

Optimización Robusta Distribucional basada en datos aplicada a la estimación de funciones de densidad de probabilidad y a optimización de portafolios

Diego Fonseca

10 de agosto de 2018

Optimización estocástica y su reformulación robusta distribucional

Determinación del parámetro bandwidth en la estimación de funciones de densidad por kernels

Optimización de portafolios desde la perspectiva de Markowitz

Un problema de optimización estocástica es de la forma

$$J^* = \min_{x \in \mathbb{X}} \mathbb{E}_{\mathbb{P}}[f(x, \xi)]$$

donde $f : \mathbb{X} \times \Xi \rightarrow \mathbb{R}$, \mathbb{X} es la región factible y ξ es un elemento aleatorio con distribución \mathbb{P} soportada en Ξ .

Para efectos de notación definimos $J(x) := \mathbb{E}_{\mathbb{P}}[f(x, \xi)]$.

Objetivo: Aproximar superiormente J^* .

Aproximación robusta distribucional J^*

Sea \mathcal{D} un conjunto de distribuciones de probabilidad de tal manera que $\mathbb{P} \in \mathcal{D}$ con alta probabilidad, entonces el problema de optimización robusta distribucional DRO que aproxima J^* superiormente con alta probabilidad es

$$\hat{J}_N := \min_{x \in \mathbb{X}} \sup_{Q \in \mathcal{D}} \mathbb{E}_Q[f(x, \xi)]. \quad (1)$$

A (1) también se le conoce como *versión robusta distribucional* de J^* .

Dada una muestra $\hat{\xi}_1, \dots, \hat{\xi}_N$ de \mathbb{P} , $\hat{\mathbb{P}}_N$ la distribución empírica de terminada por esta muestra y $\varepsilon > 0$., consideramos $\mathcal{D} = \mathcal{B}_\varepsilon(\hat{\mathbb{P}}_N)$ siendo esta la bola cerrada respecto a una métrica p -Wasserstein de radio ε y centro $\hat{\mathbb{P}}_N$.

La intención es reformular (1) como un problema de optimización semi-infinita que bajo ciertas condiciones en f se puede convertir en un problema de optimización convexo finito.

Métrica de Wasserstein

Definición 1 (Métrica de Wasserstein)

La *distancia de Wasserstein* $W_p(\mu, \nu)$ entre $\mu, \nu \in \mathcal{P}_p(\Xi)^1$ es definida por

$$W_p^p(\mu, \nu) := \inf_{\Pi \in \mathcal{P}(\Xi \times \Xi)} \left\{ \int_{\Xi \times \Xi} d^p(\xi, \zeta) \Pi(d\xi, d\zeta) \mid \begin{array}{l} \Pi(\cdot \times \Xi) = \mu(\cdot), \\ \Pi(\Xi \times \cdot) = \nu(\cdot) \end{array} \right\}$$

donde

$$\mathcal{P}_p(\Xi) := \left\{ \mu \in \mathcal{P}(\Xi) : \int_{\Xi} d^p(\xi, \zeta_0) \mu(d\xi) < \infty \text{ para algun } \zeta_0 \in \Xi \right\}$$

donde d es una métrica en Ξ .

En el contexto de $\mathcal{P}_p(\Xi)$ la bola de radio $\varepsilon > 0$ con centro en $\mu \in \mathcal{P}_p(\Xi)$ es

$$\mathcal{B}_\varepsilon^p(\mu) = \left\{ \nu \in \mathcal{P}_p(\Xi) \mid W_p^p(\mu, \nu) \leq \varepsilon^p \right\}. \quad (2)$$

¹La métrica p -Wasserstein también esta definida para distribuciones fuera de $\mathcal{P}_p(\Xi)$, lo que probablemente podría ocurrir es que ese conjunto la métrica de Wasserstein sea infinito.

Proposición 2

Sean ε y $\beta \in (0, 1)$ tales que $\mathbb{P}^N \left\{ W_p(\mathbb{P}, \hat{P}_N) \leq \varepsilon \right\} \geq 1 - \beta$. Si \hat{x}_N una solución óptima de (1) entonces

$$\mathbb{P}^N \left\{ \mathbb{E}_P [f(\hat{x}_N, \xi)] \leq \hat{J}_N \right\} \geq 1 - \beta.$$

Noten que siempre se tiene $J^* \leq \mathbb{E}_P [f(\hat{x}_N, \xi)]$.

