



**UNIVERSIDAD DE CHILE**  
**FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS**  
**DEPARTAMENTO DE INGENIERÍA MATEMÁTICA**  
**DEPARTAMENTO DE INGENIERÍA INDUSTRIAL**

**ESTUDIO DE MÉTODOS DE OPTIMIZACIÓN ROBUSTA PARA EL PROBLEMA  
DE PLANIFICACIÓN DE PRODUCCIÓN EN MINERÍA A CIELO ABIERTO**

**TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN GESTIÓN DE  
OPERACIONES**

**MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL MATEMÁTICO**

**GUIDO RENATO LAGOS BARRIOS**

**PROFESOR GUÍA:  
DANIEL ESPINOZA GONZÁLEZ**

**PROFESOR CO-GUÍA:  
EDUARDO MORENO ARAYA**

**MIEMBROS DE LA COMISIÓN:  
JORGE AMAYA ARRIAGADA  
ALEJANDRO JOFRÉ CÁCERES  
BERNARDO KULNIG PAGNONCELLI**

**SANTIAGO DE CHILE  
AGOSTO 2011**

RESUMEN DE LA TESIS PARA OPTAR AL TÍTULO DE  
INGENIERO CIVIL MATEMÁTICO Y AL GRADO DE  
MAGÍSTER EN GESTIÓN DE OPERACIONES  
POR: GUIDO LAGOS BARRIOS  
FECHA: AGOSTO DEL 2011  
PROF. GUÍA: DANIEL ESPINOZA GONZÁLEZ

## ESTUDIO DE MÉTODOS DE OPTIMIZACIÓN ROBUSTA PARA EL PROBLEMA DE PLANIFICACIÓN DE PRODUCCIÓN EN MINERÍA A CIELO ABIERTO

En el presente trabajo consideramos un problema de planificación de producción en minería a cielo abierto. Usando el *modelo de bloques*, consistente en discretizar el yacimiento minero en paralelepípedos regulares o *bloques*, el problema consiste en decidir qué bloques de la mina extraer y cuáles procesar, teniendo por objetivo maximizar el beneficio total de la explotación. Esta operación considera una capacidad limitada de extracción y procesamiento, y se realiza en un único período de tiempo, e.g. un año.

Esta tesis aborda este problema cuando la cantidad de mineral, o *ley*, de los bloques está sujeta a incertidumbre. Bajo este escenario el problema es altamente riesgoso, pues los costos de operación son elevados y la explotación sólo se puede realizar una única vez. Para enfrentar esto proponemos varios modelos de optimización bajo incertidumbre.

Proponemos primero dos enfoques de optimización aversa al riesgo, los modelos de minimización de *VaR* y minimización de *CVaR*. Presentamos también un enfoque de optimización robusta, el modelo *MCH*, y demostramos que tiene propiedades teóricas deseables de aversión al riesgo. Inspirados en esto, proponemos el modelo *MCH- $\epsilon$* , una variación del modelo *MCH*. Y finalmente, consideramos una variante *con recourse* para cada modelo, consistente en incorporar un paradigma *multi-etapa* a las decisiones de explotación.

Para representar la variabilidad de las leyes asumimos únicamente la disponibilidad de un generador de muestras *iid* para el vector de leyes. Esto permite aproximar los modelos de *VaR*, *CVaR* y *MCH- $\epsilon$*  usando la técnica *Sample Average Approximation* (SAA). Demostramos que teóricamente estos problemas aproximados convergen al problema estocástico verdadero. Sin embargo, resultados de geometría discreta sugieren que esta convergencia es muy lenta.

Finalmente, presentamos experimentos computacionales en una mina de tamaño reducido basada en datos reales. Los resultados obtenidos con la implementación usada permiten descartar el uso del modelo de *VaR*, y en contraste los modelos *CVaR*, *MCH* y *MCH- $\epsilon$*  muestran un manejo de riesgo controlable y coherente. Por otro lado, estimaciones estadísticas del gap de optimalidad evidencian una lenta velocidad de convergencia de las aproximaciones SAA. También, se constata que el uso de las variantes con recourse producen un aumento notable de los beneficios.

## AGRADECIMIENTOS

Al término de esta primera etapa de mi vida académica quisiera agradecer a todas las personas que han estimulado y apoyado mi andar en este camino.

Primeramente agradezco a mi familia, por haber definido quién soy, y por haberme dado no sólo las capacidades para poder caminar esta senda, sino que también el gusto por el conocimiento, honestidad, curiosidad, y todos esos ingredientes fundamentales de la vida académica. Agradezco especialmente a mi mamá, Silvita, pues ella siempre ha sido el pilar fundamental de mi formación, y también a mi esposa, Ani, por haber aguantado mis momentos más tensos (e.g. postulación a Ph.D., terminar tesis, etc.) y por darme la ilusión por la cual querer seguir avanzando.

En segundo lugar quisiera agradecer a los profesores con los que he trabajado cercanamente este último par de años: Daniel, Eduardo, Bernardo y Juan Pablo; por guiarme en mis primeros pasos en este oficio excitante de la investigación académica en ingeniería. A Daniel en especial, por ser un guía comprometido y exigirme siempre ir un paso más allá.

Agradezco profundamente al Departamento de Ingeniería Matemática en general, por inculcar en mí una cultura del conocimiento, trabajo y rigurosidad que me atrevo a decir es única en nuestro país. A Éterin, por ser la secretaria docente más diligente de todo el universo, y a Julie (del DII sí) por su buena disposición y voluntad en ayudar las miles de veces que la fui a molestar con mis requerimientos.

No puedo dejar de mencionar a mis grandes compañeros de batalla y aventuras. En el DIM, a mis vatos Julio, Félix, Tomás, Claudio, Calipo, Víctor y Jero, y en la adherencia, a los buenos de Itu, Héctor, Omar, Airoume y Quiñi. También a mis viejos amigos de plan común Wamaso, Latonsito, Lascani, Mondaca, Viviracha y la Roma. Y por último pero no por eso menos, a los alacranes de la risa Jano, Huerta, Pinino y, por supuesto, al increíble Patangas Fabuloso. Son re wena onda cabros, son muy pulentos, extrañaré mucho vuestra sencillez, simpatía y buen humor.

Finalmente, quisiera expresar lo afortunado que me siento de ser parte de la comunidad de la Universidad de Chile, una universidad genuinamente pluralista, laica y de excelencia, y la más grande de mi país. Espero que la tradición de esta comunidad nunca deje de empapar nuestra sociedad, y que más temprano que tarde vuelva a ser realmente pública.

Cabe mencionar que este trabajo fue financiado por el FONDEF D06-I-1031.

# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Antecedentes . . . . .	1
1.2. Literatura y estado del arte . . . . .	2
1.3. Sobre esta tesis . . . . .	5
<b>2. Descripción del Problema</b>	<b>7</b>
2.1. Problema determinista básico . . . . .	7
2.2. Enfoques de optimización bajo riesgo . . . . .	9
2.2.1. Modelo de minimización de VaR . . . . .	10
2.2.2. Modelo de minimización de CVaR . . . . .	15
2.2.3. Modelos MCH y MCH- $\epsilon$ . . . . .	22
2.2.4. Sobre el contenido de probabilidad de un polítopo . . . . .	34
<b>3. Aproximación SAA</b>	<b>38</b>
3.1. Marco teórico general . . . . .	38
3.1.1. Propiedades de estimadores SAA para Stochastic Programming Problems . . . . .	38
3.1.2. Propiedades de estimadores SAA para Chance Constrained Problems . . . . .	44
3.2. Análisis SAA para modelos sin recourse . . . . .	50
3.2.1. Modelo de minimización de CVaR . . . . .	51
3.2.2. Modelo MCH . . . . .	55
3.2.3. Modelo MCH- $\epsilon$ . . . . .	57
3.2.4. Modelo de minimización de VaR . . . . .	59
3.3. Análisis SAA para modelos con recourse . . . . .	63

---

3.3.1. Modelo de minimización de CVaR con recourse . . . . .	66
3.3.2. Modelo MCH con recourse . . . . .	68
3.3.3. Modelo MCH- $\epsilon$ con Recourse . . . . .	69
3.3.4. Modelo de minimización de VaR con recourse . . . . .	71
<b>4. Experimentos Computacionales</b>	<b>74</b>
4.1. Descripción . . . . .	74
4.2. Parámetros . . . . .	76
4.3. Resultados . . . . .	78
4.3.1. Distribución empírica de beneficios . . . . .	78
4.3.2. Convergencia . . . . .	89
4.3.3. Estudio de algunos factores que afectan convergencia . . . . .	94
<b>5. Conclusiones y Trabajo Futuro</b>	<b>97</b>
5.1. Conclusiones . . . . .	97
5.2. Trabajo futuro . . . . .	98

# Capítulo 1

## Introducción

### 1.1. Antecedentes

La minería es sin duda una de las actividades económicas más importantes en Chile. Según datos de la Comisión Chilena del Cobre, al año 2007 la minería representaba en Chile aproximadamente el 7,3 % del PIB, el 65 % de las exportaciones y el 33 % de la inversión extranjera. Más aún, en el escenario global Chile es uno de los principales actores de la minería. Según datos de la fuente anterior, al año 2007 Chile era el principal productor de cobre, nitratos, yodo y litio, poseyendo más de un tercio de las reservas mundiales de estos minerales.

La industria minera se puede clasificar en términos del tipo de excavación que realiza. Se divide así en dos grandes categorías: minería *a cielo abierto* o *de rajo abierto* si las labores se realizan por sobre la superficie terrestre, y *minería subterránea* si se realizan por debajo de la superficie. Si bien a nivel global es menos frecuente encontrar operaciones mineras de cielo abierto, Chile es uno de los países donde más se realiza este tipo de minería.

En términos de gestión y operación, la minería a cielo abierto propone una gran y compleja tarea. Particularmente, la fase de pre-factibilidad y factibilidad de este tipo de proyectos mineros involucra una amplia gama de problemas. Una familia importante de estos es la que se dedica a determinar el valor de la mina en el corto, mediano y largo plazo, para lo cual se debe realizar, entre otras tareas, una proyección de cómo se explotará el yacimiento minero.

El modelo más usual del que se sirven las planificaciones de producción en minería a cielo abierto es el *modelo de bloques*, el que consiste en discretizar el yacimiento minero en paralelepípedos, o *bloques*, del mismo tamaño. A cada bloque se le asignan varios valores representando las propiedades del macizo rocoso que contiene, por ejemplo ubicación geográfica, estimaciones de la cantidad de mineral que contiene, o *ley*, densidad del contenido, etcétera. El modelo de bloques considera además que debe haber un secuenciamiento para la extracción de los bloques. Éste establece que un bloque a mayor

profundidad puede ser extraído solamente después de haber extraído los bloques ubicados directamente sobre él, y también las paredes de la mina deben respetar una máxima pendiente para evitar derrumbes. Estas restricciones sobre el secuenciamiento de extracción se incorporan al modelo de bloques mediante las *restricciones de precedencia* para los bloques.

En el contexto de proyección de la operación minera, dos problemas que generalmente hacen uso del modelo de bloques son el *ultimate pit problem*, cuya traducción literal es “problema del foso final”, y el *production scheduling problem*, o “problema de secuenciamiento de producción”. El primer problema se realiza para valorizar una mina, determinando qué sectores del yacimiento pueden llegar a ser explotados durante el ciclo de vida de la mina. El segundo problema busca, dado un número preestablecido de períodos de tiempo, determinar qué sectores serán explotados durante qué período de tiempo. Estos dos problemas se hacen teniendo como objetivo maximizar el valor total de la operación minera. Y naturalmente, existen en la literatura y en la práctica una gran cantidad de variaciones y simplificaciones de dichos modelos.

Una característica de estos problemas es que los yacimientos mineros generalmente son de gran tamaño, por ejemplo una mina grande normalmente posee entre 3 y 10 millones de bloques. También, para que la explotación de ésta sea rentable se debe considerar un ciclo de vida largo, usualmente más de 20 años. Esto se traduce en que muchas veces el problema de optimización a resolver posea un tamaño excesivamente grande, incluso para minas medianas, haciéndolo finalmente computacionalmente intratable si es que no se recurren a métodos ad-hoc para resolverlo.

Otra complicación que aflora en estos problemas es la falta de confiabilidad de los parámetros con que se hacen las proyecciones de producción. Una fuente de incertidumbre es la duración de la explotación, que por su gran extensión hace difícil pronosticar confiablemente parámetros altamente relevantes como costos de operación y precios de los minerales. Otro parámetro que padece de imprecisión es la cantidad de mineral, o *ley*, que cada bloque posee. Esto es fruto de que la exploración del yacimiento, que determina en particular la cantidad de mineral en él, se realiza usando sondajes que poseen un elevado costo. Como resultado se debe hacer un número muy limitado de sondajes, con los que se determina las propiedades de algunos sectores de la mina, y para los sectores inexplorados sus propiedades se estiman con técnicas geo-estadísticas basadas en los sondajes.

En suma, los altos costos involucrados en la operación minera, los errores asociados a las mediciones y estimaciones, y el hecho de que la extracción del yacimiento rocoso se realiza una sola vez, hacen que finalmente toda planificación de producción posea un alto riesgo asociado.

## 1.2. Literatura y estado del arte

En el contexto de minería a cielo abierto, los problemas *ultimate pit problem* y *production scheduling problem* son los que han recibido más atención en el último par de

décadas. En cuanto al primero, si bien usualmente se aborda usando el algoritmo de Lerchs & Grossman, las maneras más eficientes para afrontarlo provienen de reformular el problema como uno de flujo máximo y usar alguno de los algoritmos más eficientes – e.g. Push-Relabel– para resolverlo (ver [UT98] y [HC00]). Y en el caso del segundo problema, *production scheduling problem*, generalmente se formula como un problema de optimización lineal entera y es tratado usando herramientas del área, por ejemplo *strong branching* (ver [CH03]) y *strengthened formulations* (ver [BBFF10]). Esfuerzos recientes para tratar este problema de manera eficiente consideran el trabajo de Bienstock et al. [BZ10], y el de Chicoisne et al. [CEG<sup>+</sup>09], donde proponen una metodología ad-hoc para el caso en que existe una única restricción de capacidad.

El problema de incertidumbre en planificación minera ha recibido gran atención en la última década. De la mano de simulaciones para el contenido mineral del yacimiento minero, llamadas *conditional simulations* (ver [EL06]), se ha mostrado que la incertidumbre en los parámetros del problema puede llevar a grandes pérdidas. Trabajos como el de Leite et al. [LD07] han propuesto maximizar el valor esperado del Valor Presente Neto (VPN) de la operación, y en cambio trabajos como el de Vielma et al. [VEM09] han propuesto un enfoque averso al riesgo usando *chance constraints*, o restricciones en probabilidad. Éste última publicación da un panorama general más detallado de los esfuerzos en esta línea en los últimos años.

En el campo de optimización bajo incertidumbre existen primordialmente dos grandes enfoques de modelamiento. Uno de ellos, frecuentemente referida como *optimización estocástica*, consiste en asumir que el parámetro incierto posee una distribución estocástica, de la cual se conoce su distribución o se tiene acceso a un generador de muestras de él. Con esto se pueden incluir en el problema de optimización restricciones probabilísticas, por ejemplo restricciones en probabilidad, en esperanza, o incluso con momentos de orden mayor. El otro enfoque de modelamiento, el de *optimización robusta*, no asume ninguna naturaleza estocástica para el parámetro incierto, si no que simplemente se remite a optimizar en la peor realización posible del parámetro dentro de algún rango predeterminado. Los siguientes párrafos se dedican a ahondar en los detalles de estas dos formas de abordar la optimización sujeta a incertidumbre.

En el caso de restricciones en probabilidad, o *chance constraints*, el estudio de éstas se inició con el trabajo seminal de Charnes et al. [CCS58]. Si bien han sido extensamente estudiadas desde entonces, dificultades como no convexidad de la región factible y la dificultad para integrar con precisión en altas dimensiones resultaron en que no se produjeran avances importantes en varias décadas. Sin embargo, el florecimiento de la optimización computacional ha producido que desde hace algo más de una década se empezaran a notar avances importantes en el tratamiento de este tipo de problemas, de la mano de aproximaciones convexas o aproximaciones en base a muestras. Para una revisión general de esta área, especialmente de la aproximación en base a muestras, llamada usualmente *sample average approximation* (SAA), recomendamos ver [PAS09]. Contribuciones importantes recientes en este ámbito se centran en el tratamiento computacional de estos problemas, vía descomposiciones ad-hoc (ver [LAN10] y [Lue10]) o vía asegurar, en probabilidad, la

satisfacción de la restricción en probabilidad (ver [CG10]).

En el caso de optimización con restricciones en valor esperado, usualmente referido simplemente como *stochastic programming*, ésta comparte un desarrollo muy similar al de las *chance constraints*; es decir, si bien este enfoque fue planteado hace varias décadas, sólo los recientes avances en computación y optimización computacional han permitido un desarrollo significativo. Especial mención merece el enfoque *sample average approximation* para estos problemas, es decir, el tratamiento de estos problemas haciendo uso de una muestra del parámetro incierto, con lo que el problema aproximado se puede ver como un estimador del problema original. Resultados cruciales en esta línea son, por ejemplo, [DW88] que asegura la consistencia del estimador en un contexto general y [KSHdM02] que garantiza convergencia del estimador a velocidad exponencial. Para una revisión amplia de esta aproximación sugerimos referirse a [LSW06].

Un enfoque relacionado al de optimización estocástica es el de *optimización aversa al riesgo*, pues generalmente asume que el parámetro sigue alguna distribución estocástica. Como su nombre sugiere, esta teoría busca dar un tratamiento averso al riesgo a la función de utilidad bajo estocacidad de un modelo. Trabajos tempranos en esta línea involucraron optimizar una combinación de esperanza y varianza de la función de utilidad (ver [Mar52]). Más recientemente, enfoques para distribuciones de utilidad asimétricas han ganado alta popularidad, como Value-at-Risk (VaR) primero, y últimamente, con fuerza aún creciente, Conditional Value-at-Risk (CVaR) (ver [RU00]). Para la masificación de este último modelo han sido determinantes la tratabilidad de su optimización (ver [RU02]), también que –a diferencia del VaR– sea una *medida de riesgo coherente* en el sentido de Artzner et al. [ADEH99], y además que sea la componente fundamental de las *medidas de riesgo de distorsión* (ver [BB09]), también conocidas como *medidas de riesgo espectrales*. No obstante, la optimización de CVaR haciendo uso de su aproximación SAA –sin duda uno de los enfoques de optimización aversa al riesgo que ha gozado de mayor popularidad– ha mostrado ser inestable si es que no se toma en cuenta la velocidad de convergencia de la aproximación, ver e.g. [LSV11]. Recomendamos el reciente libro [SDR09] como una referencia exhaustiva y unificadora para las teorías de optimización estocástica y optimización aversa al riesgo.

El enfoque de *optimización robusta*, mencionado anteriormente como una de las dos vertientes principales en la optimización bajo incertidumbre, se caracteriza por ni siquiera asumir una naturaleza estocástica para el parámetro incierto, y en vez simplemente especificar un *conjunto de robustez* para el parámetro y luego optimizar para la peor realización posible de éste en el conjunto. El primer atisbo de esta metodología se debe a Soyster [Soy73], quien argumentaba que, en optimización lineal continua, optimizar para el peor caso en un intervalo cerrado posee la ventaja de poder formularse como un problema lineal. Naturalmente optimizar para el peor caso posible probablemente resulte excesivamente conservador, afectando finalmente de sobremanera el valor objetivo del problema. Por ello se han propuesto conjuntos de robustez menos conservadores, siendo ejemplos importantes los conjuntos de robustez elipsoidales –asociados a la norma  $L^2$ –, ver e.g. [BTN00], o los asociados a un simplex generalizado –asociados a la norma  $L^1$ –

, ver [BS04]. Éste último ha gozado de gran popularidad últimamente, principalmente porque es un modelo sencillo y tratable, y también porque es rico en propiedades al ser aplicado a otras áreas, por ejemplo a optimización en redes [BS03], o a problemas clásicos del área de *management sciences* (ver [BT06]).

En optimización sujeta a incertidumbre en general existe un recurso con enorme potencial, que es permitir que algunas variables de decisión se ajusten a las realizaciones del parámetro incierto. Esto es, permitir que algunas decisiones se tomen solamente después de que el parámetro aleatorio haya “revelado” su verdadero valor. Si bien este concepto introduce gran versatilidad en el problema de optimización, posee el defecto de que generalmente el problema resultante es intratable computacionalmente. En vista de esto generalmente se propone usar simplificaciones, por ejemplo que las variables de *segunda etapa* sean lineales en el parámetro incierto. En [BTGGN04] y [BTN08] se muestran en general las características de este recurso en el contexto de optimización robusta.

Por último, resulta interesante que existen resultados que relacionan estrechamente optimización estocástica, optimización aversa al riesgo y optimización robusta. Por ejemplo [BB09] y [NPS09] ven cómo un modelo de optimización robusta se puede plantear como uno de optimización bajo riesgo, y viceversa. Y [CSS07] muestra cómo un problema de optimización estocástica se puede aproximar tratablemente usando un enfoque de optimización robusta.

### 1.3. Sobre esta tesis

En la presente tesis nos hacemos cargo del riesgo que presupone realizar planificaciones mineras en base a parámetros que están sujetos a incertidumbre. En particular, nos concentramos en el caso en que solamente la ley de mineral no se conoce con certeza. Elegimos trabajar bajo un enfoque de aversión al riesgo, pues, como se argumentó antes, el problema a tratar se caracteriza por ser de gran tamaño, poseer un horizonte de planificación extenso, tener costos operativos altos y la explotación sólo se puede realizar una vez, con lo que el problema es altamente riesgoso. En contraste, los trabajos que se han dedicado a estudiar la incertidumbre en el problema de planificación minera generalmente han usado un enfoque de *maximización de valor esperado*, el cual es un tratamiento eminentemente neutro al riesgo. Los objetivos generales de este trabajo son, primero, estudiar los principales enfoques de optimización bajo incertidumbre, segundo, aplicarlos al problema de planificación minera, y finalmente, evaluar el desempeño y resultados de los enfoques propuestos.

La presente tesis está organizada de la siguiente manera. En el Capítulo 2 presentamos el problema de planificación minera sin incertidumbre. Luego introducimos un esquema para la incertidumbre y mostramos los enfoques que proponemos para tratarla en el problema de planificación. A saber, los enfoques estudiados son minimización de Conditional Value-at-Risk, un modelo de optimización robusta propuesto, que llamamos *Modulated Convex Hull*, y minimización de Value-at-Risk. También presentamos la contraparte

*con recourse* de estos modelos. En el Capítulo 3 exponemos los resultados pertinentes a la aproximación Sample Average Approximation, la cual usamos para hacer los enfoques robustos tratables. Estos resultados consideran consistencia de la aproximación y validación estadística de soluciones. En el Capítulo 4 mostramos los resultados computacionales obtenidos con los modelos propuestos. El análisis de estos resultados considera el estudio de los planes entregados por los distintos enfoques y el estudio numérico de la convergencia de la aproximación SAA. Finalmente, en el Capítulo 5 se dan las conclusiones de nuestro trabajo y se proponen futuras líneas de investigación.

# Capítulo 2

## Descripción del Problema

### 2.1. Problema determinista básico

En esta sección introducimos el problema de planificación básico y sin incertidumbre sobre el cual trabajaremos. Esta planificación se realizará sobre un único período de tiempo, es decir, consideramos que toda la explotación se produce durante una única unidad de tiempo, e.g. un año. Consideramos esta simplificación pues el fin último de este trabajo es evaluar el desempeño de los distintos modelos aversos al riesgo para tratar la incertidumbre en los objetivos del problema; luego, en vista de que trabajar con varios períodos de tiempo hace crecer notablemente el número de variables y restricciones del problema, preferimos evitar complicaciones que no son estrictamente necesarias. Esta idea es reforzada por el hecho de que el modelo simplificado que usaremos ya es en sí mismo un modelo de gran tamaño, que requiere de técnicas de optimización ad-hoc para ser aplicado a minas a cielo abierto reales. También asumimos que la planificación considera explotar un sólo mineral, y que existe una elección de proceso, i.e. todo bloque extraído puede ser o no procesado, lo cual es decidido por el tomador de decisiones.

Consideramos un yacimiento para minería a cielo abierto, aún no explotado, del cual tenemos su modelo de bloques. Denotamos por  $\mathcal{B}$  al conjunto de bloques de la mina, y  $\mathcal{P} \subset \mathcal{B} \times \mathcal{B}$  al conjunto de restricciones de precedencia, esto es,  $(a, b) \in \mathcal{P}$  ssi para extraer el bloque  $a$  antes debe extraerse el bloque  $b$ .

Cada bloque  $b \in \mathcal{B}$  tiene 1) un costo de extracción de  $v_b^e > 0$  unidades monetarias (UM), 2) un costo de procesamiento de  $v_b^p > 0$  UM, 3) una ley de mineral –que asumimos conocida– de  $\rho_b > 0$  unidades de mineral, y 4) un beneficio de  $w_b \geq 0$  UM por unidad de mineral. Así, por ejemplo, si un bloque  $b$  es extraído y procesado se percibe una ganancia económica de  $-v_b^e + (w_b \rho_b - v_b^p)$  UM, y si un bloque es extraído solamente entonces se percibe una ganancia de  $-v_b^e$ .

Todo plan de explotación debe cumplir las restricciones de precedencia de extracción establecidas por el conjunto de precedencias  $\mathcal{P} \subset \mathcal{B} \times \mathcal{B}$ . También, un bloque puede ser

procesado solamente si antes es extraído. Y además, se debe respetar las capacidades de planta disponibles: cada bloque  $b$  consume  $c_b^e > 0$  unidades de producción (UP) para ser extraído y existe un total de  $C^e > 0$  UP disponibles para la extracción; análogamente, cada bloque  $b$  consume  $c_b^p > 0$  UP para su procesamiento, del cual existe un total de  $C^p > 0$  UP para procesamiento disponibles. Naturalmente, asumimos que la capacidad de planta es limitada, esto es,  $\sum_{b \in \mathcal{B}} c_b^e > C^e$  y  $\sum_{b \in \mathcal{B}} c_b^p > C^p$ .

Las variables de decisión del problema son las que denotan si un bloque es extraído o no, y si es procesado o no. Para cada bloque  $b \in \mathcal{B}$  la decisión de su extracción está dada por

$$x_b^e = \begin{cases} 1 & \text{si el bloque } b \text{ es extraído,} \\ 0 & \text{si el bloque } b \text{ no es extraído,} \end{cases}$$

y la decisión de procesamiento por

$$x_b^p = \begin{cases} 1 & \text{si el bloque } b \text{ es procesado,} \\ 0 & \text{si el bloque } b \text{ no es procesado.} \end{cases}$$

Con esto el costo asociado a la explotación del bloque  $b$  es de  $-v_b^e x_b^e + (w_b \rho - v_b^p) x_b^p$  UM.

Así, el problema de planificación minera básico, con elección de proceso y en un único período, es

$$\begin{aligned} \text{máx} \quad & \sum_{b \in \mathcal{B}} [-v_b^e x_b^e + (w_b \rho - v_b^p) x_b^p] \\ \text{s.a.} \quad & x_a^e \leq x_b^e & \forall (a, b) \in \mathcal{P} \\ & x_b^p \leq x_b^e & \forall b \in \mathcal{B} \\ & \sum_{b \in \mathcal{B}} c_b^e x_b^e \leq C^e \\ & \sum_{b \in \mathcal{B}} c_b^p x_b^p \leq C^p \\ & x_b^e, x_b^p \in \{0, 1\} & \forall b \in \mathcal{B}. \end{aligned} \tag{2.1}$$

En lo sucesivo denominamos a este problema como el *problema nominal* dada una ley  $\rho \in \mathbb{R}_+^{\mathcal{B}}$ .

Para alivianar la notación en las futuras secciones, preferimos usar notación matricial y definir además el *conjunto de planes factibles* como

$$X := \left\{ (x^e, x^p) \in \{0, 1\}^{\mathcal{B}} : \begin{array}{l} x_a^e \leq x_b^e \quad \forall (a, b) \in \mathcal{P} \\ x^p \leq x^e \\ (c^e)^t x^e \leq C^e \\ (c^p)^t x^p \leq C^p \end{array} \right\} \tag{2.2}$$

donde  $x^e := (x_b^e)_{b \in \mathcal{B}}$ ,  $x^p := (x_b^p)_{b \in \mathcal{B}}$ ,  $c^e := (c_b^e)_{b \in \mathcal{B}}$  y  $c^p := (c_b^p)_{b \in \mathcal{B}}$ . Adicionalmente definimos el vector de leyes conjuntas  $\rho := (\rho_b)_{b \in \mathcal{B}}$ , los vectores de costos  $v^e := (v_b^e)_{b \in \mathcal{B}}$  y  $v^p := (v_b^p)_{b \in \mathcal{B}}$ , y la matriz de beneficios por procesamiento  $W^p \in \mathbb{R}^{\mathcal{B} \times \mathcal{B}}$  tal que  $W_{a,b}^p = w_b$  si  $a = b$  y 0 si no. Dado un plan  $(x^e, x^p)$  y una ley  $\rho \in \mathbb{R}_+^{\mathcal{B}}$ , definimos también la *función de beneficio* asociada como

$$\Pi(x^e, x^p, \rho) := -(v^e)^t x^e + (W^p \rho - v^p)^t x^p. \tag{2.3}$$

Así, dado un vector de leyes  $\rho \in \mathbb{R}_+^{\mathcal{B}}$  el problema nominal (2.1) asociado es equivalente a

$$\begin{aligned} \text{máx} \quad & \Pi(x^e, x^p, \rho) \\ \text{s.a.} \quad & (x^e, x^p) \in X. \end{aligned}$$

Hacemos notar que el parámetro  $\rho$ , que en las futuras secciones consideraremos que está sujeto a incertidumbre, sólo afecta la función objetivo del problema.

Cabe destacar que en muchos modelos mineros la decisión de procesar o no un bloque se hace según una regla preestablecida, e.g. que un bloque  $b$  se procesa ssi  $\rho_b > \rho_0$ , para algún valor  $\rho_0$  predefinido, o se procesa ssi  $\rho_b > \frac{v_b^e + v_b^p}{w_b}$ , i.e. ssi el beneficio por explotación  $-v_b^e + (w_b \rho_b - v_b^p)$  es positivo. Una ventaja que esto presenta es que se puede determinar a priori qué bloques serán procesados si es que son extraídos, con lo que redefiniendo varios parámetros del problema se pueden usar solamente  $|\mathcal{B}|$  variables de decisión para resolver el problema.

En efecto, definamos  $\mathcal{B}^P \subseteq \mathcal{B}$  como el conjunto de bloques que serán procesados si es que son extraídos, que son determinados por alguna regla preestablecida, por ejemplo todos los bloques  $b \in \mathcal{B}$  tales que  $\rho_b > \frac{v_b^e + v_b^p}{w_b}$ . Definamos para todo bloque  $b \in \mathcal{B}^P$   $u_b := -v_b^e + (w_b \rho - v_b^p)$  y  $\bar{c}_b^p := c_b^p$ ; y para todo  $b \in \mathcal{B} \setminus \mathcal{B}^P$  sea  $u_b := -v_b^e$  y  $\bar{c}_b^p := 0$ . Se tiene que el problema nominal (2.1) se puede aproximar por

$$\begin{aligned} \text{máx} \quad & u^t x \\ \text{s.a.} \quad & x_a \leq x_b \quad \forall (a, b) \in \mathcal{P} \\ & (c^e)^t x \leq C^e \\ & (\bar{c}^p)^t x \leq C^p \\ & x \in \{0, 1\}^{\mathcal{B}}. \end{aligned} \tag{2.4}$$

Pese a la ventaja obvia de disminuir a la mitad el número de variables de decisión, es fácil construir ejemplos en que la aproximación anterior es sub-óptima para el problema (2.1). Esto aún si la regla para el procesamiento sea que un bloque  $b$  es procesado al ser extraído ssi  $\rho > \frac{v_b^e + v_b^p}{w_b}$ .

Otra ventaja del problema nominal (2.1) por sobre su aproximación (2.4) es que, como veremos en las secciones posteriores, en el contexto con incertidumbre podremos usar la variable de extracción  $x^e$  como una variable de *primera etapa* y la variable de procesamiento  $x^p$  como una de *segunda etapa* o *de recourse*. Veremos que este recurso permitirá obtener beneficios considerables.

En lo sucesivo nos concentramos en el problema de planificación con elección de proceso (2.1).

## 2.2. Enfoques de optimización bajo riesgo

En lo sucesivo asumimos que no se conoce con exactitud el valor del vector de leyes conjuntas  $\rho \in \mathbb{R}_+^{\mathcal{B}}$ . Nos interesa estudiar la incertidumbre que esto introduce en el prob-

lema nominal (2.1). Como se mencionó en el Capítulo 1, en la práctica los altos costos de operación y la propiedad de ser un problema “de una única oportunidad”, pues la explotación puede ser realizada sólo una vez, hacen que el riesgo asociado a la incertidumbre sea considerable.

Nuestra hipótesis clave en este ámbito es que asumimos que el vector de leyes  $\rho$  sigue alguna distribución estocástica, que no necesariamente conocemos, soportada en un conjunto  $\Xi$  compacto, que también no necesariamente conocemos, y que sin embargo tenemos acceso a un generador de muestras iid de este parámetro. Decimos que una muestra  $\rho^1, \rho^2, \dots, \rho^N$  del vector aleatorio  $\rho$  es iid ssi todos los vectores  $\rho^i$  poseen la misma distribución de probabilidad y además cada  $\rho^i$  es distribuido independientemente de los demás vectores de la muestra.

La principal ventaja de esta hipótesis es que permite separar el análisis del efecto de la incertidumbre sobre el problema de planificación (2.1), del análisis de cómo se obtienen la distribución de probabilidad del vector de leyes  $\rho$ . Esto es conveniente pues no sólo asumir que el vector  $\rho$  sigue alguna de las distribuciones de probabilidad usuales es una hipótesis fuerte, si no que además determinar la distribución de probabilidad de  $\rho$  es en sí mismo un problema complejo y, según se infiere del trabajo de Vielma et al. [VEM09], lejos de estar cerrado. Finalmente, pero no menos importante, otra ventaja de asumir que se tiene una muestra iid de  $\rho$  es que en la actualidad existen técnicas geo-estadísticas hechas precisamente para obtener una muestra de este estilo para las leyes de mineral de los bloques (ver [EL06] y [VEM09]).

### 2.2.1. Modelo de minimización de VaR

El primer modelo que consideramos es el modelo de minimización de *Value-at-Risk* (VaR). Como se verá en esta sub-sección, el gran atractivo de este enfoque averso al riesgo es que su motivación es altamente intuitiva. También veremos que este modelo equivale a uno de optimización estocástica con restricciones en probabilidad. Si bien en trabajos anteriores, por ejemplo en [VEM09], ya se ha estudiado la aplicación de esta técnica a problemas de planificación minera, este estudio no ha sido exhaustivo pues no ha considerado detalles importantes, como convergencia de la aproximación del problema o el efectivo tratamiento del riesgo en la planificación. En virtud de esto elegimos trabajar con este modelo, con la intención de completar los análisis hechos en trabajos previos.

Primeramente definimos el VaR de una variable aleatoria (v.a.) de pérdidas. Diremos en lo sucesivo que una v.a. real  $L$  es *de pérdidas* si preferimos los valores menores de ella por sobre los valores altos. Por ejemplo,  $L$  puede representar el monto de una pérdida monetaria en una apuesta. Asimismo, una v.a. real  $G$  es *de beneficios* si  $-G$  es una v.a. de pérdidas.

Dada una v.a. de pérdidas  $L : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$  y un escalar  $\epsilon \in [0, 1)$ , que llamaremos *nivel de riesgo*, se define el Value-at-Risk con nivel de riesgo  $\epsilon$  de  $L$  como

$$\text{VaR}_\epsilon(L) := \min \{z \in \mathbb{R} : \mathbb{P}(L \leq z) \geq 1 - \epsilon\}, \quad (2.5)$$

donde la probabilidad es con respecto al espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$ . Notar que equivalentemente

$$\text{VaR}_\epsilon(L) = \text{mín} \{z \in \mathbb{R} : \mathbb{P}(-L + z < 0) \leq \epsilon\}.$$

En palabras simples,  $\text{VaR}_\epsilon(L)$  es el percentil  $100(1 - \epsilon)\%$  de  $L$ . La intuición gruesa detrás de este concepto es que con probabilidad  $1 - \epsilon$ —alta si  $\epsilon$  se escoge pequeño, como es usual— las pérdidas de  $L$  no excederán el valor  $\text{VaR}_\epsilon(L)$ . Otra interpretación gruesa es que  $\text{VaR}_\epsilon(L)$  es la peor—i.e. la mayor— posible pérdida de  $L$  cuando ignoramos sus  $100\epsilon\%$  peores pérdidas. De esto es claro que el VaR es un cuantificador o medidor del riesgo de una v.a. de pérdidas, lo cual motiva su uso en nuestro problema de planificación minera bajo incertidumbre.

