso, la cota (2.18) es más informativa dado que está garantizado que la constante correspondiente $\Upsilon(\delta, \varepsilon)$ es al menos del orden $(\varepsilon - \delta)^2 / (2\sigma_{max}^2)$.

También es interesante estudiar el comportamiento del número de condición κ para un problema de optimización discreta con función objetivo lineal $G(x, W) := \sum_{i=1}^k W_i x_i$ y conjunto factible $S := \{0, 1\}^k$. En este caso, el correspondiente problema de optimización es

$$\min_{x \in \{0,1\}^k} \left\{ g(x) = \sum_{i=1}^k \bar{w}_i x_i \right\}$$

donde $\bar{w}_i := \mathbb{E}[W_i]$. Supongamos que $\bar{w}_i > 0$ para todo $i \in \{1, \dots, k\}$, y por lo tanto $S^* = \{0\}$. Sea

$$v_i^2 := \frac{Var[W_i]}{(\mathbb{E}[W_i])^2}$$

que denota el coeficiente de variación al cuadrado de W_i y sea

$$\rho_{ij} := \frac{Cov[W_i, W_j]}{\sqrt{Var[W_i]} \sqrt{Var[W_j]}}$$

el coeficiente de correlación entre W_i y W_j . Se tiene que para cualquier $x \in \{0, 1\}^k \setminus \{0\}$,

$$\frac{Var[\sum_{i=1}^k W_i x_i]}{[\sum_{i=1}^k \bar{w}_i x_i]^2} = \frac{\sum_{i=1}^k \sum_{j=1}^k \rho_{ij} v_i \bar{w}_i x_i v_j \bar{w}_j x_j}{\sum_{i=1}^k \sum_{j=1}^k \bar{w}_i x_i \bar{w}_j x_j} \leq \max_{i \in \{1, \dots, k\}} v_i^2.$$

Así,

$$\kappa = \max_{x \in \{0,1\}^k \setminus \{0\}} \frac{Var[\sum_{i=1}^k W_i x_i]}{[\sum_{i=1}^k \bar{w}_i x_i]^2} = \max_{i \in \{1, \dots, k\}} v_i^2.$$

La última igualdad se tiene debido a que el máximo se alcanza en $x_i = 1$, para el

subíndice i , para el cual W_i tiene el máximo coeficiente de variación al cuadrado v_i^2 y $x_j = 0$ para el resto de variables. Así, en este ejemplo, el número de condición κ es igual al máximo del coeficiente de variación al cuadrado de los W_i 's.

2.3. Propiedades asintóticas de la muestra de valores objetivos.

A continuación, enunciaremos varios resultados que serán necesarios para el desarrollo correcto de este apartado.

Definición 2.3.1 (*Rohatgi and Saleh* [1]) Sea $\{X_N\}$ una sucesión de variables aleatorias definidas sobre el mismo espacio de probabilidad (Ω, A, P) . Se dice que X_N converge en probabilidad a X ($X_N \xrightarrow{P} X$) si

$$\lim_{N \rightarrow \infty} P(|X_N - X| > \varepsilon) = 0 \quad \text{para todo } \varepsilon > 0.$$

Teorema 2.3.1. (*Rohatgi and Saleh* [1]) Si $X_N \xrightarrow{c.s.} X$ entonces $X_N \xrightarrow{P} X$.

Definición 2.3.2 (*Rohatgi and Saleh* [1]) Sea $\{X_N\}$ una sucesión de variables aleatorias. Se dice que X_N converge en ley a X ($X_N \xrightarrow{L} X$) si

$$\lim F_N(x) = F(x) \quad \text{para todo } x \in C_F$$

donde C_F denota el conjunto de puntos de continuidad de F .

El resultado expuesto a continuación se conoce como **Teorema Central del Límite**.

Teorema 2.3.2. (*Rohatgi and Saleh* [1]) Sea $\{X_N\}_{N \geq 1}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con $0 < Var(X_N) = \sigma^2 < \infty$

y $E[X_N] = \mu$. Entonces,

$$\sqrt{N} \frac{\bar{X}_N - \mu}{\sigma} \xrightarrow{\mathcal{L}} Z \sim N(0, 1)$$

con $\bar{X}_N = N^{-1} \sum_{i=1}^N X_i$.

En lo que sigue, estudiaremos las propiedades asintóticas del valor objetivo óptimo \hat{v}_N del problema SAA.

Para cualquier subconjunto S' de S se tiene la desigualdad $\hat{v}_N \leq \min_{x \in S'} \hat{g}_N(x)$. En particular, tomando $S' = S^*$, se tiene que $\hat{v}_N \leq \min_{x \in S^*} \hat{g}_N(x)$, y entonces

$$\mathbb{E}[\hat{v}_N] \leq \mathbb{E} \left\{ \min_{x \in S^*} \hat{g}_N(x) \right\} \leq \min_{x \in S^*} \mathbb{E}[\hat{g}_N(x)] = v^*.$$

Es decir, el estimador \hat{v}_N tiene sesgo negativo.