Momentáneamente centraremos nuestra atención en el problema de maximización interno de (1) y ya que x en los siguientes resultados solo actúa como una constante omitiremos la presencia de x en f , es decir, estudiaremos el problema de maximización

$$\sup_{Q \in \mathcal{B}_\varepsilon(\hat{\mathbb{P}}_N)} \mathbb{E}_Q[f(\xi)]. \quad (3)$$

Suposición 1: Asumimos f que alguna de las siguientes condiciones:

1. f es continua y es tal que existe $C > 0$ y $\xi_0 \in \Xi$ tal que $|f(\xi)| \leq C(1 + d^p(\xi, \xi_0))$ para todo $\xi \in \Xi$.
2. f es acotada.
3. f es máximo de funciones concavas, es decir, $f = \max_{k \leq K} f_k$ donde cada $-f_k$ es propia, convexa e inferiormente semicontinua respecto a ξ .

Para todo $x \in \mathbb{X}$.

Teorema 3 (Teorema principal)

Bajo la Suposición 1 el problema (3) se puede reformular como el problema de optimización semi-infinito

$$\left\{ \begin{array}{ll} \inf_{\lambda, s} & \lambda \varepsilon^p + \frac{1}{N} \sum_{i=1}^N s_i \\ \text{sujeto a} & \sup_{\xi \in \Xi} \left(f(\xi) - \lambda d^p(\xi, \hat{\xi}_i) \right) \leq s_i \quad \forall i \leq N. \\ & \lambda \geq 0. \end{array} \right. \quad (4)$$

Para demostrar este teorema se caracteriza (3) como un problema cónico lineal y en ese contexto se emplean resultados de dualidad fuerte para este tipo de problemas.

Corolario 4 (Kuhn [1])

Asumimos $\Xi = \{\xi \in \mathbb{R}^m : C\xi \leq g\}$ donde C es una matriz y g un vector de dimensiones apropiadas, $p = 1$ y d como una norma $\|\cdot\|$ en \mathbb{R}^m y se consideran las funciones $a_k(x, \xi) := \langle a_k x, \xi \rangle + b_k$ para todo $k \leq K$. Si $f(x, \xi) = \max_{k \leq K} a_k(x, \xi)$ entonces el valor óptimo de (1) es igual al valor óptimo de

$$\left\{ \begin{array}{ll} \inf_{\lambda, x, s, \gamma_{ik}} & \lambda \epsilon + \frac{1}{N} \sum_{i=1}^N s_i \\ \text{sujeto a} & b_k + \langle a_k x, \hat{\xi}_i \rangle + \langle \gamma_{ik}, d - C\hat{\xi}_i \rangle \leq s_i \quad \forall i \leq N \forall k \leq K. \\ & \|C^T \gamma_{ik} - a_k x\|_* \leq \gamma \quad \forall i \leq N \forall k \leq K. \\ & \gamma_{ik} \geq 0. \end{array} \right. \quad (5)$$

Determinación del parámetro bandwidth en la estimación de funciones de densidad por kernels

Sea ξ una variable aleatoria y f su función de densidad, dada una muestra $\hat{\xi}_1, \dots, \hat{\xi}_N$ este método de estimación propone el estimador de f dado por

$$\hat{f}_h(x) := \frac{1}{Nh} \sum_{i=1}^N \mathcal{K} \left(\frac{x - \hat{\xi}_i}{h} \right).$$

Este estimador depende del parámetro $h > 0$ comúnmente conocido como *ancho de banda* o *parámetro de suavidad*, también depende de \mathcal{K} la cual es una función de densidad simétrica al rededor de cero y tal que $\int x^2 \mathcal{K}(x) dx = 1$. Frecuentemente se considera \mathcal{K} como la función de densidad normal estándar, no obstante \mathcal{K} puede ser cualquier función de densidad que satisfaga las condiciones de simetría y varianza.

