

Optimización Robusta Distribucional basada en datos aplicada a optimización de portafolios

Diego Fonseca

19 de junio de 2019

Optimización estocástica y su reformulación robusta distribucional

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^* . **Yo no diría que este es el principal objetivo, lo quitaría**

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 $\hat{\mathbb{P}}_N$.

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.

Para resolver (1), momentáneamente centraremos nuestra atención en el problema de maximización interno:

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

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 2 (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.

Optimización de portafolios desde la perspectiva de Markowitz

Modelo de media-varianza de Markowitz

Sea $\xi = (\xi_1, \dots, \xi_m) \in \mathbb{R}^m$ el *vector aleatorio* con distribución \mathbb{P} que representa los retornos de m activos con riesgo, se quiere encontrar un vector de pesos $x = (x_1, \dots, x_m) \in \mathbb{R}^m$ que le garantice al inversionista un retorno esperado mínimo μ , pero con una volatilidad baja. Esta visión es representada en el siguiente modelo:

$$J := \begin{cases} \min_{x \in \mathbb{R}^m} & \text{Var}_{\mathbb{P}} [\langle x, \xi \rangle] \\ \text{sujeto a} & \mathbb{E}_{\mathbb{P}} [\langle x, \xi \rangle] \geq \mu, \\ & \sum_{i=1}^m x_i = 1. \end{cases} \quad (5)$$

Si se conociera la matriz de covarianza E y el vector de valores esperados \mathbf{m} del vector aleatorio ξ entonces (5) es equivalente al problema de optimización

$$\left\{ \begin{array}{ll} \min_{x \in \mathbb{R}^m} & x^T E x \\ \text{sujeto a} & \mathbf{m}^T x \geq \mu, \\ & \sum_{i=1}^m x_i = 1. \end{array} \right. \quad (6)$$

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 $x \in \mathbb{R}^m$ 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

$$\begin{aligned}\mathbb{X} &:= \left\{ x \in \mathbb{R}^m \left| \sum_{i=1}^m x_i = 1, \mathbb{E}_{\mathbb{Q}}[\zeta^x] \geq \mu \ \forall \mathbb{Q} \in \mathcal{B}_{\|x\|\varepsilon}(\hat{\mathbb{P}}_N^x) \right. \right\} \\ &= \left\{ x \in \mathbb{R}^m \left| \sum_{i=1}^m x_i = 1, \inf_{\mathbb{Q} \in \mathcal{B}_{\|x\|\varepsilon}(\hat{\mathbb{P}}_N^x)} \mathbb{E}_{\mathbb{Q}}[\zeta^x] \geq \mu \right. \right\}\end{aligned}\quad (7)$$

Entonces la versión Robusta Distribucional de (5) es

$$\hat{J}_N := \underset{x \in \mathbb{X}}{\text{minimizar}} \quad \sup_{\mathbb{Q} \in \mathcal{B}_{\|x\|\varepsilon}(\hat{\mathbb{P}}_N^x)} \text{Var}_{\mathbb{Q}}[\zeta^x]. \quad (8)$$

Llamaremos a este problema DR-MRK.

El **Teorema principal** permite mostrar que

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

Luego podemos reescribir \mathbb{X} como

$$\begin{aligned} \mathbb{X} &= \left\{ x \in \mathbb{R}^m \left| \sum_{i=1}^m x_i = 1, \frac{1}{N} \sum_{i=1}^N \hat{\zeta}_i^x - \varepsilon \|x\| \geq \mu \right. \right\} \\ &= \left\{ x \in \mathbb{R}^m \left| \sum_{i=1}^m x_i = 1, \frac{1}{N} \sum_{i=1}^N \langle x, \hat{\xi}_i \rangle - \varepsilon \|x\| \geq \mu \right. \right\} \quad (9) \end{aligned}$$

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 3

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

$$\left\{ \begin{array}{ll} \sup_{\mathbb{Q} \in \mathcal{B}_{\varepsilon}(\hat{\mathbb{P}}_N)} & \mathbb{E}_{\mathbb{Q}} [(\zeta - \eta)^2] \\ \text{sujeto a} & \mathbb{E}_{\mathbb{Q}} [\zeta] = \eta. \end{array} \right. = \left(\sqrt{\frac{1}{N} \sum_{i=1}^N \hat{\zeta}_i^2 - \left(\frac{1}{N} \sum_{i=1}^N \hat{\zeta}_i \right)^2} + \sqrt{\varepsilon^2 - \left(\frac{1}{N} \sum_{i=1}^N (\eta - \hat{\zeta}_i) \right)^2} \right)^2. \quad (10)$$