Se sigue de la proposición 2.1.2 que casi seguro, para N suficientemente grande, el conjunto \hat{S}_N de soluciones óptimas del problema SAA está contenido en S^* . En este caso, se tiene que

$$\hat{v}_N = \min_{x \in \hat{S}_N} \hat{g}_N(x) \geq \min_{x \in S^*} \hat{g}_N(x).$$

Dado que la desigualdad contraria se tiene siempre, se sigue que, casi seguro, $\hat{v}_N - \min_{x \in S^*} \hat{g}_N(x) = 0$ para N suficientemente grande. Multiplicando a ambos lados de la desigualdad anterior por \sqrt{N} se tiene que, casi seguro, $\sqrt{N} [\hat{v}_N - \min_{x \in S^*} \hat{g}_N(x)] = 0$ para N suficientemente grande, y entonces

$$\lim_{N \rightarrow \infty} \sqrt{N} \left[\hat{v}_N - \min_{x \in S^*} \hat{g}_N(x) \right] = 0 \quad \text{c.s.} \quad (2.31)$$

Dado que la convergencia casi segura implica convergencia en probabilidad, se tiene

de (2.31) que $\sqrt{N} [\hat{v}_N - \min_{x \in S^*} \hat{g}_N(x)]$ converge en probabilidad a 0, i.e.,

$$\hat{v}_N = \min_{x \in S^*} \hat{g}_N(x) + o_p(N^{-1/2}).$$

Además, dado que $v^* = g(x)$ para cualquier $x \in S^*$, se tiene que

$$\sqrt{N} \left[\min_{x \in S^*} \hat{g}_N(x) - v^* \right] = \sqrt{N} \min_{x \in S^*} [\hat{g}_N(x) - v^*] = \min_{x \in S^*} \left\{ \sqrt{N} [\hat{g}_N(x) - g(x)] \right\}.$$

Supongamos que para cada $x \in S$ la varianza

$$\sigma^2(x) := Var [G(x, W)] \tag{2.32}$$

existe. Entonces, por el **Teorema Central del Límite** se tiene que, para cualquier $x \in S$, $\sqrt{N} [\hat{g}_N(x) - g(x)]$ converge en ley a una variable $Z(x)$ con distribución normal de media cero y varianza $\sigma^2(x)$. Además, usando de nuevo el **Teorema Central del Límite**, la covarianza entre $Z(x)$ y $Z(x')$ es igual a la covarianza entre $G(x, W)$ y $G(x', W)$ para cualesquiera $x, x' \in S$. Por lo tanto, se obtiene el siguiente resultado.

Proposición 2.3.1. Supongamos que la varianza $\sigma^2(x)$ definida en (2.32) existe para cada $x \in S^*$. Entonces,

$$\sqrt{N} (\hat{v}_N - v^*) \xrightarrow{\mathcal{L}} \min_{x \in S^*} Z(x), \tag{2.33}$$

donde $Z(x)$ son variables aleatorias que siguen una distribución normal de media cero y covarianza la función dada por la correspondiente función covarianza de $G(x, W)$.

En particular, si $S^* = \{x^*\}$, entonces

$$\sqrt{N} (\hat{v}_N - v^*) \xrightarrow{\mathcal{L}} N(0, \sigma^2(x^*)). \tag{2.34}$$

Aunque para cualquier x el valor esperado de $Z(x)$ es cero, el valor esperado del mínimo de $Z(x)$ sobre un subconjunto S' de S puede ser negativo y tiende a ser menor cuanto mayor sea S' . Por lo tanto, de (2.33) se deduce que para problemas mal condicionados, donde el conjunto de soluciones óptimas ó casi óptimas es grande, el estimador \hat{v}_N de v^* tiende a tener mayor sesgo. Bajo determinadas condiciones, se sigue a partir de (2.33) que $\sqrt{N} [\mathbb{E}(\hat{v}_N) - v^*] \longrightarrow \mathbb{E} [\text{mín}_{x \in S^*} Z(x)]$.