El método consiste en encontrar un $h > 0$ tal que minimice el *error cuadrático integrado medio* (mean integrated squared error)

$$\text{MISE}(h) := \mathbb{E}_{\mathbb{P}^N} \left[\int \left(\hat{f}_h(x) - f(x) \right)^2 dx \right]. \quad (6)$$

La aleatoriedad en esta expresión está en el vector aleatorio $(\hat{\xi}_1, \dots, \hat{\xi}_N)$ cuya distribución es $\mathbb{P}^N = \mathbb{P} \times \dots \times \mathbb{P}$, donde \mathbb{P} es la distribución inducida por la función de densidad f , de modo que \mathbb{P} también es desconocida.

Otra opción es elegir $h > 0$ de tal manera que minimice el *error cuadrático integrado* (integrate square error)

$$\text{ISE}(h) := \int \left(\hat{f}_h(x) - f(x) \right)^2 dx. \quad (7)$$

El h producto de esta elección dependerá de la muestra.

Minimización de ISE como un DROW

Por la linealidad de la integral la expresión (7) se puede reescribir como

$$\begin{aligned}\text{ISE}(h) &:= \int \left(\hat{f}_h(x) \right)^2 dx - 2 \int \hat{f}_h(x) f(x) dx + \int (f(x))^2 dx \\ &= \int \left(\hat{f}_h(x) \right)^2 dx - 2\mathbb{E}_{\mathbb{P}} \left[\hat{f}_h(\xi) \right] + \int (f(x))^2 dx\end{aligned}$$

Considerando \bar{J} la parte que depende de h en la expresión anterior, es decir,

$$\bar{J}(h) := \int \left(\hat{f}_h(x) \right)^2 dx - 2\mathbb{E}_{\mathbb{P}} \left[\hat{f}_h(\xi) \right], \quad (8)$$

entonces se infiere que h minimiza la expresión ISE si y solo si minimiza \bar{J} , de modo que enfocamos nuestra atención en minimizar \bar{J} , minimizar \bar{J} ya es un problema de optimización estocástica.

Por un lado, el problema de interés es

$$\bar{J}^* := \min_{h \geq 0} \bar{J}(h) = \min_{h \geq 0} \left(\int \left(\hat{f}_h(x) \right)^2 dx - 2\mathbb{E}_{\mathbb{P}} \left[\hat{f}_h(\xi) \right] \right) \quad (9)$$

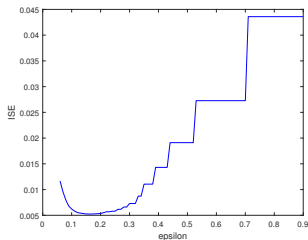
Luego, dada una muestra $\hat{\xi}_1, \dots, \hat{\xi}_N$ de ξ , tal aproximación superior es el valor óptimo del problema

$$\hat{J}_N := \min_{h \geq 0} \left(\int \left(\hat{f}_h(x) \right)^2 dx + 2 \sup_{Q \in \mathcal{B}_\epsilon(\hat{\mathbb{P}}_N)} \mathbb{E}_Q \left[-\hat{f}_h(\xi) \right] \right). \quad (10)$$

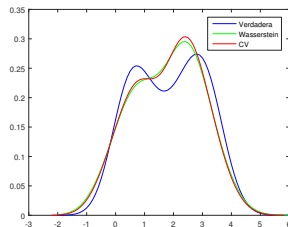
Consideraremos \mathcal{K} como una función de densidad acotada, esto implica que $-\hat{f}_h(\xi)$ es acotada, entonces satisface la Suposición 1 de modo que se satisfacen las hipótesis del *Teorema principal*, luego, considerando $\Xi = \mathbb{R}$, $p = 2$ y d como la distancia euclidiana, dicho Teorema permite reescribir (10) es equivalente al problema de optimización semi-infinito

$$\left\{ \begin{array}{l} \inf_{h, \lambda, s} \int \left(\hat{f}_h(x) \right)^2 dx + 2\lambda \varepsilon^p + \frac{2}{N} \sum_{i=1}^N s_i \\ \text{sujeto a } \sup_{\xi \in \mathbb{R}} \left(-\hat{f}_h(\xi) - \lambda \left(\xi - \hat{\xi}_i \right)^2 \right) \leq s_i \quad \forall i = 1, \dots, N, \\ \lambda \geq 0. \\ h > 0. \end{array} \right. \quad (11)$$

Progresos y limitantes numéricas



(a)



(b)

Figura: En (a) la gráfica de ε vs ISE . En (b) la representación de las funciones de densidad estimada vía Wasserstein, la verdadera y una estimada mediante validación cruzada que notamos como CV.