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 (8) es equivalente al problema de optimización:

$$\hat{J}_N = \underset{x \in \mathbb{X}}{\text{minimizar}} \left\{ \begin{array}{ll} \sup_{\eta \geq \mu} \left(\sqrt{\frac{1}{N} \sum_{i=1}^N \langle x, \hat{\xi}_i \rangle^2 - \frac{1}{N^2} \left(\sum_{i=1}^N \langle x, \hat{\xi}_i \rangle \right)^2} + \sqrt{\varepsilon^2 \|x\|^2 - \left(\mu - \frac{1}{N} \sum_{i=1}^N \langle x, \hat{\xi}_i \rangle \right)^2} \right)^2 \\ \text{sujeto a } \left(\eta - \frac{1}{N} \sum_{i=1}^N \hat{\zeta}_i^x \right)^2 \leq \varepsilon^2 \|x\|^2 \end{array} \right. \quad (11)$$

Pero el problema de maximización interno de (11) 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, (11) se puede reescribir como

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

(12)

El problema (12) se puede simplificar aun más.

Proposición 4

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

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

Ya que E es semidefinida positiva entonces semidefinida positiva de modo que existe una matriz K tal que $E = KK^T$. Entonces (12) es equivalente al problema de optimización

$$\begin{cases} \inf_{x \in \mathbb{R}^m} & (\|K^T x\| + \varepsilon \|x\|)^2 \\ \text{sujeto a} & Lx - \varepsilon \|x\| \geq \mu, \\ & \mathbf{e}^T x = 1. \end{cases} \quad (13)$$

Se debe tener en cuenta que (13) puede no ser factible para algunos valores de ε , concretamente, dada L y μ se tiene que (13) 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 $\hat{\varepsilon}_N(\mu)$.

El problema de maximizar el retorno sujeto a una varianza acotada

$$J' := \begin{cases} \max_{x \in \mathbb{R}^m} & \mathbb{E}_{\mathbb{P}} [\langle x, \xi \rangle] \\ \text{sujeto a} & \text{Var}_{\mathbb{P}} [\langle x, \xi \rangle] \leq \nu, \\ & \sum_{i=1}^m x_i = 1. \end{cases} \quad (14)$$

Siguiendo las mismas técnicas que se emplearon para proponer una versión Robusta de (5) y lograr su reformulación (13), se tiene que la versión Robusta de (14) con su reformulación es:

$$\left\{ \begin{array}{ll} \max_{x \in \mathbb{R}^m} & \inf_{Q \in \mathcal{B}_{\|x\|_{\varepsilon}(\hat{\mathbb{P}}_N^x)}^{\mathbb{P}} \mathbb{E}_Q[\zeta^x] \\ \text{sujeto a} & \sup_{Q \in \mathcal{B}_{\|x\|_{\varepsilon}(\hat{\mathbb{P}}_N^x)}^{\mathbb{P}} \text{Var}_Q[\zeta^x] \leq \nu, \\ & e^T x = 1. \end{array} \right. = \left\{ \begin{array}{ll} \max_{x \in \mathbb{R}^m} & Lx - \varepsilon \|x\| \\ \text{sujeto a} & (\|K^T x\| + \varepsilon \|x\|)^2 \leq \nu \\ & e^T x = 1. \end{array} \right. \quad (15)$$

Elección de ε para el problema de minimizar la varianza sujeto a un retorno acotado inferiormente.

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

$$V(\varepsilon) := \mathbb{E}_{\mathbb{P}^N} \left[\hat{J}_N(\varepsilon) - \text{Var}_{\mathbb{P}} [\langle \hat{x}_N(\varepsilon), \xi \rangle] \right] \geq 0. \quad (16)$$

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

$$R(\varepsilon) := \mathbb{E}_{\mathbb{P}^N} [\mathbb{E}_{\mathbb{P}} [\langle \hat{x}_N(\varepsilon), \xi \rangle]] \geq \mu. \quad (17)$$

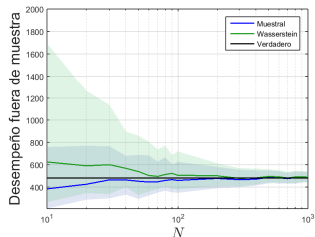
Resultados numéricos

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

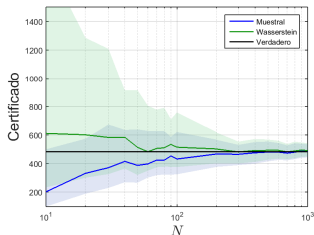
$$C = \begin{bmatrix} 185 & 86,5 & 80 & 20 \\ 86,5 & 196 & 76 & 13,5 \\ 80 & 76 & 411 & -19 \\ 20 & 13,5 & -19 & 25 \end{bmatrix} \quad \text{y} \quad \mathbf{m} = (14, 12, 15, 17).$$