3. Diseño del algoritmo.

3.1. Selección del tamaño de la muestra.

En un algoritmo, tenemos que elegir un tamaño muestral finito N o una secuencia de tamaños muestrales finitos, y el algoritmo tiene que parar tras una cantidad finita de tiempo. Una cuestión importante es cómo se deben hacer esas elecciones. La cota (2.25) da una estimación del tamaño de la muestra requerido para encontrar una solución ε -óptima con probabilidad al menos $1 - \alpha$. Esta estimación tiene dos inconvenientes para los fines del cálculo. En primer lugar, no es fácil calcular la estimación, porque σ_{max}^2 y en algunos problemas también $|S|$ pueden ser difíciles de calcular. En segundo lugar, para obtener una estimación práctica del tamaño de la muestra requerido, la estimación puede ser demasiado conservativa. Con N mayor, la función objetivo del problema SAA tiende a ser una estimación más precisa de la función objetivo del problema de minimización del coste esperado, por tanto mejora la calidad de una solución óptima del problema SAA, es decir, tiende a ser más ajustada la diferencia entre el valor objetivo de la solución aproximada y el óptimo. Sin embargo, dependiendo del problema SAA (1.2) y el método usado para la resolución del mismo, puede aumentar la complejidad de resolución del problema SAA al menos linealmente, y a menudo exponencialmente, con el tamaño de la muestra N .

3.2. Repetición.

Si aumenta más rápido que linealmente la complejidad de resolución del problema SAA en el tamaño muestral N , puede ser más eficiente elegir un tamaño muestral N más pequeño y generar y resolver varios problemas SAA con muestras independientes e idénticamente distribuidas.

Dado este enfoque, debemos abordar varias cuestiones. Una cuestión es si existe garantía de que se produzca una solución óptima (o ε -óptima) para el problema (1.1) si se resuelve un número suficiente de problemas SAA, con muestras independientes de tamaño N . Ésto se puede ver como ensayos de Bernoulli con probabilidad de éxito $p = p(N)$. El éxito sería que la solución óptima calculada del problema SAA \hat{x}_N es solución óptima del problema (1.1). De la proposición 2.1.2 se sigue que esta probabilidad p converge a 1 cuando $N \rightarrow \infty$ y, además, por la proposición 2.2.3 converge a 1 exponencialmente rápido si se cumple la hipótesis (H). Sin embargo, para N finito la probabilidad p puede ser pequeña o incluso cero. La probabilidad de producir una solución óptima del problema (1.1) al menos una vez en M ensayos es $1 - (1 - p)^M$ y esta probabilidad converge a 1 cuando $M \rightarrow \infty$, siempre que p sea positivo. Por tanto, una cuestión relevante sería si existe garantía de que p es positivo para un tamaño muestral dado N . El siguiente ejemplo muestra que el tamaño muestral requerido para que p sea positivo es un problema en sí mismo y podría ser un número arbitrariamente grande incluso para problemas con un número acotado de soluciones factibles.

Ejemplo. Supongamos que $S := \{-1, 0, 1\}$, que W puede tomar los valores w_1 y w_2 con probabilidades $1 - \gamma$ y γ , respectivamente, y que $G(-1, w_1) := -1$, $G(0, w_1) := 0$, $G(1, w_1) := 2$, $G(-1, w_2) := 2k$, $G(0, w_2) := 0$, $G(1, w_2) := -k$, donde k es un número arbitrario positivo. Sea $\gamma = 1/(k + 1)$. Entonces, $g(x) = (1 - \gamma)G(x, w_1) + \gamma G(x, w_2)$, y así $g(-1) = k/(k + 1)$, $g(0) = 0$, y $g(1) = k/(k + 1)$. Por tanto, $x^* = 0$ es la única solución óptima del problema (1.1). Si la muestra no

contiene ninguna observación w_2 , entonces $\hat{x}_N = -1 \neq x^*$. Supongamos que la muestra contiene al menos una observación w_2 . Entonces, $\hat{g}_N(1) \leq [2(N-1) - k]/N$. Así, $\hat{g}_N(1) < 0 = \hat{g}_N(0)$ si $N \leq k/2$, y $\hat{x}_N = 1 \neq x^*$. Por tanto, se necesita una muestra de tamaño $N > k/2$ de modo que sea estrictamente positiva la probabilidad p de que la solución óptima del problema SAA coincida con la del problema de minimizar el coste esperado, $x^* = 0$.

Otra cuestión que debemos abordar es la elección de M , el número de muestras que serán consideradas para resolver otros tantos problemas aproximados del tipo SAA. De forma similar a la elección del tamaño N de la muestra, el número de repeticiones M puede ser elegido de forma dinámica. A continuación se discute un enfoque para hacer esto. Supongamos que cada repetición del problema SAA produce una solución candidata, la cual puede ser una solución óptima (ε -óptima) del problema SAA. Denotemos por \hat{x}_N^m a la solución candidata producida por la m -ésima repetición del problema SAA. El gap de optimalidad $g(\hat{x}_N^m) - v^*$ puede ser estimado, como se describirá en la sección siguiente. Si se satisface un criterio de parada basado en el gap de optimalidad estimado, entonces no se llevan a cabo más repeticiones. En otro caso, se llevan a cabo las repeticiones adicionales con el mismo tamaño muestral N , o se aumenta el tamaño de la muestra. El siguiente argumento proporciona una guía sencilla en cuanto a si es probable que una repetición adicional con el mismo tamaño muestral N dé una solución mejor que la mejor solución encontrada hasta el momento.