En una sub-sección posterior, §2.2.2, formalizaremos la idea de *medida de riesgo* e introduciremos la noción de *medida de riesgo coherente* de Artzner et al. [ADEH99], que establece propiedades básicas que una medida de riesgo “razonable” debería cumplir. Ahí veremos que VaR no es una medida de riesgo coherente, pues para algunas v.a. de pérdidas  $L^1, L^2$  independientes se tiene

$$\text{VaR}_\epsilon(L^1 + L^2) > \text{VaR}_\epsilon(L^1) + \text{VaR}_\epsilon(L^2);$$

es decir, VaR considera que a veces diversificar en acciones independientes entre sí es estrictamente más riesgoso que no diversificar, lo cual atenta contra la intuición básica de riesgo. Pese a esta debilidad teórica del VaR como medida de riesgo, consideramos de todas maneras válido probar su desempeño práctico en nuestro problema de planificación minera.

Consideramos ahora un plan de extracción y procesamiento  $(x^e, x^p) \in X$  cualquiera, donde el conjunto de planes factibles  $X$  es como se definió en (2.2). Como el vector de leyes  $\rho$  es una v.a. entonces el beneficio  $\Pi((x^e, x^p), \rho)$  del plan  $(x^e, x^p)$  es también una v.a. Luego, mirando  $-\Pi((x^e, x^p), \rho)$  como una v.a. de pérdidas y eligiendo un nivel de riesgo  $\epsilon \in [0, 1)$ , el valor  $\text{VaR}_\epsilon[-\Pi((x^e, x^p), \rho)]$  es una cuantificación o valorización del riesgo del plan  $(x^e, x^p)$ . Naturalmente es deseable que el riesgo de nuestro plan  $(x^e, x^p)$  sea el menor posible. Eligiendo entonces un nivel de riesgo  $\epsilon \in [0, 1)$ , el modelo de *minimización de VaR* es

$$\text{mín}_{x^e, x^p} \text{VaR}_\epsilon[-\Pi((x^e, x^p), \rho)] \quad \text{s.a.} \quad (x^e, x^p) \in X. \quad (2.6)$$

Recordando la definición (2.5) de VaR podemos reformular el problema de minimización de VaR de la siguiente manera equivalente:

$$\begin{aligned} & \text{mín} \quad z \\ & \text{s.a.} \quad \mathbb{P}(\Pi((x^e, x^p), \rho) + z < 0) \leq \epsilon \\ & \quad (x^e, x^p) \in X, \quad z \in \mathbb{R}. \end{aligned}$$

En palabras, la solución de este problema nos da un plan que tiene el mejor—el más bajo—percentil  $100(1 - \epsilon)\%$  de pérdidas.

Sin embargo, este problema de optimización adolece de varias inconveniencias. Primeramente, no es realista asumir que se conoce la distribución de  $\rho$ , con lo que no se puede calcular a exactitud la probabilidad conjunta  $\mathbb{P}(\Pi((x^e, x^p), \rho) + z < 0)$  en el vector aleatorio  $\rho$ . Más aún, aún sabiendo la distribución de  $\rho$  el cálculo exacto de dicha probabilidad involucra integración en varias dimensiones, lo cual puede ser impracticable de incorporar en una restricción de un problema de optimización. Otra dificultad no menor es que, aún conociendo con exactitud la distribución de  $\rho$ , el problema (2.6) puede no ser convexo, pues el conjunto

$$\{(x^e, x^p, z) \in \mathbb{R}^B \times \mathbb{R}^B \times \mathbb{R} : \mathbb{P}(\Pi((x^e, x^p), \rho) + z < 0) \leq \epsilon\}$$

puede ser no convexo.

Para sortear al mismo tiempo todas estas dificultades hacemos uso de la técnica *Sample Average Approximation*, que consiste en aproximar la restricción de probabilidad usando una muestra iid del parámetro incierto. Esto es, asumimos que tenemos una muestra iid  $\rho^1, \dots, \rho^N$  de  $\rho$  y utilizamos la aproximación

$$\mathbb{P}(\Pi((x^e, x^p), \rho) + z < 0) \approx \frac{1}{N} \sum_{k=1}^N \mathbf{1}_{(-\infty, 0)}(\Pi((x^e, x^p), \rho^k) + z),$$

donde para  $I \subseteq \mathbb{R}$ ,  $\mathbf{1}_I(x)$  es la función indicatriz de  $I$ :

$$\mathbf{1}_I(x) := \begin{cases} 1 & \text{si } x \in I \\ 0 & \text{si no.} \end{cases}$$

La ventaja de realizar esta aproximación es que el conjunto

$$\left\{ (x^e, x^p, z) \in \mathbb{R}^B \times \mathbb{R}^B \times \mathbb{R}^B : \frac{1}{N} \sum_{k=1}^N \mathbf{1}_{(-\infty, 0)}(\Pi((x^e, x^p), \rho^k) + z) \leq \epsilon \right\}$$

sí es convexo, y también se puede reformular linealmente como

$$\left\{ (x^e, x^p, z) : \exists \zeta \in \{0, 1\}^N \text{ tal que } \begin{cases} \Pi(x^e, x^p, \rho^k) + z + \zeta_k M^k \geq 0 & k = 1, \dots, N \\ \sum_{k=1}^N \zeta_k \leq \lfloor \epsilon N \rfloor \end{cases} \right\}$$

donde  $M^k$ ,  $k = 1, \dots, N$ , son parámetros “gran-M” suficientemente grandes, e.g.

$$M^k := \max_{i=1, \dots, N} \max_{x \in X} \Pi(x, \rho^i) - \min_{x \in X} \Pi(x, \rho^k) \quad \forall k = 1, \dots, N.$$

Hemos obtenido así la *aproximación SAA* del problema de minimización de VaR, dada una muestra iid  $\rho^1, \dots, \rho^N$  de  $\rho$  y un nivel de riesgo  $\epsilon \in [0, 1)$ :

$$\begin{aligned} & \text{mín } z \\ & \text{s.a. } \begin{cases} \Pi(x^e, x^p, \rho^k) + z + \zeta_k M^k \geq 0 & k = 1, \dots, N \\ \sum_{k=1}^N \zeta_k \leq \lfloor \epsilon N \rfloor \\ \zeta \in \{0, 1\}^N \\ (x^e, x^p) \in X, \quad z \in \mathbb{R}. \end{cases} \end{aligned} \quad (2.7)$$

El plan  $(x^{e*}, x^{p*})$  solución de este problema será un candidato a solución óptima del problema “verdadero” (2.6); i.e. será un plan candidato a minimizar el VaR con nivel de riesgo  $\epsilon$  del beneficio de la explotación minera.

Destacamos que la aproximación SAA (2.7) es un problema lineal, con variables binarias y continuas, y convexo. No obstante, esta formulación con parámetros gran-M es sumamente ineficiente desde un punto de vista computacional. Esto es, usar una formulación como ésta causa que, al tratar de resolver una instancia de este problema con un *solver* computacional típico de *Branch & Bound & Cut*, e.g. CPLEX, sepamos a priori que con alta probabilidad la resolución computacional demorará una gran cantidad de tiempo, que posiblemente exceda con creces un tiempo de cómputo razonable. El argumento que explica esto es clásico del área de *programación entera*, y es que la relajación lineal del conjunto factible de la formulación (2.7) puede ser, debido a los parámetros  $M^k$ , muy poco ajustada a la envoltura convexa del conjunto factible de (2.7).

### Contraparte con recourse

Consideramos ahora un ligero cambio de paradigma en el planteamiento del problema de optimización. Explotamos el hecho de que no conocemos a priori el verdadero vector de leyes  $\rho$  y nos preguntamos si podemos postergar alguna decisión para después de que se haya “revelado” la verdadera ley  $\rho$ . La elección natural es tomar la elección del plan de extracción  $x^e$  como una variable a decidir “aquí y ahora”, i.e. sin esperar a que se revelen las verdaderas leyes  $\rho$ , y tomar la elección de qué bloques procesar,  $x^p$ , como una variable *ajustable* o *de recourse* que se ajusta a la realización final de  $\rho$ . Esta nomenclatura es propia del área de optimización robusta (ver [BTGGN04]). Sin embargo esta idea viene originalmente de optimización estocástica multi-etapa, donde la terminología sería que la decisión  $x^e$  es una variable *de primera etapa* y la decisión  $x^p$  *de segunda etapa*.

Para introducir el problema de planificación minera con recourse supongamos momentáneamente que efectivamente una vez que elegimos el plan de extracción  $x^e$  se revela la verdadera ley  $\rho$ , y podemos tomar la decisión de recourse  $x^p$ . Bajo este esquema el problema de optimizar el beneficio de la explotación es:

$$\begin{aligned} \text{máx}_{x^e} \quad & -(v^e)^t x^e + \mathcal{Q}(x^e, \rho) \\ \text{s.a.} \quad & x_a^e \leq x_b^e \quad \forall (a, b) \in \mathcal{P} \\ & (c^e)^t x^e \leq B^e \\ & x^e \in \{0, 1\}^{\mathcal{B}} \end{aligned} \tag{2.8}$$

donde

$$\begin{aligned} \mathcal{Q}(x^e, \rho) := \text{máx}_{x^p} \quad & (W^p \rho - v^p)^t x^p \\ \text{s.a.} \quad & x^p \leq x^e \\ & (c^p)^t x^p \leq B^p \\ & x^p \in \{0, 1\}^{\mathcal{B}}. \end{aligned}$$

Notar que si  $\rho$  es determinista, i.e. se conoce de antemano el verdadero valor de las leyes, entonces este problema es equivalente al problema (2.1).

Si este esquema para la realización de  $\rho$  fuera realista entonces obtendríamos que la decisión óptima  $x^p$  es en realidad una función del espacio de todas las posibles leyes,  $\Xi^1$ , al espacio de todos los planes de procesamiento factibles. No obstante, asumimos al inicio de esta sección que no necesariamente conoceríamos la distribución de  $\rho$ , y en vez sólo tendríamos acceso a un generador de muestras iid de  $\rho$ . Con esto el enfoque anterior sería impracticable.

Una opción posible es modelar el problema bajo riesgo como

$$\begin{aligned} & \text{máx}_{x^e} \quad -(v^e)^t x^e + \mathbb{E}[\mathcal{Q}(x^e, \rho)] \\ & \text{s.a.} \quad x_a^e \leq x_b^e \quad \forall (a, b) \in \mathcal{P} \\ & \quad (c^e)^t x^e \leq B^e \\ & \quad x^e \in \{0, 1\}^{\mathcal{B}} \end{aligned}$$

y resolver este problema estocástico usando una aproximación SAA, i.e. aproximarlo usando una muestra iid de  $\rho$ . Sin embargo habíamos establecido que el tratamiento adecuado para el problema de planificación bajo incertidumbre debía ser averso al riesgo, y minimizar esperanza es, por excelencia, un enfoque neutro al riesgo.

Lo que proponemos es aplicar este esquema, de usar variables con recourse, a los modelos aversos al riesgo que usamos en este trabajo. Para introducir el modelo de minimización de VaR con recourse, y también con el fin de alivianar la formulación, definimos primero la función de beneficio con recourse  $\Pi^R$  como

$$\begin{aligned} \Pi^R(x^e, \rho) := & -(v^e)^t x^e + \text{máx}_{x^p} \quad (W^p \rho - v^p)^t x^p \\ & \text{s.a.} \quad x^p \leq x^e \\ & \quad (c^p)^t x^p \leq B^p \\ & \quad x^p \in \{0, 1\}^{\mathcal{B}} \end{aligned} \quad (2.9)$$

y también definimos el conjunto  $X^R$  de planes (únicamente) de extracción factibles como:

$$X^R := \left\{ x^e \in \{0, 1\}^{\mathcal{B}} : \begin{array}{l} x_a^e \leq x_b^e \quad \forall (a, b) \in \mathcal{P} \\ (c^e)^t x^e \leq B^e \end{array} \right\}. \quad (2.10)$$

Dado un nivel de riesgo  $\epsilon \in [0, 1)$ , el problema de *minimización de VaR con recourse* es

$$\begin{aligned} & \text{mín}_{x^e} \quad z \\ & \text{s.a.} \quad \mathbb{P}(\Pi^R(x^e, \rho) + z < 0) \leq \epsilon \\ & \quad x^e \in X^R, \quad z \in \mathbb{R}. \end{aligned} \quad (2.11)$$

Como en el caso del problema de minimización de VaR sin recourse, en este problema también tenemos dificultades con la integración multidimensional que requiere la restricción en probabilidad, y también esta restricción puede hacer que el problema no sea convexo. Adicionalmente, es un problema cómo se tratará la variable con recourse  $x^p$ , que

<sup>1</sup>Recordamos al lector que asumimos que el parámetro  $\rho$  está soportado en el conjunto  $\Xi$

en optimalidad, como se mencionó antes, no es un único vector, si no que es una función sobre el espacio  $\Xi$  de todas las posibles realizaciones de  $\rho$ .

Nuevamente, como en el problema de minimización de VaR sin recourse, sorteamos al mismo tiempo todas estas dificultades para el problema (2.11) haciendo uso de su aproximación SAA. Esto es, tomamos una muestra iid  $\rho^1, \dots, \rho^N$  de  $\rho$  y aproximamos la probabilidad por la probabilidad frecuentista en la muestra. Es decir, usamos en el problema (2.11) la aproximación

$$\mathbb{P}(\Pi^R(x^e, \rho) + z < 0) \approx \frac{1}{N} \sum_{k=1}^N \mathbf{1}_{(-\infty, 0)}(\Pi^R(x^e, \rho^k) + z).$$

No es difícil verificar que la siguiente es una formulación lineal para la *aproximación SAA del problema de minimización de VaR con recourse*:

$$\begin{aligned} & \text{mín } z \\ & \text{s.a. } \Pi(x^e, x^{p,k}, \rho^k) + z + \zeta_k M^k \geq 0 \quad k = 1, \dots, N \\ & \quad \sum_{k=1}^N \zeta_k \leq \lfloor \epsilon N \rfloor \\ & \quad \zeta \in \{0, 1\}^N \\ & \quad (x^e, x^{p,k}) \in X \quad k = 1, \dots, N \\ & \quad z \in \mathbb{R} \end{aligned} \tag{2.12}$$

donde  $\Pi$  y  $X$  son del modelo sin recourse, i.e. como se definieron en (2.3) y (2.2) respectivamente.

Hacemos notar que la única diferencia entre las aproximaciones SAA para los problemas de VaR con y sin recourse es que, en el caso con recourse, tenemos una decisión de procesamiento,  $x^{p,k}$ , por cada escenario  $\rho^k$ . Esto está en perfecta sintonía con la afirmación que hicimos antes, de que en el modelo con recourse la variable con recourse óptima es en realidad una función de todas las realizaciones del parámetro incierto. En efecto, en el caso de la aproximación SAA lo que estamos haciendo es reducir el espacio para  $\rho$  desde su verdadero soporte  $\Xi$  a la muestra  $\{\rho^1, \dots, \rho^N\}$ ; luego la decisión óptima del problema aproximado es una función de la muestra  $\{\rho^1, \dots, \rho^N\}$  al espacio de los planes de procesamiento factibles, i.e. es un plan  $x^{p,k}$  por cada escenario  $\rho^k$ .

### 2.2.2. Modelo de minimización de CVaR

El segundo modelo que consideramos en nuestro estudio es el de minimización de *Conditional Value-at-Risk* (CVaR). Veremos que el CVaR es una medida de riesgo inspirada en el VaR, cuya finalidad es cuantificar el riesgo de una v.a. de pérdidas. Sin embargo, a diferencia del VaR, el CVaR poseerá varias propiedades teóricas deseables en una medida de riesgo, y más aún, será la componente fundamental de la familia de *medidas de riesgo de distorsión*. El modelo de minimización de CVaR que propondremos en esta sub-sección, veremos, consiste en elegir el plan que minimice el CVaR de las pérdidas.

Primeramente, definamos el CVaR de una v.a. de pérdidas<sup>2</sup>. Consideramos un espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$  y una v.a. de pérdidas  $L$  sobre dicho espacio. Sea  $\epsilon \in (0, 1]$ , y supongamos momentáneamente que  $L$  no tiene átomos de probabilidad, i.e.  $\mathbb{P}(L = p) = 0$  para todo  $p \in \mathbb{R}$ . En este caso, el CVaR con nivel de riesgo  $\epsilon$  de  $L$  se define como

$$\text{CVaR}_\epsilon(L) := \mathbb{E}[L | L \geq \text{VaR}_\epsilon(L)], \quad (2.13)$$

donde, recordemos, el VaR con nivel de riesgo  $\epsilon$  de  $L$  es  $\text{VaR}_\epsilon(L) := \min\{z : \mathbb{P}(L \leq z) \geq 1 - \epsilon\}$ . Notar que si  $\epsilon = 1$  entonces  $\text{CVaR}_\epsilon(L) = \mathbb{E}(L)$ .

Dijimos en §2.2.1 que, de manera gruesa,  $\text{VaR}_\epsilon(L)$  es el peor –el más alto– valor que puede tener  $L$  cuando desechamos sus  $100\epsilon\%$  peores pérdidas. Luego, intuitivamente,  $\text{CVaR}_\epsilon(L)$  es el promedio de las  $100\epsilon\%$  peores –mayores– pérdidas posibles de  $L$ .

Rockafellar et al. [RU00] muestran que, para distribuciones sin átomos, el CVaR de una posición  $L$  se puede obtener minimizando una función real, convexa y continuamente diferenciable, lo cual es una ventaja considerable para poder optimizar el CVaR de una variable afecta a riesgo. Para conservar esta propiedad al pasar a distribuciones generales, posteriormente en [RU02] también Rockafellar et al. generalizan la definición de CVaR de la siguiente manera: dada una v.a. de pérdidas  $L : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$  y  $\epsilon \in (0, 1]$ , se define el CVaR con nivel de riesgo  $\epsilon$  de  $L$  como

$$\text{CVaR}_\epsilon(L) := (1 - \epsilon^-) \text{VaR}_\epsilon(L) + \frac{\epsilon^-}{\epsilon} \text{CVaR}_\epsilon^+(L), \quad (2.14)$$

donde  $\epsilon^- := \mathbb{P}(L > \text{VaR}_\epsilon(L))$  y  $\text{CVaR}_\epsilon^+(L) := \mathbb{E}[L | L > \text{VaR}_\epsilon(L)]$ . Se verifica fácilmente que si  $L$  no tiene átomos entonces esta definición es equivalente a la definición (2.13). Además, también se mantiene que  $\text{CVaR}_\epsilon(L) = \mathbb{E}(L)$  si  $\epsilon = 1$ .

El resultado principal de [RU02], como se insinuó antes, es que en cualquier caso – para v.a. de pérdidas con o sin átomos– el CVaR con nivel de riesgo  $\epsilon$  de una v.a.  $L$  se puede escribir como

$$\text{CVaR}_\epsilon(L) = \min_{t \in \mathbb{R}} \left[ t + \frac{1}{\epsilon} \mathbb{E}[(L - t)^+] \right], \quad (2.15)$$

donde  $\alpha^+ := \alpha \mathbf{1}_{\{\alpha \geq 0\}}$  para todo  $\alpha \in \mathbb{R}$ . En ese trabajo se muestra que el lado derecho de (2.15) es convexo y continuamente diferenciable en  $t$ , lo cual es una ventaja considerable si se desea optimizar el CVaR de las pérdidas. Esto es, supongamos que  $L$  depende continuamente en una decisión  $x \in X$  y es medible en un parámetro incierto  $\omega \in \Omega$ , i.e.  $L = L(x, \omega)$  es de Carathéodory. Si  $L$  es convexa en  $x$  entonces  $t + \frac{1}{\epsilon} \mathbb{E}[(L(x, \omega) - t)^+]$  es convexa en  $(x, t)$  y se tiene

$$\min_{x \in X} \text{CVaR}_\epsilon(L(x, \omega)) = \min_{x \in X, t \in \mathbb{R}} \left\{ t + \frac{1}{\epsilon} \mathbb{E}[(L(x, \omega) - t)^+] \right\}.$$

<sup>2</sup>Recordamos que  $L : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$  es una v.a. de pérdidas si preferimos sus valores bajos por sobre los altos.

Para más detalles referimos al lector a [RU02], y para un enfoque alternativo para llegar a la misma conclusión recomendamos [SDR09, §6.2.4].

Si bien de lo expuesto se vislumbra claramente cuál será el modelo de minimización de CVaR que aplicaremos a nuestro problema de planificación minera, preferimos introducir previamente las nociones básicas de teoría de riesgo. Esto no sólo servirá para ahondar en las buenas propiedades teóricas que posee el CVaR –las cuales carece el VaR– si no que además nos será útil para discutir las propiedades del modelo MCH, el cual presentaremos posteriormente en §2.2.3. La teoría de riesgo que sigue corresponde a la iniciada por el trabajo de Artzner et al. [ADEH99], y la idea fundamental es cuantificar el riesgo de un costo que está sujeto a variabilidad.

### Introducción a teoría de riesgo

Consideremos un espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$ . Sea  $\mathcal{Z} \subseteq L^\infty(\Omega, \mathcal{F}, \mathbb{P})$  un conjunto de v.a. de pérdidas.

**Definición 2.2.1.** *Sea una función  $\mu : \mathcal{Z} \rightarrow \mathbb{R}$ , consideremos las siguientes propiedades para  $\mu$ , para todo  $X, Y \in \mathcal{Z}$ :*

1. *Monotonía:* Si  $X \leq Y$ , esto es  $X(\omega) \leq Y(\omega)$  para todo  $\omega \in \Omega$ , entonces  $\mu(X) \leq \mu(Y)$ .
2. *Equivarianza a traslación:* Si  $c \in \mathbb{R}$  entonces  $\mu(X - c) = \mu(X) - c$ .

Diremos que  $\mu$  es medida de riesgo si satisface 1. y 2.

3. *Convexidad:* Si  $\lambda \in [0, 1]$  entonces  $\mu(\lambda X + (1 - \lambda)Y) \leq \lambda\mu(X) + (1 - \lambda)\mu(Y)$ .
4. *Homogeneidad positiva:* Si  $\lambda \geq 0$ , entonces  $\mu(\lambda X) = \lambda\mu(X)$ .

Se dirá que  $\mu$  es medida de riesgo coherente si cumple 1.–4.

5. *Comonotonía:* Si  $X$  e  $Y$  son comonotónicas, i.e.  $(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \geq 0$  para todo  $\omega, \omega' \in \Omega$ , entonces  $\mu(X + Y) = \mu(X) + \mu(Y)$ .
6. *Ley-invarianza:* Si  $X$  e  $Y$  tienen la misma distribución bajo  $\mathbb{P}$  entonces  $\mu(X) = \mu(Y)$ .

Por último, se dirá que  $\mu$  es medida de riesgo de distorsión, o espectral, si satisface 1.–6.

Para  $\mu$  medida de riesgo y  $X$  v.a. de pérdidas,  $\mu(X)$  puede interpretarse como la mínima recompensa, o disminución de pérdidas, necesaria para hacer “aceptable” a  $X$ . Por convención, diremos que  $X$  es “aceptable” si  $\mu(X) \leq 0$ . Así, para una medida de riesgo la monotonía nos asegura que si entre dos posiciones riesgosas la primera, digamos,

proporciona consistentemente menores pérdidas que la segunda, entonces la primera es menos riesgosa. Asimismo, la interpretación a equivarianza a traslación es que si disminuimos nuestras pérdidas en una cantidad  $c$  garantizada, entonces nuestra recompensa por riesgo debe disminuir en la misma cantidad.

Una medida de riesgo coherente corresponde a una medida de riesgo con una estructura particular para lidiar con la diversificación de riesgo. Fundamentalmente, convexidad exige que diversificar posiciones nunca aumentará el riesgo. Adicionalmente, homogeneidad positiva exige que el riesgo escale linealmente con el tamaño de la posición. Hacemos notar también que, cuando se tiene homogeneidad positiva, convexidad equivale a *subaditividad*:

$$\mu(X + Y) \leq \mu(X) + \mu(Y) \quad \text{para todo } X, Y \in \mathcal{Z}.$$

Las medidas de riesgo coherentes proveen ventajas adicionales en optimización. Primeramente, debido a la propiedad de convexidad, optimizar el riesgo de una posición usando una medida de riesgo coherente no es un impedimento para que el problema sea convexo. Y segundo, existe una conexión con optimización robusta, pues una medida de riesgo es coherente si y sólo si puede ser expresada como el peor valor esperado sobre una familia de distribuciones, como muestra el siguiente resultado:

**Teorema 2.2.2** (Teorema de representación de medidas de riesgo coherentes). *Una función  $\mu : \mathcal{Z} \rightarrow \mathbb{R}$  es una medida de riesgo coherente ssi existe una familia  $\mathcal{Q}$  de medidas de probabilidad sobre  $(\Omega, \mathcal{F})$  satisfaciendo*

- $\mathbb{Q} \ll \mathbb{P}$  para todo  $\mathbb{Q} \in \mathcal{Q}$ , i.e.  $\mathbb{Q}(A) = 0$  para todo  $A \in \mathcal{F}$  tal que  $\mathbb{P}(A) = 0$ ; y
- para todo  $\mathbb{Q} \in \mathcal{Q}$  y todo  $X \in \mathcal{Z}$  se tiene

$$\mu(X) = \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[X],$$

donde  $\mathbb{E}_{\mathbb{Q}}$  es la esperanza según la medida de probabilidad  $\mathbb{Q}$ .

Para una demostración detallada de este resultado referimos al lector a la demostración del Teorema 6.4 de [SDR09] y la discusión que le sigue.

Para el caso de medidas de riesgo de distorsión, la propiedad de comonotonía exige que si dos posiciones “se mueven juntas” entonces diversificar no disminuye el riesgo conjunto. En otras palabras, cuando se tiene convexidad, comonotonía exige que el riesgo conjunto de dos posiciones riesgosas que se mueven juntas sea el mayor posible. La propiedad de ley-invarianza, por su parte, es una propiedad básica para poder estimar consistentemente la medida de riesgo de una posición a partir de datos. Para ver esto, consideremos  $X$  e  $Y$  con la misma distribución y  $\mu$  medida de riesgo no ley-invariante tal que  $\mu(X) \neq \mu(Y)$ ; consideremos también muestras iid  $\{X_i\}_{i=1, \dots, N}$  y  $\{Y_i\}_{i=1, \dots, N}$  de  $X$  e  $Y$ , respectivamente, y un estimador  $\hat{\mu}_N : \mathbb{R}^N \rightarrow \mathbb{R}$  de  $\mu$  que sólo depende de la distribución de la muestra. Al hacer  $N \rightarrow \infty$  se tiene que las muestras son asintóticamente idénticas y

entonces  $\lim_N \widehat{\mu}_N(X_1, \dots, X_N) = \lim_N \widehat{\mu}_N(Y_1, \dots, Y_N)$ ; sin embargo,  $\mu(X) \neq \mu(Y)$ , y entonces al menos alguno de los dos estimadores debe ser incorrecto asintóticamente.

A continuación damos algunos ejemplos de distintos tipos de medidas de riesgo:

- **Value-at-Risk:** el VaR con nivel de riesgo  $\epsilon \in [0, 1]$ ,  $\text{VaR}_\epsilon(L) = \min\{z \in \mathbb{R} : \mathbb{P}(L \leq z) \geq 1 - \epsilon\}$  es una medida de riesgo, sin embargo no satisface subaditividad –y entonces convexidad tampoco–, por lo que no es una medida de riesgo coherente.
- **Momentos uni-laterales:** las medidas de riesgo

$$\mu_{r,\alpha}(L) = \mathbb{E}[X] + \alpha \sigma_{r,-}(L),$$

donde  $r \geq 1$ ,  $\alpha \in [0, 1]$  y

$$\sigma_{r,-}(L) := [\mathbb{E}((L - \mathbb{E}[L])^+)^r]^{1/r},$$

son medidas de riesgo coherentes (ver [BB09]).

- **Conditional Value-at-Risk:** no es difícil verificar que el CVaR con nivel de riesgo  $\epsilon \in (0, 1]$ ,  $\text{CVaR}_\epsilon(L) := \min_{t \in \mathbb{R}} [t + \frac{1}{\epsilon} \mathbb{E}(L - t)^+]$ , es una medida de riesgo de distorsión.

Para una lista extensa de distintas medidas de riesgo referimos al lector a [SDR09, §6.3.2].

Por último, mostramos el siguiente resultado que establece que, en espacios de probabilidad discretos y equiprobables, toda medida de riesgo de distorsión es una combinación lineal convexa de medidas de riesgo CVaR.

**Teorema 2.2.3.** *Sea  $(\Omega_N, \mathcal{P}, \mathbb{P}_N)$  un espacio de probabilidad finito y equiprobable, i.e.  $\Omega_N = \{\omega^1, \dots, \omega^N\}$ ,  $\mathcal{P}$  es el conjunto de las partes de  $\Omega_N$  y  $\mathbb{P}_N(\{\omega^i\}) = 1/N$  para todo  $i$ . Se tiene que  $\mu$  medida de riesgo es medida de riesgo de distorsión ssi existe función  $\nu : [0, 1] \rightarrow [0, 1]$  satisfaciendo  $\int_{\alpha=0}^1 \nu(d\alpha) = 1$  y*

$$\mu(X) = \int_{\alpha=0}^1 \text{CVaR}_\alpha(X) \nu(d\alpha).$$

Para la demostración de este resultado remitimos al lector a la demostración del Lema 4.1 de [BB09]. Este resultado es válido también para distribuciones sin átomos en espacios de probabilidad generales, para lo cual referimos al lector también a la demostración del lema ya citado de la referencia anterior. Así, en estos casos podemos decir que CVaR es la componente fundamental de las medidas de riesgo de distorsión.

En nuestro caso, donde asumimos que sólo tenemos una muestra iid del parámetro incierto  $\rho$ , es de especial relevancia el caso de espacio de probabilidad discreto equiprobable. Esto, pues nuestro foco es aproximar el espacio de probabilidad “real” de  $\rho$  haciendo uso de un sample iid de  $\rho$ , y estimar en este espacio la medida de riesgo de distorsión a estudiar. Hacemos notar que esto es exactamente lo que se hizo para el caso de minimización de VaR, y es el enfoque natural a seguir en el modelo de minimización de CVaR, que presentamos en la siguiente sub-sección.

### Modelo sin recourse

Sea  $(x^e, x^p) \in X$  un plan de extracción y procesamiento factible. Asumiendo que el vector de leyes conjuntas  $\rho$ —el parámetro incierto— pertenece a un espacio de probabilidad  $(\Xi, \mathcal{F}, \mathbb{P})$ , se tiene que  $-\Pi((x^e, x^p), \rho)$  es una v.a. de pérdidas. Dado un nivel de riesgo  $\epsilon \in (0, 1]$ , es una cuantificación del riesgo del plan  $(x^e, x^p)$

$$\text{CVaR}_\epsilon [-\Pi((x^e, x^p), \rho)] = \min_{t \in \mathbb{R}} \left\{ t + \frac{1}{\epsilon} \mathbb{E}[-\Pi((x^e, x^p), \rho) - t]^+ \right\}.$$

El modelo que proponemos es escoger el plan que minimice el riesgo cuantificado con el CVaR.

Formalmente, dado un nivel de riesgo  $\epsilon \in (0, 1]$ , el *modelo de minimización de CVaR* es

$$\min_x \text{CVaR}_\epsilon [-\Pi((x^e, x^p), \rho)] \quad \text{s.a.} \quad x \in X. \quad (2.16)$$

Se verifica fácilmente que  $\Pi((x^e, x^p), \rho)$  es lineal, y entonces convexa, en  $(x^e, x^p)$ , y también que es lineal afín en  $\rho$ , con lo que  $\Pi$  es Carathéodory si la  $\sigma$ -álgebra asociada a  $\rho$  es boreliana. Luego, como se dijo en el inicio de esta sub-sección, la función  $t + \frac{1}{\epsilon} \mathbb{E}[-\Pi((x^e, x^p), \rho) - t]^+$  es convexa en  $(x^e, x^p, t)$ . Concluimos que si  $\mathcal{F}$ , la  $\sigma$ -álgebra asociada a  $\rho$ , es boreliana entonces se tiene la siguiente formulación equivalente para el problema (2.16):

$$\min_{x,t} t + \frac{1}{\epsilon} \mathbb{E}[-\Pi((x^e, x^p), \rho) - t]^+ \quad \text{s.a.} \quad x \in X, t \in \mathbb{R}. \quad (2.17)$$

No obstante, como sucedió en el modelo de minimización de VaR en §2.2.1, esta formulación requiere el cálculo de una esperanza, la cual involucra integrar en varias dimensiones. Adicionalmente, dijimos en §2.2 que era irreal en nuestro problema asumir conocida la distribución conjunta del vector de leyes  $\rho \in \mathbb{R}_+^B$ , lo cual es un escollo extra para el cálculo de la esperanza en (2.17). Para abordar este problema nuevamente recurrimos a la aproximación SAA de este problema. Esto es, tomamos una muestra iid  $\rho^1, \dots, \rho^N$  de  $\rho$  y aproximamos

$$\mathbb{E}[-\Pi((x^e, x^p), \rho) - t]^+ \approx \frac{1}{N} \sum_{k=1}^N [-\Pi((x^e, x^p), \rho^k) - t]^+.$$

Así, la *aproximación SAA del modelo de minimización de CVaR* es

$$\min_{x,t} t + \frac{1}{\epsilon N} \sum_{k=1}^N [-\Pi((x^e, x^p), \rho^k) - t]^+ \quad \text{s.a.} \quad x \in X, t \in \mathbb{R}, \quad (2.18)$$

donde  $\rho^1, \dots, \rho^N$  es una muestra iid de  $\rho$ . Más aún, este problema se puede reformular como un problema lineal a variables binarias y continuas de la siguiente manera:

$$\begin{aligned}
\text{mín} \quad & t + \frac{1}{\epsilon N} \sum_{k=1}^N \zeta^k \\
\text{s.a.} \quad & \zeta^k \geq -\Pi((x^e, x^p), \rho^k) - t \quad \forall k = 1, \dots, N \\
& \zeta^k \geq 0 \quad \forall k = 1, \dots, N \\
& (x^e, x^p) \in X, \quad t \in \mathbb{R}.
\end{aligned} \tag{2.19}$$

En síntesis, el modelo presentado busca escoger un plan averso al riesgo al elegir el plan que minimiza su CVaR, una medida de riesgo con numerosas propiedades teóricas. El problema “verdadero” (2.16) no es tratable pues requiere el cálculo de una esperanza usando una distribución que resulta poco realista asumir conocida. Por ello, nos remitimos a aproximar el problema verdadero usando la aproximación SAA (2.18), el cual requiere conocer solamente una muestra iid del parámetro incierto  $\rho$ . A continuación adaptamos este modelo al cambio de paradigma presentado en §2.2.1, en el cual asumimos que la decisión  $x^p$  es una variable de segunda etapa.

### Contraparte con recourse

Buscamos ahora aplicar el modelo de minimización de CVaR al esquema de decisión multi-etapa presentado en §2.2.1. Bajo este paradigma alternativo asumimos que debemos elegir “aquí y ahora” la decisión  $x^e$  de qué bloques extraer, posteriormente, después de realizarse la extracción, podemos ver cuál era la verdadera ley  $\rho$  de los bloques, y sólo entonces debemos decidir qué bloques procesar. Así, la decisión  $x^e$  es una variable *de primera etapa* o “aquí y ahora”, y la decisión  $x^p$  es una variable *de segunda etapa* o *de recourse*.

Para incorporar esta regla de decisión a nuestro modelo basta redefinir apropiadamente la función de beneficio  $\Pi$ , haciendo que la decisión  $x^p$  dependa de la decisión de extracción  $x^e$  y la ley  $\rho$ . Esto ya fue hecho en §2.2.1, donde, recordemos, se definió el *beneficio con recourse*  $\Pi^R$  en (2.9) como

$$\begin{aligned}
\Pi^R(x^e, \rho) := & -(v^e)^t x^e + \max_{x^p} (W^p \rho - v^p)^t x^p \\
\text{s.a.} \quad & x^p \leq x^e \\
& (c^p)^t x^p \leq B^p \\
& x^p \in \{0, 1\}^B.
\end{aligned}$$

Usaremos a continuación también la notación (2.10) para el conjunto de planes de extracción factibles:

$$X^R := \left\{ x^e \in \{0, 1\}^B : \begin{array}{l} x_a^e \leq x_b^e \quad \forall (a, b) \in \mathcal{P} \\ (c^e)^t x^e \leq B^e \end{array} \right\}.$$

Dado un nivel de riesgo  $\epsilon \in (0, 1]$ , el *modelo de minimización de CVaR con recourse* es

$$\text{mín}_{x^e} \quad \text{CVaR}_\epsilon [-\Pi^R(x^e, \rho)] \quad \text{s.a.} \quad x^e \in X^R. \tag{2.20}$$

En la Propiedad 3.3.1, en §3.3 del siguiente capítulo, mostraremos que  $-\Pi^R(\cdot, \rho)$  se puede extender convexamente desde  $\{0, 1\}^B$  a  $[0, 1]^B$ , y además esta extensión es Carathéodory si la  $\sigma$ -álgebra asociada a  $\rho$  es borealiana. Asumiendo esto último, el cual es un supuesto razonable, podemos reformular equivalentemente este problema como

$$\min_{x^e, t} \quad t + \frac{1}{\epsilon} \mathbb{E} [-\Pi^R(x^e, \rho) - t]^+ \quad \text{s.a.} \quad x^e \in X^R, t \in \mathbb{R}. \quad (2.21)$$

Nuevamente, la esperanza en la formulación anterior presenta una dificultad mayor en su tratabilidad. Usamos de nuevo entonces el método de aproximación *Sample Average Approximation* para sortear este problema: tomamos una muestra iid  $\rho^1, \dots, \rho^N$  del vector de leyes  $\rho$  y aproximamos la esperanza por el promedio dentro de la muestra. La *aproximación SAA del modelo de minimización de CVaR con recourse* queda

$$\min_{x^e, t} \quad t + \frac{1}{\epsilon N} \sum_{k=1}^N [-\Pi^R(x^e, \rho^k) - t]^+ \quad \text{s.a.} \quad x^e \in X^R, t \in \mathbb{R}. \quad (2.22)$$

Se verifica fácilmente que este problema se puede reformular linealmente como

$$\begin{aligned} \min \quad & t + \frac{1}{\epsilon N} \sum_{k=1}^N \zeta^k \\ \text{s.a.} \quad & \zeta^k \geq -\Pi(x^e, x^{p,k}, \rho^k) - t \quad \forall k = 1, \dots, N \\ & \zeta^k \geq 0 \quad \forall k = 1, \dots, N \\ & (x^e, x^{p,k}) \in X \quad \forall k = 1, \dots, N \\ & t \in \mathbb{R}, \end{aligned} \quad (2.23)$$

donde  $X$  y  $\Pi$  son las versiones sin recourse del conjunto de planes factibles y beneficio, respectivamente. Remarcamos finalmente que este procedimiento es absolutamente análogo al hecho para el caso del modelo de VaR con recourse de §2.2.1.