Trabajo futuro y en progreso

El problema (11) es de optimización semi-infinita y el algoritmo empleado hasta el momento es costoso computacionalmente y demanda un tiempo considerable, siendo esta la principal limitante, una asignatura pendiente es solucionar o aproximar la solución de dicho problema de optimización de manera eficiente.

Optimización de portafolios desde la perspectiva de Markowitz

Modelo de media-varianza de Markowitz

Un inversionista tiene un monto de dinero que desea invertir en m bienes, denotamos por ξ_i el retorno del bien i -ésimo y x_i la proporción del monto inicial invertida en el bien i -ésimo, en este sentido, dado que los retornos de cada bien son en la practica aleatorios entonces $\xi = (\xi_1, \dots, \xi_m) \in \mathbb{R}^m$ es un *vector aleatorio*, adicionalmente, $x = (x_1, \dots, x_m) \in \mathbb{R}^m$ es un vector conocido como el vector de pesos, este satisface la relación $\sum_{i=1}^m x_i = 1$. El conjunto de los m bienes se le conoce como el *portafolio* y la expresión $R := \sum_{i=1}^m x_i \xi_i$ es el *retorno total del portafolio*, considerando $\langle \cdot, \cdot \rangle$ el producto interno usual de \mathbb{R}^m se infiere que $R = \langle x, \xi \rangle$.

Dado que ξ es un vector aleatorio con distribución \mathbb{P} entonces los siguientes conceptos son los que tomaran relevancia en esta sección:

Retorno esperado $:= \mathbb{E}_{\mathbb{P}} [\langle x, \xi \rangle]$.

Volatilidad $:= \text{Var}_{\mathbb{P}} [\langle x, \xi \rangle] = \mathbb{E}_{\mathbb{P}} [(\langle x, \xi \rangle - \mathbb{E}_{\mathbb{P}} [\langle x, \xi \rangle])^2]$.

Lo ideal es encontrar un vector de pesos x que le garantice un retorno esperado alto pero con una volatilidad baja, en tal virtud el inversionista establece un nivel de retorno esperado mínimo μ . Esta visión es representada en el siguiente modelo:

$$J := \begin{cases} \min_{x \in \mathbb{R}^m} & \text{Var}_{\mathbb{P}} [\langle x, \xi \rangle] \\ \text{sujeto a} & \mathbb{E}_{\mathbb{P}} [\langle x, \xi \rangle] \geq \mu, \\ & \sum_{i=1}^m x_i = 1. \end{cases} \quad (12)$$

Si se conociera la matriz de covarianza E y el vector de valores esperados \mathbf{m} del vector aleatorio ξ entonces (12) es equivalente al problema de optimización

$$\left\{ \begin{array}{ll} \min_{x \in \mathbb{R}^m} & x^T E x \\ \text{sujeto a} & \mathbf{m}^T x \geq \mu, \\ & \sum_{i=1}^m x_i = 1. \end{array} \right. \quad (13)$$

Pero en la practica E y \mathbf{m} no son conocidos, ante esta situación es común considerar E y \mathbf{m} como las versiones muestrales. De modo que la aproximación Robusta Distribucional es una opción viable.

Versión Robusta Distribucional:

Fijando $x \in \mathbb{X}$ definimos $\zeta^x := \langle x, \xi \rangle$ la cual es una variable aleatoria, llamamos \mathbb{P}^x su distribución la cual depende de \mathbb{P} , luego, dada $\hat{\xi}_1, \dots, \hat{\xi}_N$ una muestra de \mathbb{P} , entonces $\hat{\zeta}_1^x, \dots, \hat{\zeta}_N^x$ definida por $\hat{\zeta}_i^x := \langle x, \hat{\xi}_i \rangle$ es una muestra de ζ^x , esto permite definir la distribución empírica $\hat{\mathbb{P}}_N^x$ asociada a ζ^x la cual es dada por

$$\hat{P}_N^x := \sum_{i=1}^N \delta_{\hat{\zeta}_i^x}.$$

En el espacio de las distribuciones se considera la noción de distancia determinada por la métrica de 2-Wasserstein denotada por W_2 , esto permite considerar el conjunto

$$\mathcal{B}_{\varepsilon \|x\|}(\hat{\mathbb{P}}_N^x) = \left\{ \mathbb{Q} \in \mathcal{P}(\mathbb{R}^m) \mid W(\mathbb{Q}, \hat{\mathbb{P}}_N^x) \leq \varepsilon \|x\| \right\}.$$

Este conjunto es la bola respecto a la métrica 2-Wasserstein con centro en $\hat{\mathbb{P}}_N^x$ y radio $\varepsilon \|x\|$.