Hay que tener en cuenta que, por construcción, las variables aleatorias $g(\hat{x}_N^m)$, $m = 1, \dots$, son *i.i.d.*, y su distribución de probabilidad tiene un soporte finito ya que el conjunto S es finito. Supongamos que se han realizado hasta el momento M repeticiones de tamaño muestral N . Si la distribución de probabilidad de $g(\hat{x}_N)$ fuese continua, entonces la probabilidad de que la $(M+1)$ -ésima repetición de mismo tamaño muestral produjese una solución mejor que la mejor de las soluciones pro-

ducidas por las M repeticiones anteriores sería igual a $1/(M + 1)$. Dado que $g(\hat{x}_N)$ tiene una distribución discreta, esta probabilidad es menor o igual que $1/(M + 1)$. Así, cuando $1/(M + 1)$ llega a ser suficientemente pequeño, es probable que las repeticiones adicionales con el mismo tamaño muestral no merezcan la pena, y, o bien debería aumentarse el tamaño de la muestra, o bien debería detenerse el proceso.

3.3. Estimación del gap para el diseño de un test de parada.

Con el fin de ayudar en la toma de decisiones de parada nos gustaría calcular el gap de optimalidad $g(\hat{x}) - v^*$ para una solución dada $\hat{x} \in S$. Desafortunadamente, la propuesta descrita en este documento implica que son difíciles de calcular ambos términos del gap de optimalidad.

$$\hat{g}_{N'}(\hat{x}) := \frac{1}{N'} \sum_{j=1}^{N'} G(\hat{x}, W^j)$$

es un estimador insesgado de $g(\hat{x})$, y la varianza de $\hat{g}_{N'}(\hat{x})$ puede ser estimada por $S_{N'}^2(\hat{x})/N'$, donde $S_{N'}^2(\hat{x})$ es la varianza muestral de $G(\hat{x}, W^j)$, tomando una muestra de tamaño N' .

Un estimador de v^* viene dado por

$$\bar{v}_N^M := \frac{1}{M} \sum_{m=1}^M \hat{v}_N^m$$

donde \hat{v}_N^m denota al valor objetivo óptimo de la m -ésima repetición del problema SAA. Debemos tener en cuenta que $\mathbb{E}[\bar{v}_N^M] = \mathbb{E}[\hat{v}_N]$ y entonces el estimador \bar{v}_N^M tiene el mismo sesgo negativo que \hat{v}_N . La proposición 2.3.1 indica que ese sesgo tiende a ser mayor en problemas mal condicionados con conjuntos grandes de soluciones óptimas. Consideremos el correspondiente estimador $\hat{g}_{N'}(\hat{x}) - \bar{v}_N^M$ del gap de optimalidad

$g(\hat{x}) - v^*$ en el punto \hat{x} . Dado que

$$\mathbb{E} [\hat{g}_{N'}(\hat{x}) - \bar{v}_N^M] = g(\hat{x}) - \mathbb{E} [\hat{v}_N] \geq g(\hat{x}) - v^*, \quad (3.1)$$

se deduce que, en media, el estimador anterior sobreestima el gap de optimalidad $g(\hat{x}) - v^*$. Es posible demostrar que el sesgo $v^* - \mathbb{E} [\hat{v}_N]$ es monótonamente decreciente en el tamaño de la muestra N .

La varianza de \bar{v}_N^M puede ser estimada por

$$\frac{S_M^2}{M} = \frac{1}{M(M-1)} \sum_{m=1}^M (\hat{v}_N^m - \bar{v}_N^M)^2. \quad (3.2)$$

Si las M muestras, de tamaño N , y la muestra de evaluación, de tamaño N' , son independientes, entonces la varianza del estimador del gap de optimalidad $\hat{g}_{N'}(\hat{x}) - \bar{v}_N^M$ puede ser estimada por $S_{N'}^2(\hat{x})/N' + S_M^2/M$.

Un estimador del gap de optimalidad $g(\hat{x}) - v^*$ con varianza posiblemente más pequeña es $\bar{g}_N^M(\hat{x}) - \bar{v}_N^M$, donde

$$\bar{g}_N^M(\hat{x}) := \frac{1}{M} \sum_{m=1}^M \hat{g}_N^m(\hat{x})$$

y

$$\hat{g}_N^m(\hat{x}) := \frac{1}{N} \sum_{j=1}^N G(\hat{x}, W^{mj}).$$

Aquí, la varianza de $\bar{g}_N^M(\hat{x}) - \bar{v}_N^M$ puede ser estimada por

$$\frac{\bar{S}_M^2}{M} = \frac{1}{M(M-1)} \sum_{m=1}^M [(\hat{g}_N^m(\hat{x}) - \hat{v}_N^m) - (\bar{g}_N^M(\hat{x}) - \bar{v}_N^M)]^2.$$