### 2.2.3. Modelos MCH y MCH- $\epsilon$

El tercer modelo que consideramos en este trabajo está inspirado en optimización robusta. El modelo *Modulated Convex-Hull* (MCH) consistirá en asumir que el vector de leyes  $\rho$  se encuentra dentro de un conjunto que se obtendrá a partir de una muestra iid. Se obtendrá una planificación aversa al riesgo al tomar el plan que optimice la “peor” pérdida cuando el vector  $\rho$  se encuentra en este conjunto. Veremos que este enfoque será equivalente a minimizar una medida de riesgo de distorsión de las pérdidas. Propondremos además una versión multi-etapa o con recourse. Asimismo propondremos una versión alternativa del modelo MCH, el modelo *MCH- $\epsilon$* , que tendrá propiedades de convergencia, a diferencia del modelo MCH.

El enfoque robusto que se propone en este modelo es asumir simplemente que el parámetro incierto  $\rho$  pertenece a un conjunto  $\mathcal{U} \subset \mathbb{R}_+^B$ , el cual considerará todas las realizaciones “razonables” para  $\rho$ . A continuación, escogemos el plan  $(x^{e*}, x^{p*})$  que minimice

la peor realización de las pérdidas cuando  $\rho \in \mathcal{U}$ , esto es, el plan solución de

$$\min_{(x^e, x^p) \in X} \max_{\rho \in \mathcal{U}} -\Pi((x^e, x^p), \rho). \quad (2.24)$$

En principio, el plan  $(x^{e*}, x^{p*})$  será robusto pues para cualquier otro plan existirá una realización razonable de  $\rho$ , i.e.  $\rho \in \mathcal{U}$ , que será peor que la peor realización del primero.

Acto seguido debemos decidir el conjunto  $\mathcal{U}$  de realizaciones “razonables” para  $\rho$ . Una opción posible es tomar  $\mathcal{U} = \mathbb{R}_+^{\mathcal{B}}$ , sin embargo la solución de (2.24) probablemente sea no hacer nada,  $x^e = x^p = 0$ , una solución extremadamente conservadora. Otra posibilidad es tomar la ley  $\rho_b$  de cada bloque  $b \in \mathcal{B}$  variando en un intervalo; o es más, establecer un “presupuesto de robustez”  $\Gamma \in [0, |\mathcal{B}|]$  y especificar que no más de  $\Gamma$  bloques tengan su peor realización en el intervalo. Una tercera opción es considerar las leyes  $\rho$  que estén a una distancia euclídeana no mayor que, digamos, algún  $\Omega > 0$ ; o mejor aún, usar una distancia euclídeana modulada en varias direcciones, i.e.  $\|A^{-1} \cdot\|_2$  donde  $A$  es matriz simétrica definida positiva. Estos últimos dos enfoques son los que por su buena tratabilidad computacional gozan de mayor popularidad en optimización robusta, para el primero ver [BS04] y el segundo, [BTN00]. Sin embargo requieren decidir varios parámetros que frecuentemente no son claros de establecer, e.g. el valor de  $\Gamma$  o  $\Omega$ .

Al inicio de esta sección, en §2.2, establecimos que nuestro modelo de incerteza para el parámetro  $\rho$  sería asumir que podríamos obtener una muestra iid de él tan grande como queramos. Basados en esto proponemos tomar un conjunto de incerteza basados en una muestra iid  $\rho^1, \dots, \rho^N$ . Una primera aproximación sería tomar  $\mathcal{U}$  como  $\text{conv}(\{\rho^1, \dots, \rho^N\})$ , la envoltura convexa de la muestra. Pero con el fin de tener un control de cuán conservador será el plan resultante proponemos *modular* la envoltura convexa en torno al promedio de la muestra.

Formalmente, dado  $\epsilon \in [0, 1]$  y una muestra iid  $\rho^1, \dots, \rho^N$  del vector de leyes  $\rho$ , definimos

$$\mathcal{U}_\epsilon = \mathcal{U}_\epsilon(\rho^1, \dots, \rho^N) := \epsilon \bar{\rho} + (1 - \epsilon) \text{conv}(\{\rho^1, \dots, \rho^N\}), \quad (2.25)$$

donde  $\bar{\rho} := \frac{1}{N} \sum_{k=1}^N \rho^k$ . Definimos el modelo *Modulated Convex Hull* (MCH) con nivel de riesgo  $\epsilon$  como

$$\min_{(x^e, x^p) \in X} \max_{\rho \in \mathcal{U}_\epsilon} -\Pi((x^e, x^p), \rho). \quad (2.26)$$

Es fácil verificar que como  $\Pi((x^e, x^p), \cdot)$  es lineal afín para todo  $(x^e, x^p)$  y  $\mathcal{U}_\epsilon$  es polígono entonces

$$\max_{\rho \in \mathcal{U}_\epsilon} -\Pi((x^e, x^p), \rho) = \max_{k=1, \dots, N} -\Pi((x^e, x^p), \rho^{(k)}),$$

donde  $\rho^{(k)} := \epsilon \bar{\rho} + (1 - \epsilon) \rho^k$ . Notar que  $\{\rho^{(1)}, \dots, \rho^{(N)}\}$  contiene a todos los vértices de  $\mathcal{U}_\epsilon$ . Luego, el problema MCH (2.26) se puede formular como un problema lineal a variables binarias y continuas como

$$\begin{aligned} \min \quad & \zeta \\ \text{s.a.} \quad & \zeta \geq -\Pi((x^e, x^p), \rho^{(k)}) \quad \forall k = 1, \dots, N \\ & (x^e, x^p) \in X. \end{aligned} \quad (2.27)$$

Nos dedicamos ahora a demostrar que este modelo se puede ver como uno de optimización aversa al riesgo. Esto es, es equivalente a un problema en el que se minimiza una medida de riesgo de las pérdidas del plan. Más aún, esta medida de riesgo será de distorsión. Usaremos el Teorema 4.2 de [BB09] para esta demostración, y con el fin de hacer un desarrollo claro de esta idea, a continuación hacemos un resumen de dicho teorema y de su contexto.

### Conexión con optimización aversa al riesgo y optimización robusta

Consideremos un espacio de probabilidad  $(\Omega_N, \mathcal{P}, \mathbb{P}_N)$  finito y equiprobable. Es decir,  $\Omega_N = \{\omega^1, \dots, \omega^N\}$ ,  $\mathcal{P}$  es la  $\sigma$ -álgebra de las partes de  $\Omega_N$ , y  $\mathbb{P}_N(\omega = \omega^k) = 1/N$  para todo  $k = 1, \dots, N$ . En este contexto, una v.a.  $X : (\Omega_N, \mathcal{P}, \mathbb{P}_N) \rightarrow \mathbb{R}$  es representable por un vector  $x \in \mathbb{R}^N$ ,  $x_k = X(\omega^k)$ ; una medida de probabilidad  $\mathbb{Q}$  sobre  $(\Omega_N, \mathcal{P})$  se puede representar por  $q \in \Delta^N$ ,  $q_k = \mathbb{Q}(\omega = \omega^k)$ , donde

$$\Delta^N := \left\{ p \in \mathbb{R}_+^N : \sum_{k=1}^N p_k = 1 \right\}$$

es el *símplex*  $N$ -dimensional; y se tiene  $\mathbb{E}_{\mathbb{Q}}(X) = q^t x$  donde  $\mathbb{E}_{\mathbb{Q}}$  es la esperanza con respecto a la medida de probabilidad  $\mathbb{Q}$ . Definimos además el *símplex*  $N$ -dimensional restringido como

$$\widehat{\Delta}^N := \{ p \in \Delta^N : p_1 \geq \dots \geq p_N \},$$

y para  $X$  v.a. representada por  $x \in \mathbb{R}^N$  notaremos  $x_{[\cdot]} := (x_{[1]}, \dots, x_{[N]})^t$  donde  $x_{[1]} \geq \dots \geq x_{[N]}$ . Por último, asumiremos que toda v.a. sobre  $(\Omega_N, \mathcal{P}, \mathbb{P}_N)$  es una v.a. de pérdidas.

Por el Teorema 2.2.3 sabemos que toda medida de riesgo de distorsión es igual a una combinación lineal convexa de medidas de CVaR. El siguiente teorema establece ese hecho en el contexto presentado aquí, explotando la estructura particular que otorga el espacio de probabilidad finito equiprobable. Para la demostración de este teorema remitimos al lector a [BB09].

**Teorema 2.2.4** (Teorema 4.2 de [BB09]). *Sea  $\mu : L^\infty(\Omega_N, \mathcal{P}, \mathbb{P}_N) \rightarrow \mathbb{R}$  una medida de riesgo.  $\mu$  es medida de riesgo de distorsión ssi existe una medida de probabilidad  $q \in \widehat{\Delta}_N$  sobre  $(\Omega_N, \mathcal{P})$  tal que para todo  $X \in L^\infty(\Omega_N, \mathcal{P}, \mathbb{P}_N)$  se tiene*

$$\mu(X) = \sum_{k=1}^N q_k x_{[k]}.$$

Más aún, cada uno de estos  $q \in \widehat{\Delta}_N$  se puede escribir de la forma

$$q = \sum_{k=1}^N \lambda_k \widehat{q}^k,$$

donde  $\lambda \in \Delta^N$  y  $\hat{q}^k \in \hat{\Delta}^N$  es la medida de probabilidad asociada a la medida de riesgo de distorsión CVaR con nivel de riesgo  $\epsilon = k/N$ ,  $k = 1, \dots, N$ . Estas medidas  $\hat{q}^k$  son

$$\hat{q}_i^k = \begin{cases} 1/k & \text{si } i = 1, \dots, k \\ 0 & \text{si } i = k+1, \dots, N. \end{cases}$$

Notar que, en particular, en el espacio  $L^\infty(\Omega_N, \mathcal{P}, \mathbb{P}_N)$   $\text{CVaR}_{1/N}$  mide la peor pérdida:

$$\text{CVaR}_{1/N}(X) = (\hat{q}^1)^t x_{[\cdot]} = x_{[1]} = \max_{k=1, \dots, N} X(\omega^k),$$

y  $\text{CVaR}_1$  mide la pérdida esperada:

$$\text{CVaR}_1(X) = (\hat{q}^N)^t x_{[\cdot]} = \frac{1}{N} \sum_{k=1}^N x_{[k]} = \frac{1}{N} \sum_{k=1}^N X(\omega^k).$$

Además, usando el Teorema 2.2.2 y el hecho de que en  $(\Omega_N, \mathcal{P}, \mathbb{P}_N)$  la familia de medidas que generan  $\text{CVaR}_\epsilon$ , para  $\epsilon \in (0, 1]$ , es  $\mathcal{Q}_\epsilon := \{q \in \Delta^N : q_k \leq \frac{1}{\epsilon N}\}$  (ver [BB09]), se tiene que la medida  $\hat{q}^{\epsilon N} \in \Delta^N$  definida como

$$\hat{q}_k^{\epsilon N} := \frac{1}{\epsilon N} \begin{cases} 1 & \text{si } k \leq \lfloor \epsilon N \rfloor \\ \epsilon N - \lfloor \epsilon N \rfloor & \text{si } k = \lfloor \epsilon N \rfloor + 1 \\ 0 & \text{si no,} \end{cases} \quad (2.28)$$

es la medida de probabilidad que genera  $\text{CVaR}_\epsilon$ . Esto es, dada  $X \in L^\infty(\Omega_N, \mathcal{P}, \mathbb{P}_N)$  v.a. de pérdidas representada por  $x \in \mathbb{R}^N$ , se satisface

$$\text{CVaR}_\epsilon(X) = \sum_{k=1}^N \hat{q}_k^{\epsilon N} x_{[k]}. \quad (2.29)$$

Notar que cuando  $\epsilon = j/N$  con  $j \in \{1, \dots, N\}$  se tiene que, por definición de  $\hat{q}^{\epsilon N}$ ,  $\hat{q}^{\epsilon N} = \hat{q}^j$ , y entonces también se satisface la igualdad (2.29).

Consideremos nuevamente el espacio finito equiprobable  $(\Omega_N, \mathcal{P}, \mathbb{P}_N)$  y tomemos una v.a.  $a \in \mathbb{R}^d$  sobre este espacio de probabilidad, i.e.  $a : (\Omega_N, \mathcal{P}, \mathbb{P}_N) \rightarrow \mathbb{R}^d$ . Por simplicidad, notemos  $a^k := a(\omega^k)$  para  $k = 1, \dots, N$  y  $\mathcal{A} := \{a^1, \dots, a^N\}$ . Queremos usar un esquema averso al riesgo para hacer que se cumpla la restricción

$$z \in \mathbb{R}^d : a^t z \leq c,$$

donde  $c \in \mathbb{R}$  es dado, cuando  $a$  sigue la distribución mencionada.

Consideremos para ello una medida de riesgo de distorsión  $\mu : L^\infty(\Omega_N, \mathcal{P}, \mathbb{P}_N) \rightarrow \mathbb{R}$  representada por  $q \in \hat{\Delta}^N$ . Es decir, consideremos la restricción aversa al riesgo

$$z \in \mathbb{R}^d : \mu(z^t a - c) \leq 0.$$

Se tiene que ésta es equivalente a la restricción robusta

$$z \in \mathbb{R}^d : z^t a' \leq c \quad \forall a' \in \Pi_q(\mathcal{A}),$$

donde

$$\Pi_q(\{a^1, \dots, a^N\}) := \text{conv}\left(\left\{\sum_{k=1}^N q_{\sigma(k)} a^k : \sigma \text{ permutación}\right\}\right) \quad (2.30)$$

es el  $q$ -permutahull de  $\mathcal{A}$ .

En efecto, basta notar que

$$\begin{aligned} \mu(a^t z - c) &= \mu(z^t a) - c \\ &= \sum_{k=1}^N q_k (z^t a)_{[k]} - c \\ &= \text{máx}_{\sigma \text{ permutación}} \left\{ \sum_{k=1}^N q_{\sigma(k)} (z^t a)_k \right\} - c \\ &= \text{máx}_{\sigma \text{ permutación}} \left\{ \sum_{k=1}^N q_{\sigma(k)} z^t a^k \right\} - c \\ &= \text{máx}_{\sigma \text{ permutación}} \left\{ z^t \left( \sum_{k=1}^N q_{\sigma(k)} a^k \right) \right\} - c \\ &= \text{máx}_{a' \in \Pi_q(\mathcal{A})} \{ z^t a' \} - c, \end{aligned}$$

donde  $(z^t a)_k := z^t a^k$ , la segunda igualdad viene del Teorema 2.2.4, y la tercera igualdad viene de que  $q \in \widehat{\Delta}^N$ . La cuarta desigualdad viene de que la función objetivo  $z^t a'$  es lineal.

Notemos que un corolario de este sencillo resultado es que, si  $a \in \mathbb{R}^d$  es un vector aleatorio sobre  $(\Omega_N, \mathcal{P}, \mathbb{P}_N)$ , entonces el problema de optimización aversa al riesgo

$$\min_{z \in Z} \mu(z^t a)$$

es equivalente al problema robusto

$$\min_{z \in Z} \max_{a' \in \Pi_q(\mathcal{A})} z^t a',$$

donde  $Z \subseteq \mathbb{R}^d$  es un subconjunto cualquiera.

Volviendo al modelo MCH, queremos aplicar esta colección de corolarios del Teorema 2.2.4 a nuestro modelo. La idea fundamental es tomar el espacio de probabilidad finito equiprobable que induce la muestra iid  $\rho^1, \dots, \rho^N$  de  $\rho$ , y definir una medida de riesgo de distorsión sobre este espacio.

**Proposición 2.2.5.** *Sea  $\rho^1, \dots, \rho^N$  una muestra iid del vector de leyes  $\rho$ . Consideremos  $(\Omega_N, \mathcal{P}, \mathbb{P}_N)$  el espacio equiprobable finito que la muestra iid define, es decir  $\Omega_N = \{\rho^1, \dots, \rho^N\}$ ,  $\mathcal{P}$  es el conjunto de las partes de  $\Omega_N$  y  $\mathbb{P}_N(\rho = \rho^k) = 1/N$  para todo  $k = 1, \dots, N$ . Sea también  $\epsilon \in [0, 1]$  un nivel de riesgo.*

1. Definamos  $\mu_\epsilon$  como

$$\mu_\epsilon(L) := \epsilon \text{CVaR}_1(L) + (1 - \epsilon) \text{CVaR}_{1/N}(L)$$

para toda v.a. de pérdidas  $L \in L^\infty(\Omega_N, \mathcal{P}, \mathbb{P}_N)$ . Se tiene que  $\mu_\epsilon$  es una medida de riesgo de distorsión.

2. El problema (2.26) es equivalente al problema

$$\min_{(x^e, x^p) \in X} \mu_\epsilon(-\Pi((x^e, x^p), \rho)),$$

donde  $\mu_\epsilon$  opera en el espacio  $L^\infty(\Omega_N, \mathcal{P}, \mathbb{P}_N)$ .

*Demostración.* El punto 1. es directo del Teorema 2.2.4. Para 2. tomemos el conjunto

$$X_Z := \{((x^e, x^p), \zeta) \in X \times \mathbb{R} : -\Pi((x^e, x^p), \rho) \leq \zeta \quad \forall \rho \in \mathcal{U}_\epsilon\}.$$

Es fácil ver que al tomar  $\rho \in (\Omega_N, \mathcal{P}, \Omega_N)$  se tiene

$$X_Z = \{((x^e, x^p), \zeta) \in X \times \mathbb{R} : \mu_\epsilon(-\Pi((x^e, x^p), \rho)) \leq \zeta\}. \quad (2.31)$$

En efecto, primero notemos que

$$\begin{aligned} \text{CVaR}_{1/N}(-\Pi((x^e, x^p), \rho)) &= (c^e)^t x^e + \text{CVaR}_{1/N}(-(W^p \rho)^t x^p) + (c^p)^t x^p \\ &= (c^e)^t x^e + (\max_{k=1, \dots, N} -(W^p \rho)^t x^p) + (c^p)^t x^p \\ &= \max_{k=1, \dots, N} -\Pi((x^e, x^p), \rho). \end{aligned}$$

Luego,

$$\begin{aligned} \mu_\epsilon(-\Pi((x^e, x^p), \rho)) &= \epsilon \mathbb{E}[-\Pi((x^e, x^p), \rho)] + (1 - \epsilon) \text{CVaR}_{1/N}(-\Pi((x^e, x^p), \rho)) \\ &= -\epsilon \Pi((x^e, x^p), \bar{\rho}) + (1 - \epsilon) \max_{k=1, \dots, N} -\Pi((x^e, x^p), \rho^k) \\ &= \max_{k=1, \dots, N} -[\epsilon \Pi((x^e, x^p), \bar{\rho}) + (1 - \epsilon) \Pi((x^e, x^p), \rho^k)] \\ &= \max_{k=1, \dots, N} -\Pi((x^e, x^p), \epsilon \bar{\rho} + (1 - \epsilon) \rho^k) \\ &= \max_{\rho \in \mathcal{U}_\epsilon} -\Pi((x^e, x^p), \rho), \end{aligned}$$

y entonces son equivalentes

$$\begin{aligned} \mu_\epsilon(-\Pi((x^e, x^p), \rho)) &\leq \zeta \\ \Leftrightarrow \max_{\rho \in \mathcal{U}_\epsilon} -\Pi((x^e, x^p), \rho) &\leq \zeta \\ \Leftrightarrow -\Pi((x^e, x^p), \rho) &\leq \zeta \quad \forall \rho \in \mathcal{U}_\epsilon, \end{aligned}$$

lo que demuestra la igualdad de conjuntos (2.31). Finalmente, se concluye notando que el problema MCH (2.26) es equivalente a

$$\min \zeta \quad \text{s.a.} \quad ((x^e, x^p), \zeta) \in X_Z,$$

y entonces también es equivalente a

$$\min_{(x^e, x^p) \in X} \mu_\epsilon(-\Pi((x^e, x^p), \rho)).$$

□

Cabe hacer notar que  $\mu_\epsilon$  es la medida de riesgo coherente sobre  $L^\infty(\Omega_N, \mathcal{P}, \mathbb{P}_N)$  formada por la familia de medidas de probabilidad  $\mathcal{Q}_\epsilon := \epsilon \hat{q}^N + (1 - \epsilon) \Delta^N$  sobre  $(\Omega_N, \mathcal{P}, \mathbb{P}_N)$ , donde  $\hat{q}^N = (1/N, \dots, 1/N)^t$  es la medida de probabilidad asociada a  $\text{CVaR}_1(\cdot)$  sobre  $L^\infty(\Omega_N, \mathcal{P}, \mathbb{P}_N)$ . Esto es, si  $\mu_{\mathcal{Q}_\epsilon}$  es la medida de riesgo coherente inducida por la familia  $\mathcal{Q}_\epsilon$  como en el Teorema 2.2.2, entonces  $\mu_{\mathcal{Q}_\epsilon} = \mu_\epsilon$ .

A modo de síntesis, se tiene que, dada una muestra iid  $\rho^1, \dots, \rho^N$  y un nivel de riesgo  $\epsilon \in [0, 1]$ , el modelo MCH (2.26) es equivalente al enfoque averso al riesgo

$$\begin{aligned} \text{mín} \quad & \epsilon \mathbb{E}[-\Pi((x^e, x^p), \rho)] + (1 - \epsilon) \text{CVaR}_{1/N}[-\Pi((x^e, x^p), \rho)] \\ \text{s.a.} \quad & (x^e, x^p) \in X, \end{aligned} \tag{2.32}$$

donde  $\mathbb{E}$  y  $\text{CVaR}_{1/N}$  se toman con respecto al espacio de probabilidad finito y equiprobable inducido por la muestra iid. Destacamos que en este espacio  $\text{CVaR}_{1/N}$  es simplemente el *peor escenario*, i.e. para una v.a. de pérdidas  $L$ ,

$$\text{CVaR}_{1/N}(L) = \max_{k=1, \dots, N} L(\rho^k).$$

Así, en cierto sentido, el modelo MCH consiste en optimizar una combinación lineal convexa de la peor pérdida y la pérdida esperada de la mina. En particular, si elegimos  $\epsilon = 0$  en el modelo MCH entonces estamos optimizando el peor escenario, que es el enfoque más conservador de todos; y si elegimos  $\epsilon = 1$  entonces optimizamos la pérdida esperada, que es el enfoque neutro al riesgo. Luego, la nomenclatura “nivel de riesgo” para  $\epsilon$  es coherente, pues efectivamente controla cuán conservadores o aversos al riesgo queremos realizar la planificación.

### Modelo MCH- $\epsilon$

Nos detenemos ahora a reflexionar en torno a una posible convergencia, en algún sentido, de este modelo. En los modelos de minimización de VaR y CVaR asumimos que  $\rho$  seguía una distribución estocástica y propusimos un modelo de optimización aversa al riesgo, llamémosle *problema verdadero*. No obstante, a falta del conocimiento exacto de la distribución de  $\rho$  preferimos aproximar el problema verdadero con una muestra iid, obteniendo un *problema aproximado*. Es de esperarse que al aumentar el tamaño de la muestra ésta representará más fielmente la distribución estocástica original, traduciéndose en que el problema aproximado provea soluciones cada vez más cercanas, en algún sentido, a las del problema verdadero. Y en efecto, en el siguiente capítulo veremos que, bajo ciertas condiciones suaves, en los modelos de minimización de VaR y CVaR –en sus versiones con y sin recourse– el problema aproximado converge al problema verdadero, tanto en valor objetivo como en solución, al hacer crecer el tamaño  $N$  de la muestra.

En contraste, el modelo MCH se basa desde un principio en una muestra de  $\rho$ . En particular, no existe un *problema verdadero* y una *aproximación SAA* para tal problema. Si bien la formulación (2.32) del modelo MCH nos puede inducir a pensar que estamos tomando  $\mathbb{E}$  y  $\text{CVaR}_{1/N}$  con respecto a un hipotético “verdadero” espacio de probabilidad

$(\Xi, \mathcal{F}, \mathbb{P})$  para  $\rho$ , esto no es correcto porque estas medidas están definidas sobre el espacio  $(\Omega_N, \mathcal{P}, \mathbb{P}_N)$  definido por la muestra equiprobable de  $\rho$ .

Otra opción posible es que el *problema verdadero* sea el problema (2.32) con respecto a la distribución *verdadera* de  $\rho$ , y luego tomar su aproximación SAA con una muestra iid de tamaño  $N$ , obteniendo el problema MCH (2.26). Sin embargo este enfoque sería sub-óptimo, pues el problema verdadero dependería desde un inicio de  $N$  –pues estamos tomando  $\text{CVaR}_{1/N}$ – y para tomar su aproximación SAA tomaríamos una muestra de tamaño  $N$ . O en otras palabras, bajo este esquema si no tomamos una muestra de tamaño  $N$  entonces la aproximación SAA del problema original no sería equivalente al problema MCH (2.26). Y peor aún, al hacer crecer  $N$ , i.e. tomar un tamaño de muestra mayor, estaríamos tomando un problema verdadero distinto pues éste también depende de  $N$ .

Mencionamos anteriormente que el modelo MCH consiste en optimizar, en el espacio de probabilidad que define la muestra iid, una combinación lineal convexa de la peor pérdida y la pérdida esperada de la mina. Luego, es esperable que al hacer crecer el tamaño  $N$  de la muestra tanto la pérdida esperada como la peor pérdida converjan a sus verdaderos valores. Sin embargo el análisis de convergencia para un enfoque de este estilo se sale de la teoría de *Sample Average Approximation* para optimización estocástica (ver [SDR09, Capítulo 5]), pues esta teoría no considera una metodología para analizar la convergencia del mejor peor caso al aumentar el tamaño de la muestra.

La conclusión de esta discusión es que no hay una manera clara de realizar un análisis de convergencia para el problema MCH al hacer crecer el tamaño de la muestra. Para corregir esta situación proponemos usar adicionalmente en el problema de planificación minera bajo riesgo el modelo *MCH- $\epsilon$* . Este enfoque se inspira en la equivalencia entre el modelo MCH (2.26) y el problema de optimización aversa al riesgo (2.32), demostrada por la Proposición 2.2.5.

Dado un nivel de riesgo  $\epsilon \in (0, 1]$ , definimos el *modelo MCH- $\epsilon$*  como

$$\begin{aligned} \text{mín} \quad & \epsilon \mathbb{E} [-\Pi((x^e, x^p), \rho)] + (1 - \epsilon) \text{CVaR}_\epsilon [-\Pi((x^e, x^p), \rho)] \\ \text{s.a.} \quad & (x^e, x^p) \in X, \end{aligned} \quad (2.33)$$

donde  $\mathbb{E}$  y  $\text{CVaR}_\epsilon$  están tomados con respecto al espacio de probabilidad  $(\Xi, \mathcal{F}, \mathbb{P})$  *verdadero* para  $\rho$ .

Como en el modelo de minimización de CVaR, si la  $\sigma$ -álgebra  $\mathcal{F}$  para  $\rho$  es boreliana entonces el problema (2.33) es equivalente a

$$\begin{aligned} \text{mín} \quad & \epsilon \mathbb{E} [-\Pi((x^e, x^p), \rho)] + (1 - \epsilon) \left( t + \frac{1}{\epsilon} \mathbb{E} [-\Pi((x^e, x^p), \rho) - t]^+ \right) \\ \text{s.a.} \quad & (x^e, x^p) \in X, \quad t \in \mathbb{R}. \end{aligned} \quad (2.34)$$

Para poder hacer tratables las esperanzas de este problema aplicamos la técnica SAA y aproximamos la esperanza usando una muestra iid  $\rho^1, \dots, \rho^N$  de  $\rho$ . La *aproximación*

SAA del problema MCH- $\epsilon$  queda

$$\begin{aligned} \text{mín} \quad & \epsilon \frac{1}{N} \sum_{k=1}^N [-\Pi((x^e, x^p), \rho^k)] + (1 - \epsilon) \left( t + \frac{1}{\epsilon N} \sum_{k=1}^N [-\Pi((x^e, x^p), \rho^k) - t]^+ \right) \\ \text{s.a.} \quad & (x^e, x^p) \in X, \quad t \in \mathbb{R}, \end{aligned} \quad (2.35)$$

o en una versión linealizada,

$$\begin{aligned} \text{mín} \quad & \epsilon \frac{1}{N} \sum_{k=1}^N [-\Pi((x^e, x^p), \rho^k)] + (1 - \epsilon) \left( t + \frac{1}{\epsilon N} \sum_{k=1}^N \zeta_k \right) \\ \text{s.a.} \quad & \zeta_k \geq -\Pi((x^e, x^p), \rho^k) - t \quad \forall k = 1, \dots, N \\ & \zeta \in \mathbb{R}_+^N, \quad (x^e, x^p) \in X, \quad t \in \mathbb{R}. \end{aligned} \quad (2.36)$$

El modelo MCH- $\epsilon$  aventaja al modelo MCH en que el modelo *verdadero* (2.33) no depende de  $N$ , por lo que es absolutamente razonable preguntarse si su aproximación SAA (2.35) convergerá al problema verdadero, en algún sentido, al hacer crecer el tamaño de muestra  $N$ . Veremos en el siguiente capítulo que, bajo hipótesis bastante razonables sobre el espacio de probabilidad  $(\Xi, \mathcal{F}, \mathbb{P})$  de  $\rho$ , en este modelo y en su variante con recourse –que veremos a continuación– hay convergencia, en el sentido que tanto el valor objetivo como la solución de la aproximación SAA convergerán a sus respectivas contrapartes en el problema verdadero.

Mostramos por último una similitud o analogía adicional entre el modelo MCH y el modelo MCH- $\epsilon$ .

**Proposición 2.2.6.** *Consideremos  $\rho^1, \dots, \rho^N$  una muestra iid de  $\rho$  y sea  $(\Omega_N, \mathcal{P}, \mathbb{P}_N)$  el espacio finito equiprobable que ésta define. El problema MCH- $\epsilon$  aproximado (2.35) es equivalente al problema*

$$\text{mín}_{(x^e, x^p) \in X} \quad \text{máx}_{\rho \in \mathcal{V}_\epsilon} \quad -\Pi((x^e, x^p), \rho), \quad (2.37)$$

donde  $\mathcal{V}_\epsilon := \epsilon \bar{\rho} + (1 - \epsilon) \Pi_{\hat{q}^{\epsilon N}}(\{\rho^1 \dots \rho^N\})$  es un conjunto de robustez para  $\rho$ , y donde  $\Pi_{\hat{q}^{\epsilon N}}(\{\rho^1, \dots, \rho^N\})$  es como en (2.30). Aquí  $\hat{q}^{\epsilon N} \in \Delta^N$  es el vector definido en (2.28), es una medida de probabilidad sobre  $(\Omega_N, \mathcal{P})$ , y representa a  $\text{CVaR}_\epsilon : L^\infty(\Omega_N, \mathcal{P}, \mathbb{P}_N) \rightarrow \mathbb{R}$  de la manera que establece el Teorema 2.2.4.

*Demostración.* Por la representación que establece el Teorema 2.2.4, y en particular usando la igualdad (2.29), se tiene

$$\begin{aligned} \text{CVaR}_\epsilon(-(W^p \rho)^t x^p) &= \text{máx}_{\sigma \text{ permutación}} \sum_{k=1}^N \hat{q}_{\sigma(k)}^{\epsilon N} (-(W^p \rho^k)^t x^p) \\ &= \text{máx}_{\sigma \text{ permutación}} - \left( W^p \left( \sum_{k=1}^N \hat{q}_{\sigma(k)}^{\epsilon N} \rho^k \right) \right)^t x^p \\ &= \text{máx}_{\rho \in \Pi_{\hat{q}^{\epsilon N}}} -(W^p \rho)^t x^p, \end{aligned}$$

donde tomamos  $\text{CVaR}_\epsilon$  sobre  $L^\infty(\Omega_N, \mathcal{P}, \mathbb{P}_N)$ . Luego,

$$\begin{aligned} \text{CVaR}_\epsilon(-\Pi((x^e, x^p), \rho)) &= (c^e)^t x^e + \text{CVaR}_\epsilon(-(W^p \rho)^t x^p) + (c^p)^t x^p \\ &= (c^e)^t x^e + \left( \text{máx}_{\rho \in \Pi_{\hat{q}^{\epsilon N}}} -(W^p \rho)^t x^p \right) + (c^p)^t x^p \\ &= \text{máx}_{\rho \in \Pi_{\hat{q}^{\epsilon N}}} -\Pi((x^e, x^p), \rho), \end{aligned}$$

con lo que

$$\begin{aligned} \max_{\rho \in \mathcal{V}_\epsilon} -\Pi((x^e, x^p), \rho) &= \max_{\rho \in \Pi_{\bar{q}^\epsilon N}} -[\epsilon \Pi((x^e, x^p), \bar{\rho}) + (1 - \epsilon) \Pi((x^e, x^p), \rho)] \\ &= \epsilon [-\Pi((x^e, x^p), \bar{\rho})] + (1 - \epsilon) \max_{\rho \in \Pi_{\bar{q}^\epsilon N}} -\Pi((x^e, x^p), \rho) \\ &= \epsilon \mathbb{E} [-\Pi((x^e, x^p), \rho)] + (1 - \epsilon) \text{CVaR}_\epsilon [-\Pi((x^e, x^p), \rho)], \end{aligned}$$

lo que concluye la demostración.  $\square$

### Contrapartes con recourse

Adaptamos ahora los modelos MCH y MCH- $\epsilon$  para que incorporen el paradigma multi-etapa presentado en §2.2.1. En él se deben elegir “aquí y ahora” qué bloques extraer; posteriormente, después de realizarse la extracción, se puede ver cuál era la verdadera ley  $\rho$  de los bloques, y sólo entonces se debe decidir qué bloques procesar. Como se hizo antes, para incorporar este modelo de decisión a nuestro paradigma basta redefinir el beneficio  $\Pi$ , haciendo que la decisión  $x^p$  dependa de la decisión de extracción  $x^e$  y la ley  $\rho$ . Recordamos entonces la definición (2.9) de *beneficio con recourse*

$$\begin{aligned} \Pi^R(x^e, \rho) &:= -(v^e)^t x^e + \max_{x^p} (W^p \rho - v^p)^t x^p \\ \text{s.a. } &x^p \leq x^e \\ &(c^p)^t x^p \leq B^p \\ &x^p \in \{0, 1\}^B, \end{aligned}$$

y también recordamos la notación (2.10) para el conjunto de planes de extracción factibles:

$$X^R := \left\{ x^e \in \{0, 1\}^B : \begin{array}{l} x_a^e \leq x_b^e \quad \forall (a, b) \in \mathcal{P} \\ (c^e)^t x^e \leq B^e \end{array} \right\}.$$

Primeramente presentamos esta variante para el modelo MCH- $\epsilon$ . Dado un nivel de riesgo  $\epsilon \in (0, 1]$ , el *modelo MCH- $\epsilon$  con recourse* es

$$\begin{aligned} \min & \epsilon \mathbb{E} [-\Pi^R(x^e, \rho)] + (1 - \epsilon) \text{CVaR}_\epsilon [-\Pi^R(x^e, \rho)] \\ \text{s.a. } & x^e \in X^R. \end{aligned} \tag{2.38}$$

Como se argumentó para el modelo de minimización de CVaR con recourse, si la  $\sigma$ -álgebra  $\mathcal{F}$  para  $\rho$  es boreliana entonces el problema (2.38) es equivalente a

$$\begin{aligned} \min & \epsilon \mathbb{E} [-\Pi^R(x^e, \rho)] + (1 - \epsilon) \left( t + \frac{1}{\epsilon} \mathbb{E} [-\Pi^R(x^e, \rho) - t]^+ \right) \\ \text{s.a. } & x^e \in X^R, \quad t \in \mathbb{R}. \end{aligned} \tag{2.39}$$

Una vez más, para sortear el problema de modelamiento y tratabilidad computacional que las esperanzas en este problema presentan, recurrimos a una aproximación SAA. Dada una muestra iid  $\rho^1, \dots, \rho^N$  de  $\rho$ , la *aproximación SAA del modelo MCH- $\epsilon$  con recourse* es

$$\begin{aligned} \min & \epsilon \frac{1}{N} \sum_{k=1}^N [-\Pi^R(x^e, \rho^k)] + (1 - \epsilon) \left( t + \frac{1}{\epsilon N} \sum_{k=1}^N [-\Pi^R(x^e, \rho^k) - t]^+ \right) \\ \text{s.a. } & x^e \in X^R, \quad t \in \mathbb{R}, \end{aligned} \tag{2.40}$$

o equivalentemente, el problema linealizado

$$\begin{aligned}
\text{mín} \quad & \epsilon \frac{1}{N} \sum_{k=1}^N [-\Pi((x^e, x^{p,k}), \rho^k)] + (1 - \epsilon) \left( t + \frac{1}{\epsilon N} \sum_{k=1}^N \zeta_k \right) \\
\text{s.a.} \quad & \zeta_k \geq -\Pi((x^e, x^{p,k}), \rho^k) - t \quad \forall k = 1, \dots, N \\
& (x^e, x^{p,k}) \in X \quad \forall k = 1, \dots, N \\
& \zeta \in \mathbb{R}_+^N, \quad t \in \mathbb{R},
\end{aligned} \tag{2.41}$$

donde  $\Pi$  y  $X$  son el beneficio y conjunto factibles, respectivamente, sin recourse.