A partir de lo anterior definimos el conjunto

$$\begin{aligned} \mathbb{X} &:= \left\{ x \in \mathbb{R}^m \left| \sum_{i=1}^m x_i = 1, \mathbb{E}_{\mathbb{Q}}[\zeta^x] \geq \mu \ \forall \mathbb{Q} \in \mathcal{B}_{\|x\|\varepsilon}(\hat{\mathbb{P}}_N^x) \right. \right\} \\ &= \left\{ x \in \mathbb{R}^m \left| \sum_{i=1}^m x_i = 1, \inf_{\mathbb{Q} \in \mathcal{B}_{\|x\|\varepsilon}(\hat{\mathbb{P}}_N^x)} \mathbb{E}_{\mathbb{Q}}[\zeta^x] \geq \mu \right. \right\} \end{aligned} \quad (14)$$

Entonces la versión Robusta Distribucional de (12) es

$$\hat{J}_N := \underset{x \in \mathbb{X}}{\text{minimizar}} \quad \sup_{\mathbb{Q} \in \mathcal{B}_{\|x\|\varepsilon}(\hat{\mathbb{P}}_N^x)} \text{Var}_{\mathbb{Q}}[\zeta^x]. \quad (15)$$

Llamaremos a este problema DR-MRK.

Pero por el **Teorema principal** se tiene

$$\inf_{\mathbb{Q} \in \mathcal{B}_{\|x\| \varepsilon}(\hat{\mathbb{P}}_N^x)} \mathbb{E}_{\mathbb{Q}}[\zeta^x] = \frac{1}{N} \sum_{i=1}^N \hat{\zeta}_i^x - \varepsilon \|x\|.$$

Lo que permite reescribir \mathbb{X} como

$$\begin{aligned} \mathbb{X} &= \left\{ x \in \mathbb{R}^m \left| \sum_{i=1}^m x_i = 1, \frac{1}{N} \sum_{i=1}^N \hat{\zeta}_i^x - \varepsilon \|x\| \geq \mu \right. \right\} \\ &= \left\{ x \in \mathbb{R}^m \left| \sum_{i=1}^m x_i = 1, \frac{1}{N} \sum_{i=1}^N \langle x, \hat{\xi}_i \rangle - \varepsilon \|x\| \geq \mu \right. \right\} \quad (16) \end{aligned}$$

Por otro lado, introduciendo una variable de holgura se obtiene

$$\sup_{\mathbb{Q} \in \mathcal{B}_{\|x\| \varepsilon}(\hat{\mathbb{P}}_N^x)} \text{Var}_{\mathbb{Q}}[\zeta^x] = \sup_{\substack{\eta \geq \mu, \\ \left(\eta - \frac{1}{N} \sum_{i=1}^N \hat{\zeta}_i^x\right)^2 \leq \varepsilon^2 \|x\|^2}} \begin{cases} \sup_{\mathbb{Q} \in \mathcal{B}_{\|x\| \varepsilon}(\hat{\mathbb{P}}_N^x)} \text{Var}_{\mathbb{Q}}[\zeta^x] \\ \text{sujeto a } \mathbb{E}_{\mathbb{Q}}[\zeta^x] = \eta. \end{cases}$$

La primer tarea es reescribir el problema interno

$$\begin{cases} \sup_{\mathbb{Q} \in \mathcal{B}_{\|x\| \varepsilon}(\hat{\mathbb{P}}_N^x)} \text{Var}_{\mathbb{Q}}[\zeta^x] \\ \text{sujeto a } \mathbb{E}_{\mathbb{Q}}[\zeta^x] = \eta. \end{cases}$$

En ese sentido es importante el siguiente Teorema.