En cuanto a la cuestión de qué estimador del gap de optimalidad tiene menor varianza, depende de la correlación entre $\hat{g}_N^m(\hat{x})$ y \hat{v}_N^m , así como de los tamaños muestrales

N , N' , y M . **El Teorema Central del Límite** puede ser aplicado a los estimadores del gap de optimalidad $\hat{g}_{N'}(\hat{x}) - \bar{v}_N^M$ y $\bar{g}_N^M(\hat{x}) - \bar{v}_N^M$, de modo que la precisión de un estimador del gap de optimalidad puede ser corregido por la adición al estimador en cuestión de un múltiplo z_α de su desviación estándar estimada. $z_\alpha := \Phi^{-1}(1 - \alpha)$, donde $\Phi(z)$ es la función de distribución acumulativa de la distribución normal estándar. Por ejemplo, si $\hat{x} \in S$ denota a la solución candidata con el mejor valor de $\hat{g}_{N'}(\hat{x})$ encontrado tras M repeticiones, entonces, teniendo en cuenta la precisión, un estimador del gap de optimalidad viene dado por

$$\hat{g}_{N'}(\hat{x}) - \bar{v}_N^M + z_\alpha \left(\frac{S_{N'}^2(\hat{x})}{N'} + \frac{S_M^2}{M} \right)^{1/2}$$

ó

$$\bar{g}_N^M(\hat{x}) - \bar{v}_N^M + z_\alpha \frac{\bar{S}_M}{\sqrt{M}}.$$

Para el control del algoritmo, es útil separar un estimador del gap de optimalidad en sus componentes elementales. Por ejemplo,

$$\begin{aligned} & \hat{g}_{N'}(\hat{x}) - \bar{v}_N^M + z_\alpha \left(\frac{S_{N'}^2(\hat{x})}{N'} + \frac{S_M^2}{M} \right)^{1/2} = & (3.3) \\ & = (\hat{g}_{N'}(\hat{x}) - g(\hat{x})) + (g(\hat{x}) - v^*) + (v^* - \bar{v}_N^M) + z_\alpha \left(\frac{S_{N'}^2(\hat{x})}{N'} + \frac{S_M^2}{M} \right)^{1/2}. \end{aligned}$$

En los cuatro términos del lado derecho de la ecuación anterior, el primer término tiene valor esperado cero; el segundo término es el gap de optimalidad; el tercer término es el término de sesgo, que tiene un valor esperado positivo que disminuye con el tamaño de la muestra N ; y el cuarto término es el término de precisión, que decrece en el número de repeticiones M y el tamaño muestral N' . Así, una desventaja de esos estimadores del gap de optimalidad es que pueden ser grandes si M , N , ó N' es pequeño, incluso si \hat{x} es una solución óptima, i.e., $g(\hat{x}) = v^*$.

3.4. Algoritmo.

A continuación proponemos un algoritmo para el problema de optimización estocástica discreta.

ALGORITMO SAA PARA OPTIMIZACIÓN ESTOCÁSTICA DISCRETA.

1. Elija los tamaños muestrales iniciales N y N' , una regla de decisión para determinar el número M de repeticiones del problema SAA (posiblemente con un número máximo M' de repeticiones SAA con el mismo tamaño muestral, tal que $1/(M' + 1)$ es suficientemente pequeño), una regla de decisión para aumentar los tamaños de la muestra N y N' si fuese necesario, y tolerancia ε .
2. Para $m = 1, \dots, M$ realice los pasos de 2.1 a 2.3.
 - 2.1 Genere una muestra de tamaño N y resuelva el problema SAA (1.2) con valor objetivo \hat{v}_N^m y solución ε -óptima \hat{x}_N^m .
 - 2.2 Estime el gap de optimalidad $g(\hat{x}_N^m) - v^*$ y la varianza del estimador.
 - 2.3 Si el gap de optimalidad y la varianza del estimador son suficientemente pequeñas, realice el paso 4.
3. Si el gap de optimalidad o la varianza del estimador son demasiado grandes, aumente los tamaños muestrales N y/o N' , y vuelva al paso 2.
4. Elija la mejor solución \hat{x} entre todas las soluciones candidatas \hat{x}_N^m producidas, utilizando un procedimiento de detección y selección. Parar.

4. Aplicaciones.

En esta sección describimos varias aplicaciones del método SAA a problemas de optimización.