Discutimos ahora sobre una posible variante con recourse para el modelo MCH. Recordemos que, dado un nivel de riesgo  $\epsilon \in [0, 1]$  y una muestra  $\rho^1, \dots, \rho^N$  del parámetro  $\rho$ , el problema MCH sin recourse “original” (2.26) es

$$\text{mín}_{(x^e, x^p) \in X} \quad \text{máx}_{\rho \in \mathcal{U}_\epsilon} \quad -\Pi((x^e, x^p), \rho),$$

donde  $\mathcal{U}_\epsilon := \bar{\rho} + (1 - \epsilon) (\text{conv}(\{\rho^1, \dots, \rho^N\}) - \bar{\rho})$  y  $\bar{\rho} := \frac{1}{N} \sum_{k=1}^N \rho^k$ . Un atractivo de este enfoque es que como la función de beneficio sin recourse  $\Pi((x^e, x^p), \rho)$  es lineal-afín en  $\rho$ , y  $\mathcal{U}_\epsilon$  es un polítopo del cual conocemos sus puntos extremos, entonces el problema MCH (2.26) se puede formular como el problema lineal (2.27).

Volviendo a la variante multi-etapa, si replicáramos el esquema anterior la contraparte con recourse directa del modelo MCH sería

$$\text{mín}_{x^e \in X^R} \quad \text{máx}_{\rho \in \mathcal{U}_\epsilon} \quad -\Pi^R(x^e, \rho), \tag{2.42}$$

donde  $\Pi^R$  es la función de beneficio con recourse, definida en (2.9). Sin embargo  $\Pi^R$  no es lineal ni lineal-afín en el parámetro  $\rho$ . Veamos que esto conduce finalmente a que no exista una formulación sencilla y tratable computacionalmente del problema (2.42), candidato a ser contraparte con recourse del problema MCH.

Primero notemos que el problema (2.42) es equivalente a

$$\begin{aligned}
\text{mín} \quad & \zeta \\
\text{s.a.} \quad & -\Pi^R(x^e, \rho) \leq \zeta \quad \forall \rho \in \mathcal{U}_\epsilon \\
& x^e \in X^R.
\end{aligned}$$

Definiendo

$$Y = Y(x^e) := \{x^p \in \{0, 1\}^B : x^p \leq x^e, (c^p)^t x^p \leq B^p\}$$

se tiene que

$$\begin{aligned}
-\text{máx}_{x^p \in Y} (W^p \rho - v^p)^t x^p &= \text{mín}_{x^p \in Y} -(W^p \rho - v^p)^t x^p \\
&= \text{mín}_{x^p \in \text{conv}(Y)} -(W^p \rho - v^p)^t x^p,
\end{aligned}$$

donde  $\text{conv}(Y)$  es la envoltura convexa de  $Y$ , y la segunda igualdad viene de que  $Y$  es acotado y la función objetivo  $(W^p \rho - v^p)^t x^p$  es lineal en  $x^p$ . Por otro lado, son equivalentes las restricciones

$$\begin{aligned}
& -\Pi^R(x^e, \rho) \leq \zeta \quad \forall \rho \in \mathcal{U}_\epsilon \\
\Leftrightarrow \quad & \text{máx}_{\rho \in \mathcal{U}_\epsilon} [(v^e)^t x^e - \text{máx}_{x^p \in Y} (W^p \rho - v^p)^t x^p] \leq \zeta \\
\Leftrightarrow \quad & \text{máx}_{\rho \in \mathcal{U}_\epsilon} \text{mín}_{x^p \in \text{conv}(Y)} -(W^p \rho - v^p)^t x^p \leq \zeta - (v^e)^t x^e.
\end{aligned}$$

Y, como  $\mathcal{U}_\epsilon$  y  $\text{conv}(Y)$  son compactos convexos, y la función objetivo  $-(W^p \rho - v^p)^t x^p$  es lineal en  $x^p$  y lineal-afín en  $\rho$ , entonces podemos intercambiar el orden en que se toma el mínimo y el máximo, ver e.g. teorema Min-Max de Sion [Sio58]. Luego, se puede reescribir la restricción de las siguientes maneras equivalentes

$$\begin{aligned}
& \max_{\rho \in \mathcal{U}_\epsilon} \min_{x^p \in \text{conv}(Y)} -(W^p \rho - v^p)^t x^p \leq \zeta - (v^e)^t x^e \\
\Leftrightarrow & \min_{x^p \in \text{conv}(Y)} \max_{\rho \in \mathcal{U}_\epsilon} -(W^p \rho - v^p)^t x^p \leq \zeta - (v^e)^t x^e \\
\Leftrightarrow & \exists x^p \in \text{conv}(Y) \text{ t.q. } \max_{\rho \in \mathcal{U}_\epsilon} -(W^p \rho - v^p)^t x^p \leq \zeta - (v^e)^t x^e \\
\Leftrightarrow & \exists x^p \in \text{conv}(Y) \text{ t.q. } \max_{k=1, \dots, N} -(W^p \rho^{(k)} - v^p)^t x^p \leq \zeta - (v^e)^t x^e \\
\Leftrightarrow & \exists x^p \in \text{conv}(Y) \text{ t.q. } \forall k = 1, \dots, N \quad -(W^p \rho^{(k)} - v^p)^t x^p \leq \zeta - (v^e)^t x^e \\
\Leftrightarrow & \exists x^p \in \text{conv}(Y) \text{ t.q. } \forall k = 1, \dots, N \quad -\Pi((x^e, x^p), \rho^{(k)}) \leq \zeta,
\end{aligned}$$

donde  $\rho^{(k)} := (1 - \epsilon)\rho^k + \epsilon\bar{\rho}$ ,  $k = 1, \dots, N$ , y donde la tercera equivalencia viene de que  $\{\rho^{(1)}, \dots, \rho^{(N)}\}$  contiene a los puntos extremos de  $\mathcal{U}_\epsilon$ .

De esto concluimos que el problema (2.42) es equivalente a

$$\begin{aligned}
& \min \quad \zeta \\
& \text{s.a.} \quad -\Pi((x^e, x^p), (1 - \epsilon)\rho^k + \epsilon\bar{\rho}) \leq \zeta \quad \forall k = 1, \dots, N \\
& \quad \quad x^e \in X^R \\
& \quad \quad x^p \in \text{conv}(Y),
\end{aligned}$$

donde  $\Pi$  es el beneficio sin recourse.

Sin embargo, esta formulación no es tratable tal como está. En primer lugar, el conjunto  $Y$ , de planes de procesamiento  $x^p$  factibles, depende explícitamente de  $x^e$ . Y segundo, este conjunto es básicamente el conjunto factible de un problema *knapsack*, que como es ampliamente conocido es un problema NP-completo. Estas dos características de esta formulación hacen que encontrar una reformulación tratable sea una tarea compleja.

Como conclusión de este análisis sacamos que el problema (2.42), que sería la contraparte con recourse directa del enfoque MCH, no posee una formulación tratable evidente. No obstante, inspirándonos en la Proposición 2.2.5 proponemos una definición alternativa para el problema MCH con recourse.

Dado  $\epsilon \in [0, 1]$ , definimos el *modelo MCH con recourse con nivel de riesgo  $\epsilon$*  como

$$\min_{x^e \in X^R} \left\{ \epsilon \mathbb{E} [-\Pi^R(x^e, \rho)] + (1 - \epsilon) \text{CVaR}_{1/N} [-\Pi^R(x^e, \rho)] \right\}, \quad (2.43)$$

donde  $\Pi^R$  es el beneficio con recourse y  $X^R$  es el conjunto de planes de extracción factibles.

A partir de aquí el tratamiento es exactamente análogo al hecho con el modelo MCH- $\epsilon$  con recourse. Primero, si la  $\sigma$ -álgebra  $\mathcal{F}$  asociada a la hipotética distribución *verdadera*  $(\Xi, \mathcal{F}, \mathbb{P})$  de  $\rho$  es boreliana, entonces el problema (2.43) es equivalente a

$$\min_{x^e \in X^R, t \in \mathbb{R}} \left\{ \epsilon \mathbb{E} [-\Pi^R(x^e, \rho)] + (1 - \epsilon) \left( t + \frac{1}{\epsilon} \mathbb{E} [-\Pi^R(x^e, \rho) - t]^+ \right) \right\}. \quad (2.44)$$

Luego, en nuestro enfoque usual, para abordar el problema de modelamiento y tratabilidad que presentan las esperanzas usamos la aproximación SAA de este problema. Dada una muestra iid  $\rho^1, \dots, \rho^N$  la *aproximación SAA del problema MCH con recourse* es

$$\min_{x^e \in X^R, t \in \mathbb{R}} \left\{ \epsilon \frac{1}{N} \sum_{k=1}^N [-\Pi^R(x^e, \rho^k)] + (1 - \epsilon) \left( t + \frac{1}{\epsilon N} \sum_{k=1}^N [-\Pi^R(x^e, \rho^k) - t]^+ \right) \right\}, \quad (2.45)$$

o en una formulación linealizada,

$$\begin{aligned} \min \quad & \epsilon \frac{1}{N} \sum_{k=1}^N [-\Pi((x^e, x^{p,k}), \rho^k)] + (1 - \epsilon) \left( t + \frac{1}{\epsilon N} \sum_{k=1}^N \zeta_k \right) \\ & -\Pi((x^e, x^{p,k}), \rho^k) - t \geq 0 \quad \forall k = 1, \dots, N \\ & (x^e, x^{p,k}) \in X \quad \forall k = 1, \dots, N \\ & t \in \mathbb{R}, \end{aligned} \quad (2.46)$$

donde  $\Pi$  y  $X$  son las variantes sin recourse del beneficio y conjunto de planes factibles, respectivamente.

#### 2.2.4. Sobre el contenido de probabilidad de un polítopo

En las secciones §2.2.2 y §2.2.3 hemos ido presentando resultados que muestran la rica estructura de las medidas de riesgo de distorsión. Particularmente, hemos visto cómo las aproximaciones SAA de los modelos de minimización de CVaR y MCH- $\epsilon$ , y el modelo MCH, todos en su variante sin recourse, pueden formularse indistintamente como problemas de optimización aversa al riesgo o como problemas de optimización robusta. En esta sección nos concentramos en esta última representación y nos preguntamos cuán representativo de la distribución original de  $\rho$  puede ser el conjunto de robustez en cada modelo. Veremos que esto tendrá implicancias importantes en la convergencia de las aproximaciones SAA, tema que discutiremos en el capítulo siguiente.

Sea  $\epsilon \in (0, 1)$  un nivel de riesgo cualquiera y consideremos una muestra iid  $\rho^1, \dots, \rho^N$  de  $\rho$ . Recordemos primeramente que el problema MCH sin recourse (2.26) es

$$\min_{(x^e, x^p) \in X} \max_{\rho \in \mathcal{U}_\epsilon} -\Pi((x^e, x^p), \rho),$$

donde

$$\mathcal{U}_\epsilon := \epsilon \bar{\rho} + (1 - \epsilon) \text{conv}(\{\rho^1, \dots, \rho^N\})$$

y

$$\bar{\rho} := \frac{1}{N} \sum_{k=1}^N \rho^k.$$

También, del Teorema 2.2.4 y la discusión que le sigue, obtenemos fácilmente que la aproximación SAA (2.18) del problema de minimización de CVaR sin recourse es equivalente al problema

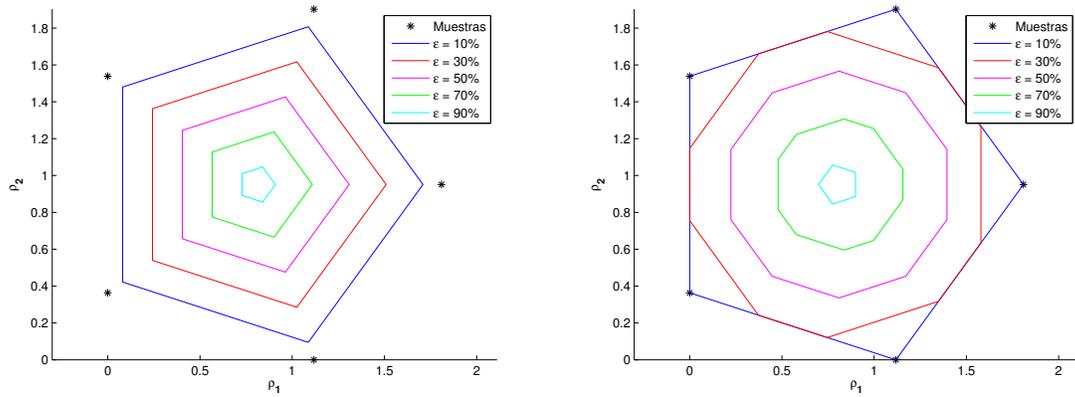
$$\min_{(x^e, x^p) \in X} \max_{\rho \in \Pi_{\bar{q}\epsilon N}(\{\rho^1 \dots \rho^N\})} -\Pi((x^e, x^p), \rho), \quad (2.47)$$

donde  $\hat{q}^{\epsilon N}$  es como en la definición (2.28) y  $\Pi_{\hat{q}^{\epsilon N}}(\{\rho^1 \dots \rho^N\})$  es como en (2.30).

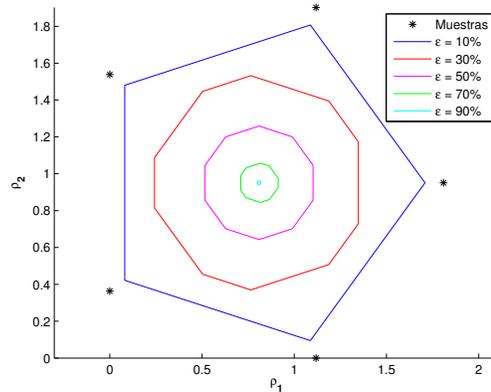
Y de manera análoga, la Proposición 2.37 afirma que la aproximación SAA del problema MCH- $\epsilon$  es equivalente al problema

$$\min_{(x^e, x^p) \in X} \max_{\rho \in \mathcal{V}_\epsilon} -\Pi((x^e, x^p), \rho),$$

donde  $\mathcal{V}_\epsilon := \epsilon \bar{\rho} + (1 - \epsilon) \Pi_{\hat{q}^{\epsilon N}}(\{\rho^1 \dots \rho^N\})$ . En síntesis, las tres aproximaciones SAA se formulan como un problema de optimización robusta donde el único cambio es la elección del conjunto de robustez.



(a) Conjunto  $\mathcal{U}_\epsilon$  de modelo MCH, para varios  $\epsilon$       (b) Conjunto  $\Pi_{\hat{q}^{\epsilon N}}(\{\rho^1 \dots \rho^N\})$  de modelo CVaR, para varios  $\epsilon$



(c) Conjunto  $\mathcal{V}_\epsilon$  de modelo MCH- $\epsilon$ , para varios  $\epsilon$

Figura 2.1: Conjuntos de robustez para modelos MCH, CVaR aproximado y MCH- $\epsilon$  aproximado, todos sin recourse, para  $N = 5$  muestras de  $\rho \in \mathbb{R}_+^2$  y varios niveles de riesgo  $\epsilon$

En las Figuras 2.1(b), 2.1(a) y 2.1(c) mostramos ejemplos de los conjuntos de robustez  $\mathcal{U}_\epsilon$ ,  $\Pi_{\hat{q}^{\epsilon N}}(\{\rho^1 \dots \rho^N\})$  y  $\mathcal{V}_\epsilon$ , respectivamente. Para estos ejemplos consideramos una

mina con solamente  $|\mathcal{B}| = 2$  bloques y tomamos  $N = 5$  muestras del vector de leyes  $\rho \in \mathbb{R}_+^2$ . Las tres figuras están hechas usando las mismas cinco muestras de  $\rho$ , y se consideran los mismos niveles de riesgo  $\epsilon$ : 10 %, 30 %, 50 %, 70 % y 90 %. Este ejemplo resume la dinámica entre el nivel de riesgo, el conjunto de robustez y la conservatividad de estos modelos: a menor nivel de riesgo  $\epsilon$  tomamos un mayor conjunto de robustez, lo cual significa planificar para la peor ley  $\rho$  de entre un mayor número de leyes posibles, obteniéndose un plan más robusto y más averso al riesgo.

Una manera de establecer qué modelo puede ser mejor que otro sería evaluar para cuál de estos tres conjuntos de robustez el vector  $\rho$  tiene mayor probabilidad de estar en él. Es decir, calcular el *contenido de probabilidad*  $\mathbb{P}(\rho \in \mathcal{U})$  del conjunto de robustez  $\mathcal{U}$  de cada modelo. La primera dificultad de este enfoque es práctica, pues el conjunto  $\mathcal{U}$  depende de  $N$  muestras independientes de  $\rho$ , lo cual agrega complejidad al cálculo. Y la segunda es de modelamiento, pues hemos asumido desde un principio que desconocemos la verdadera distribución de probabilidad de  $\rho$ , con lo cual no podríamos calcular la probabilidad para cada modelo.

A falta de resultados exactos nos apoyamos en el siguiente resultado, que nos permitirá hacer algunas observaciones cualitativas importantes sobre el contenido de probabilidad de un polítopo.

**Teorema 2.2.7** (Teorema 13.2.1 de [Mat02]). *Sea  $B^d$  la bola unitaria en  $\mathbb{R}^d$  y sea  $\mathcal{U} \subseteq B^d$  un polítopo convexo contenido en  $B^d$ . Si  $\mathcal{U}$  tiene a lo más  $N$  vértices entonces*

$$\frac{\text{vol}(\mathcal{U})}{\text{vol}(B^d)} \leq \left( \frac{C \ln \left( \frac{N}{d} + 1 \right)}{d} \right)^{d/2} \quad (2.48)$$

donde  $\text{vol}(\cdot)$  es el volumen en  $\mathbb{R}^d$  y  $C$  es una constante independiente de  $d$  y  $N$ .

Supongamos momentáneamente que el soporte de la distribución de  $\rho$  es la bola unitaria en  $\mathbb{R}^d$ , y que la distribución es uniforme en este conjunto. En tal caso el término de la izquierda de (2.48) es exactamente  $\mathbb{P}(\rho \in \mathcal{U})$ . En consecuencia, a menos que el número de vértices de  $\mathcal{U}$  sea exponencial en  $d$ , esta probabilidad es despreciable. O en otras palabras, se necesita que el número de vértices  $N$  sea exponencial en la dimensión  $d$  del espacio para que la probabilidad tenga alguna posibilidad de no ser cercana a 0.

Las implicancias de este resultado son fuertes, pues en nuestro caso la dimensión  $d$  del espacio es la cantidad de bloques,  $|\mathcal{B}|$ . Como se ha mencionado en varias ocasiones, una característica principal del problema de planificación minera es el gran número de bloques que tienen los modelos de bloques de minas de tamaño normal. Y veremos en el Capítulo 4 que para una mina de casi 20.000 bloques no se pueden tomar más de 400 muestras.

Como conclusión de esta sección tenemos que la cantidad de bloques, es decir la dimensión del vector  $\rho$ , es un factor a tomar en cuenta para las aproximaciones SAA. En el

siguiente capítulo ahondamos en la discusión que se tuvo al presentar el modelo MCH- $\epsilon$ , en la cual dijimos que sería esperable que, al aumentar el tamaño de la muestra, la aproximación SAA de cada modelo convergiera al problema verdadero. Veremos que esto es cierto, y más aún presentaremos un algoritmo que nos permitirá obtener soluciones “mejores” en un sentido estadístico. Esto por un lado refleja que al hacer crecer a infinito  $N$ , el tamaño de la muestra, el contenido de probabilidad de su envoltura convexa será 1. Sin embargo, la discusión de esta última sección nos dice que el tamaño de muestra necesario para lograr esto posiblemente tendrá que ser más que exponencial en la cantidad de bloques de la mina.

# Capítulo 3

## Aproximación SAA

### 3.1. Marco teórico general

#### 3.1.1. Propiedades de estimadores SAA para Stochastic Programming Problems

En la presente sub-sección se muestran las principales propiedades estadísticas de los problemas de optimización estocástica resueltos usando la técnica *Sample Average Approximation* (SAA). Esta sub-sección se basa en las secciones §5.1.1, §5.1.2 y §5.6.1 de [SDR09].

Consideramos el siguiente problema de optimización estocástica:

$$\vartheta^* := \min_{x \in X} \{f(x) := \mathbb{E}[F(x, \xi)]\} \quad (3.1)$$

donde  $X \subseteq \mathbb{R}^n$  es un conjunto no-vacío cerrado,  $\xi$  es un vector aleatorio cuya distribución de probabilidad  $\mathbb{P}$  está soportada en el conjunto  $\Xi \subset \mathbb{R}^d$ , y  $F : X \times \Xi \rightarrow \mathbb{R}$  es una función de Carathéodory, i.e., continua en  $x$  y medible en  $\xi$ . Asumimos además que la esperanza  $f(x)$  está bien definida y es finita para cada  $x \in X$ .

Suponemos ahora que se tiene una muestra  $\xi^1, \dots, \xi^N$  de  $N$  realizaciones del vector estocástico  $\xi$ . Para cada  $x \in X$  podemos estimar el valor esperado  $f(x)$  promediando los valores  $F(x, \xi^j)$ ,  $j = 1, \dots, N$ . Esto lleva a la, así llamada, aproximación *Sample Average Approximation* (SAA)

$$\hat{\vartheta}_N := \min_{x \in X} \left\{ \hat{f}_N(x) := \frac{1}{N} \sum_{j=1}^N F(x, \xi^j) \right\} \quad (3.2)$$

del problema “verdadero” (3.1). Asumimos siempre que cada muestra  $\xi^j$  tiene la misma distribución (marginal)  $\mathbb{P}$  que el vector estocástico  $\xi$ . Si, más aún, cada  $\xi^j$ ,  $j = 1, \dots, N$ ,

es distribuido independientemente de los demás vectores de la muestra, decimos que la muestra es *independiente e idénticamente distribuida* (iid).

Por la Ley de los Grandes Números clásica (LGN) se tiene que, bajo algunas condiciones de regularidad,  $\widehat{f}_N(x)$  converge puntualmente a  $f(x)$  c.s. cuando  $N \rightarrow \infty$ . En particular, por la LGN clásica esto se tiene si la muestra es iid. También se tiene que  $\mathbb{E}[\widehat{f}_N(x)] = f(x)$ , i.e.,  $\widehat{f}_N(x)$  es estimador insesgado de  $f(x)$ . Luego, es natural esperar que el valor óptimo y soluciones óptimas del problema SAA (3.2) converjan a sus contrapartes del problema verdadero (3.1) cuando  $N \rightarrow \infty$ . En adelante denotamos por  $S$  al conjunto de soluciones óptimas del problema verdadero (3.1) y por  $\widehat{S}_N$  al conjunto de soluciones óptimas del problema SAA (3.2).

A continuación discutimos las propiedades estadísticas de los estimadores SAA  $\widehat{\vartheta}_N$  y  $\widehat{S}_N$ .

### Consistencia

Esta sección se basa íntegramente en [SDR09, §5.1.1 Consistency of SAA Estimators]. Buscamos estudiar las propiedades de convergencia de los estimadores SAA  $\widehat{\vartheta}_N$  y  $\widehat{S}_N$ . Se dice que un estimador  $\widehat{\vartheta}_N$  de un parámetro  $\vartheta$  es *consistente* si  $\widehat{\vartheta}_N$  converge c.s. a  $\vartheta$  cuando  $N \rightarrow \infty$ . Consideremos primero la consistencia del estimador SAA  $\widehat{\vartheta}_N$  del valor óptimo  $\vartheta^*$ . Tenemos que para todo  $x \in X$  fijo,  $\widehat{\vartheta}_N \leq \widehat{f}_N(x)$ , y entonces si se tiene puntualmente la LGN, entonces

$$\limsup_{N \rightarrow \infty} \widehat{\vartheta}_N \leq \lim_{N \rightarrow \infty} \widehat{f}_N(x) = f(x) \quad \text{c.s.}$$

Sigue que si se tiene puntualmente la LGN, entonces

$$\limsup_{N \rightarrow \infty} \widehat{\vartheta}_N \leq \vartheta^* \quad \text{c.s.} \quad (3.3)$$

Sin algunas condiciones adicionales, la desigualdad puede ser estricta.

**Proposición 3.1.1** (Proposition 5.2, [SDR09]). *Supongamos que  $\widehat{f}_N(x)$  converge c.s. a  $f(x)$ , cuando  $N \rightarrow \infty$ , uniformemente en  $X$ . Entonces  $\widehat{\vartheta}_N$  converge a  $\vartheta^*$  c.s. cuando  $N \rightarrow \infty$ .*

Para establecer la consistencia de los estimadores SAA de soluciones óptimas necesitamos condiciones ligeramente más fuertes. Denotamos por  $\mathbb{D}$  la distancia usual entre conjuntos,  $\mathbb{D}(A, B) := \sup_{x \in A} \text{dist}(x, B)$ , donde  $\text{dist}(x, B) := \inf_{y \in B} \|x - y\|$  y  $\|\cdot\|$  es la norma euclideana en  $\mathbb{R}^n$ .

**Teorema 3.1.2** (Theorem 5.3, [SDR09]). *Supongamos que existe un conjunto compacto  $C \subset \mathbb{R}^n$  tal que:*

1. *el conjunto  $S$  de soluciones óptimas del problema verdadero es no vacío y está contenido en  $C$ ,*

2. la función  $f(x)$  es a valores finitos y continua en  $C$ ,
3.  $\widehat{f}_N(x)$  converge c.s. a  $f(x)$ , cuando  $N \rightarrow \infty$ , uniformemente en  $x \in C$ , y
4. con probabilidad 1 para  $N$  suficientemente grande el conjunto  $\widehat{S}_N$  es no vacío y  $\widehat{S}_N \subset X$ .

Entonces  $\widehat{\vartheta}_N \rightarrow \vartheta^*$  y  $\mathbb{D}(\widehat{S}_N, S) \rightarrow 0$  c.s. cuando  $N \rightarrow \infty$ .

El siguiente resultado da condiciones suficientes para satisfacer los dos resultados anteriores.

**Teorema 3.1.3** (Theorem 7.48, [SDR09]). *Sea  $X$  un subconjunto compacto y no vacío de  $\mathbb{R}^n$ , y supongamos que:*

1. para todo  $x \in X$ , la función  $F(\cdot, \xi)$  es continua en  $x$ , para casi todo  $\xi \in \Xi$ ,
2. para todo  $x \in X$ ,  $F(x, \cdot)$  está dominada por una función integrable, y
3. la muestra de  $\widehat{f}_N$  es iid.

Entonces la función  $f(x)$  tiene valores finitos y es continua en  $X$ , y  $\widehat{f}_N(x)$  converge a  $f(x)$ , con probabilidad 1, uniformemente sobre  $X$ .

### Resultados asintóticos

Esta sección se basa íntegramente en [SDR09, §5.1.2 Asymptotics of the SAA Optimal Value]. Supongamos que la muestra es iid y consideremos una solución  $x \in X$ . Entonces sabemos que el estimador SAA  $\widehat{f}_N(x) = \frac{1}{N} \sum_{j=1}^N F(x, \xi^j)$ , de  $f(x) = \mathbb{E}(F(x, \xi))$ , es insesgado y tiene varianza  $\sigma^2(x)/N$ , donde asumimos  $\sigma^2(x) := \text{Var}(F(x, \xi))$  finito. Más aún, por el Teorema Central Límite se tiene que

$$N^{1/2} \left[ \widehat{f}_N(x) - f(x) \right] \xrightarrow{\mathcal{D}} Y_x, \quad (3.4)$$

donde  $\xrightarrow{\mathcal{D}}$  denota convergencia en distribución, y  $Y_x$  tiene distribución normal con media 0 y varianza  $\sigma^2(x)$ , notado  $Y_x \sim \mathcal{N}(0, \sigma^2(x))$ . Esto es,  $\widehat{f}_N(x)$  tiene *asintóticamente* distribución normal, i.e., para  $N$  grande,  $\widehat{f}_N(x)$  tiene aproximadamente distribución normal con media  $f(x)$  y varianza  $\sigma^2(x)/N$ .

Esto lleva al siguiente intervalo de confianza (aproximado) con confianza  $100(1-\alpha)\%$  para  $f(x)$ :

$$\left[ \widehat{f}_N(x) - z_{\alpha/2} \frac{\widehat{\sigma}(x)}{\sqrt{N}}, \widehat{f}_N(x) + z_{\alpha/2} \frac{\widehat{\sigma}(x)}{\sqrt{N}} \right] \quad (3.5)$$

donde  $z_{\alpha/2} := \Phi^{-1}(1 - \alpha/2)$  y

$$\widehat{\sigma}^2(x) := \frac{1}{N-1} \sum_{j=1}^N \left[ F(x, \xi^j) - \widehat{f}_N(x) \right]^2 \quad (3.6)$$

es el estimador de la varianza muestral de  $\sigma^2(x)$ . Esto es, el error de la estimación de  $f(x)$  es (estocásticamente) de orden  $O_p(N^{-1/2})$ .

Consideramos ahora el valor óptimo  $\widehat{\vartheta}_N$  del problema SAA (3.2). Claramente para todo  $x' \in X$  se tiene  $\widehat{f}_N(x') \geq \inf_{x \in X} \widehat{f}_N(x)$ . Tomando valor esperado en ambos lados de la desigualdad y minimizando en el lado izquierdo sobre todo  $x' \in X$ , obtenemos

$$\inf_{x \in X} \mathbb{E} \left[ \widehat{f}_N(x) \right] \geq \mathbb{E} \left[ \inf_{x \in X} \widehat{f}_N(x) \right] \quad (3.7)$$

Notar que la desigualdad (3.7) se tiene aún si  $f(x) = +\infty$  o  $f(x) = -\infty$  para algún  $x \in X$ . Como  $\mathbb{E} \left[ \widehat{f}_N(x) \right] = f(x)$ , sigue que  $\vartheta^* \geq \mathbb{E} \left[ \widehat{\vartheta}_N \right]$ . De hecho, típicamente  $\mathbb{E} \left[ \widehat{\vartheta}_N \right]$  es estrictamente menor que  $\vartheta^*$ , i.e.,  $\widehat{\vartheta}_N$  es estimador *sesgado hacia abajo* de  $\vartheta^*$ . El siguiente resultado muestra que el sesgo decrece monótonamente con el aumento del tamaño  $N$  de la muestra.

**Proposición 3.1.4** (Proposition 5.6, [SDR09]). *Sea  $\widehat{\vartheta}_N$  el valor óptimo del problema SAA (3.2) y suponer que la muestra es iid. Entonces  $\mathbb{E} \left[ \widehat{\vartheta}_N \right] \leq \mathbb{E} \left[ \widehat{\vartheta}_{N+1} \right] \leq \vartheta^*$  para todo  $N \in \mathbb{N}$ .*

**Teorema 3.1.5** (Theorem 5.7, [SDR09]). *Sea  $\widehat{\vartheta}_N$  el valor óptimo del problema SAA (3.2). Supongamos que la muestra es iid,  $X$  es compacto y se tienen las siguientes hipótesis:*

(A1) *Para algún  $\bar{x} \in X$  la esperanza  $\mathbb{E} [F(\bar{x}, \xi)^2]$  es finita.*

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

$$|f(x) - f(x')| \leq C(\xi) \|x - x'\| \quad (3.8)$$

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

*Entonces se tiene:*

$$\widehat{\vartheta}_N = \inf_{x \in S} \widehat{f}_N(x) + o_P(N^{-1/2}) \quad (3.9)$$

$$N^{1/2} \left( \widehat{\vartheta}_N - \vartheta^* \right) \xrightarrow{\mathcal{D}} \inf_{x \in S} Y(x) \quad (3.10)$$

*Más aún, si  $S = \{\bar{x}\}$  es un singleton, entonces*

$$N^{1/2} \left( \widehat{\vartheta}_N - \vartheta^* \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, \sigma^2(\bar{x}) \right) \quad (3.11)$$

### Análisis de validación

Esta sub-sección se basa íntegramente en [SDR09, §5.6.1 Estimation of the Optimality Gap]. Consideramos el problema de, dado un punto factible  $\bar{x} \in X$  candidato a ser solución óptima del problema real (3.1), estimar su gap de optimalidad

$$\text{gap}(\bar{x}) := f(\bar{x}) - \vartheta^*. \quad (3.12)$$

Claramente, para cada  $\bar{x} \in X$  factible,  $\text{gap}(\bar{x})$  es no-negativo y  $\text{gap}(\bar{x}) = 0$  ssi  $\bar{x}$  es una solución óptima del problema real.

**Cota Inferior.** Consideramos el valor óptimo del problema SAA (3.2). Por (3.1.4) se tiene que  $\vartheta^* \geq \mathbb{E}[\hat{\vartheta}_N]$ . Esto significa que  $\hat{\vartheta}_N$  proporciona una *cota inferior estadística* válida del valor óptimo  $\vartheta^*$  del problema real. La esperanza  $\mathbb{E}[\hat{\vartheta}_N]$  puede estimarse promediando. Esto es, resolver  $M$  veces el problema SAA (3.2) con muestras iid de tamaño  $N$  distintas. Sean  $\hat{\vartheta}_N^1, \dots, \hat{\vartheta}_N^M$  los valores óptimos de estos problemas SAA. Entonces

$$\bar{v}_{N,M} := \frac{1}{M} \sum_{m=1}^M \hat{\vartheta}_N^m \quad (3.13)$$

es un estimador insesgado de  $\mathbb{E}[\hat{\vartheta}_N]$ . Como las muestras son iid, y entonces  $\hat{\vartheta}_N^1, \dots, \hat{\vartheta}_N^M$  también, se tiene  $\text{Var}[\bar{v}_{N,M}] = \frac{1}{M} \text{Var}[\hat{\vartheta}_N]$ , y entonces podemos estimar la varianza de  $\bar{v}_{N,M}$  por

$$\hat{\sigma}_{N,M}^2 := \frac{1}{M} \left[ \frac{1}{M-1} \sum_{m=1}^M (\hat{\vartheta}_N^m - \bar{v}_{N,M})^2 \right] \quad (3.14)$$

En general, la variable aleatoria  $\hat{\vartheta}_N$ , y entonces sus replicaciones  $\hat{\vartheta}_N^j$ , no siguen una distribución normal, incluso aproximadamente (ver (3.1.5)). Sin embargo, por el Teorema Central Límite, la distribución de probabilidad del promedio  $\bar{v}_{N,M}$  se hace aproximadamente normal cuando  $M$  crece. En consecuencia, podemos usar

$$L_{N,M} := \bar{v}_{N,M} - t_{\alpha, M-1} \hat{\sigma}_{N,M}, \quad (3.15)$$

como una cota inferior aproximada con confianza  $(1 - \alpha)$  de la esperanza  $\mathbb{E}[\hat{\vartheta}_N]$ , donde  $t_{\alpha, \nu}$  es el valor crítico  $\alpha$  de una distribución t-student con  $\nu$  grados de libertad.

**Cota Superior.** También podemos estimar  $f(\bar{x})$  con muestreo. Esto es, sea  $\hat{f}_{N'}(\bar{x})$  el estimador muestral promedio de  $f(\bar{x})$ , basado en una muestra de tamaño  $N'$  generado independiente de las muestras involucradas en computar  $\bar{x}$ . Sea  $\hat{\sigma}_{N'}^2$  el estimador de la varianza de  $\hat{f}_{N'}(\bar{x})$ :

$$\hat{\sigma}_{N'}^2(\bar{x}) := \frac{1}{N'(N'-1)} \sum_{j=1}^{N'} \left[ F(\bar{x}, \xi^j) - \hat{f}_{N'}(\bar{x}) \right]^2 \quad (3.16)$$

Entonces

$$U_{N'}(\bar{x}) := \widehat{f}_{N'}(\bar{x}) + z_\alpha \widehat{\sigma}_{N'}(\bar{x}) \quad (3.17)$$

da una cota superior aproximada con confianza  $(1 - \alpha)$  para  $f(\bar{x})$ , donde  $z_\alpha := \Phi^{-1}(1 - \beta)$  para  $\Phi$  función de distribución acumulada de la normal estándar. Notar que como  $N'$  puede ser grande entonces usamos el valor crítico  $z_\alpha$  de la distribución normal estándar en vez de una  $t$ -distribución.