Teorema 5

Sea ζ un a variable aleatoria con distribución \mathbb{P} y tal que se conoce su esperanza, es decir, se sabe que $\mathbb{E}_{\mathbb{P}}[\zeta] = \eta$, además, sea $\hat{\zeta}_1, \dots, \hat{\zeta}_N$ una muestra de ζ y $\varepsilon^2 \geq \left(\frac{1}{N} \sum_{i=1}^N (\eta - \hat{\zeta}_i) \right)^2$. Considerando $\mathcal{B}_{\varepsilon}(\hat{\mathbb{P}}_N)$ como la bola respecto a la métrica 2-Wasserstein de radio ε centrada en $\hat{\mathbb{P}}_N$ la distribución empírica respecto a la muestra anterior, entonces

$$\left\{ \begin{array}{ll} \sup_{\mathbb{Q} \in \mathcal{B}_{\varepsilon}(\hat{\mathbb{P}}_N)} & \mathbb{E}_{\mathbb{Q}} [(\zeta - \eta)^2] \\ \text{sujeto a} & \mathbb{E}_{\mathbb{Q}} [\zeta] = \eta. \end{array} \right. = \left(\sqrt{\frac{1}{N} \sum_{i=1}^N \hat{\zeta}_i^2 - \left(\frac{1}{N} \sum_{i=1}^N \hat{\zeta}_i \right)^2} + \sqrt{\varepsilon^2 - \left(\frac{1}{N} \sum_{i=1}^N (\eta - \hat{\zeta}_i) \right)^2} \right)^2. \quad (17)$$

Por lo tanto

$$\left\{ \begin{array}{ll} \sup_{\mathbb{Q} \in \mathcal{B}_{\|x\|, \varepsilon}(\hat{\mathbb{P}}_N^x)} \text{Var}_{\mathbb{Q}}[\zeta^x] \\ \text{sujeto a } \mathbb{E}_{\mathbb{Q}}[\zeta^x] = \eta. \end{array} \right. = \left(\sqrt{\frac{1}{N} \sum_{i=1}^N \langle x, \hat{\xi}_i \rangle^2 - \frac{1}{N^2} \left(\sum_{i=1}^N \langle x, \hat{\xi}_i \rangle \right)^2} + \sqrt{\varepsilon^2 \|x\|^2 - \left(\eta - \frac{1}{N} \sum_{i=1}^N \langle x, \hat{\xi}_i \rangle \right)^2} \right)^2$$

Así pues, por este teorema se infiere que (15) es equivalente al problema de optimización:

$$\hat{J}_N = \underset{x \in \mathbb{X}}{\text{minimizar}} \left\{ \begin{array}{ll} \sup_{\eta \geq \mu} & \left(\sqrt{\frac{1}{N} \sum_{i=1}^N \langle x, \hat{\xi}_i \rangle^2 - \frac{1}{N^2} \left(\sum_{i=1}^N \langle x, \hat{\xi}_i \rangle \right)^2} + \sqrt{\varepsilon^2 \|x\|^2 - \left(\mu - \frac{1}{N} \sum_{i=1}^N \langle x, \hat{\xi}_i \rangle \right)^2} \right)^2 \\ \text{sujeto a } & \left(\eta - \frac{1}{N} \sum_{i=1}^N \hat{\zeta}_i^x \right)^2 \leq \varepsilon^2 \|x\|^2 \end{array} \right. \quad (18)$$

Pero el problema de maximización interno de (18) puede solucionarse explícitamente, en realidad dicho problema alcanza su valor óptimo en $\eta^* = \frac{1}{N} \sum_{i=1}^N \hat{\zeta}_i^x$, por lo tanto, (18) se puede reescribir como

$$\hat{J}_N = \underset{x \in \mathbb{X}}{\text{minimizar}} \left(\sqrt{\frac{1}{N} \sum_{i=1}^N \langle x, \hat{\xi}_i \rangle^2 - \frac{1}{N^2} \left(\sum_{i=1}^N \langle x, \hat{\xi}_i \rangle \right)^2} + \varepsilon \|x\| \right)^2$$

$$= \left\{ \begin{array}{l} \underset{x \in \mathbb{R}^m}{\text{minimizar}} \left(\sqrt{\frac{1}{N} \sum_{i=1}^N \langle x, \hat{\xi}_i \rangle^2 - \frac{1}{N^2} \left(\sum_{i=1}^N \langle x, \hat{\xi}_i \rangle \right)^2} + \varepsilon \|x\| \right)^2 \\ \text{sujeto a} \quad \sum_{i=1}^N \langle x, \hat{\xi}_i \rangle - \varepsilon \|x\| \geq \mu, \\ \sum_{i=1}^m x_i = 1. \end{array} \right. \quad (19)$$

El problema (19) se puede simplificar aun más.