4.1. Problema de asignación de recursos.

Aplicamos el método al siguiente problema de asignación de recursos. Un decisor tiene que seleccionar un subconjunto de entre k proyectos conocidos a priori. Para este fin, tiene disponible una cantidad de recurso q de relativamente bajo costo. Puede obtenerse cualquier cantidad adicional de recursos requerida a un costo adicional conocido de c unidades por unidad de recurso. En el momento de la toma de decisiones no se conoce la cantidad W_i de recursos requeridos por cada proyecto i , pero suponemos que el decisor tiene una estimación de la distribución de probabilidad de $W = (W_1, \dots, W_k)$. Cada proyecto i tiene una media del beneficio neto (ingresos menos costes) a la que llamamos r_i . Entonces, se puede formular el problema de optimización como sigue:

$$\max_{x \in \{0,1\}^k} \left\{ \sum_{i=1}^k r_i x_i - c \mathbb{E} \left[\sum_{i=1}^k W_i x_i - q \right]^+ \right\}, \quad (4.1)$$

donde $[x]^+ := \max\{x, 0\}$. Este problema también puede ser descrito como un problema de la mochila, donde se tiene que elegir un subconjunto de k elementos, dada una mochila de tamaño q en la que introducir dichos elementos. El tamaño W_i de cada elemento i es aleatorio, y cuando se excede la capacidad de la mochila hay que pagar una cantidad c por unidad. Por esta razón el problema se denota por SSKP (del inglés, *static stochastic knapsack problem*).

Se elige este problema por varias razones. Primero, se producen valores esperados similares a los de la función objetivo de (4.1) en muchos problemas de optimización estocástica interesantes, por ejemplo en el problema del camino más corto, como ya vimos en la introducción. Una segunda razón para elegir el SSKP es que las funciones objetivos con términos como $\mathbb{E} \left[\sum_{i=1}^k W_i x_i - q \right]^+$ son interesantes por la siguiente razón. Se pueden obtener buenas soluciones para muchos problemas de optimización estocástica reemplazando las variables aleatorias W por sus medias y

resolviendo el resultante problema de optimización determinista $\max_x G(x, E[W])$, llamado problema valor esperado. Es fácil ver que este caso no puede darse si el objetivo contiene un término valor esperado como en (4.1). Para una solución dada x , este término puede ser muy grande, pero puede llegar a ser pequeño si W_1, \dots, W_k son reemplazados por sus medias. En tal caso, el problema valor esperado obtenido puede producir muy malas soluciones para el correspondiente problema de optimización estocástica.

El problema SAA del SSKP se puede escribir con la siguiente fórmula de programación lineal entera:

$$\begin{aligned}
& \max_{x,z} \quad \sum_{i=1}^k r_i x_i - \frac{c}{N} \sum_{j=1}^N z_j \\
\text{sujeto a:} \quad & z_j \geq \sum_{i=1}^k W_i^j x_i - q, \quad j = 1, \dots, N, \\
& x_i \in \{0, 1\}, \quad i = 1, \dots, k, \\
& z_j \geq 0, \quad j = 1, \dots, N.
\end{aligned} \tag{4.2}$$

Resultados numéricos:

En la Figura 1 se ha representado un conjunto de instancias del SSKP con 20 variables de decisión. Para esta gráfica, el número de condición κ es 954000, se conoce el valor de v^* que es 96,5 y $g(\bar{x})$ el valor de la solución óptima \bar{x} del problema $\max_x G(x, \mathbb{E}[W])$ es 75,9. La Figura 1 muestra la probabilidad estimada de que una solución óptima \hat{x}_N del problema SAA tenga valor objetivo $g(\hat{x}_N)$ dentro de la tolerancia relativa d del valor óptimo v^* , i.e., $\hat{P}[v^* - g(\hat{x}_N) \leq dv^*]$, como función del tamaño de la muestra N , para distintos valores de d . Para su representación se llevó a cabo un experimento en el que se generaron, para cada N , 1000 repeticiones independientes del problema SAA, calculando en cada caso la solución óptima \hat{x}_N^m , $m = 1, \dots, 1000$ y su valor objetivo $g(\hat{x}_N^m)$, y contando el número de veces M_d que