**Gap de optimalidad.** Tenemos que

$$\mathbb{E} \left[ \widehat{f}_{N'}(\bar{x}) - \bar{v}_{N,M} \right] = f(\bar{x}) - \mathbb{E}[\widehat{\vartheta}_N] = \text{gap}(\bar{x}) + \vartheta^* - \mathbb{E}[\widehat{\vartheta}_N] \geq \text{gap}(\bar{x}) \quad (3.18)$$

i.e.,  $\widehat{f}_{N'}(\bar{x}) - \bar{v}_{N,M}$  es un estimador sesgado del  $\text{gap}(\bar{x})$ . Además, la varianza de este estimador es igual a la suma de las varianzas de  $\widehat{f}_{N'}(\bar{x})$  y  $\bar{v}_{N,M}$ , y entonces

$$\widehat{f}_{N'}(\bar{x}) - \bar{v}_{N,M} + z_\alpha \sqrt{\widehat{\sigma}_{N'}^2(\bar{x}) + \widehat{\sigma}_{N,M}^2} \quad (3.19)$$

da una cota superior conservadora de  $\text{gap}(\bar{x})$  con confianza  $(1 - \alpha)$ . Decimos que esta cota superior es *conservadora* porque de hecho da, con confianza  $(1 - \alpha)$ , una cota superior para  $\text{gap}(\bar{x}) + \vartheta^* - \mathbb{E}[\widehat{\vartheta}_N]$ , y tenemos  $\vartheta^* - \mathbb{E}[\widehat{\vartheta}_N] \geq 0$ .

### Un algoritmo de búsqueda y validación de soluciones SAA

La sub-sección anterior sugiere directamente un procedimiento para encontrar soluciones SAA con menor gap de optimalidad. Consideramos el problema (3.1), es decir,

$$\vartheta^* := \min_{x \in X} \{f(x) := \mathbb{E}[F(x, \xi)]\}$$

Dados tamaño de muestra  $N$  y  $N'$ , número de iteraciones  $M$  y nivel de confianza  $1 - \alpha \in [0, 1]$ , el procedimiento es el siguiente.

1. Con  $M$  muestras de tamaño  $N$ ,  $\xi^{1,m}, \dots, \xi^{N,m}$ ,  $m = 1, \dots, M$ , iid todos entre sí, resolver  $M$  problemas SAA de tamaño  $N$ :

$$\widehat{\vartheta}_N^m := \min_{x \in X} \left\{ \widehat{f}_N(x) := \frac{1}{N} \sum_{k=1}^N F(x, \xi^{k,m}) \right\}, \quad m = 1, \dots, M$$

Obtenemos  $M$  candidatos a solución  $\bar{x}^m$ ,  $m = 1, \dots, M$ .

2. Cálculo de cota inferior para  $\vartheta^*$ . Calculamos

$$\begin{aligned} \bar{v}_{N,M} &:= \frac{1}{M} \sum_{m=1}^M \widehat{\vartheta}_N^m \\ \widehat{\sigma}_{N,M}^2 &:= \frac{1}{M} \left[ \frac{1}{M-1} \sum_{m=1}^M (\widehat{\vartheta}_N^m - \bar{v}_{N,M})^2 \right] \end{aligned}$$

Se tiene que

$$L_{N,M} := \bar{v}_{N,M} - t_{\alpha, M-1} \widehat{\sigma}_{N,M} \quad (3.20)$$

es cota inferior inferior de  $\vartheta^*$ , con confianza  $(1 - \alpha)$ .

3. Cálculo de cota superior para  $\vartheta^*$ . Sabemos que dado un punto  $x \in X$  factible,  $f(x)$  es cota superior de  $\vartheta^*$ ; y del primer paso tenemos  $M$  candidatos  $\bar{x}^m \in X$ ,  $m = 1, \dots, M$ , a óptimos del problema real (3.1), que en particular son factibles.

Tomamos una muestra iid de tamaño  $N'$ ,  $\xi^1, \dots, \xi^{N'}$ , del parámetro incierto  $\xi$ . Adicionalmente esta muestra es independiente de la muestra tomada en el paso 1. Para cada  $\bar{x}^m$ ,  $m = 1, \dots, M$ , calculamos

$$\begin{aligned} \widehat{f}_{N'}(\bar{x}^m) &:= \frac{1}{N'} \sum_{k=1}^{N'} F(\bar{x}^m, \xi^k) \\ \widehat{\sigma}_{N'}^2(\bar{x}^m) &:= \frac{1}{N'} \left[ \frac{1}{N'-1} \sum_{k=1}^{N'} (F(\bar{x}^m, \xi^k) - \widehat{f}_{N'}(\bar{x}^m))^2 \right] \end{aligned}$$

En particular tenemos que, para cada  $m = 1, \dots, M$ ,

$$\widehat{f}_{N'}(\bar{x}^m) + z_\alpha \widehat{\sigma}_{N'}(\bar{x}^m) \quad (3.21)$$

es una cota superior para  $f(\bar{x}^m)$ , y entonces para  $\vartheta^*$  también, con confianza  $(1 - \alpha)$ .

4. Cota superior para el gap de optimalidad  $f(\bar{x}^m) - \vartheta^*$ . Para cada  $\bar{x}^m$  factible,  $m = 1, \dots, M$ , obtenemos la siguiente cota superior para el gap de optimalidad  $f(\bar{x}^m) - \vartheta^*$ :

$$\widehat{f}_{N'}(\bar{x}^m) - \bar{v}_{N,M} + z_\alpha \sqrt{\widehat{\sigma}_{N'}^2(\bar{x}^m) + \widehat{\sigma}_{N,M}^2} \quad (3.22)$$

Usamos como mejor candidato a solución óptima del problema real la solución  $\bar{x}^m$  que de la menor cota superior para el gap de optimalidad.

### 3.1.2. Propiedades de estimadores SAA para Chance Constrained Problems

Consideramos ahora problemas con restricciones de probabilidad, o *chance constrained problems*, y su resolución usando la técnica Sample Average Approximation. Esta subsección se basa íntegramente en [SDR09, §5.7 Chance Constrained Problems].

Consideramos un chance constrained problem de la forma

$$\min_{x \in X} f(x) \text{ s.t. } p(x) \leq \alpha \quad (3.23)$$

donde  $X \subset \mathbb{R}^n$  es un conjunto cerrado,  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  es función continua,  $\alpha \in (0, 1)$  es un nivel de riesgo dado, y

$$p(x) := \mathbb{P}[C(x, \xi) > 0] \quad (3.24)$$

es la probabilidad de violación de la restricción en el punto  $x \in X$ . Asumimos que  $\xi$  es un vector aleatorio, cuya distribución de probabilidad  $\mathbb{P}$  está soportada en el conjunto

$\Xi \subset \mathbb{R}^d$ , y la función  $C : \mathbb{R}^n \times \Xi \rightarrow \mathbb{R}$  es una función de Carathéodory conocida explícitamente.

Podemos escribir la probabilidad  $p(x)$  en forma de esperanza,

$$p(x) = \mathbb{E} \left[ \mathbf{1}_{(0,\infty)} (C(x, \xi)) \right], \quad (3.25)$$

y, dada una muestra  $\xi^1, \dots, \xi^N$  del parámetro incierto  $\xi$ , podemos estimar esta probabilidad con la aproximación SAA correspondiente

$$\widehat{p}_N(x) := \frac{1}{N} \sum_{j=1}^N \mathbf{1}_{(0,\infty)} (C(x, \xi^j)). \quad (3.26)$$

Así,  $\widehat{p}_N(x)$  es igual a la proporción de veces que en la muestra se cumple la condición  $C(x, \xi) > 0$ . Consecuentemente podemos escribir la contraparte SAA del problema (3.23) como

$$\min_{x \in X} f(x) \text{ s.t. } \widehat{p}_N(x) \leq \alpha \quad (3.27)$$

### Consistencia

**Proposición 3.1.6** (Proposition 5.29, [SDR09]). *Sea  $C(x, \xi)$  una función de Carathéodory. Entonces la función  $p(x)$  y  $\widehat{p}_N(x)$  son semi-continuas inferiores, esto es, para todo  $x$  se tiene que  $p(x) \leq \liminf_{\bar{x} \rightarrow x} p(\bar{x})$  y análogamente para  $\widehat{p}_N$ . Suponer además que la muestra es iid. Entonces  $\widehat{p}_N$  epi-converge c.s. a  $p$  cuando  $N \rightarrow \infty$ , notado  $\widehat{p}_N \xrightarrow{e} p$  c.s.; esto es, c.s. se tiene que  $\widehat{p}_N$  epi-converge a  $p$ . Más aún, suponer que para todo  $x \in X$  se tiene*

$$\mathbb{P}(C(x, \xi) = 0) = 0, \quad (3.28)$$

*i.e.  $C(x, \xi) \neq 0$  c.s. Entonces la función  $p(x)$  es continua en  $X$  y  $\widehat{p}_N(x)$  converge a  $p(x)$  c.s. uniformemente sobre compactos de  $X$ .*

Para una definición exhaustiva de epi-convergencia de funciones reales extendidas y sus propiedades ver [SDR09, §7.1.6 Epiconvergence].

Como la función  $p(x)$  es semi-continua inferior y el conjunto  $X$  es cerrado, sigue que el conjunto factible del problema (3.23) es cerrado. Si más aún es no-vacío y acotado entonces el problema (3.23) tiene un conjunto no-vacío  $S$  de soluciones óptimas (recordar que la función objetivo  $f(x)$  se asume continua). Lo mismo aplica al problema SAA (3.27). Tenemos aquí la siguiente propiedad de consistencia para el valor óptimo  $\widehat{v}_N$  y el conjunto  $\widehat{S}_N$  de soluciones óptimas del problema SAA (3.27).

**Proposición 3.1.7** (Proposition 5.30, [SDR09]). *Supongamos que el conjunto  $X$  es compacto, la función  $f$  es continua,  $C(x, \xi)$  es función de Carathéodory, la muestra es iid, y se tiene lo siguiente:*

(a) existe solución óptima  $\bar{x}$  al problema verdadero tal que para todo  $\epsilon > 0$  existe  $x \in X$  con  $\|x - \bar{x}\| \leq \epsilon$  y  $p(x) < \alpha$

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

La condición (a) anterior es esencial para la consistencia de  $\hat{\vartheta}_N$  y  $\hat{S}_N$ . Pensar, por ejemplo, en alguna situación donde la condición  $p(x) \leq \alpha$  define un único punto  $\bar{x}$  tal que  $p(\bar{x}) = \alpha$ . Entonces cambios arbitrariamente pequeños en la restricción  $\hat{p}_N(x) \leq \alpha$  puede resultar en que el conjunto factible del problema SAA correspondiente se vuelve vacío.

Hasta aquí tenemos que bajo condiciones suaves de regularidad se tiene que el valor óptimo y soluciones óptimas del problema SAA (3.27) convergen c.s., cuando  $N \rightarrow \infty$ , a sus contrapartes respectivas del problema verdadero (3.23). Sin embargo, existen varios potenciales problemas con el enfoque aquí presentado. Por ejemplo, para que  $\hat{p}_N(x)$  sea un estimador razonablemente preciso de  $p(x)$ , el tamaño  $N$  de la muestra debe ser significativamente mayor a  $\alpha^{-1}$ . Para  $\alpha$  pequeño esto puede resultar en un tamaño de muestra muy grande. Otro problema es que típicamente la función  $\hat{p}_N(x)$  es discontinua en  $x$  y entonces el problema SAA (3.27) es un problema combinatorial difícil de resolver.

### Validación de una solución óptima

Consideramos ahora una técnica práctica para validar un candidato  $\bar{x} \in X$  a solución óptima del problema verdadero (3.23). Esta tarea es doble, pues se necesita verificar la factibilidad y optimalidad de  $\bar{x}$ . Por supuesto, si un punto  $\bar{x}$  es factible para el problema verdadero entonces  $\vartheta^* \leq f(\bar{x})$ , i.e.,  $f(\bar{x})$  da una cota superior para el valor óptimo  $\vartheta^*$ .

**Cota Superior.** Comencemos verificando la factibilidad del punto  $\bar{x}$ . Para ello necesitamos estimar la probabilidad  $p(x) = \mathbb{P}(C(\bar{x}, \xi) > 0)$ . Procedemos usando las técnicas de muestreo Monte Carlo. Para una muestra iid  $\xi^1, \dots, \xi^N$  definamos  $\kappa$  como el número de veces que la restricción  $C(\bar{x}, \xi) \leq 0$  es violada en la muestra, i.e.,

$$\kappa := \sum_{j=1}^N \mathbf{1}_{(0, \infty)}(C(\bar{x}, \xi^j)). \quad (3.29)$$

Entonces  $\hat{p}_N(\bar{x}) = \kappa/N$  es un estimador insesgado de  $p(\bar{x})$ , y  $\kappa$  tiene distribución binomial  $B(p(\bar{x}), N)$ . En adelante usamos la notación  $\mathbf{b}(k; \epsilon, N) := \sum_{j=1}^k \binom{N}{j} \epsilon^j (1 - \epsilon)^{N-j}$ .

Para  $\beta \in (0, 1)$  dado consideramos

$$U_{\beta, N}(\bar{x}) := \sup_{\rho \in [0, 1]} \{\rho : \mathbf{b}(\kappa; \rho, N) \geq \beta\} \quad (3.30)$$

Como  $U_{\beta, N}(\bar{x})$  es función de  $\kappa$  entonces es una variable aleatoria. Notar que  $\mathbf{b}(\kappa; \rho, N)$  es continua y monótonamente decreciente en  $\rho \in (0, 1)$ . Luego, de hecho, el supremo

de la mano derecha de (3.30) es alcanzado, y  $U_{\beta,N}(\bar{x})$  es igual al  $\bar{p} \in (0, 1)$  tal que  $\mathbf{b}(\kappa; \bar{p}, N) = \beta$ . Denotando  $V := \mathbf{b}(\kappa; p(\bar{x}), N)$  tenemos que

$$\begin{aligned} \mathbb{P}(p(\bar{x}) < U_{\beta,N}(\bar{x})) &= \mathbb{P}\left(V > \overbrace{\mathbf{b}(\kappa; \bar{p}, N)}^{\beta}\right) \\ &= 1 - \mathbb{P}(V \leq \beta) \\ &= 1 - \sum_{k=0}^N \mathbb{P}(V \leq \beta : \kappa = k) \mathbb{P}(\kappa = k) \end{aligned} \quad (3.31)$$

Como

$$\mathbb{P}(V \leq \beta : \kappa = k) = \begin{cases} 1 & \text{si } \mathbf{b}(\kappa; \bar{p}, N) \leq \beta \\ 0 & \text{si no} \end{cases} \quad (3.32)$$

y  $\mathbb{P}(\kappa = k) = \binom{N}{k} p(\bar{x})^k (1 - p(\bar{x}))^{N-k}$ , sigue que

$$\sum_{k=0}^N \mathbb{P}(V \leq \beta : \kappa = k) \mathbb{P}(\kappa = k) \leq \beta \quad (3.33)$$

y entonces

$$\mathbb{P}(p(\bar{x}) < U_{\beta,N}(\bar{x})) \geq 1 - \beta \quad (3.34)$$

Esto es,  $p(\bar{x}) < U_{\beta,N}(\bar{x})$  con probabilidad al menos  $1 - \beta$ . Luego, podemos tomar  $U_{\beta,N}(\bar{x})$  como una cota superior de  $p(\bar{x})$  con confianza  $(1 - \beta)$ .

Obtenemos que si  $U_{\beta,N}(\bar{x}) \leq \alpha$  entonces  $\bar{x}$  es solución factible del problema verdadero (3.23) con probabilidad al menos  $1 - \beta$ . En tal caso, podemos usar  $f(\bar{x})$  como una cota superior, con confianza  $(1 - \beta)$ , del valor óptimo  $\vartheta^*$  del problema verdadero (3.27). Como este procedimiento involucra sólo cálculo de  $C(\bar{x}, \xi^j)$ , entonces puede ser realizado para un tamaño de muestra  $N$  grande, con lo que la factibilidad de  $\bar{x}$  puede ser verificada con alta precisión, siempre que  $\alpha$  no sea muy pequeño.

**Cota Inferior.** Para enteros positivos  $N$  y  $M$ , y una constante  $\gamma \in [0, 1)$ , generamos  $M$  muestras independientes  $\xi^{1,m}, \dots, \xi^{N,m}$ ,  $m = 1, \dots, M$ , cada una de tamaño  $N$ , del vector aleatorio  $\xi$ . Para cada muestra, resolvemos el problema SAA con nivel de significancia  $\gamma$

$$\widehat{\vartheta}_{\gamma,N}^m := \min_{x \in X} f(x) \text{ s.t. } \sum_{j=1}^N \mathbf{1}_{(0,\infty)}(C(x, \xi^{j,m})) \leq \gamma N. \quad (3.35)$$

Si el problema es (3.35) es infactible o no-acotado asignamos su valor óptimo como  $\infty$  o  $-\infty$ , respectivamente. Podemos ver  $\widehat{\vartheta}_{\gamma,N}^m$ ,  $m = 1, \dots, M$ , como una muestra iid de la variable aleatoria  $\widehat{\vartheta}_{\gamma,N}$ , donde  $\widehat{\vartheta}_{\gamma,N}$  es el valor óptimo del respectivo problema SAA con nivel de significancia  $\gamma$ . A continuación ordenamos los valores en orden no-decreciente,  $\widehat{\vartheta}_{\gamma,N}^{(1)} \leq \dots \leq \widehat{\vartheta}_{\gamma,N}^{(M)}$ . Queremos escoger un entero  $L \in \{1, \dots, M\}$  y usar la cantidad aleatoria  $\widehat{\vartheta}_{\gamma,N}^{(L)}$  como una cota inferior del valor óptimo verdadero  $\vartheta^*$ .

Analizamos el procedimiento resultante para acotar. Sea  $\bar{x} \in X$  un punto factible para el problema, i.e., satisface

$$\mathbb{P}(C(\bar{x}, \xi) > 0) \leq \alpha. \quad (3.36)$$

Como  $\sum_{j=1}^N \mathbf{1}_{(0, \infty)}(C(x, \xi^{j,m}))$  tiene distribución binomial con probabilidad de éxito igual a la probabilidad del evento  $\{C(\bar{x}, \xi) > 0\}$ , sigue que  $\bar{x}$  es factible para el problema (3.35) con probabilidad al menos

$$\sum_{i=0}^{\lfloor \gamma N \rfloor} \binom{N}{i} \alpha^i (1 - \alpha)^{N-i} = \mathbf{b}(\lfloor \gamma N \rfloor; \alpha, N) =: \theta_N. \quad (3.37)$$

Cuando  $\bar{x}$  es factible para (3.35) por supuesto se tiene  $\widehat{\vartheta}_{\gamma, N}^m \leq f(\bar{x})$ . Sea  $\epsilon > 0$  una constante arbitraria y  $\bar{x}$  un punto factible para el problema verdadero tal que  $f(\bar{x}) \leq \vartheta^* + \epsilon$ . Entonces para cada  $m \in \{1, \dots, M\}$  tenemos

$$\theta := \mathbb{P}\left(\widehat{\vartheta}_{\gamma, N}^m \leq \vartheta^* + \epsilon\right) \geq \mathbb{P}\left(\widehat{\vartheta}_{\gamma, N}^m \leq f(\bar{x})\right) \geq \theta_N \quad (3.38)$$

Ahora, en el caso de  $\widehat{\vartheta}_{\gamma, N}^{(L)} > \vartheta^* + \epsilon$ , la correspondiente realización de la secuencia aleatoria  $\widehat{\vartheta}_{\gamma, N}^1, \dots, \widehat{\vartheta}_{\gamma, N}^M$  contiene menos de  $L$  elementos que son menores o iguales que  $\vartheta^* + \epsilon$ . Como los elementos de la secuencia aleatoria son independientes, la probabilidad del último evento es  $\mathbf{b}(L-1; \theta, M)$ . Como  $\theta \geq \theta_N$ , tenemos que  $\mathbf{b}(L-1; \theta, M) \leq \mathbf{b}(L-1; \theta_N, M)$ . Entonces,  $\mathbb{P}\left(\widehat{\vartheta}_{\gamma, N}^{(L)} > \vartheta^* + \epsilon\right) \leq \mathbf{b}(L-1; \theta_N, M)$ . Como la desigualdad resultante es válida para todo  $\epsilon > 0$  llegamos a la cota

$$\mathbb{P}\left(\widehat{\vartheta}_{\gamma, N}^{(L)} > \vartheta^*\right) \leq \mathbf{b}(L-1; \theta_N, M). \quad (3.39)$$

Obtenemos el siguiente resultado.

**Proposición 3.1.8** (Proposition 5.33, [SDR09]). *Dados  $\beta \in (0, 1)$  y  $\gamma \in [0, 1)$ , elijamos enteros positivos  $M, N$  y  $L$  tales que*

$$\mathbf{b}(L-1; \theta_N, M) \leq \beta \quad (3.40)$$

donde  $\theta_N := \mathbf{b}(\lfloor \gamma N \rfloor; \alpha, N)$ . Entonces

$$\mathbb{P}\left(\widehat{\vartheta}_{\gamma, N}^{(L)} > \vartheta^*\right) \leq \beta. \quad (3.41)$$

Para  $N$  y  $M$  dados, es mejor tomar el mayor entero  $L \in \{1, \dots, M\}$  satisfaciendo la condición (3.40). Esto es, para

$$L^* := \max_{1 \leq L \leq M} \{L : \mathbf{b}(L-1; \theta_N, M) \leq \beta\}, \quad (3.42)$$

se tiene que la cantidad aleatoria  $\widehat{\vartheta}_{\gamma, N}^{(L^*)}$  da una cota inferior para el valor óptimo verdadero  $\vartheta^*$  con probabilidad al menos  $1 - \beta$ . Si no existe  $L \in \{1, \dots, M\}$  satisfaciendo (3.40), la cota inferior, por definición, es  $-\infty$ .

La pregunta que surge en relación al esquema de acotamiento señalado es cómo escoger  $M$ ,  $N$  y  $\gamma$ . En el caso en que el conjunto  $X$  y las funciones  $f(\cdot)$  y  $C(\cdot, \xi)$ ,  $\xi \in \Xi$ , son convexas es ventajoso tomar  $\gamma = 0$ , porque entonces necesitamos resolver el problema convexo siguiente

$$\min_{x \in X} f(x) \text{ s.t. } C(x, \xi^j) \leq 0, \quad j = 1, \dots, N, \quad (3.43)$$

en vez del problema combinatorial (3.35). Notar que para  $\gamma = 0$  tenemos que  $\theta_N = (1 - \alpha)^N$  y la cota (3.40) toma la forma

$$\sum_{k=0}^{L-1} \binom{M}{k} (1 - \alpha)^{Nk} [1 - (1 - \alpha)^N]^{M-k} \leq \beta \quad (3.44)$$

En la práctica,  $M$  está acotado por el esfuerzo computacional requerido para resolver  $M$  problemas de la forma (3.35). Notar que el esfuerzo por problema es mayor mientras el tamaño  $N$  de la muestra es mayor. Para  $L = 1$  (que es el menor valor de  $L$ ) y  $\gamma = 0$ , el lado izquierdo de (3.44) es igual a  $[1 - (1 - \alpha)^N]^M$ . Notar que  $(1 - \alpha)^N \approx e^{-\alpha N}$  para  $\alpha > 0$  cercano a 0. Luego, si  $\alpha N$  es grande, uno necesitará un gran  $M$  para hacer  $[1 - (1 - \alpha)^N]^M$  menor que, por ejemplo,  $\beta = 0,01$ , y entonces obtener una cota inferior significativa. Por ejemplo, para  $\alpha N = 7$  tenemos que  $e^{-\alpha N} = 0,0009$ , y necesitaríamos  $M > 5000$  para hacer  $[1 - (1 - \alpha)^N]^M$  menor a 0,01. Por ello, para  $\gamma = 0$  es recomendable tomar  $N$  no más grande que, por ejemplo,  $2/\alpha$ .

### Un algoritmo de búsqueda y validación de soluciones SAA

La sub-sección anterior sugiere directamente un procedimiento para encontrar soluciones SAA con menor gap de optimalidad. Consideramos el problema (3.23) es decir,

$$\vartheta^* := \min_{x \in X} f(x) \quad \text{s.t.} \quad p(x) := \mathbb{P}[C(x, \xi) > 0] \leq \alpha$$

Dados tamaño de muestra  $N$  y  $N'$ , número de iteraciones  $M$ ,  $\gamma \in (0, 1)$  y nivel de confianza  $1 - \beta \in [0, 1]$ , el procedimiento es el siguiente.

1. Con  $M$  muestras de tamaño  $N$ ,  $\xi^{1,m}, \dots, \xi^{N,m}$ ,  $m = 1, \dots, M$ , iid todos entre sí, resolver  $M$  problemas SAA de tamaño  $N$ :

$$\widehat{\vartheta}_{\gamma, N}^m := \min_{x \in X} f(x) \quad \text{s.t.} \quad \sum_{j=1}^N \mathbf{1}_{(0, \infty)}(C(x, \xi^{j,m})) \leq \gamma N$$

Obtenemos  $M$  candidatos a solución  $\bar{x}^m$ ,  $m = 1, \dots, M$ .

2. Cálculo de cota inferior para  $\vartheta^*$ . Ordenamos los valores  $\widehat{\vartheta}_{\gamma,N}^m$ ,  $m = 1, \dots, M$ , de menor a mayor,  $\widehat{\vartheta}_{\gamma,N}^{(1)} \leq \dots \leq \widehat{\vartheta}_{\gamma,N}^{(M)}$ . Definimos

$$\begin{aligned}\theta_N &:= \mathbf{b}(\lfloor \gamma N \rfloor; \alpha, N) \\ L &:= \max\{L \in \mathbb{N} : \mathbf{b}(L-1; \theta_N, M) \leq \beta\}\end{aligned}$$

Se tiene que  $\widehat{\vartheta}_{\gamma,N}^{(L)}$  es cota inferior de  $\vartheta^*$ , con confianza  $(1 - \beta)$ .

3. Cálculo de cota superior para  $\vartheta^*$ . Tomamos una muestra iid de tamaño  $N'$ ,  $\xi^1, \dots, \xi^{N'}$ , del parámetro incierto  $\xi$ . Adicionalmente esta muestra es independiente de la muestra tomada en el paso 1. Para cada  $\bar{x}^m$ ,  $m = 1, \dots, M$ , calculamos

$$U_{\beta,N'}(\bar{x}^m) := \sup_{\rho \in [0,1]} \{\rho : \mathbf{b}(\kappa(\bar{x}^m); \rho, N') \geq \beta\} \quad (3.45)$$

donde  $\kappa(\bar{x}^m) := \sum_{j=1}^{N'} \mathbf{1}_{(0,\infty)}(C(\bar{x}^m, \xi^j))$ . Se tiene que  $U_{\beta,N'}(\bar{x}^m)$  es cota superior, con confianza  $(1 - \beta)$ , de  $p(\bar{x}^m)$ . Luego, si  $U_{\beta,N'}(\bar{x}^m) \leq \alpha$ , entonces  $\bar{x}^m$  es punto factible para el problema verdadero (3.23) y entonces  $f(\bar{x}^m)$  es cota superior para  $\vartheta^*$ , todo con confianza  $(1 - \beta)$ .

4. Cota superior para el gap de optimalidad  $f(\bar{x}^m) - \vartheta^*$ . Para cada  $\bar{x}^m$  factible obtenido en el paso 3.  $f(\bar{x}^m) - \widehat{\vartheta}_{\gamma,N}^{(L)}$  es cota superior, con confianza  $(1 - \beta)$ , para el gap de optimalidad  $f(\bar{x}^m) - \vartheta^*$ .

Usamos como mejor candidato a solución óptima del problema real (3.23) la solución  $\bar{x}^m$  que sea factible (paso 3.) y que de la menor cota superior para el gap de optimalidad.

## 3.2. Análisis SAA para modelos sin recourse

Consideramos el problema de planificación minera con elección de proceso, sin recourse, definido en (2.1):

$$\begin{aligned}\text{máx} & -(v^e)^t x^e + (W^p \rho - v^p)^t x^p \\ \text{s.a.} & x_a^e \leq x_b^e \quad \forall (a, b) \in \mathcal{P} \\ & x^p \leq x^e \\ & (c^e)^t x^e \leq B^e \\ & (c^p)^t x^p \leq B^p \\ & x^e, x^p \in \{0, 1\}^{\mathcal{B}}.\end{aligned}$$

Denotamos  $\Pi$  a la función de beneficio sin recourse:

$$\Pi(x^e, x^p, \rho) := -(v^e)^t x^e + (W^p \rho - v^p)^t x^p$$

y  $X$  al conjunto de planes de extracción y procesamiento factibles:

$$X := \left\{ x = (x^e, x^p) : \begin{array}{l} x_a^e \leq x_b^e \quad \forall (a, b) \in \mathcal{P} \\ x^p \leq x^e \\ (c^e)^t x^e \leq B^e \\ (c^p)^t x^p \leq B^p \\ x^e, x^p \in \{0, 1\}^{\mathcal{B}}. \end{array} \right\}$$

Notar que solamente la función objetivo  $\Pi$  depende del parámetro  $\rho \in \mathbb{R}_+^{\mathcal{B}}$ ,  $\Pi(\cdot, \rho)$  es lineal para toda ley  $\rho$ , y  $\Pi(x, \cdot)$  es lineal-afín para todo plan  $x$  factible. En particular,  $\Pi(\cdot, \cdot)$  es continua.

Consideramos ahora que el parámetro  $\rho \in \mathbb{R}_+^{\mathcal{B}}$  está sujeto a incertidumbre. Asumimos que  $\rho$  sigue una distribución estocástica  $\mathbb{P}$ , no necesariamente conocida, y que la  $\sigma$ -álgebra asociada  $\mathcal{F}$  es boreliana, con lo que  $\Pi(\cdot, \cdot)$  es Carathéodory pues  $\Pi(\cdot, \cdot)$  es continua. Podemos asumir también que el vector de leyes  $\rho$  está soportado en un conjunto  $\Xi \subset \mathbb{R}_+^{\mathcal{B}}$  y, más aún,  $\Xi$  es compacto, con lo que  $\Pi(x, \tilde{\rho}) \in \mathcal{L}^\infty(\Xi, \mathcal{F}, \mathbb{P})$  para todo plan  $x$  factible.

### 3.2.1. Modelo de minimización de CVaR

Recordemos el modelo de minimización de CVaR y su aproximación SAA, la cual presentamos en §2.2.2. Dado un nivel de riesgo  $\epsilon \in (0, 1]$ , el modelo de minimización de CVaR es

$$\vartheta^* := \min_{x \in X, t \in \mathbb{R}} \mathbb{E} \left[ t + \frac{1}{\epsilon} (-\Pi(x, \rho) - t)^+ \right] \quad (3.46)$$

Y dada una muestra del parámetro incierto  $\rho^1, \dots, \rho^N$ , la aproximación SAA del modelo es

$$\widehat{\vartheta}_N := \min_{x \in X, t \in \mathbb{R}} \frac{1}{N} \sum_{k=1}^N \left[ t + \frac{1}{\epsilon} (-\Pi(x, \rho^k) - t)^+ \right]. \quad (3.47)$$

A continuación llamamos al problema (3.46) *problema verdadero* y al problema (3.47) su *aproximación SAA*.

Usamos la notación de la sección §3.1.1 y notamos

$$F(x, t, \rho) := t + \frac{1}{\epsilon} [-\Pi(x, t, \rho) - t]^+,$$

$$f(x, t) := \mathbb{E} [F(x, t, \rho)],$$

y, dada una muestra iid  $\rho^1, \dots, \rho^N$  del parámetro  $\rho$ ,

$$\widehat{f}_N(x, t) := \frac{1}{N} \sum_{k=1}^N F(x, t, \rho^k).$$

Así, el problema de minimización de CVaR (3.46) es equivalente a

$$\vartheta^* = \min_{x,t} f(x,t) \quad \text{s.t.} \quad x \in X, t \in \mathbb{R},$$

y su aproximación SAA (3.47) es, equivalentemente,

$$\widehat{\vartheta}_N = \min_{x,t} \widehat{f}_N(x,t) \quad \text{s.t.} \quad x \in X, t \in \mathbb{R}.$$

Destacamos que para todo plan  $x \in X$  se tiene que la siguiente expresión para el CVaR, con riesgo  $\epsilon$ , de la variable aleatoria de pérdidas  $-\Pi(x, \rho)$ :

$$\text{CVaR}_\epsilon(-\Pi(x, \rho)) = \min_{t \in \mathbb{R}} f(x, t).$$

Definimos  $S$  como el conjunto de soluciones óptimas del problema verdadero (3.46) en  $X \times \mathbb{R}$ , y  $\widehat{S}_N$  el conjunto de soluciones óptimas de la aproximación SAA (3.47), también siguiendo la notación de §3.1.1.

### Consistencia

Para analizar la convergencia de los estimadores SAA  $\widehat{\vartheta}_N$  y  $\widehat{S}_N$  a sus respectivos contrapartes  $\vartheta^*$  y  $S$  aplicamos los resultados de consistencia de §3.1.1. Primeramente, para aplicar el Teorema 3.1.3, consideramos el conjunto  $C := X \times [-\bar{\pi}, -\underline{\pi}]$ , donde

$$\underline{\pi} := \min_{x \in X} \min_{\rho \in \Xi} \Pi(x, \rho)$$

y

$$\bar{\pi} := \max_{x \in X} \max_{\rho \in \Xi} \Pi(x, \rho).$$

Estos valores son finitos, y entonces  $[-\bar{\pi}, -\underline{\pi}]$  es compacto. En efecto, como  $\Xi$  es compacto y  $\Pi(x, \rho)$  es lineal afín en  $\rho$ , entonces  $\max_{\rho \in \Xi} \Pi(x, \rho)$  y  $\min_{\rho \in \Xi} \Pi(x, \rho)$  son finitos, para todo  $x \in X$ , y se concluye porque  $X \subset \{0, 1\}^B$  contiene un número finito de puntos.

Notar que podemos asumir que la solución  $(x^*, t^*)$  de (3.46) está en  $C$ , y así mismo, dada cualquier muestra  $\rho^1, \dots, \rho^N$  distribuido según  $\mathbb{P}$ , la solución  $(\widehat{x}_N, \widehat{t}_N)$  del problema SAA (3.47) puede asumirse también contenida en  $C$ . En efecto, sean  $x \in X$  y  $\rho \in \Xi$  cualesquiera. Se tiene para todo  $t \leq -\bar{\pi}$  que

$$F(x, t, \rho) = -\left(\frac{1}{\epsilon} - 1\right)t - \frac{1}{\epsilon}\Pi(x, \rho) \geq \left(\frac{1}{\epsilon} - 1\right)\bar{\pi} - \frac{1}{\epsilon}\Pi(x, \rho) = F(x, -\bar{\pi}, \rho)$$

y para todo  $t \geq -\underline{\pi}$  que

$$F(x, t, \rho) = t \geq -\underline{\pi} = F(x, -\underline{\pi}, \rho).$$

Sigue que

$$\begin{aligned} \mathbb{E}[F(x, -\bar{\pi}, \rho)] &\leq \mathbb{E}[F(x, t, \rho)] && \forall x \in X, t \leq -\bar{\pi}, \\ \mathbb{E}[F(x, -\underline{\pi}, \rho)] &\leq \mathbb{E}[F(x, t, \rho)] && \forall x \in X, t \geq -\underline{\pi}, \end{aligned}$$

y

$$\begin{aligned} \frac{1}{N} \sum_{k=1}^N F(x, -\bar{\pi}, \rho) &\leq \frac{1}{N} \sum_{k=1}^N F(x, t, \rho) && \forall x \in X, t \leq -\bar{\pi}, \\ \frac{1}{N} \sum_{k=1}^N F(x, -\underline{\pi}, \rho) &\leq \frac{1}{N} \sum_{k=1}^N F(x, t, \rho) && \forall x \in X, t \geq -\underline{\pi} \end{aligned}$$

para cualquier muestra  $\rho^1, \dots, \rho^N$  de  $\rho$  en  $\Xi$ . Así, es claro que

$$\vartheta^* = \min_{(x,t) \in C} f(x, t)$$

y

$$\widehat{\vartheta}_N = \min_{(x,t) \in C} \widehat{f}_N(x, t).$$

Asumiendo que la muestra de  $\rho$  es iid, aplicamos el Teorema 3.1.3 usando el conjunto compacto no vacío  $C$ , pues trivialmente  $F(\cdot, \cdot, \rho)$  es continua en todo  $(x, t) \in C$ , para todo  $\rho \in \Xi$ , y también  $F(x, t, \cdot)$  es integrable, para todo  $(x, t) \in C$ . Obtenemos así que  $f(x, t)$  es finita y continua en  $C$ , y que  $\widehat{f}_N(x, t)$  converge a  $f(x, t)$ , c.s., uniformemente en  $C$ , con lo que la Proposición 3.1.1 y Teorema 3.1.2 nos aseguran la consistencia del problema SAA. Hemos demostrado el siguiente resultado.