Proposición 6

Sea M la matriz de tamaño $m \times N$ cuyas columnas son los vectores de la muestra $\hat{\xi}_1, \dots, \hat{\xi}_N$ y sean $\mathbf{0}, \mathbf{e} \in \mathbb{R}^N$ los vectores columna de ceros y unos respectivamente. A partir de estas convenciones se definen las matrices

$$E := \frac{1}{N} M M^T - \frac{1}{N^2} (M \mathbf{e})(M \mathbf{e})^T \quad \text{y} \quad L := \frac{1}{N} (M \mathbf{e})^T.$$

Ya que E es semidefinida positiva entonces semidefinida positiva de modo que existe una matriz K tal que $E = K K^T$. Entonces (19) es equivalente al problema de optimización

$$\begin{cases} \inf_{x \in \mathbb{R}^m} & (\|K^T x\| + \varepsilon \|x\|)^2 \\ \text{sujeto a} & Lx - \varepsilon \|x\| \geq \mu, \\ & \mathbf{e}^T x = 1. \end{cases} \quad (20)$$

Se debe tener en cuenta que (20) puede no ser factible para algunos valores de ε , concretamente, dada L y μ se tiene que (20) es factible si

$$\varepsilon < \hat{\varepsilon}_N(\mu) := \begin{cases} \sup_{x \in \mathbb{R}^m} \frac{Lx - \mu}{\|x\|} \\ \text{sujeto a } \sum_{i=1}^m x_i = 1. \end{cases}$$

La dependencia de N en $\hat{\varepsilon}_N(\mu)$ se debe a que L depende de la muestra. En adelante llamaremos *radio extremo factible* a la expresión $\hat{\varepsilon}_N(\mu)$.

Elección de ε

Priorizando el Riesgo: Este es ε_{var} el menor valor posible de ε que satisface

$$V(\varepsilon) := \mathbb{E}_{\mathbb{P}^N} \left[\hat{J}_N(\varepsilon) - \text{Var}_{\mathbb{P}} [\langle \hat{x}_N(\varepsilon), \xi \rangle] \right] \geq 0. \quad (21)$$

Priorizando rentabilidad: Este es ε_{ret} el menor valor de ε que satisface

$$R(\varepsilon) := \mathbb{E}_{\mathbb{P}^N} [\mathbb{E}_{\mathbb{P}} [\langle \hat{x}_N(\varepsilon), \xi \rangle]] \geq \mu. \quad (22)$$

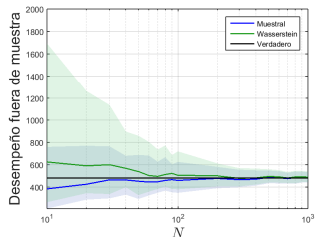
Resultados numéricos

Los resultados numéricos que se traducen en las siguiente gráficas originadas por simulaciones realizadas para un portafolio compuesto de cuatro bienes, es decir, $m = 4$, la distribución de ξ es multinormal con matriz de covarianza C y vector de medias \mathbf{m} dados por

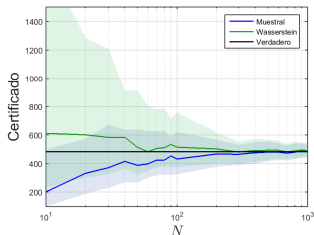
$$C = \begin{bmatrix} 185 & 86,5 & 80 & 20 \\ 86,5 & 196 & 76 & 13,5 \\ 80 & 76 & 411 & -19 \\ 20 & 13,5 & -19 & 25 \end{bmatrix} \quad \text{y} \quad \mathbf{m} = (14, 12, 15, 17).$$

Para los siguientes experimentos se toma $\mu = 20$.