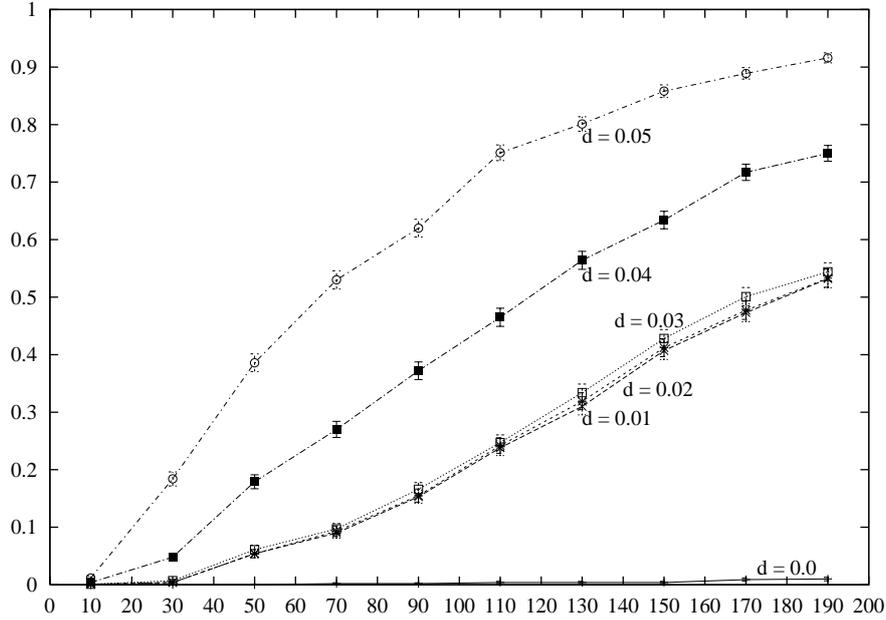


Figura 1: $N \mapsto \hat{P}[v^* - g(\hat{x}_N) \leq dv^*]$

$v^* - g(\hat{x}_N^m) \leq dv^*$. Así, la probabilidad fue estimada por $\hat{P}[v^* - g(\hat{x}_N) \leq dv^*] = \frac{M_d}{1000}$.

4.2. El comportamiento empírico de los métodos de muestreo para la programación estocástica.

(*J. Linderoth, A. Shapiro and S. Wright [3]*)

En este caso, los autores estudian la calidad de las soluciones obtenidas por del método de aproximación por media muestral aplicado a programas lineales estocásticos de dos etapas con recurso (o rectificación).

El programa lineal estocástico de dos etapas con recurso se puede escribir como sigue:

$$\begin{aligned}
 & \min_x c^T x + Q(x) \\
 & \text{sujeto a: } Ax = b \\
 & \quad x \geq 0,
 \end{aligned} \tag{4.3}$$

donde $Q(x) := \mathbb{E}[Q(x, \xi(w))]$ y $Q(x, \xi)$ es el valor óptimo del problema de la segunda etapa

$$\begin{aligned} & \underset{y}{\text{mín}} && q^T y \\ \text{sujeto a: } & && Tx + Wy = h \\ & && y \geq 0. \end{aligned} \tag{4.4}$$

En la matriz $\xi = (q, h, T, W)$ uno o más elementos serán aleatorios, es decir, dependerán de w . En problemas con recurso fijo, la matriz W es fija (determinista).

Se consideran problemas donde la probabilidad de distribución de $\xi(w)$ tiene un número finito de realizaciones posibles, llamadas escenarios $\{\xi_1, \xi_2, \dots, \xi_K\}$ con probabilidades $p_k \in (0, 1)$, $k = 1, 2, \dots, K$. Para tales problemas, la función $Q(x)$ se puede escribir como

$$Q(x) = \sum_{k=1}^K p_k Q(x, \xi_k)$$

donde $Q(x, \xi)$ es el valor óptimo de (4.4) y $\xi_k = (q_k, h_k, T_k, W)$, $k = 1, 2, \dots, K$. Los problemas (4.3) y (4.4) se pueden combinar en un solo programa lineal como sigue

$$\begin{aligned} & \underset{x, y_1, y_2, \dots, y_k}{\text{mín}} && c^T x + \sum_{k=1}^K p_k q_k^T y_k \\ \text{sujeto a: } & && Ax = b, x \geq 0, \\ & && Wy_k = h_k - T_k x, y_k \geq 0, k = 1, 2, \dots, K. \end{aligned}$$

Como se sabe, la solución óptima de un problema de Programación Lineal se alcanza en uno de los puntos extremos del conjunto poliédrico de soluciones factibles. Por tanto, la formulación anterior se corresponde con el problema (1.1) que se ha estudiado en secciones precedentes en el que el conjunto factible se puede considerar finito, reduciéndonos a los puntos extremos.

Los autores han aplicado el método SAA a 5 programas lineales estocásticos de

dos etapas con recurso, que varían en dificultad, y resolviendo para distintos tamaños muestrales, han obtenido intervalos de confianza para los valores óptimos de esos 5 problemas. Tras esto, han observado que obtienen buenas soluciones aproximadas con tamaños muestrales pequeños en unos casos y grandes en otros. También, han llevado a cabo un estudio de la convergencia del conjunto de soluciones obtenidas con el método SAA al conjunto de soluciones del problema original haciendo uso de la proposición 2.2.3.

El mayor descubrimiento que los autores hacen con dicho estudio es que en todos los problemas examinados es posible calcular "buenas" soluciones del problema original tras un esfuerzo computacional razonable. También este estudio pone de manifiesto que, como ya dijimos en secciones precedentes, los problemas mal condicionados son más difíciles de resolver.