**Proposición 3.2.1.** *Consideramos, para un nivel de riesgo dado  $\epsilon \in (0, 1]$ , los problemas de minimización de CVaR verdadero (3.46) y su aproximación SAA (3.47). Asumiendo que el vector de leyes  $\rho$  se distribuye según la distribución  $\mathbb{P}$  soportada en el compacto  $\Xi \subset \mathbb{R}_+^B$ , que la  $\sigma$ -álgebra asociada  $\mathcal{F}$  es boreliana, y que la muestra  $\rho^1, \dots, \rho^k$  de  $\rho$  es iid, se tiene que*

$$\widehat{\vartheta}_N \longrightarrow \vartheta^*$$

y

$$\mathbb{D}(\widehat{S}_N, S) \longrightarrow 0$$

cuando  $N \rightarrow \infty$ .

### Análisis estadístico

Buscamos aplicar el algoritmo de la sección §3.1.1 para encontrar, con el problema SAA (3.47), soluciones que minimicen el gap de optimalidad del problema de minimización de CVaR verdadero (3.46). El algoritmo mencionado es directamente aplicable, sin embargo se puede ser un poco más perspicaz y notar que, para un  $x$  fijo, el problema  $\min_{t \in \mathbb{R}} \widehat{f}_N(x, t)$  admite tomar un tamaño de muestra considerablemente mayor al tamaño de muestra que admite el problema  $\min_{x \in X, t \in \mathbb{R}} \widehat{f}_N(x, t)$ , i.e. el problema SAA (3.47). Esto, pues el último problema tiene una gran cantidad de variables de decisión binarias y restricciones; en cambio el primer problema tiene una única variable de decisión, que es continua, y  $N$  restricciones, asociadas a la linealización del problema. La idea de fondo es con el problema SAA (3.47) encontrar un plan  $\bar{x}$  candidato, y luego obtener un mejor candidato  $\bar{t}$  resolviendo el problema  $\min_{t \in \mathbb{R}} \widehat{f}_{N'}(\bar{x}, t)$  con una muestra de tamaño  $N'$  considerablemente mayor. De esta manera se puede obtener un mejor candidato  $(\bar{x}, \bar{t})$  a solución del problema real (3.46). Se propone explotar este recurso siguiendo la siguiente modificación del algoritmo de la sección §3.1.1.

**Algoritmo de búsqueda estadística de soluciones**

0. Definir parámetros de tamaños de muestra  $N$  y  $N'$ ,  $N' \gg N$ , número de repeticiones  $M$ , y nivel de confianza  $(1 - \alpha) \in (0, 1)$ .
1. Tomar  $M$  colecciones de muestras iid de tamaño  $N$ :  $\rho^{1,m}, \dots, \rho^{N,m}$ ,  $m = 1, \dots, M$ , todos iid entre sí. Resolver  $M$  aproximaciones SAA (3.47) con las  $M$  colecciones de muestras iid de tamaño  $N$ :

$$\hat{\vartheta}_N^m := \min_{x \in X, t \in \mathbb{R}} \hat{f}_N(x, t),$$

$m = 1, \dots, M$ . Obtenemos en particular  $\bar{x}^1, \dots, \bar{x}^M$ ,  $M$  candidatos a minimizadores para el problema verdadero (3.46).

2. Cálculo de cota inferior para  $\vartheta^* = \min_{x \in X, t \in \mathbb{R}} f(x, t)$ , el valor objetivo del problema verdadero (3.46). Calculamos

$$\begin{aligned} \bar{v}_{N,M} &:= \frac{1}{M} \sum_{m=1}^M \hat{\vartheta}_N^m \\ \hat{\sigma}_{N,M}^2 &:= \frac{1}{M} \left[ \frac{1}{M-1} \sum_{m=1}^M (\hat{\vartheta}_N^m - \bar{v}_{N,M})^2 \right] \end{aligned}$$

Se tiene que

$$L_{N,M} := \bar{v}_{N,M} - t_{\alpha, M-1} \hat{\sigma}_{N,M}$$

es cota inferior inferior de  $\vartheta^*$ , con confianza  $(1 - \alpha)$ .

3. Cálculo de cota superior para  $\vartheta^*$ . Tomamos una muestra iid de tamaño  $N'$ ,  $\rho^1, \dots, \rho^{N'}$ , del parámetro incierto  $\rho$ , que además sea independiente de las muestras del paso 1. Para cada plan  $\bar{x}^m$ ,  $m = 1, \dots, M$ , realizamos lo siguiente.

- 3.1 Obtenemos  $\bar{t}^m$  de resolver el problema

$$\min_{t \in \mathbb{R}} \hat{f}_{N'}(\bar{x}^m, t).$$

- 3.2 Calculamos

$$\begin{aligned} \hat{f}_{N'}(\bar{x}^m, \bar{t}^m) &:= \frac{1}{N'} \sum_{k=1}^{N'} F(\bar{x}^m, \bar{t}^m, \rho^k) \\ \hat{\sigma}_{N'}^2(\bar{x}^m, \bar{t}^m) &:= \frac{1}{N'} \left[ \frac{1}{N'-1} \sum_{k=1}^{N'} \left( F(\bar{x}^m, \bar{t}^m, \rho^k) - \hat{f}_{N'}(\bar{x}^m, \bar{t}^m) \right)^2 \right] \end{aligned}$$

Se tiene que

$$\hat{f}_{N'}(\bar{x}^m, \bar{t}^m) + z_{\alpha} \hat{\sigma}_{N'}(\bar{x}^m, \bar{t}^m) \quad (3.48)$$

es cota superior, con confianza  $(1 - \alpha)$ , de  $\min_{t \in \mathbb{R}} f(\bar{x}^m, t)$ , y entonces de  $\vartheta^*$  también.

4. Elección de solución con menor cota para gap de optimalidad  $f$ . Para cada candidato  $\bar{x}^m$ ,  $m = 1, \dots, M$ , es cota superior, con confianza  $(1 - \alpha)$ , para el gap de optimalidad  $\text{Gap}(x) := \min_{t \in \mathbb{R}} f(x, t) - \vartheta^*$ :

$$\widehat{f}_{N'}(\bar{x}^m, \bar{t}^m) - \bar{v}_{N,M} + z_\alpha \sqrt{\widehat{\sigma}_{N'}^2(\bar{x}^m, \bar{t}^m) + \widehat{\sigma}_{N,M}^2} \quad (3.49)$$

Usamos como mejor candidato a solución óptima del problema verdadero (3.46) la solución  $\bar{x}^m$  que de la menor cota superior estadística para el gap de optimalidad.

Como se verá en las sucesivas secciones, tienen una estructura análoga a la del problema de minimización de CVaR los modelos MCH y MCH- $\epsilon$ , así como también las versiones con recourse de los modelos de CVaR, MCH y MCH- $\epsilon$ . Esto permitirá que el algoritmo anterior pueda ser usado también para encontrar soluciones en dichos modelos, cambiando obviamente los problemas “verdaderos” y su aproximación SAA de manera acorde a cada enfoque. Así, con el fin de no ser innecesariamente repetitivos, haremos referencia a este algoritmo en las sucesivas secciones, dejando en claro siempre cuáles parámetros del algoritmo hay que modificar para hacerlo aplicable al modelo en cuestión

Hacemos notar que se podría pensar que si en una iteración  $m^* \in \{1, \dots, M\}$  se minimiza la cota superior (3.48) entonces en tal iteración también se minimiza la cota (3.49). Sin embargo esto no es cierto si  $\widehat{\sigma}_{N,M} > 0$ , lo cual es esperable. En efecto, la pregunta a responder básicamente es si, dados  $x, y, a \in \mathbb{R}$ ,  $u, v, b, \delta > 0$  cualesquiera, se tiene que si  $x - y \leq \delta(v - u)$  entonces  $x - y \leq \delta(\sqrt{v^2 + b^2} - \sqrt{u^2 + b^2})$ . Esto no es en general cierto, pues es falso que para todo  $u', v', b' > 0$  se tenga  $v' - u' \leq \sqrt{v'^2 + b'^2} - \sqrt{u'^2 + b'^2}$ .

### 3.2.2. Modelo MCH

Recordemos que, dado un nivel de riesgo  $\epsilon \in [0, 1]$  y una muestra  $\rho^1, \dots, \rho^N$  del parámetro  $\rho$ , el problema MCH “original” es

$$\widehat{\vartheta}_N := \min_{x \in X} \max_{\rho \in \mathcal{U}_\epsilon} -\Pi(x, \rho), \quad (3.50)$$

donde  $\mathcal{U}_\epsilon := \bar{\rho} + (1 - \epsilon) (\text{conv}(\{\rho^1, \dots, \rho^N\}) - \bar{\rho})$  y  $\bar{\rho} := \frac{1}{N} \sum_{k=1}^N \rho^k$ . Por Proposición 2.2.5 tenemos que este problema original es equivalente a

$$\widehat{\vartheta}_N = \min_{x \in X, t \in \mathbb{R}} \frac{1}{N} \sum_{k=1}^N \left[ \epsilon (-\Pi(x, \rho)) + (1 - \epsilon) \left( t + \frac{1}{1/N} [-\Pi(x, \rho) - t]^+ \right) \right]. \quad (3.51)$$

Podemos ver este último problema como la aproximación SAA, al usar la muestra  $\rho^1, \dots, \rho^N$ , del problema

$$\vartheta_N^* := \min_{x \in X, t \in \mathbb{R}} \mathbb{E} \left[ \epsilon (-\Pi(x, \rho)) + (1 - \epsilon) \left( t + \frac{1}{1/N} [-\Pi(x, \rho) - t]^+ \right) \right], \quad (3.52)$$

el cual, notemos, depende del tamaño  $N$  de la muestra. Inspirados en esto llamaremos al problema (3.52) *problema MCH “verdadero”* y al problema (3.51) *aproximación SAA del problema verdadero*.

El presente enfoque se propone para poder usar los resultados de la sección §3.1.1. Si bien éste es bastante limitado –estamos proponiendo aproximar un problema estocástico, que depende de un parámetro  $N$ , usando una muestra de tamaño  $N$ –, esto nos permitirá obtener finalmente un criterio de selección de soluciones “estadísticamente mejores” para el problema MCH original.

Usamos la notación de la sección §3.1.1 y notamos

$$F(x, t, \rho) := \epsilon (-\Pi(x, \rho)) + (1 - \epsilon) \left( t + \frac{1}{1/N} [-\Pi(x, \rho) - t]^+ \right),$$

$$f(x, t) := \mathbb{E} [F(x, t, \rho)],$$

y, dada una muestra  $\rho^1, \dots, \rho^N$  del parámetro  $\rho$ ,

$$\widehat{f}_N(x, t) := \frac{1}{N} \sum_{k=1}^N F(x, t, \rho^k).$$

Así, el problema MCH verdadero (3.52) es equivalente a

$$\vartheta_N^* = \min_{x, t} f(x, t) \quad \text{s.t.} \quad x \in X, t \in \mathbb{R},$$

y su aproximación SAA (3.51) es, equivalentemente,

$$\widehat{\vartheta}_N = \min_{x, t} \widehat{f}_N(x, t) \quad \text{s.t.} \quad x \in X, t \in \mathbb{R}.$$

Denotamos también como  $S$  al conjunto de soluciones óptimas del problema verdadero (3.52) y  $\widehat{S}_N$  al conjunto de soluciones óptimas del problema SAA (3.51).

**Consistencia:** para el modelo MCH obviamos el análisis de convergencia del problema SAA (3.51) al problema verdadero (3.52) al hacer crecer el tamaño de muestra  $N$ , pues tanto  $\widehat{\vartheta}_N$  como  $\vartheta_N^*$  dependen de  $N$ . Por otro lado, una pregunta razonable a hacerse es si, en algún sentido, la interpretación geométrica del problema MCH original (3.50) converge; sin embargo esto sale del alcance de la teoría y práctica de la técnica SAA para problemas de optimización estocástica.

### Análisis estadístico

Buscamos aplicar el algoritmo de la sección §3.1.1 para encontrar, usando la aproximación SAA (3.51), candidatos a soluciones para el problema verdadero (3.52) que minimicen, estadísticamente, el gap de optimalidad entre la aproximación SAA y el valor

óptimo del problema verdadero. Como se mencionó antes, este enfoque es bastante limitado, pues estamos aproximando el problema verdadero, que depende de  $N$ , usando una muestra de tamaño  $N$ . No obstante su obvia sub-optimalidad, igualmente es un enfoque válido, y en particular también será válido aplicar el algoritmo de §3.1.1 para seleccionar sus estadísticamente mejores soluciones.

Al adaptar el algoritmo de §3.1.1 al modelo MCH también podemos separar la elección del mejor candidato a plan  $\bar{x}$  de la elección del mejor candidato  $\bar{t}$ , de manera análoga a lo hecho en el análisis estadístico del modelo de CVaR, en §3.2.1. En consecuencia, el algoritmo de §3.2.1 es perfectamente aplicable al modelo MCH, siempre que se redefinan en dicho algoritmo los siguientes términos: el problema verdadero ahora corresponderá al problema (3.52), la aproximación SAA al problema (3.51), y usaremos  $\vartheta^*$ ,  $\hat{\vartheta}_N$ ,  $F$ ,  $f$ ,  $\hat{f}_N$  según la definición que se dio para el enfoque MCH, al inicio de esta sub-sección.

Hacemos notar finalmente que cuando  $\epsilon = 1$  el problema verdadero (3.52) no depende de  $N$ , pues es simplemente minimizar las pérdidas esperadas. Particularmente, el algoritmo de selección de soluciones entregará la solución que, estadísticamente, sea el mejor minimizador en el problema de minimizar la pérdida esperada.

### 3.2.3. Modelo MCH- $\epsilon$

Recordemos el modelo MCH- $\epsilon$  y su aproximación SAA, ambos introducidos en §2.2.3. Dado un nivel de riesgo  $\epsilon \in (0, 1]$  el problema MCH- $\epsilon$  era

$$\vartheta^* := \min_{x \in X, t \in \mathbb{R}} \mathbb{E} \left[ \epsilon (-\Pi(x, \rho)) + (1 - \epsilon) \left( t + \frac{1}{\epsilon} [-\Pi(x, \rho) - t]^+ \right) \right] \quad (3.53)$$

Y dada una muestra de tamaño  $N$  del parámetro incierto  $\rho, \rho^1, \dots, \rho^N$ , la aproximación SAA del problema MCH- $\epsilon$  era

$$\hat{\vartheta}_N := \min_{x \in X, t \in \mathbb{R}} \frac{1}{N} \sum_{k=1}^N \left[ \epsilon (-\Pi(x, \rho)) + (1 - \epsilon) \left( t + \frac{1}{\epsilon} [-\Pi(x, \rho) - t]^+ \right) \right]. \quad (3.54)$$

Nuevamente, usamos la notación de la sección §3.1.1:

$$F(x, t, \rho) := \epsilon (-\Pi(x, \rho)) + (1 - \epsilon) \left( t + \frac{1}{\epsilon} [-\Pi(x, \rho) - t]^+ \right),$$

$$f(x, t) := \mathbb{E} [F(x, t, \rho)],$$

y, dada una muestra  $\rho^1, \dots, \rho^N$  del parámetro  $\rho$ ,

$$\hat{f}_N(x, t) := \frac{1}{N} \sum_{k=1}^N F(x, t, \rho^k).$$

Así, el problema MCH- $\epsilon$  verdadero (3.53) es equivalente a

$$\vartheta^* = \min_{x,t} f(x,t) \quad \text{s.t.} \quad x \in X, t \in \mathbb{R},$$

y su aproximación SAA (3.54) es, equivalentemente,

$$\hat{\vartheta}_N = \min_{x,t} \hat{f}_N(x,t) \quad \text{s.t.} \quad x \in X, t \in \mathbb{R}.$$

Denotamos, nuevamente,  $S$  al conjunto de soluciones óptimas del problema verdadero (3.53) y  $\hat{S}_N$  al conjunto de soluciones óptimas del problema SAA (3.54).

### Consistencia

Queremos verificar la consistencia de los estimadores SAA  $\hat{\vartheta}_N$  y  $\hat{S}_N$ . De manera análoga a lo hecho para la consistencia del modelo de minimización de CVaR en §3.2.1, aquí buscamos aplicar primeramente el Teorema 3.1.3, para lo cual consideramos el conjunto  $C := X \times [-\bar{\pi}, -\underline{\pi}]$ , donde

$$\underline{\pi} := \min_{x \in X} \min_{\rho \in \Xi} \Pi(x, \rho)$$

y

$$\bar{\pi} := \max_{x \in X} \max_{\rho \in \Xi} \Pi(x, \rho).$$

Como se argumentó en dicha sub-sección,  $\bar{\pi}$  y  $\underline{\pi}$  son finitos, esencialmente porque el vector de leyes  $\rho$  está soportado en el conjunto  $\Xi$  compacto y  $X$  es de cardinal finito.

Veamos primeramente que toda solución óptima al problema verdadero (3.53) y al problema SAA (3.54) está contenida en  $C$ , cuando la muestra de  $\rho$  se distribuye según la distribución  $\mathbb{P}$ . En efecto, procedemos como se razonó en §3.2.1, y consideramos primeramente  $N > 1$  entero,  $x \in X$  y  $\rho \in \Xi$  cualesquiera. Recordando que

$$F(x, t, \rho) := -\epsilon \Pi(x, \rho) + (1 - \epsilon) \left( t + \frac{1}{\epsilon} [-\Pi(x, \rho) - t]^+ \right)$$

se tiene para todo  $t \leq -\bar{\pi}$  y  $\rho \in \Xi$  que  $F(x, t, \rho) \geq F(x, -\bar{\pi}, \rho)$  y para todo  $t \geq -\underline{\pi}$  y  $\rho \in \Xi$  que  $F(x, t, \rho) \geq F(x, -\underline{\pi}, \rho)$ . Sigue que

$$\begin{aligned} \mathbb{E}[F(x, -\bar{\pi}, \rho)] &\leq \mathbb{E}[F(x, t, \rho)] & \forall x \in X, t \leq -\bar{\pi}, \\ \mathbb{E}[F(x, -\underline{\pi}, \rho)] &\leq \mathbb{E}[F(x, t, \rho)] & \forall x \in X, t \geq -\underline{\pi}, \end{aligned}$$

y

$$\begin{aligned} \frac{1}{N} \sum_{k=1}^N F(x, -\bar{\pi}, \rho_k) &\leq \frac{1}{N} \sum_{k=1}^N F(x, t, \rho_k) & \forall x \in X, t \leq -\bar{\pi}, \\ \frac{1}{N} \sum_{k=1}^N F(x, -\underline{\pi}, \rho_k) &\leq \frac{1}{N} \sum_{k=1}^N F(x, t, \rho_k) & \forall x \in X, t \geq -\underline{\pi} \end{aligned}$$

para cualquier muestra  $\rho^1, \dots, \rho^N$  de  $\rho$  en  $\Xi$ . Luego, claramente

$$\vartheta^* = \min_{(x,t) \in C} \mathbb{E}[F(x, t)]$$

y

$$\hat{\vartheta}_N = \min_{(x,t) \in C} \frac{1}{N} \sum_{k=1}^N F(x, t, \rho^k).$$

El argumento sigue idéntico al usado en la sección de consistencia del modelo de CVaR, §3.2.1. Asumimos que la muestra de  $\rho$  es iid y aplicamos el Teorema 3.1.3 usando el conjunto compacto no vacío  $C$ . En efecto, trivialmente  $F(\cdot, \cdot, \rho)$  es continua en todo  $(x, t) \in C$ , para todo  $\rho \in \Xi$ , y también  $F(x, t, \cdot)$  es integrable, para todo  $(x, t) \in C$ . Obtenemos así que  $f(x, t)$  es finita y continua en  $C$ , y que  $\hat{f}_N(x, t)$  converge a  $f(x, t)$ , c.s., uniformemente en  $C$ , con lo que la Proposición 3.1.1 y Teorema 3.1.2 nos aseguran la consistencia del problema SAA para el modelo MCH- $\epsilon$ . Esto demuestra la siguiente proposición.

**Proposición 3.2.2.** *Consideramos, para un nivel de riesgo dado  $\epsilon \in (0, 1]$ , los problemas MCH- $\epsilon$  verdadero (3.53) y su aproximación SAA (3.54). Asumiendo que el vector de leyes  $\rho$  se distribuye según la distribución  $\mathbb{P}$  con soporte compacto  $\Xi \subset \mathbb{R}_+^B$  y  $\sigma$ -álgebra boreliana, y además que la muestra  $\rho^1, \dots, \rho^k$  de  $\rho$  es iid, se tiene que*

$$\hat{\vartheta}_N \longrightarrow \vartheta^*$$

y

$$\mathbb{D}(\hat{S}_N, S) \longrightarrow 0$$

cuando  $N \rightarrow \infty$ .

**Análisis Estadístico:** Para realizar el análisis estadístico de búsqueda de soluciones que minimicen el gap de optimalidad reutilizamos el algoritmo de §3.2.1 del modelo de CVaR, al igual que lo se hizo con el modelo MCH. Para que dicho algoritmo encuentre soluciones en el caso del modelo MCH- $\epsilon$ , debemos reinterpretar varios parámetros del algoritmo: el problema verdadero ahora corresponderá al problema (3.53), la aproximación SAA al problema (3.54), y usaremos  $\vartheta^*$ ,  $\hat{\vartheta}_N$ ,  $F$ ,  $f$ ,  $\hat{f}_N$  según la definición que se dio al inicio de esta sub-sección.

Se obtiene así un algoritmo para seleccionar, con el problema SAA (3.54), una solución al problema verdadero (3.53) que minimiza hasta cierto punto el gap de optimalidad, y también cotas superiores e inferiores estadísticas para  $\vartheta^*$ .

### 3.2.4. Modelo de minimización de VaR

Dado un nivel de riesgo  $\epsilon \in (0, 1)$  el modelo de minimización de VaR, recordemos, era

$$\vartheta^* := \min_{x,z} z \quad \text{s.t.} \quad \mathbb{P}(\Pi(x, \rho) + z < 0) \leq \epsilon, \quad x \in X. \quad (3.55)$$

Y dada una muestra  $\rho^1, \dots, \rho^N$  del parámetro incierto  $\rho$ , la aproximación SAA del modelo es

$$\widehat{\vartheta}_N := \min_{x,z} z \quad \text{s.t.} \quad \frac{1}{N} \sum_{j=1}^N \mathbf{1}_{(-\infty,0)}(\Pi(x, \rho) + z) \leq \epsilon, \quad x \in X. \quad (3.56)$$

Siguiendo la notación de la sección §3.1.2, definimos

$$\begin{aligned} f(x, z) &:= z \\ C(x, z, \rho) &:= -\Pi(x, \rho) - z \\ p(x, z) &:= \mathbb{P}(C(x, z, \rho) > 0) \\ \widehat{p}_N(x, z) &:= \frac{1}{N} \sum_{j=1}^N \mathbf{1}_{(0,\infty)}(C(x, z, \rho^j)) \end{aligned}$$

donde  $\rho^1, \dots, \rho^N$  es una muestra del parámetro  $\rho$ . Con esto, el problema (3.55) de minimización de VaR con nivel de riesgo  $\epsilon$  es equivalente a

$$\vartheta^* = \min_{x,z} f(z) \quad \text{s.t.} \quad p(x, z) \leq \epsilon, \quad x \in X,$$

y su aproximación SAA (3.56) es equivalente a

$$\widehat{\vartheta}_N = \min_{x,z} f(z) \quad \text{s.t.} \quad \widehat{p}_N(x, z) \leq \epsilon, \quad x \in X.$$

Definimos también como  $S$  al conjunto de soluciones óptimas del problema verdadero (3.55) y  $\widehat{S}_N$  al conjunto de soluciones óptimas del problema SAA (3.56).

### Consistencia

Para asegurar la consistencia del problema SAA basta aplicar la Proposición 3.1.7. En nuestro caso esta última proposición queda de la siguiente manera.

**Proposición 3.2.3.** *Supongamos que el vector de leyes  $\rho$  se distribuye según la distribución  $\mathbb{P}$  con soporte compacto  $\Xi \subset \mathbb{R}_+^{\mathcal{B}}$  y  $\sigma$ -álgebra boreliana. Si la muestra  $\rho^1, \dots, \rho^N$  del vector de leyes  $\rho \in \mathbb{R}^{\mathcal{B}}$  es iid y se cumple la siguiente condición*

(a') *existe solución óptima  $x^* \in X$  al problema verdadero (3.55) tal que  $F_{-\Pi(x^*, \rho)}$ , la función de distribución acumulada de la variable aleatoria  $-\Pi(x^*, \rho)$ , es tal que el conjunto  $F_{-\Pi(x^*, \rho)}^{-1}(\{1 - \epsilon\})$  tiene interior vacío*

*entonces  $\widehat{\vartheta}_N \rightarrow \vartheta^*$  y  $\mathbb{D}(\widehat{S}_N, S) \rightarrow 0$  cuando  $N \rightarrow \infty$ .*

*Demostración.* Primeramente notemos que trivialmente en nuestro problema la función objetivo es continua y  $C$  es Carathéodory en  $(x, z, \rho)$  si la  $\sigma$ -álgebra asociada a  $\rho$  es boreliana. Ahora, análogo a lo hecho para la consistencia del modelo de minimización de CVaR en §3.2.1, consideremos el conjunto  $K := X \times [-\bar{\pi}, -\underline{\pi}]$ , donde

$$\underline{\pi} := \min_{x \in X} \min_{\rho \in \Xi} \Pi(x, \rho)$$

y

$$\bar{\pi} := \max_{x \in X} \max_{\rho \in \Xi} \Pi(x, \rho).$$

Se tiene que  $K$  es compacto, pues  $X$  es de cardinal finito y el vector de leyes  $\rho$  está soportado en el conjunto  $\Xi$  compacto, con lo que  $\bar{\pi}$  y  $\underline{\pi}$  son finitos. Sigue que en los problemas (3.55) y (3.56) nos podemos restringir a  $K$ , cuando la muestra de  $\rho$  se distribuye según la distribución  $\mathbb{P}$ . En efecto, para todo  $x \in X$  y todo  $z < -\bar{\pi}$  se tiene  $p(x, z) = \hat{p}_N(x, z) = 1 > \epsilon$ , con lo que  $(-\infty, -\bar{\pi})$  no es parte del conjunto factible para  $z$  en los problemas (3.55) y (3.56); y también, para todo  $x \in X$  y todo  $z > -\underline{\pi}$ , se tiene  $p(x, z) = \hat{p}_N(x, z) = 0 < \epsilon$ , entonces, como estamos minimizando  $z$ , basta restringir  $z$  al conjunto  $[-\bar{\pi}, -\underline{\pi}]$ .

Por último, es fácil verificar que la hipótesis (a) de la Proposición 3.1.7 es equivalente a la hipótesis (a') en nuestro caso. Primero, como  $X \subseteq \{0, 1\}^{2\mathcal{B}}$  entonces (a) es equivalente a que exista  $x^*$  solución óptima al problema verdadero (3.55) tal que para todo  $\delta > 0$  existe  $z \in \mathbb{R}$  tal que  $|z - z^*| \leq \delta$  y tal que  $\mathbb{P}(\Pi(x^*, \rho) + z < 0) < \epsilon$ . Como se tiene que  $z^* = \min\{z : \mathbb{P}(\Pi(x^*, \rho) + z < 0) \leq \epsilon\}$  y que  $\mathbb{P}(\Pi(x^*, \rho) + z < 0)$  es decreciente como función de  $z$ , entonces la condición (a) es equivalente a que para todo  $\delta > 0$  se tenga  $\mathbb{P}(\Pi(x^*, \rho) + z^* + \delta < 0) < \epsilon$ , o equivalentemente,  $\mathbb{P}(-\Pi(x^*, \rho) \leq z^* + \delta) > 1 - \epsilon$ . Esto es,  $F_{-\Pi(x^*, \rho)}(z^* + \delta) > 1 - \epsilon$  para todo  $\delta > 0$ , de lo que se concluye que las condiciones (a) y (a') son equivalentes.  $\square$

Hacemos notar que lo que exige la condición (a') es que la función de distribución acumulada de  $-\Pi(x^*, \rho)$  no tenga una “meseta” o “planicie” en el nivel  $1 - \epsilon$ .

### Análisis estadístico

Buscamos aplicar el algoritmo de la sección §3.1.2 de búsqueda soluciones óptimas que minimicen el gap de optimalidad entre el problema aproximado (3.56) y el problema real (3.55).

Notemos que la naturaleza de nuestro problema nos permite tomar una cota superior estadística más ajustada que la que se calcula en el paso 4. del algoritmo de §3.1.2. Para ello consideramos el siguiente resultado.

**Lema 3.2.4.** *Sea  $x \in X$  un plan,  $\rho^1, \dots, \rho^{N'}$  una muestra iid de  $\rho$  y  $1 - \beta$  un nivel de confianza. Definamos  $\pi_j := \Pi(x, \rho^j)$ ,  $j = 1, \dots, N'$ . Consideremos el problema de,*

dada la muestra de tamaño  $N'$ , calcular la menor cota superior estadística, con confianza  $1 - \beta$ , para  $\text{VaR}_\epsilon(x) = \min_{z \in \mathbb{R}} z$  s.t.  $\mathbb{P}(\Pi(x, \rho) + z < 0) \leq \epsilon$ :

$$\begin{aligned} z^* &:= \inf_{z, \kappa} z \\ \text{s.t. } \kappa &= \sum_{j=1}^{N'} \mathbf{1}_{(0, \infty)}(-\Pi(x, \rho^j) - z) \\ \sup_{\nu \in [0, 1]} \left\{ \nu : \sum_{j=0}^{\kappa} \nu^j (1 - \nu)^{N'-j} \geq \beta \right\} &\leq \epsilon. \end{aligned} \quad (3.57)$$

Se tiene que el valor objetivo de (3.57) es

$$z^* = -\pi_{(\kappa^*+1)},$$

donde

$$\kappa^* := \max\{ \kappa \in \mathbb{N} : \sum_{j=0}^{\kappa} \epsilon^j (1 - \epsilon)^{N'-j} \leq \beta \}$$

y  $\pi_{(j)}$ ,  $j = 1, \dots, N'$ , son los estadísticos de orden de  $\pi_1, \dots, \pi_{N'}$ , i.e.  $\pi_{(1)} \leq \dots \leq \pi_{(N')}$ .

*Demostración.* Primeramente, notemos que, dado  $\kappa \geq 0$ , la restricción

$$\sup_{\nu \in [0, 1]} \left\{ \nu : \sum_{j=0}^{\kappa} \binom{N'}{j} \nu^j (1 - \nu)^{N'-j} \geq \beta \right\} \leq \epsilon$$

es equivalente a

$$\sum_{j=0}^{\kappa} \binom{N'}{j} \epsilon^j (1 - \epsilon)^{N'-j} \leq \beta, \quad (3.58)$$

pues  $\Psi_{\kappa, N'}(\nu) := \sum_{j=0}^{\kappa} \nu^j (1 - \nu)^{N'-j}$  es estrictamente decreciente en  $\nu \in [0, 1]$ , salvo si  $\kappa = N'$ . El caso  $\kappa = N'$  no nos interesa pues no es una solución factible del problema (3.57).

Segundo, como

$$\begin{aligned} \kappa = \kappa(z) &:= \sum_{j=1}^{N'} \mathbf{1}_{(0, \infty)}(-\Pi(x, \rho^j) - z) \\ &= \#\{ j \in \{1, \dots, N'\} : -z > \pi_{(j)} \} \end{aligned}$$

es decreciente en  $z$  entonces queremos encontrar el máximo  $\kappa$  tal que se satisfaga (3.58). Sea  $\kappa^*$  tal  $\kappa$ , i.e.  $\kappa^* := \max\{ \kappa \in \mathbb{N} : \sum_{j=0}^{\kappa} \binom{N'}{j} \epsilon^j (1 - \epsilon)^{N'-j} \leq \beta \}$ .

Consideremos primero el caso  $\pi_{(\kappa^*)} < \pi_{(\kappa^*+1)}$ . Se tiene que  $z$  satisface  $\kappa(z) = \kappa^*$  ssi  $z \in [-\pi_{(\kappa^*+1)}, -\pi_{(\kappa^*)}]$ , luego  $z^* = -\pi_{(\kappa^*+1)}$ . Consideremos ahora el caso  $\pi_{(\kappa^*)} = \pi_{(\kappa^*+1)}$ . Notemos que  $z := -\pi_{(\kappa^*+1)}$  es factible para (3.57), pues

$$\kappa(-\pi_{(\kappa^*+1)}) = \#\{ j \in \{1, \dots, N'\} : \pi_{(\kappa^*+1)} > \pi_{(j)} \} < \kappa^*$$

y entonces, por definición de  $\kappa^*$ ,  $\kappa(-\pi_{(\kappa^*+1)})$  satisface (3.58). Más aún,  $z$  es óptimo, pues  $\forall \delta > 0$  se tiene que  $\kappa(-\pi_{(\kappa^*+1)} - \delta) = \#\{ j \in \{1, \dots, N'\} : \pi_{(\kappa^*+1)} + \delta > \pi_{(j)} \} \geq \kappa^* + 1$ , con lo que no satisface (3.58).  $\square$

Aprovechando esta ventaja estructural que posee nuestro problema, se propone el siguiente algoritmo de búsqueda y selección de soluciones  $x \in X$  que minimicen el gap de optimalidad del modelo de minimización de VaR con nivel de riesgo  $\epsilon$ .

0. Definir parámetros de tamaños de muestra  $N$  y  $N'$ , con  $N' \gg N$ , número de repeticiones  $M$  y nivel de confianza  $(1 - \beta) \in (0, 1)$ .
1. Tomar  $M$  colecciones de muestras iid de tamaño  $N$ :  $\rho^{1,m}, \dots, \rho^{N,m}$ ,  $m = 1, \dots, M$ , todos iid entre sí. Resolver  $M$  problemas SAA (3.56) de tamaño  $N$ :

$$\widehat{v}_{\epsilon, N}^m := \min_{x \in X, z \in \mathbb{R}} z \quad \text{s.t.} \quad \sum_{j=1}^N \mathbf{1}_{(-\infty, 0)}(\Pi(x, \rho^{j,m}) + z) \leq \epsilon N,$$

para  $m = 1, \dots, M$ . Obtenemos, en particular,  $M$  candidatos a solución:  $\bar{x}^1, \dots, \bar{x}^M$ .

2. Cálculo de cota inferior para  $v^*$ , objetivo de (3.55). Ordenamos los valores  $\widehat{v}_{\epsilon, N}^m$ ,  $m = 1, \dots, M$ , de menor a mayor,  $\widehat{v}_{\epsilon, N}^{(1)} \leq \dots \leq \widehat{v}_{\epsilon, N}^{(M)}$ . Definimos

$$\begin{aligned} \theta_N &:= \mathbf{b}(\lfloor \epsilon N \rfloor; \epsilon, N) \\ L &:= \max\{L \in \mathbb{N} : \mathbf{b}(L - 1; \theta_N, M) \leq \beta\} \end{aligned}$$

Se tiene que  $\widehat{v}_{\epsilon, N}^{(L)}$  es cota inferior de  $v^*$ , con confianza  $(1 - \beta)$ .

3. Cálculo de cota superior para  $v^*$ . Tomamos una muestra iid de tamaño  $N'$ ,  $\rho^1, \dots, \rho^{N'}$ , del parámetro  $\rho$ . Adicionalmente esta muestra es independiente de las muestras tomadas en el paso 1. Para cada  $\bar{x}^m$ ,  $m = 1, \dots, M$ , calculamos la solución  $\bar{z}^m$  de (3.57) según el lema 3.2.4. Se tiene que  $\bar{z}^m$  es cota superior para  $\text{VaR}_\epsilon(\bar{x}^m)$ , y entonces para  $v^*$  también, con confianza  $(1 - \beta)$ .
4. Cota superior para el gap de optimalidad  $\text{VaR}_\epsilon(\bar{x}^m) - v^*$ . Para cada iteración  $m = 1, \dots, M$ ,  $\bar{z}^m - \widehat{v}_{\epsilon, N}^{(L)}$  es cota superior, con confianza  $(1 - \beta)$ , para el gap de optimalidad  $\text{VaR}_\epsilon(\bar{x}^m) - v^*$ .

Usamos como mejor candidato a solución óptima del problema de minimización de  $\text{VaR}_\epsilon$  real (3.55) la solución  $\bar{x}^m$  que de la menor cota superior para el gap de optimalidad.

### 3.3. Análisis SAA para modelos con recourse

Recordamos el modelo básico de planificación con recourse introducido en §2.2.1:

$$\begin{aligned} \text{máx}_{x^e} \quad & -(v^e)^t x^e + Q(x^e, \rho) \\ \text{s.a.} \quad & x_a^e \leq x_b^e \quad \forall (a, b) \in \mathcal{P} \\ & (c^e)^t x^e \leq B^e \\ & x^e \in \{0, 1\}^{\mathcal{B}} \end{aligned} \tag{3.59}$$

donde

$$\begin{aligned} \mathcal{Q}(x^e, \rho) &:= \max_{x^p} (W^p \rho - v^p)^t x^p \\ \text{s.a. } &x^p \leq x^e \\ &(c^p)^t x^p \leq B^p \\ &x^p \in \{0, 1\}^{\mathcal{B}} \end{aligned}$$

Como dijimos en dicha sección, en este paradigma multi-etapa el plan de procesamiento  $x^p \in \{0, 1\}^{\mathcal{B}}$  se ajusta para maximizar el beneficio que supone elegir el plan de extracción  $x^e \in \{0, 1\}^{\mathcal{B}}$ . Es decir, la variable de segunda etapa  $x^p$  es un *recurso*, o “*recourse*” en inglés, de la variable de primera etapa  $x^e$ .