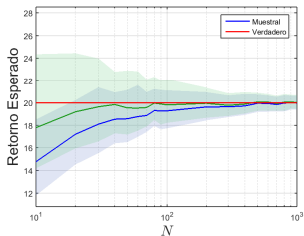
Priorizando el riesgo



(a)



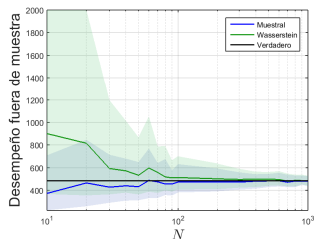
(b)



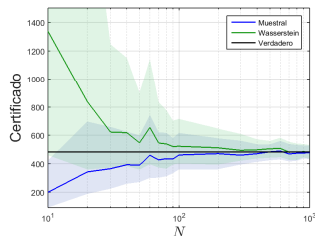
(c)

(a) Desempeño fuera de muestra $\text{Var}_{\mathbb{P}}[\langle \hat{x}_N, \xi \rangle]$ (línea verde y área sombreada verde) y $\text{Var}_{\mathbb{P}}[\langle \hat{x}_N^{\text{muest}}, \xi \rangle]$ (línea azul y área sombreada azul) donde \hat{x}_N^{muest} es la solución muestral. (b) certificado \hat{J}_N (línea verde y área sombreada verde) y $\text{Var}_{\hat{\mathbb{P}}_N}[\langle \hat{x}_N^{\text{muest}}, \xi \rangle]$ (línea azul y área sombreada azul). (c) Retorno esperado $\mathbb{E}_{\mathbb{P}}[\langle \hat{x}_N, \xi \rangle]$ (línea verde y área sombreada verde) y $\mathbb{E}_{\mathbb{P}}[\langle \hat{x}_N^{\text{muest}}, \xi \rangle]$ (línea azul y área sombreada azul). Se tomó $\varepsilon = \varepsilon_{\text{var}}$.

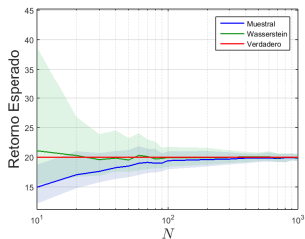
Priorizando el retorno



(a)



(b)



(c)

(a) Desempeño fuera de muestra $\text{Var}_{\mathbb{P}}[\langle \hat{x}_N, \xi \rangle]$ (línea verde y área sombreada verde) y $\text{Var}_{\mathbb{P}}[\langle \hat{x}_N^{\text{muestral}}, \xi \rangle]$ (línea azul y área sombreada azul) donde $\hat{x}_N^{\text{muestral}}$ es la solución muestral. (b) certificado \hat{J}_N (línea verde y área sombreada verde) y $\text{Var}_{\hat{\mathbb{P}}_N}[\langle \hat{x}_N^{\text{muestral}}, \xi \rangle]$ (línea azul y área sombreada azul). (c) Retorno esperado $\mathbb{E}_{\mathbb{P}}[\langle \hat{x}_N, \xi \rangle]$ (línea verde y área sombreada verde) y $\mathbb{E}_{\mathbb{P}}[\langle \hat{x}_N^{\text{muestral}}, \xi \rangle]$ (línea azul y área sombreada azul). Se tomó $\varepsilon = \varepsilon_{\text{ret}}$.

Trabajo futuro y en progreso

Una asignatura pendiente es comparar el desempeño de la Versión Robusta Distribucional de Markowitz respecto a otros modelos diseñados para mejorar el modelo de Markowitz, por ejemplo el Modelo Black-Litterman.

Otro aspecto es hacer mas robusto el modelo robusto distribucional.

Gracias por su atención.



Esfahani, PM. y Kuhn, D. A.

Data-driven Distributionally Robust Optimization Using the Wasserstein Metric: Performance Guarantees and Tractable Reformulations.

arXiv preprint arXiv:1505.05116v2, 2016.



Lee, C. y Mehrotra, S.

A distributionally-Robust Optimization approach for finding support vector machines.

Optimization Online, 2015.



Pflug, G. y Wozabal, D.

Ambiguity in portfolio selection.

Quantitative finance, 435-442, 2006.



Luo, F. y Mehrotra, S

Decomposition Algorithm for Distributionally Robust Optimization using Wasserstein Metric.

preprint arXiv:1610.05627v2, 2017.



Shapiro, A. and Dentcheva, D.

Lectures on Stochastic programming: modeling and theory.
SIAM, 2016.