4.3. Optimización de estrategias de detección de virus en redes sociales.

(J. Lee, J. J. Hasenbein and D. P. Morton [4])

Con el objetivo de representar la dinámica de un virus que se propaga en una red social (whatsapp, facebook, twitter,...) se han desarrollado diferentes modelos atendiendo a las características de la red. Basándose en dicha dinámica, se presentan modelos de optimización para detectar con rapidez el virus. Se persiguen dos objetivos: maximizar la probabilidad de detectar el virus en un umbral de tiempo prefijado y minimizar el tiempo esperado para la detección del mismo.

Nos vamos a centrar en detectar virus mediante el control de teléfonos móviles que presentan un comportamiento anómalo.

Consideremos la red social como un grafo $G = (V, E)$ donde V denota el conjunto de nodos del grafo y E el conjunto de aristas. Sea $S \subseteq V$ un conjunto de nodos

llamados *detectores* que debe ser fijado de antemano, y el virus será detectado cuando infecte a un nodo en S .

A continuación, vamos a plantear el modelo SAA que se usa para maximizar la probabilidad de detectar un virus en un umbral de tiempo determinado, bajo la restricción de que el número de detectores que se pueden instalar en la red es k . Consideremos N observaciones *i.i.d.* de la propagación del virus hasta el momento t_0 . Para cada observación denotamos por $P_l \subseteq V$, $l = 1, \dots, N$, al conjunto de nodos que hayan resultado infectados. Con esas observaciones, el modelo SAA correspondiente es el siguiente:

$$\begin{aligned} \hat{v}_N = \max_{x,y} \quad & \frac{1}{N} \sum_{l=1}^N y_l \\ \text{sujeto a:} \quad & x \in X \\ & y_l \leq \sum_{i \in P_l} x_i, \quad l = 1, \dots, N, \\ & y_l \in \{0, 1\}, \quad l = 1, \dots, N. \end{aligned}$$

$X = \{x : \sum_{i \in V} x_i = k, x_i \in \{0, 1\}, i \in V\}$, la variable y_l indica si la realización l contiene un detector en su conjunto de nodos, P_l , y el conjunto de variables x_i , $i \in V$, determina el conjunto de detectores. De acuerdo con la formulación del problema SAA, el objetivo tomará el máximo valor posible de 1 si es posible seleccionar al menos un detector en cada uno de los conjuntos de nodos infectados. Se trataría por tanto de un problema de cubrimiento.

Una importante propiedad de este modelo es que el hecho de que X es un conjunto finito hace que se cumplan las 2 propiedades de la Proposición 2.1.2.

4.4. Un método SAA para el problema de balance de línea de desmontaje bajo incertidumbre.

(*M. L. Bentaha, O. Battaïa and A. Dolgui [5]*)

Consideramos un problema de balance o equilibrado de la carga de trabajo de varias líneas de desmontaje de productos, cuya funcionalidad ha finalizado, bajo el alto grado de incertidumbre que suponen la estructura y calidad de estos productos, debido a las condiciones variables de los mismos. Se supone que los tiempos de realización de la tarea de desmontaje son variables aleatorias con distribución de probabilidad conocida. Los autores proponen una solución para este problema utilizando un modelo de programación estocástica. A la vez, elige la mejor alternativa de desmontaje para el producto y asigna las tareas de desmontaje correspondientes a las estaciones de trabajo con el objetivo de minimizar el coste de la línea de desmontaje. Este último, incluye los costos de operación de las estaciones de trabajo, así como los costes de penalización generados por violaciones de las restricciones de tiempo. También pueden haber restricciones de precedencia entre tareas. El algoritmo SAA es capaz de proporcionar soluciones aproximadas de buena calidad incluso en aquellos casos en los que el problema es a gran escala, para ello es necesario que el tamaño de la muestra N cumpla la cota (2.25).

Referencias

- [1] V. K. Rohatgi and A. K. MD. Ehsanes Saleh. An introduction to probability and statistics. John Wiley and Sons, (2001).
- [2] A. J. Kleywegt, A. Shapiro and T. Homem-de-Mello. The sample average approximation method for stochastic discrete optimization. *Siam Journal on Optimization*, 12(2):479–502, (2001).
- [3] J. Linderoth, A. Shapiro and S. Wright. The empirical behavior of sampling methods for stochastic programming. *Annals of Operations Research*, 142:215–241, (2006).
- [4] J. Lee, J. J. Hasenbein and D. P. Morton. Optimization of stochastic virus detection in contact networks. *Operations Research Letters*, 43(1):59–64, (2015).
- [5] M. L. Bentaha, O. Battaïa and A. Dolgui. A sample average approximation method for disassembly line balancing problem under uncertainty. *Computers and Operations Research*, 51:111–122, (2014).