Recordemos las definiciones de §2.2.1, de función de beneficio con recourse  $\Pi^R$  hecha en (2.9):

$$\begin{aligned} \Pi^R(x^e, \rho) &:= -(v^e)^t x^e + \max_{x^p} (W^p \rho - v^p)^t x^p \\ \text{s.a. } &x^p \leq x^e \\ &(c^p)^t x^p \leq B^p \\ &x^p \in \{0, 1\}^{\mathcal{B}} \end{aligned}$$

y del conjunto  $X^R$  de planes (únicamente) de extracción factibles hecha en (2.10):

$$X^R := \left\{ x^e \in \{0, 1\}^{\mathcal{B}} : \begin{array}{l} x_a^e \leq x_b^e \quad \forall (a, b) \in \mathcal{P} \\ (c^e)^t x^e \leq B^e \end{array} \right\}.$$

Notamos que solamente la función objetivo  $\Pi^R$  depende del parámetro  $\rho \in \mathbb{R}_+^{\mathcal{B}}$ , y que ahora  $\Pi^R(\cdot, \rho)$  y  $\Pi^R(x, \cdot)$  ya no son necesariamente lineales o lineal-afines.

Introducimos ahora incertidumbre asumiendo que el parámetro  $\rho \in \mathbb{R}_+^{\mathcal{B}}$ , de ley de los bloques, no es determinista. Nuevamente asumimos que  $\rho$  sigue una distribución estocástica  $\mathbb{P}$ , no necesariamente conocida, soportada en un conjunto compacto  $\Xi \subset \mathbb{R}_+^{\mathcal{B}}$ . Además asumiremos que la  $\sigma$ -álgebra  $\mathcal{F}$  asociada es boreliana. Los siguientes resultados nos permitirán posteriormente establecer la consistencia de los enfoques de optimización bajo riesgo en este nuevo contexto de beneficio con recourse.

**Propiedad 3.3.1.** *Consideremos  $\Pi^R$  definido como en (2.9).*

1. *Si el soporte de la distribución de  $\rho$  es un compacto  $\Xi \subset \mathbb{R}_+^{\mathcal{B}}$  y la  $\sigma$ -álgebra  $\mathcal{F}$  es boreliana entonces  $\Pi^R(x^e, \cdot) \in L^\infty(\Xi, \mathcal{F}, \mathbb{P})$  para todo  $x^e \in [0, 1]^{\mathcal{B}}$ , donde  $(\Xi, \mathcal{F}, \mathbb{P})$  es el espacio de probabilidad asociado a  $\rho$ .*
2.  *$\Pi^R(\cdot, \rho) : X^R \rightarrow \mathbb{R}$  se puede extender continua y cóncavamente a  $[0, 1]^{\mathcal{B}}$ , para todo  $\rho \in \mathbb{R}^{\mathcal{B}}$ .*

*Demostración.* Para 1. como  $\Xi$  es compacto y  $\mathcal{F}$  boreliana entonces basta demostrar que  $\Pi^R(x^e, \cdot)$  es continua en  $\mathbb{R}^{\mathcal{B}}$ . En efecto, fijemos  $x^e \in \{0, 1\}^{\mathcal{B}}$  cualquiera y definamos

$$Y = \{x^p \in \{0, 1\}^{\mathcal{B}} : x^p \leq x^e, (c^p)^t x^p \leq B^p\}.$$

Se tiene que

$$\begin{aligned}\Pi^R(x^e, \rho) &= -(v^e)^t x^e + \max_{x^p \in Y} (W^p \rho - v^p)^t x^p \\ &= -(v^e)^t x^e + \max_{x^p \in \text{conv}(Y)} (W^p \rho - v^p)^t x^p\end{aligned}$$

donde  $\text{conv}(Y)$  es la envoltura convexa de  $Y$ , y la última igualdad viene de que  $(W^p \rho - v^p)^t x^p$  es lineal en  $x^p$  e  $Y$  es finito. Sigue que por dualidad  $\max_{x^p \in \text{conv}(Y)} (W^p \rho - v^p)^t x^p$  es continua en  $\rho$ , con lo que en particular  $\Pi^R(x^e, \cdot)$  es continua.

Para 2. reconsideramos el conjunto  $X$  de planes de extracción y procesamiento factibles,

$$X = \left\{ (x^e, x^p) \in \{0, 1\}^{2B} : \begin{array}{l} x_a^e \leq x_b^e \quad \forall (a, b) \in \mathcal{P} \\ (c^e)^t x^e \leq B^e \\ x^p \leq x^e \\ (c^p)^t x^p \leq B^p \end{array} \right\}$$

y  $\bar{X} := \text{conv}(X)$  su envoltura convexa. Veamos que

$$\bar{\Pi}^R(x^e, \rho) := -(v^e)^t x^e + \max_{(x^e, x^p) \in \bar{X}} (W^p \rho - v^p)^t x^p$$

es una extensión continua y cóncava de  $\Pi^R$  a  $[0, 1]^B$ .

En efecto, primero notemos que  $\Pi^R(x^e, \rho) = -(v^e)^t x^e + \max_{(x^e, x^p) \in X} (W^p \rho - v^p)^t x^p$  para  $x^e \in X^R$ . También, dado  $x^e \in X^R$ ,

$$\{y : (x^e, y) \in \bar{X}\} = \text{conv}(\{y : (x^e, y) \in X\})$$

con lo que  $\max_{(x^e, x^p) \in X} (W^p \rho - v^p)^t x^p = \max_{(x^e, x^p) \in \bar{X}} (W^p \rho - v^p)^t x^p$ , y entonces  $\Pi^R(x^e, \rho) = \bar{\Pi}^R(x^e, \rho)$  para todo  $x^e \in X^R$ .

Ahora veamos que  $\bar{\Pi}^R(\cdot, \rho) : [0, 1]^B \rightarrow \mathbb{R}$  es continua y cóncava. Sea  $x^e \in [0, 1]^B$  y tomemos  $w \in \mathbb{R}^B$  cualquiera. Como  $X$  es polítopo, se tiene que  $\max_{(x^e, x^p) \in \bar{X}} w^t x^p$  es igual a

$$\begin{array}{ll}\max_{\alpha, x^p} & w^t x^p \\ \text{s.a.} & \alpha \in \mathbb{R}_+^n \\ & \sum_{i=1}^n \alpha_i = 1 \\ & \sum_{i=1}^n \alpha_i x^i = x^e \\ & \sum_{i=1}^n \alpha_i y^i = x^p\end{array}$$

donde  $\{(x^i, y^i)\}_{i=1, \dots, n}$  es la colección de vértices de  $X$ . Este último problema es equivalente a

$$\begin{array}{ll}\max_{\alpha} & \sum_{i=1}^N \alpha_i (w^t y^i) \\ \text{s.a.} & \alpha \in \mathbb{R}_+^N \\ & \sum_{i=1}^N \alpha_i = 1 \\ & \sum_{i=1}^N \alpha_i x^i = x^e\end{array}$$

que es un problema lineal continuo con solución finita. Por dualidad, claramente este último problema es continuo y cóncavo en el parámetro  $x^e$ . Luego, en particular para  $w := W^p \rho - v^p$ ,  $\max_{(x^e, x^p) \in \bar{X}} (W^p \rho - v^p)^t x^p$  es continuo y cóncavo en  $x^e \in [0, 1]^B$ , con lo que  $\bar{\Pi}^R(\cdot, \rho)$  también.  $\square$

La propiedad anterior nos da el siguiente resultado como corolario, el cual será esencial para la consistencia de los modelos de optimización bajo riesgo y con recourse.

**Corolario 3.3.2.** *Si la  $\sigma$ -álgebra  $\mathcal{F}$  es boreliana entonces la función de beneficio con recourse  $\Pi^R : X^R \times \Xi \rightarrow \mathbb{R}$  puede ser extendida a una función de Carathéodory en  $[0, 1]^B \times \Xi$ .*

Finalmente definimos un conjunto que usaremos frecuentemente.

**Propiedad 3.3.3.** *Definiendo el conjunto  $K := X^R \times [-\bar{\pi}^R, -\underline{\pi}^R]$ , donde*

$$\bar{\pi}^R := \max_{x^e \in X^R} \max_{\rho \in \Xi} \Pi^R(x^e, \rho)$$

y

$$\underline{\pi}^R := \min_{x^e \in X^R} \min_{\rho \in \Xi} \Pi^R(x^e, \rho),$$

se tiene que éste es compacto.

*Demostración.* Viene de notar que  $\bar{\pi}^R$  y  $\underline{\pi}^R$  son finitos, pues  $\Xi$  es compacto y  $\Pi^R(x, \cdot)$  es continua –ver demostración de Propiedad 3.3.1 parte 1.–, y además  $X^R$  es finito.  $\square$

### 3.3.1. Modelo de minimización de CVaR con recourse

Dado un nivel de riesgo  $\epsilon \in (0, 1]$ , recordemos el modelo de minimización de CVaR con recourse, el cual introdujimos en §2.2.2:

$$\vartheta^* := \min_{x^e \in X^R, t \in \mathbb{R}} \mathbb{E} \left[ t + \frac{1}{\epsilon} (-\Pi^R(x^e, \rho) - t)^+ \right]. \quad (3.60)$$

Además, dada una muestra del parámetro incierto  $\rho^1, \dots, \rho^N$ , su aproximación SAA es

$$\widehat{\vartheta}_N := \min_{x^e \in X^R, t \in \mathbb{R}} \frac{1}{N} \sum_{k=1}^N \left[ t + \frac{1}{\epsilon} (-\Pi^R(x^e, \rho^k) - t)^+ \right]. \quad (3.61)$$

Usamos la notación de la sección §3.1.1 y notamos

$$F(x, t, \rho) := t + \frac{1}{\epsilon} [-\Pi^R(x, t, \rho) - t]^+,$$

$$f(x, t) := \mathbb{E}[F(x, t, \rho)],$$

y, dada una muestra iid  $\rho^1, \dots, \rho^N$  del parámetro  $\rho$ ,

$$\widehat{f}_N(x, t) := \frac{1}{N} \sum_{k=1}^N F(x, t, \rho^k).$$

Bajo esta notación el problema de minimización de CVaR con recourse (3.60) es

$$\vartheta^* = \min_{x^e, t} f(x^e, t) \quad \text{s.t.} \quad x^e \in X^R, t \in \mathbb{R},$$

y su aproximación SAA (3.61) es

$$\widehat{\vartheta}_N = \min_{x^e, t} \widehat{f}_N(x^e, t) \quad \text{s.t.} \quad x^e \in X^R, t \in \mathbb{R}.$$

Definimos  $S$  como el conjunto de soluciones óptimas de (3.60) en  $X^R \times \mathbb{R}$  y  $\widehat{S}_N$  el conjunto de soluciones óptimas del problema SAA (3.61).

### Consistencia

**Proposición 3.3.4.** *Consideramos, para un nivel de riesgo dado  $\epsilon \in (0, 1]$ , los problemas de minimización de CVaR con recourse verdadero (3.60) y su aproximación SAA (3.61). Asumiendo que el vector de leyes  $\rho$  se distribuye según la distribución  $\mathbb{P}$ , soportada en un compacto  $\Xi \subset \mathbb{R}_+^B$  y con  $\sigma$ -álgebra  $\mathcal{F}$  boreliana, si la muestra  $\rho^1, \dots, \rho^k$  de  $\rho$  es iid, entonces*

$$\widehat{\vartheta}_N \longrightarrow \vartheta^*$$

y

$$\mathbb{D}(\widehat{S}_N, S) \longrightarrow 0$$

cuando  $N \rightarrow \infty$ .

*Demostración.* La demostración es análoga a la hecha para consistencia del modelo CVaR sin recourse, es decir, aplicamos los resultados de consistencia de §3.1.1. Primeramente, es fácil notar que podemos considerar tanto el problema verdadero (3.60) como su aproximación SAA (3.61) como usando la función de beneficio de Carathéodory del Corolario 3.3.2 y restringida al compacto  $K \subset X^R \times \mathbb{R}$ . Usando Teorema 3.1.3 obtenemos que  $f$  es finita y continua sobre  $K$ , y que  $\widehat{f}_N$  converge, con probabilidad 1, a  $f$  uniformemente en  $K$ . Concluimos usando la Proposición 3.1.1 y Teorema 3.1.2.  $\square$

### Análisis estadístico

Usamos el mismo algoritmo de §3.2.1, del modelo de CVaR sin recourse, para encontrar soluciones estadísticamente mejores y también una estimación del gap estadístico entre el problema SAA (3.61) y el problema verdadero (3.60). El algoritmo del modelo

CVaR sin recourse es aplicable a la versión con recourse pues éste no requería que la función de beneficio  $\Pi^R$  tuviera alguna estructura en particular, o siquiera que el problema de optimización fuera computacionalmente tratable. Y naturalmente, si bien en §3.2.1  $\Pi$  y  $X$  eran los beneficios y planes sin recourse, en esta sección  $\Pi^R$  y  $X^R$  son sus contrapartes con recourse, i.e. como se definieron en (2.9) y (2.10) respectivamente. Para que dicho algoritmo sea aplicable al presente modelo, en el algoritmo se debe usar  $F$ ,  $f$  y  $\widehat{f}_N$  como se definieron al inicio de esta sección, y también ahora el problema verdadero será el problema (3.60), y la aproximación SAA será el problema (3.61).

### 3.3.2. Modelo MCH con recourse

Dados una muestra iid  $\rho^1, \dots, \rho^N$  de  $\rho$  y un nivel de riesgo  $\epsilon \in [0, 1]$ , consideramos el problema MCH con recourse, introducido en §2.2.3:

$$\widehat{\vartheta}_N := \min_{x^e \in X^R, t \in \mathbb{R}} (1 - \epsilon) \left( t + \frac{1}{N\epsilon} \sum_{k=1}^N [-\Pi^R(x^e, \rho^k) - t]^+ \right) + \epsilon \frac{1}{N} \sum_{k=1}^N -\Pi^R(x^e, \rho^k). \quad (3.62)$$

Podemos ver este problema como la aproximación SAA, usando la muestra  $\rho^1, \dots, \rho^N$ , del problema para  $\rho$  en el espacio de probabilidad  $(\Xi, \mathcal{F}, \mathbb{P})$

$$\vartheta_N^* := \min_{x^e \in X^R} (1 - \epsilon) \text{CVaR}_{1/N}(-\Pi^R(x^e, \rho)) + \epsilon \mathbb{E}(-\Pi^R(x^e, \rho)). \quad (3.63)$$

Llamamos al problema (3.63) *problema MCH con recourse verdadero*, y al problema (3.62) *aproximación SAA del problema verdadero*.

Análogo a lo argumentado en §3.2.2 para el modelo MCH sin recourse, este enfoque es bastante limitado pues estamos aproximando el problema estocástico (3.63), que claramente depende de  $N$ , usando una muestra de tamaño  $N$ . Sin embargo, lo usamos pues nos permitirá obtener finalmente un criterio de selección de soluciones “estadísticamente mejores”.

Usando notación de §3.1.1, definimos

$$F(x, t, \rho) := \epsilon (-\Pi^R(x, \rho)) + (1 - \epsilon) \left( t + \frac{1}{1/N} [-\Pi^R(x, \rho) - t]^+ \right),$$

$$f(x, t) := \mathbb{E}[F(x, t, \rho)],$$

y, dada una muestra  $\rho^1, \dots, \rho^N$ ,

$$\widehat{f}_N(x, t) := \frac{1}{N} \sum_{k=1}^N F(x, t, \rho^k).$$

Así, el problema MCH verdadero (3.63) es equivalente a

$$\vartheta_N^* = \min_{x^e, t} f(x^e, t) \quad \text{s.t.} \quad x^e \in X^R, t \in \mathbb{R},$$

y su aproximación SAA (3.62) es, equivalentemente,

$$\widehat{\vartheta}_N = \min_{x^e, t} \widehat{f}_N(x^e, t) \quad \text{s.t.} \quad x^e \in X^R, t \in \mathbb{R}.$$

Denotamos como  $S$  al conjunto de soluciones óptimas del problema verdadero y  $\widehat{S}_N$  al conjunto de soluciones óptimas de su aproximación SAA.

**Consistencia:** Análogo a lo sucedido para el modelo MCH sin recourse, obviamos el análisis de convergencia del problema SAA (3.62) al problema verdadero (3.63), pues tanto  $\widehat{\vartheta}_N$  como  $\vartheta_N^*$  dependen de  $N$ . Carece de todo sentido entonces preguntarse si  $\widehat{\vartheta}_N$  converge a  $\vartheta_N^*$  cuando  $N \rightarrow \infty$ .

### Análisis estadístico

Usamos el algoritmo de §3.2.1 para encontrar soluciones al problema MCH con recourse que, estadísticamente, minimicen el gap de optimalidad entre el problema verdadero y su aproximación SAA. Debemos redefinir en dicho algoritmo  $F$ ,  $f$  y  $\widehat{f}_N$  como se definieron al inicio de esta sub-sección, y también el problema verdadero ahora corresponderá al problema (3.63) y la aproximación SAA al problema (3.62).

### 3.3.3. Modelo MCH- $\epsilon$ con Recourse

Dado un nivel de riesgo  $\epsilon \in (0, 1]$  el problema MCH- $\epsilon$  con Recourse es

$$\vartheta^* := \min_{x^e \in X^R, t \in \mathbb{R}} \mathbb{E} \left[ \epsilon (-\Pi^R(x^e, \rho)) + (1 - \epsilon) \left( t + \frac{1}{\epsilon} [-\Pi^R(x^e, \rho) - t]^+ \right) \right] \quad (3.64)$$

Dada una muestra del parámetro incierto  $\rho$  de tamaño  $N$ ,  $\rho^1, \dots, \rho^N$ , la aproximación SAA del problema MCH- $\epsilon$  con Recourse es

$$\widehat{\vartheta}_N := \min_{x^e \in X^R, t \in \mathbb{R}} \frac{1}{N} \sum_{k=1}^N \left[ \epsilon (-\Pi^R(x^e, \rho^k)) + (1 - \epsilon) \left( t + \frac{1}{\epsilon} [-\Pi^R(x^e, \rho^k) - t]^+ \right) \right]. \quad (3.65)$$

Nuevamente, usamos la notación de §3.1.1:

$$F(x, t, \rho) := \epsilon (-\Pi^R(x, \rho)) + (1 - \epsilon) \left( t + \frac{1}{\epsilon} [-\Pi^R(x, \rho) - t]^+ \right),$$

$$f(x, t) := \mathbb{E}[F(x, t, \rho)],$$

y, dada una muestra  $\rho^1, \dots, \rho^N$ ,

$$\widehat{f}_N(x, t) := \frac{1}{N} \sum_{k=1}^N F(x, t, \rho^k).$$

Así, el problema MCH- $\epsilon$  verdadero (3.64) equivale a

$$\vartheta^* = \min_{x^e, t} f(x^e, t) \quad \text{s.t.} \quad x^e \in X^R, t \in \mathbb{R},$$

y su aproximación SAA (3.65) a

$$\widehat{\vartheta}_N = \min_{x^e, t} \widehat{f}_N(x^e, t) \quad \text{s.t.} \quad x^e \in X^R, t \in \mathbb{R}.$$

Denotamos  $S$  al conjunto de soluciones óptimas de (3.64) y  $\widehat{S}_N$  al conjunto de soluciones óptimas de (3.65).

### Consistencia

**Proposición 3.3.5.** *Consideramos, para un nivel de riesgo dado  $\epsilon \in (0, 1]$ , los problemas MCH- $\epsilon$  con Recourse verdadero (3.64) y su aproximación SAA (3.65). Asumiendo que el vector de leyes  $\rho$  se distribuye según la distribución  $\mathbb{P}$ , soportada en el compacto  $\Xi \subset \mathbb{R}_+^B$  y con  $\sigma$ -álgebra  $\mathcal{F}$  boreliana, si la muestra  $\rho^1, \dots, \rho^k$  de  $\rho$  es iid entonces se tiene que*

$$\widehat{\vartheta}_N \longrightarrow \vartheta^*$$

y

$$\mathbb{D}(\widehat{S}_N, S) \longrightarrow 0$$

cuando  $N \rightarrow \infty$ .

*Demostración.* El argumento es el mismo que el dado para la Proposición 3.3.4, de consistencia de la aproximación SAA del modelo CVaR con recourse.  $\square$

**Análisis Estadístico:** Para seleccionar en el modelo MCH- $\epsilon$  con recourse soluciones que, estadísticamente, minimizan el gap de optimalidad entre el problema verdadero y su aproximación SAA, usamos una vez más el algoritmo de §3.2.1. Para hacer dicho algoritmo aplicable al presente modelo tomamos en él las funciones  $F$ ,  $f$  y  $\widehat{f}_N$  como se definieron al inicio de esta sub-sección, y también ahora el problema verdadero será el problema (3.64) y su aproximación SAA el problema (3.65).

### 3.3.4. Modelo de minimización de VaR con recourse

Dado un nivel de riesgo  $\epsilon \in (0, 1)$ , recordemos el modelo de minimización de VaR con recourse introducido en §2.2.1:

$$\vartheta^* := \min_{x^e, z} z \quad \text{s.t.} \quad \mathbb{P}(\Pi^R(x^e, \rho) + z < 0) \leq \epsilon, \quad x^e \in X^R. \quad (3.66)$$

Además, dada una muestra  $\rho^1, \dots, \rho^N$  de  $\rho$  su aproximación SAA era

$$\widehat{\vartheta}_N := \min_{x^e, z} z \quad \text{s.t.} \quad \frac{1}{N} \sum_{j=1}^N \mathbf{1}_{(-\infty, 0)}(\Pi^R(x^e, \rho) + z) \leq \epsilon, \quad x^e \in X^R. \quad (3.67)$$

Siguiendo la notación de §3.1.2 definimos

$$\begin{aligned} f(x, z) &:= z \\ C(x, z, \rho) &:= -\Pi^R(x, \rho) - z \\ p(x, z) &:= \mathbb{P}(C(x, z, \rho) > 0) \\ \widehat{p}_N(x, z) &:= \frac{1}{N} \sum_{j=1}^N \mathbf{1}_{(0, \infty)}(C(x, z, \rho^j)) \end{aligned}$$

donde  $\rho^1, \dots, \rho^N$  es una muestra del parámetro  $\rho$ . Así, el problema verdadero (3.66) es equivalente a

$$\vartheta^* = \min_{x^e, z} f(z) \quad \text{s.t.} \quad p(x^e, z) \leq \epsilon, \quad x^e \in X^R,$$

y su aproximación SAA (3.67) equivalente a

$$\widehat{\vartheta}_N = \min_{x^e, z} f(z) \quad \text{s.t.} \quad \widehat{p}_N(x^e, z) \leq \epsilon, \quad x^e \in X^R.$$

Definimos también como  $S$  al conjunto de soluciones óptimas del problema verdadero (3.66) y  $\widehat{S}_N$  al conjunto de soluciones óptimas del problema SAA (3.67).

#### Consistencia

Como lo realizado en el caso del modelo VaR sin recourse, la consistencia viene de aplicar la Proposición 3.1.7. Obtenemos un resultado absolutamente análogo al de consistencia para VaR sin recourse en §3.2.4.

**Proposición 3.3.6.** *Consideremos el problema verdadero de minimización de VaR con recourse (3.66) y su aproximación SAA (3.67), donde  $\Pi^R$  y  $X^R$  son como en (2.9) y (2.10), respectivamente. Supongamos que  $\rho$  está soportado en un compacto  $\Xi$  y su  $\sigma$ -álgebra asociada  $\mathcal{F}$  es boreliana. Si la muestra  $\rho^1, \dots, \rho^N$  es iid y se cumple la siguiente condición*

(a'') existe solución óptima  $x^{e,*} \in X^R$  al problema verdadero (3.66) tal que  $F_{-\Pi^R(x^{e,*}, \rho)}$ , la función de distribución acumulada de la variable aleatoria  $-\Pi^R(x^{e,*}, \rho)$ , es tal que el conjunto  $F_{-\Pi^R(x^{e,*}, \rho)}^{-1}(\{1 - \epsilon\})$  tiene interior vacío

entonces  $\widehat{\vartheta}_N \rightarrow \vartheta^*$  y  $\mathbb{D}(\widehat{S}_N, S) \rightarrow 0$  cuando  $N \rightarrow \infty$ .

Hacemos notar que la condición (a'') pide únicamente que la función de distribución acumulada de  $-\Pi^R(x^{e,*}, \rho)$  no tenga una “planicie” en el nivel  $1 - \epsilon$ .

*Demostración.* Primeramente, es fácil ver que podemos restringir tanto el problema verdadero (3.66) como el aproximado (3.67) al conjunto compacto  $K$ , definido en (3.3.3). También, claramente  $f(x, z) = z$  es continua, y  $C(x, z) = -\Pi^R(x, \rho) - z$  es de Carathéodory pues por Proposición 3.3.1  $\Pi^R$  es Carathéodory. Luego, basta verificar que la condición (a) de Proposición 3.3.1 es equivalente a la condición (a''), y la demostración de esto es totalmente análoga a la hecha para verificar la equivalencia entre las condiciones (a) y (a') en Proposición 3.2.3.  $\square$

### Análisis estadístico

Queremos aplicar el algoritmo de la sección §3.1.2 de búsqueda soluciones óptimas que minimicen el gap de optimalidad entre el problema aproximado (3.67) y el problema verdadero (3.66). Hacemos notar que, pese a haber cambiado el conjunto factible  $X^R$  y la función de beneficio  $\Pi^R$  con respecto al enfoque sin recourse, igualmente estamos estructuralmente en las mismas condiciones que en el caso de VaR sin recourse: podemos restringir el problema verdadero (3.66) y el aproximado (3.67) al compacto  $K$ , y  $C$  es Carathéodory. Luego todos los resultados de la sub-sección §3.2.4 son igualmente válidos para el caso de VaR con recourse, tomando en cuenta siempre la redefinición de  $X^R$  y  $\Pi^R$ . En particular, tanto el Lema 3.2.4 de §3.2.4 como el algoritmo de selección de mejores soluciones estadísticas de dicha sección siguen siendo válidos para el caso con recourse, con  $X^R$  y  $\Pi^R$  definidos según (2.10) y (2.9) respectivamente.

Se propone entonces el siguiente algoritmo para la búsqueda y selección de soluciones  $x^e \in X^R$  que minimicen el gap de optimalidad del modelo de minimización de VaR con recourse con nivel de riesgo  $\epsilon$ .

0. Definir parámetros de tamaños de muestra  $N$  y  $N'$ , con  $N' \gg N$ , número de repeticiones  $M$  y nivel de confianza  $(1 - \beta) \in (0, 1)$ .
1. Tomar  $M$  colecciones de muestras iid de tamaño  $N$ :  $\rho^{1,m}, \dots, \rho^{N,m}$ ,  $m = 1, \dots, M$ , todos iid entre sí. Resolver  $M$  problemas SAA (3.67) de tamaño  $N$ :

$$\widehat{\vartheta}_{\epsilon, N}^m := \min_{x^e \in X^R, z \in \mathbb{R}} z \quad \text{s.t.} \quad \sum_{j=1}^N \mathbf{1}_{(-\infty, 0)}(\Pi^R(x^e, \rho^{j,m}) + z) \leq \epsilon N,$$

para  $m = 1, \dots, M$ . Obtenemos, en particular,  $M$  candidatos a solución:  $\bar{x}^{e,1}, \dots, \bar{x}^{e,M}$ .

2. Cálculo de cota inferior para  $\vartheta^*$ , objetivo de (3.66). Ordenamos los valores  $\widehat{\vartheta}_{\epsilon, N}^m$ ,  $m = 1, \dots, M$ , de menor a mayor,  $\widehat{\vartheta}_{\epsilon, N}^{(1)} \leq \dots \leq \widehat{\vartheta}_{\epsilon, N}^{(M)}$ . Definimos

$$\begin{aligned}\theta_N &:= \mathbf{b}(\lfloor \epsilon N \rfloor; \epsilon, N) \\ L &:= \text{máx}\{L \in \mathbb{N} : \mathbf{b}(L - 1; \theta_N, M) \leq \beta\}\end{aligned}$$

Se tiene que  $\widehat{\vartheta}_{\epsilon, N}^{(L)}$  es cota inferior de  $\vartheta^*$ , con confianza  $(1 - \beta)$ .

3. Cálculo de cota superior para  $\vartheta^*$ . Tomamos una muestra iid de tamaño  $N'$ ,  $\rho^1, \dots, \rho^{N'}$ , del parámetro  $\rho$ . Adicionalmente esta muestra es independiente de las muestras tomadas en el paso 1. Para cada  $\bar{x}^{e, m}$ ,  $m = 1, \dots, M$ , calculamos la solución  $\bar{z}^m$  del problema (3.57) según el Lema 3.2.4 en el contexto con recourse, i.e. tomando en Lema 3.2.4  $X^R$  en vez de  $X$  y  $\Pi^R$  en vez de  $\Pi$ . Se tiene que  $\bar{z}^m$  es cota superior para  $\text{VaR}_\epsilon(\bar{x}^{e, m})$ , y entonces para  $\vartheta^*$  también, con confianza  $(1 - \beta)$ .
4. Cota superior para el gap de optimalidad  $\text{VaR}_\epsilon(\bar{x}^{e, m}) - \vartheta^*$ . Para cada iteración  $m = 1, \dots, M$ ,  $\bar{z}^m - \widehat{\vartheta}_{\epsilon, N}^{(L)}$  es cota superior, con confianza  $(1 - \beta)$ , para el gap de optimalidad  $\text{VaR}_\epsilon(\bar{x}^{e, m}) - \vartheta^*$ .

Usamos como mejor candidato a solución óptima del problema verdadero de minimización de  $\text{VaR}_\epsilon$  con recourse (3.66) la solución  $\bar{x}^m$  que de la menor cota superior para el gap de optimalidad.

# Capítulo 4

## Experimentos Computacionales

### 4.1. Descripción

Nos proponemos ahora comparar los resultados que dan los distintos modelos de optimización bajo riesgo propuestos. En primera instancia, nos gustaría que esta comparación respondiera preguntas básicas como qué modelo da mejores beneficios, qué modelo provee una mejor “protección” ante la incertidumbre, qué diferencias existen entre usar un modelo neutro al riesgo –e.g. minimizar la pérdida esperada– y usar un modelo averso al riesgo, qué compensación hay entre disminuir el beneficio y disminuir el riesgo, etcétera. Sin embargo, todas estas preguntas no serán fáciles de responder pues no existe una metodología clara y concluyente para hacer este tipo de evaluaciones. El problema principal en este ámbito es cuál es realmente la ley de los bloques. Si finalmente no tenemos manera de conocerla, podemos tratar de responder estas preguntas asumiendo que sigue alguna distribución. No obstante, esta solución es bastante limitada pues evalúa nuestras distintas soluciones solamente en el marco que establece la distribución que usemos.

La comparación que nos proponemos hacer tendrá el escollo adicional de que estamos haciendo una aproximación de cada modelo, pues estamos aproximando la distribución original del parámetro  $\rho$  con una muestra de él. Así, en primer lugar, nuestro problema aproximado dependerá del modelo que elijamos para obtener una muestra de  $\rho$ . Éste no es un problema trivial pues, como dijimos al inicio de §2.2, elegimos usar la aproximación SAA pues es poco realista asumir que se conoce la verdadera distribución de  $\rho$ , y sin embargo para obtener una muestra se estarán haciendo necesariamente algunas suposiciones de cuál es la distribución real. En segundo lugar, la convergencia del problema aproximado al problema verdadero puede no ser rápida, traduciéndose en que el comportamiento impuesto en la muestra no se traduzca fielmente a la distribución verdadera. En definitiva, apoyarnos en la aproximación SAA nos obligará a evaluar también cómo convergen los distintos modelos propuestos y cuál es el comportamiento que impone cada modelo aproximado.

Para hacernos cargo de la mayoría de las incógnitas y dificultades expuestas, usaremos muestras del parámetro  $\rho$  tanto para hacer la aproximación SAA como para representar su verdadera distribución. Como hemos discutido aquí, toda evaluación que realicemos con este método será restringida, pues 1) estaremos haciendo inherentemente una hipótesis sobre cuál es la distribución que sigue realmente  $\rho$ , y 2) estaremos evaluando solamente cuán buena es la aproximación SAA al modelo asumido para  $\rho$ .

Para los experimentos computacionales tomamos un yacimiento muy pequeño, de 17 por 26 por 15 bloques, y donde las dimensiones de cada bloque es  $10 \times 10 \times 10$  metros. El argumento para tomar una mina pequeña es que ya sin incertidumbre el problema de planificación minera es altamente complejo, debido a la gran cantidad de variables que tiene y el tipo de restricciones que usa, y requiere el uso de técnicas ad-hoc para ser abordado eficientemente. Luego, preferimos tomar en nuestra evaluación una mina de tamaño reducido para evitar las complicaciones que surgen cuando el problema tiene un gran número de bloques. Esta mina simula un yacimiento de cobre consistente de vetas mineralizadas, y está basada en datos reales. Elegimos tomar una mina de tipo veta pues en éstas el mineral está depositado de manera discontinua; por ejemplo, una mina sencilla de este tipo sería una donde el mineral está totalmente concentrado en un plano que corta el yacimiento. Usamos una mina con esta característica pues ésta provocará, en principio, los resultados más adversos en presencia de incertidumbre, con lo que sería esperable que este tipo de mina provea los datos más desafiantes para los modelos bajo riesgo propuestos.

Para esta mina contamos con 20.000 escenarios para la distribución conjunta de ley de los bloques. Esto es, cada escenario representa un vector  $\rho \in \mathbb{R}^B$  de leyes. Estos escenarios están basados en la técnica geo-estadística de *Conditional Simulation*, y fueron obtenidos usando el procedimiento expuesto en [EL06]. Podemos asumir que estos 20.000 escenarios son una muestra iid del vector  $\rho$  de leyes.

En esta mina todos los bordes poseen algún escenario con un bloque con ley positiva, luego es posible que alguna solución contemple extraer bloques en el borde. Recordemos que las restricciones de precedencia del modelo de bloques requiere que para extraer un bloque se deben extraer también todos los bloques ubicados en un cono –precisado por las restricciones de precedencia– arriba de él. Sin embargo si un plan establece extraer un bloque del borde entonces solamente se requerirían extraer los bloques en el cono truncado que está arriba de él. Esto sería un error de modelamiento, pues en la práctica si se desea extraer un bloque en la frontera del yacimiento igualmente se deberían extraer todos los bloques en el cono encima de él, produciéndose un consumo de capacidad de extracción no previsto por el modelo. Prevemos esta situación y agregamos varias corridas de bloques extras, con ley nula, en los bordes. Agregamos todos los bloques necesarios para que extraer un bloque del borde requiera extraer un cono completo sobre él.

Para hacer las pruebas computacionales usamos el algoritmo de selección de soluciones presentado para cada modelo en §3.2 y §3.3. Más formalmente, seguiremos el siguiente procedimiento general.

0. Dado un modelo, en su variante con o sin recourse, y un nivel de riesgo  $\epsilon \in [0, 1]$

–con  $\epsilon \neq 0$  salvo si el modelo es MCH–, definimos tamaños de muestra  $N$  y  $N'$ ,  $N' \gg N$ . Tomamos también  $M$ , el número de repeticiones para las pruebas, y  $(1 - \beta) \in (0, 1)$ , el nivel de confianza para el análisis estadístico.

1. Tomamos, de entre las primeras 20000 –  $N'$  muestras del total de 20000,  $M$  colecciones, sin repeticiones, de muestras de tamaño  $N$ :  $\rho^{1,m}, \dots, \rho^{N,m}$ ,  $m = 1, \dots, M$ . Para cada colección  $m = 1, \dots, M$  resolvemos un problema SAA de tamaño  $N$  usando la colección  $m$ -ésima:  $\rho^{1,m}, \dots, \rho^{N,m}$ . Obtenemos, en particular,  $M$  valores objetivos  $\hat{v}_{\epsilon,N}^1, \dots, \hat{v}_{\epsilon,N}^M$  y  $M$  candidatos a solución,  $\bar{x}^1, \dots, \bar{x}^M$ .
2. Calculamos la cota inferior estadística, con confianza  $(1 - \beta)$ , para el valor objetivo del problema verdadero,  $v^*$ . Esto se hace usando  $\hat{v}_{\epsilon,N}^1, \dots, \hat{v}_{\epsilon,N}^M$  en el procedimiento ad-hoc del modelo.
3. Tomando las últimas  $N'$  muestras de las 20000 calculamos la cota superior estadística, con confianza  $(1 - \beta)$ , para  $v^*$  usando el procedimiento de cada modelo.
4. Calculamos, para cada candidato  $\bar{x}^1, \dots, \bar{x}^M$ , la cota superior estadística, con confianza  $(1 - \beta)$ , del gap de optimalidad. Diremos que la solución  $\bar{x}^m$  que minimice esta cota es la solución *estadísticamente óptima* para este modelo y estos parámetros.

Usando este procedimiento para encontrar la, estadísticamente, mejor solución de cada modelo y combinación de parámetros, nos concentramos principalmente en estudiar los siguientes aspectos.