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

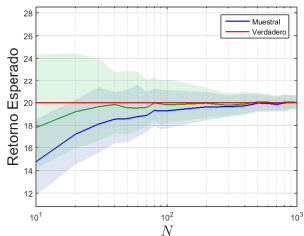
Priorizando el riesgo



(a)



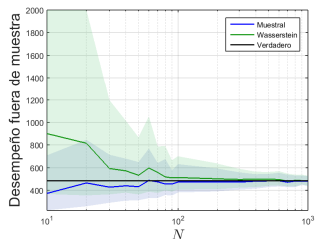
(b)



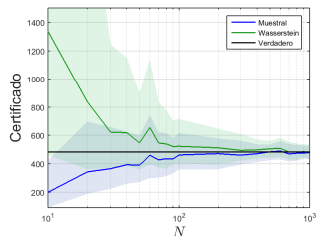
(c)

(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 $\varepsilon = \varepsilon_{\text{var}}$.

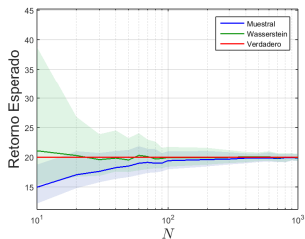
Priorizando el retorno



(a)



(b)

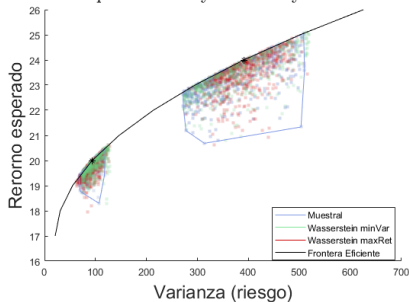


(c)

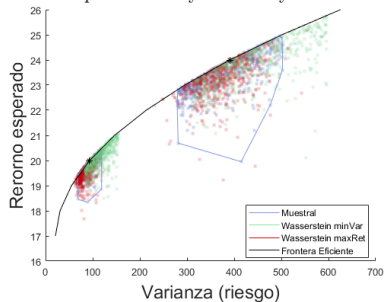
(a) Desempeño fuera de muestra $\text{Var}_{\mathbb{P}}[\langle \hat{x}_N, \xi \rangle]$ (línea verde y área sombreada verde) y $\text{Var}_{\mathbb{P}}[\langle \hat{x}_N^{\text{muestral}}, \xi \rangle]$ (línea azul y área sombreada azul) donde $\hat{x}_N^{\text{muestral}}$ es la solución muestral. (b)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

Desempeño en la frontera eficiente respecto a ε

Estimacion para $\varepsilon = 0.2$ y $N = 100$ y 1000 remuestreos

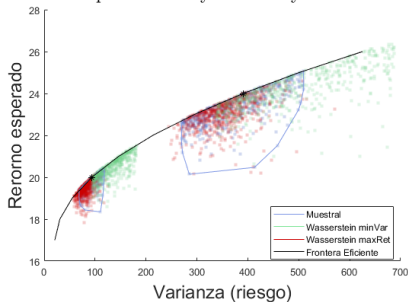


Estimacion para $\varepsilon = 0.4$ y $N = 100$ y 1000 remuestreos

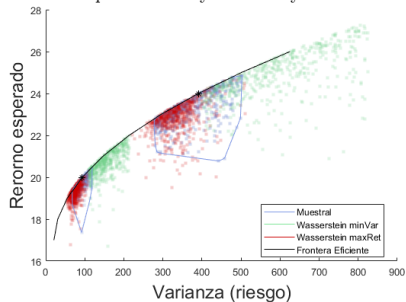


Desempeño en la frontera eficiente respecto a ε

Estimación para $\varepsilon = 0.6$ y $N = 100$ y 1000 remuestreos



Estimación para $\varepsilon = 0.8$ y $N = 100$ y 1000 remuestreos



Trabajo en progreso y futuro Se esta trabajando en un método iterativo que combina los problemas (8) y (15) con el fin de estimar puntos en la frontera eficiente. La idea de este enfoque surge de los resultados evidenciados en las imágenes anteriores.

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.