- Cómo converge el modelo con el tamaño  $N$  de la muestra.
- Cuál es el perfil de beneficios que entrega cada modelo.
- Comportamiento del modelo al variar el nivel de riesgo  $\epsilon$ , y comparación con minimizar pérdida esperada y minimizar peor pérdida.
- Para cada modelo, comparación de resultados entre las versiones con y sin recourse.

## 4.2. Parámetros

Recordamos que usaremos una mina que simula un yacimiento de cobre consistente de vetas mineralizadas. Las dimensiones de cada bloque son  $10 \times 10 \times 10$  metros, y el yacimiento es una grilla de  $17 \times 26 \times 15$  bloques, i.e. 6.630 bloques. Para no complicar el problema consideramos que para extraer cada bloque también se deben extraer los 5 bloques en la cruz inmediatamente encima de él. Definido esto, se procedió a agregar bloques de ley nula en los bordes del yacimiento, según se argumentó antes. El modelo resultante consta de 19.720 bloques y de 86.450 restricciones de precedencia.

Cada bloque consume una unidad de capacidad de extracción y de procesamiento, y la capacidad total de extracción y procesamiento es de 9.860. Es decir, no se pueden extraer ni procesar más de la mitad del total de los bloques de la mina. El costo por extraer cada bloque es de 1 UM, el costo por procesarlo 5 UM, y el beneficio por bloque por unidad de mineral es 25 UM.

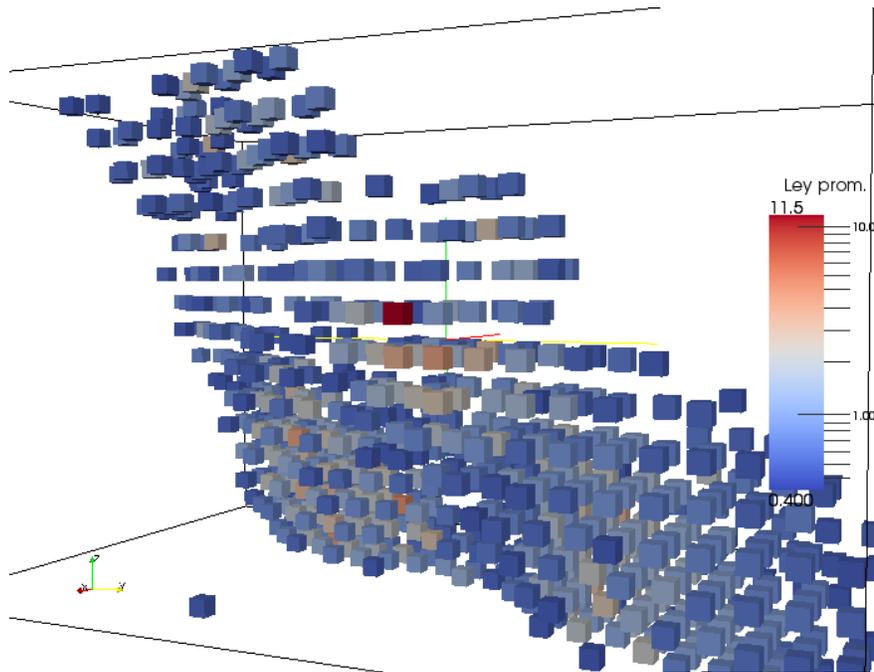


Figura 4.1: Diagrama de bloques con leyes promedio superior a 0.4

En cuanto a las leyes de la mina, como ya se dijo, se cuenta con 20.000 escenarios para el vector de leyes  $\rho$ . En la Figura 4.1 se muestra un diagrama de la posición topológica de los bloques para los cuales la ley promedio es mayor o igual a 0.4. Como la leyenda de la figura muestra, los bloques están coloreados según su ley promedio. Por otro lado, en la Figura 4.2 mostramos histogramas de las leyes promedio, para el yacimiento original de 6.630 bloques y para el yacimiento completado de 19.720 bloques. Cada columna de cada gráfico representa cuántos bloques tienen ley promedio en ese rango. La columna con asterisco de más a la derecha de cada histograma representa la cantidad de bloques que tienen ley promedio en, o mayor, al rango especificado.

Para todos los modelos, con nivel de riesgo  $\epsilon$  y tamaño de muestra  $N$  dados, tomamos siempre los siguientes parámetros para el algoritmo de selección de soluciones estadísticamente óptimas:

- nivel de confianza  $(1 - \beta) = 95\%$ ,
- $M = 30$  repeticiones de cada prueba, y
- $N' = 10000$  muestras para computar la cota superior.

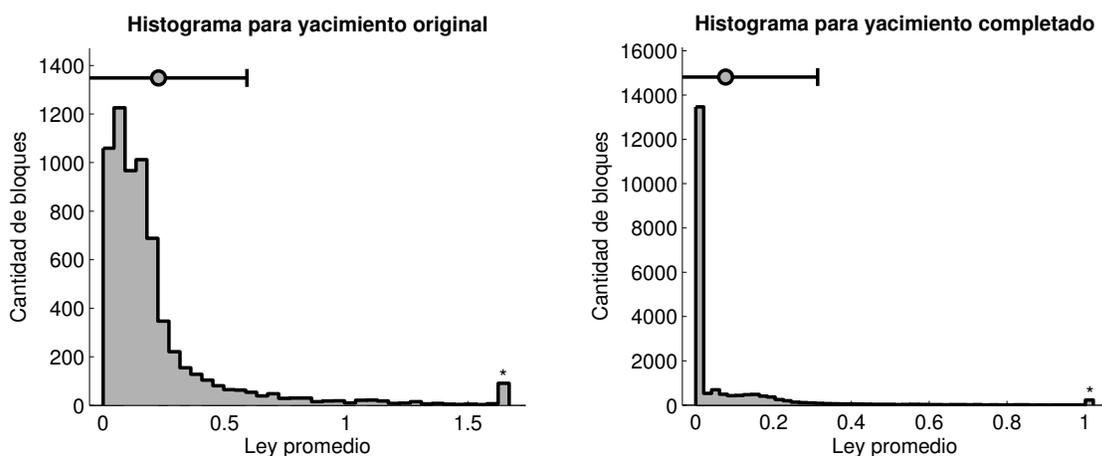


Figura 4.2: Histograma de leyes promedio

Con estos parámetros podremos calcular las cotas superiores e inferiores estadísticas del valor objetivo del problema verdadero, y cota superior estadística del gap de optimalidad.

Adicionalmente, estimaremos la pérdida esperada y CVaR –para varios niveles de riesgo– de cada plan estadísticamente óptimo, usando las mismas  $N' = 10000$  muestras del cálculo de la cota superior. Para el caso de pérdida esperada, la Ley de los Grandes Números nos asegura que ésta será una buena estimación de la verdadera pérdida esperada del plan. En el caso del CVaR, el argumento en [SDR09, §6.5.1] nos asegura la convergencia del estimador, y entonces también será una buena estimación.

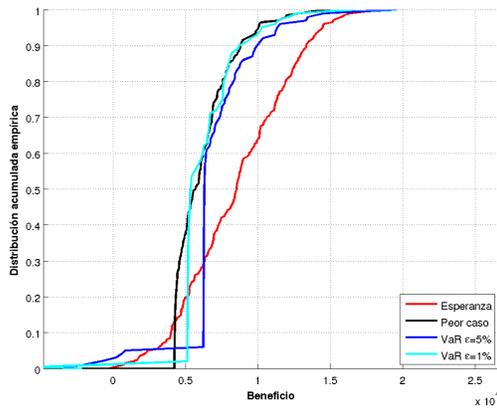
En la siguiente sección mostramos los resultados más relevantes de los experimentos computacionales hechos en esta mina, con la metodología presentada.

## 4.3. Resultados

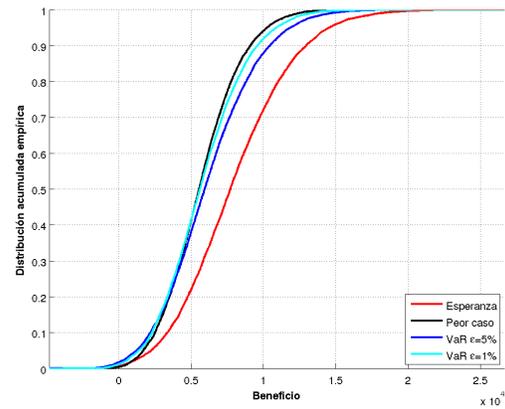
### 4.3.1. Distribución empírica de beneficios

#### Modelo minimización de VaR

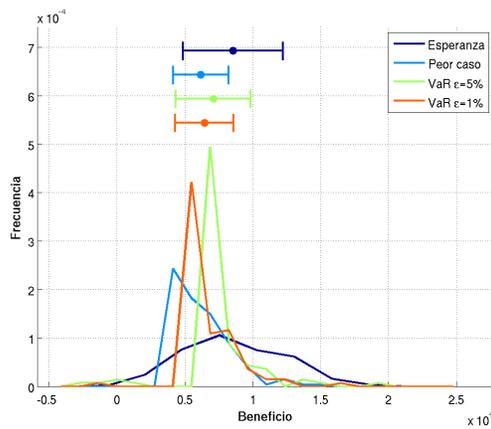
En general, los experimentos numéricos para el modelo de minimización de VaR se caracterizaron por ser sumamente ineficientes en su resolución numérica. En circunstancias en que todos los otros modelos sin recourse y el mismo tamaño  $N$  de muestra requerían un tiempo razonable de resolución, digamos no más que un par de horas, el modelo de VaR demoraba varios días. Peor aún, la mayoría de las pruebas con este modelo no pudieron ser terminadas, pues el consumo excesivo de memoria que realizaba hacía que el nodo de cómputo donde se ejecutaba detuviera el trabajo. Esta ineficiencia computacional probablemente se deba en gran parte a la formulación con “gran-M” que usamos.



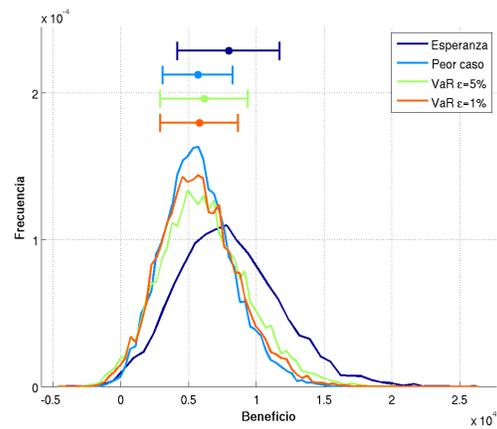
(a) Dist. acumulada, en muestra de planificación



(b) Dist. acumulada, en muestra de evaluación



(c) Histograma, en muestra de planificación



(d) Histograma, en muestra de evaluación

Figura 4.3: Distribución empírica de beneficios, modelo minimización de VaR, para planes estadísticamente óptimos obtenidos resolviendo  $M = 30$  iteraciones del problema SAA con  $N = 100$  muestras.

El mayor tamaño de muestra en la que pudieron terminar las  $M = 30$  repeticiones de la prueba fue  $N = 100$  muestras, y solamente para algunos niveles de riesgo  $\epsilon$ . En la Figura 4.3 se muestran, en dos tipos de gráfico, la distribución empírica que tuvieron las soluciones estadísticamente óptimas para el modelo VaR para  $\epsilon = 1\%$  y  $\epsilon = 3\%$ .

Los gráficos 4.3(a) y 4.3(b) muestran la función de distribución acumulada de los beneficios empíricos al evaluar el plan estadísticamente óptimo en las muestras de  $\rho$  de *planificación* y de *evaluación*. La *muestra de planificación* es la colección  $m$ -ésima de  $N$  muestras de  $\rho$  en la que se obtuvo la planificación estadísticamente óptima para una prueba en particular. En cambio, la *muestra de evaluación* es la colección de  $N' = 10000$  muestras de  $\rho$  donde se trata de evaluar el *desempeño real* de la solución estadísticamente óptima de cada prueba. Así, dada una solución estadísticamente óptima  $x^*$  para alguna

prueba, obtenida, digamos, en la iteración  $m$ -ésima del problema, la *distribución empírica en la muestra de evaluación* es la distribución de la colección de beneficios

$$\Pi(x^*, \rho^{1,m}), \dots, \Pi(x^*, \rho^{N,m})$$

y la *distribución empírica en la muestra de planificación* es la distribución de

$$\Pi(x^*, \rho^1), \dots, \Pi(x^*, \rho^{N'}).$$

Cabe resaltar que la muestra de evaluación no contiene a la muestra de planificación. También, la muestra de evaluación siempre es la misma para todas las pruebas, no así la muestra de planificación, pues cada una de las  $m = 1, \dots, 30$  colecciones de muestras de tamaño  $N$  está tomada al azar, como ya mencionamos.

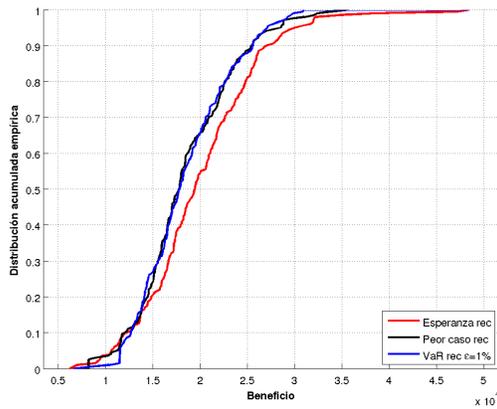
En los gráficos 4.3(c) y 4.3(d) se muestran histogramas de la distribución empírica de los beneficios, en la muestra de planificación el primero y en la de evaluación el segundo. En los histogramas las barras sobre las curvas muestran el promedio y desviación estándar de la distribución empírica. Se incluyen además en todos los gráficos los resultados de las pruebas de minimizar la pérdida esperada –notada *Esperanza* en la leyenda– y minimizar peor pérdida –notada *Peor caso* en la leyenda– para hacer una comparación de los resultados del modelo con estos dos enfoques extremos en cuanto al tratamiento de riesgo. Estas últimas dos pruebas se resolvieron usando  $N = 200$  muestras y usando también el algoritmo de selección de soluciones estadísticamente óptimas.

En los gráficos 4.3(a) y 4.3(c) se refleja cuál es el comportamiento que está tratando de imponer el modelo de VaR: maximizar el percentil  $\epsilon$  de los beneficios –o equivalentemente, minimizar el percentil  $(1 - \epsilon)$  de las pérdidas. Sin embargo este comportamiento no se logra traducir a la muestra de evaluación. Es más, pese a que los niveles de riesgo son bajos, las distribuciones resultantes en la muestra de evaluación son peores que las pruebas *Esperanza* y *Peor caso*: en la parte de menores beneficios las pruebas VaR dan mayor peso, y tienen un beneficio esperado muy cercano al del enfoque *Peor caso*, el más conservador.

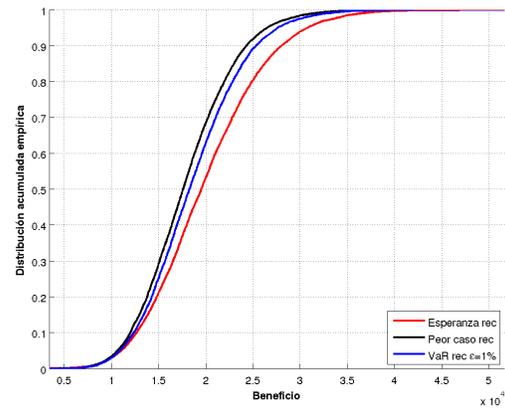
Hacemos notar también la *coherencia* que muestran las pruebas de minimizar pérdidas y minimizar peor pérdida: mientras la solución a la primera prueba efectivamente da el beneficio esperado más alto, dando en desmedro peso a los peores beneficios, la solución de la segunda prueba da el menor peso a los peores beneficios en detrimento de dar menos peso a los beneficios más altos y entonces logrando un menor beneficio esperado.

### Modelo minimización de VaR con recourse

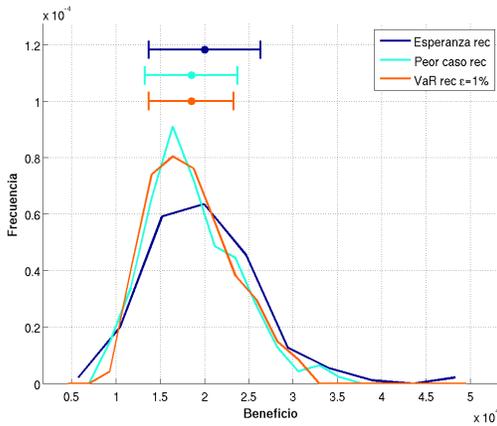
Para el modelo VaR con recourse, en la Figura 4.4 mostramos la distribución empírica de los beneficios para la prueba con  $\epsilon = 1\%$  y tamaño de muestra de planificación  $N = 200$ . Además se muestran los resultados de las pruebas de minimizar pérdida esperada –leyenda *Esperanza rec*– y minimizar peor pérdida –leyenda *Peor caso rec*–, ambas en su variante con recourse.



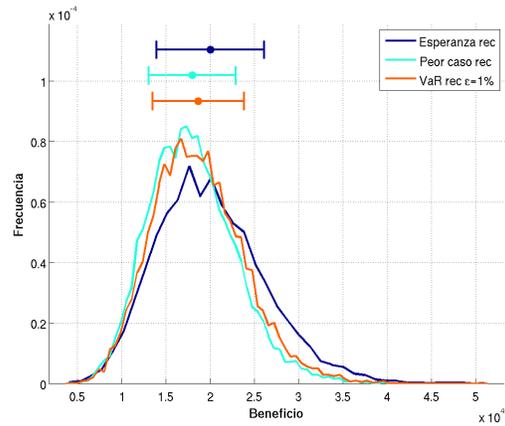
(a) Dist. acumulada, en muestra de planificación



(b) Dist. acumulada, en muestra de evaluación



(c) Histograma, en muestra de planificación



(d) Histograma, en muestra de evaluación

Figura 4.4: Distribución empírica de beneficios, modelo minimización de VaR con recourse, para planes estadísticamente óptimos obtenidos resolviendo  $M = 30$  iteraciones del problema SAA con  $N = 200$  muestras.

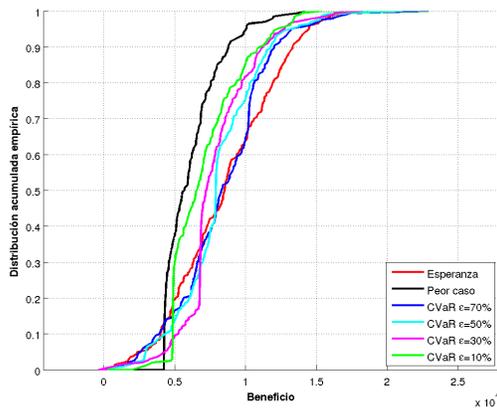
En primer lugar, se observa que todas las variantes con recourse entregan beneficios considerablemente mayores a la variante sin recourse. Además, ninguna prueba entrega beneficios negativos, o pérdidas. Se verificará en todos los modelos que los planes obtenidos con la variante con recourse entregan beneficios mucho mayores y no entregan beneficios negativos.

Vemos también que el comportamiento que se trata de imponer en la muestra de planificación no es tan claro como en el caso sin recourse. Efectivamente, se ve que el modelo en la muestra de planificación trata de maximizar el percentil  $\epsilon$  de los beneficios, sin embargo no se observa que esto sea impuesto de manera clara. Al pasar a la muestra de evaluación, el modelo VaR entrega una solución con riesgo intermedio entre minimizar esperanza y minimizar peor caso. Esto contrasta con la variante sin recourse, donde vimos

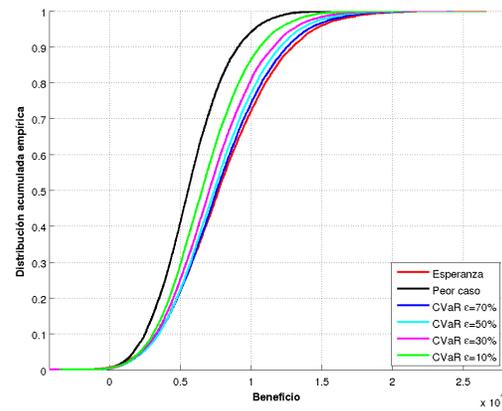
que la solución obtenida daba, a modo grueso, un riesgo peor a la de estos dos modelos.

Por otro lado, en el ámbito computacional, la ejecución de la variante con recourse del modelo VaR fue más eficiente, permitiendo ejecutar la prueba para un tamaño de muestra  $N$  mayor. Esto puede ser debido a que los parámetros gran- $M$  usados era los mismos para las pruebas con y sin recourse, y en el caso con recourse estos parámetros eran más ajustados a los valores que la variable  $\zeta$  tomaba en optimalidad. No obstante, esta prueba pudo ser ejecutada solamente para  $\epsilon$  muy bajo, 1 %, y para  $\epsilon$  mayor la prueba siempre fue botada por el nodo de cómputo por falta de memoria.

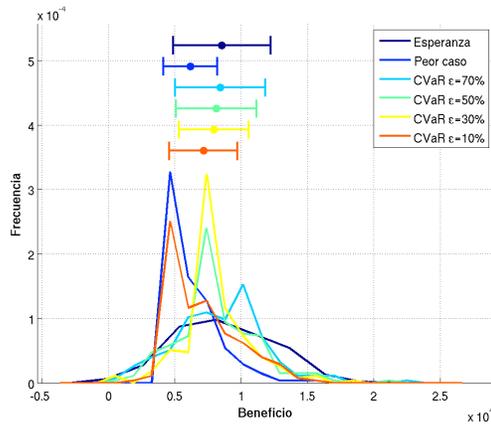
### Modelos CVaR, MCH y MCH- $\epsilon$ sin recourse



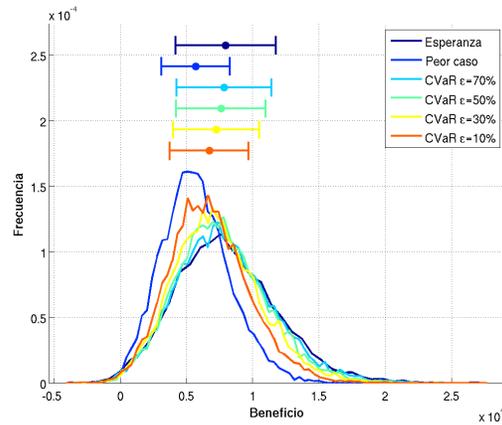
(a) Dist. acumulada, en muestra de planificación



(b) Dist. acumulada, en muestra de evaluación

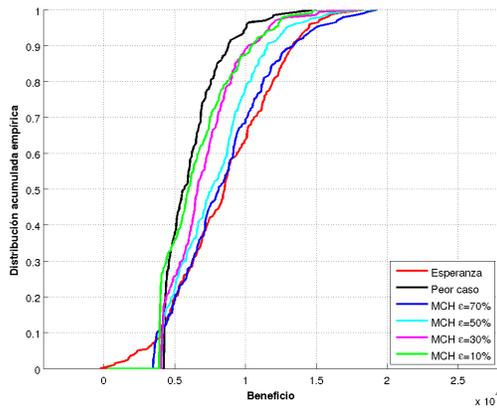


(c) Histograma, en muestra de planificación

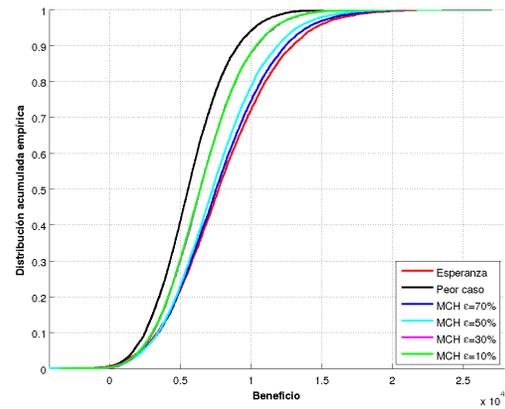


(d) Histograma, en muestra de evaluación

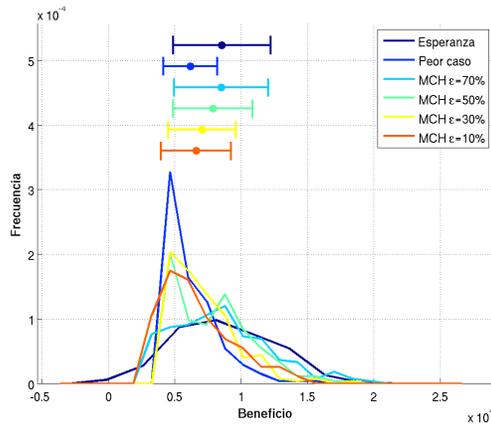
Figura 4.5: Distribución empírica de beneficios, modelo minimización de CVaR, para planes estadísticamente óptimos obtenidos resolviendo  $M = 30$  iteraciones del problema SAA con  $N = 200$  muestras.



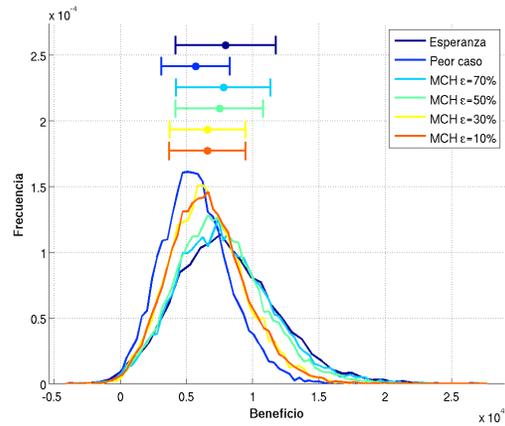
(a) Dist. acumulada, en muestra de planificación



(b) Dist. acumulada, en muestra de evaluación



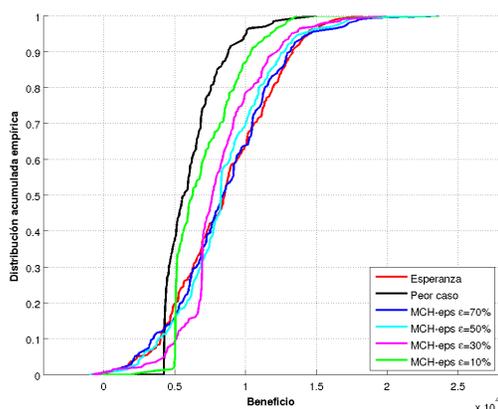
(c) Histograma, en muestra de planificación



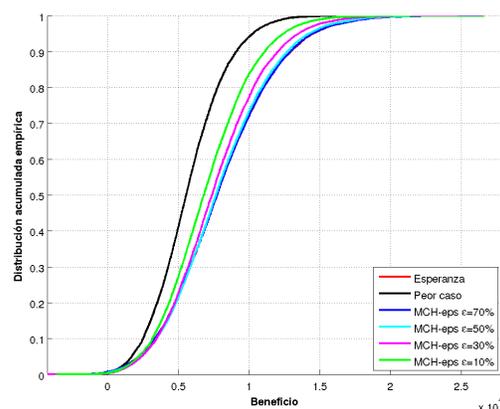
(d) Histograma, en muestra de evaluación

Figura 4.6: Distribución empírica de beneficios, modelo MCH, para planes estadísticamente óptimos obtenidos resolviendo  $M = 30$  iteraciones del problema SAA con  $N = 200$  muestras.

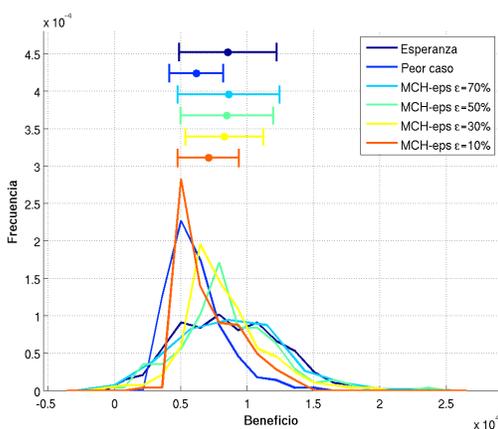
Mostramos en las Figuras 4.5, 4.6 y 4.7 las distribuciones empíricas del beneficio para las pruebas de minimización de CVaR MCH y MCH- $\epsilon$ , respectivamente. Hacemos notar que en todos estos gráficos se tiene el mismo rango para el eje de beneficios. En todas estas pruebas siempre se tomó un tamaño de muestra de  $N = 200$ , tamaño que permitió una eficiente ejecución de las pruebas. Así, por ejemplo, el peor tiempo de ejecución fue el de la prueba de peor pérdida –i.e. MCH con  $\epsilon = 0$ – que requirió en promedio 7,5 horas en completarse, y el segundo peor tiempo fue menor a 3 horas. El siguiente tamaño de muestra tomado fue  $N = 400$ , pero tuvo mala ejecución y rara vez se pudieron completar las pruebas, pues los trabajos necesitaba mucha memoria y eran botados por el nodo donde se estaban computando. También, con  $N = 200$  muestras se pudo ejecutar, para los tres modelos, las pruebas para varios niveles de riesgo  $\epsilon$ : 1, 2, 3, 5, 10, 30, 50 y 70 %. No obstante, en los gráficos no mostramos los resultados para niveles



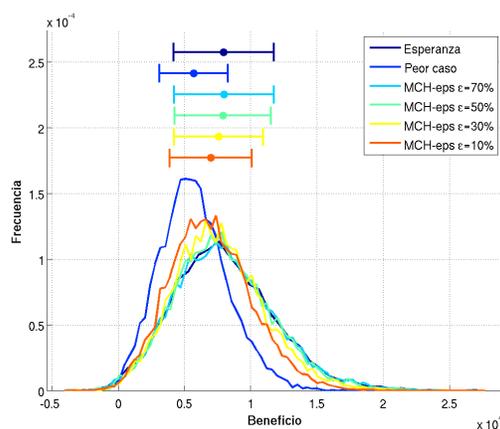
(a) Dist. acumulada, en muestra de planificación



(b) Dist. acumulada, en muestra de evaluación



(c) Histograma, en muestra de planificación



(d) Histograma, en muestra de evaluación

Figura 4.7: Distribución empírica de beneficios, modelo MCH- $\epsilon$ , para planes estadísticamente óptimos obtenidos resolviendo  $M = 30$  iteraciones del problema SAA con  $N = 200$  muestras.

de riesgo bajos pues siempre los resultados no eran significativamente distintos a los de la prueba de minimizar peor pérdida. Cabe destacar que las pruebas para estos tres modelos siempre fueron realizadas con las mismas  $M = 30$  colecciones de muestras de tamaño  $N = 200$ , i.e. se usaron solamente estas 30 colecciones de 200 muestras de  $\rho$  para correr las pruebas para los tres modelos, para todos los niveles de riesgo.

En general, para los tres modelos se observa que al disminuir  $\epsilon$  se hace variar gradualmente el riesgo de las distribuciones empíricas, cambiando progresivamente desde el resultado de minimizar esperanza al de minimizar peor caso. En efecto, al ir disminuyendo el nivel de riesgo disminuye en la muestra de planificación la probabilidad de tener beneficios bajos, disminuye la desviación estándar, y todo esto a costo de obtener un menor beneficio promedio. Este comportamiento es usual en los métodos de optimización

robusta y optimización aversa al riesgo, y puede interpretarse como que el aumento de certidumbre en el resultado final generalmente se “paga” teniendo un menor beneficio esperado.

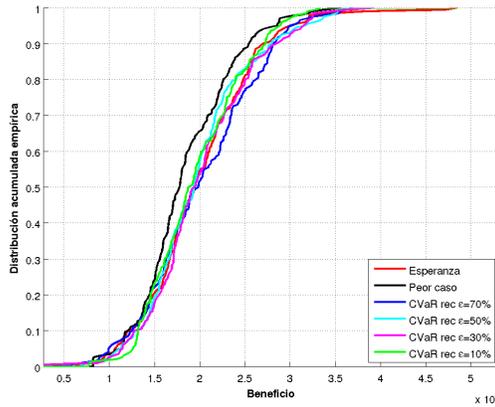
Estas tres componentes del comportamiento impuesto por los modelos se mantienen, de manera gruesa, al pasar a la muestra de evaluación. No obstante, en esta última muestra es bastante poco claro el efecto de tener menos probabilidad de tener bajos beneficios. Esto puede ser debido a que el CVaR, componente fundamental de los modelos de minimización de CVaR, MCH y MCH- $\epsilon$ , hace énfasis en los peores beneficios, lo cual exige que la muestra sea particularmente representativa de las leyes que dan los peores beneficios. E intuitivamente, para tener una muestra que represente fielmente los casos extremos, se necesita que ésta tenga gran tamaño. Podemos conjeturar entonces que un tamaño de muestra de  $N = 200$  aún no es suficientemente grande para resolver estos modelos. Esta idea es reforzada por el hecho de que, en los tres modelos, el comportamiento en general en la muestra de planificación no concuerda totalmente con el comportamiento en la muestra de evaluación.

Ahora, siendo más detallista, se aprecia en los gráficos de distribución acumulada que los modelos CVaR y MCH- $\epsilon$  se diferencian, para un mismo nivel de riesgo, solamente en los beneficios medios y altos; en los beneficios bajos los dos modelos dan el mismo resultado, tanto en la muestra de planificación –comparar Figuras 4.7(a) y 4.5(a)– como en la de evaluación –compara Figuras 4.7(b) y 4.5(b). En los rangos medios y altos, el modelo MCH- $\epsilon$  da más peso a tener mayores beneficios, comparado con el modelo MCH. En palabras simples, los resultados del modelo MCH- $\epsilon$  en el rango de beneficios bajos son casi idénticos a los del modelo CVaR, y para el rango de beneficios medios y altos el modelo MCH- $\epsilon$  tiene a parecerse más a la solución de minimizar pérdida esperada, especialmente para niveles  $\epsilon$  más altos. De nuevo, esto tanto en la muestra de planificación como en la de evaluación. Esto se puede explicar porque, para un mismo  $\epsilon$ , el modelo CVaR es minimizar  $\text{CVaR}_\epsilon$  de las pérdidas, y el modelo MCH- $\epsilon$  es minimizar  $\epsilon \mathbb{E} + (1 - \epsilon) \text{CVaR}_\epsilon$  de las pérdidas.

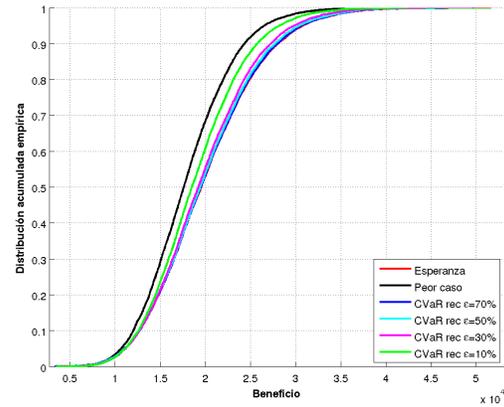
Por otro lado, el comportamiento impuesto por el modelo MCH en la muestra de planificación se aleja bastante del impuesto por los modelos CVaR y MCH- $\epsilon$ , y se caracteriza por ser sumamente conservador en el rango de las peores pérdidas. En este rango el modelo es muy similar al de minimizar peor caso (ver e.g. Figura 4.6(a)). Sin embargo, al pasar a la muestra de evaluación este comportamiento se pierde, y el desempeño es casi idéntico al que impone el modelo CVaR con el mismo  $\epsilon$  (ver e.g. las Figuras 4.6(b) y 4.5(b)).

### Modelos CVaR, MCH y MCH- $\epsilon$ con recourse

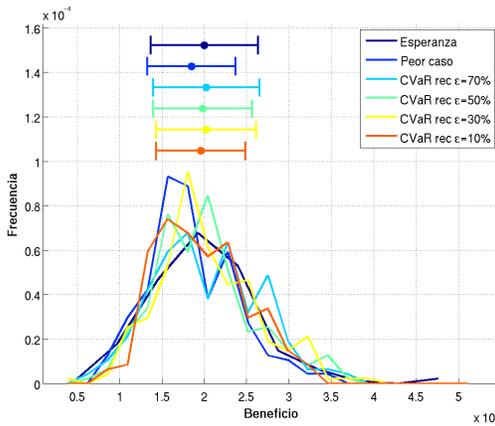
En las Figuras 4.8, 4.9 y 4.10 se muestran las distribuciones empíricas que dan los modelos de minimización de CVaR, MCH y MCH- $\epsilon$ , los tres en su variante con recourse. Destacamos que en todos estos gráficos el eje de beneficios tiene el mismo rango, para facilitar la comparación. También se muestra el desempeño de las soluciones a minimizar pérdida esperada y minimizar peor pérdida, ambas también en su variante con recourse.



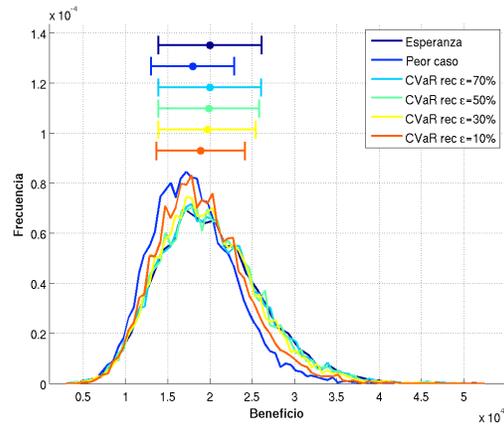
(a) Dist. acumulada, en muestra de planificación



(b) Dist. acumulada, en muestra de evaluación



(c) Histograma, en muestra de planificación

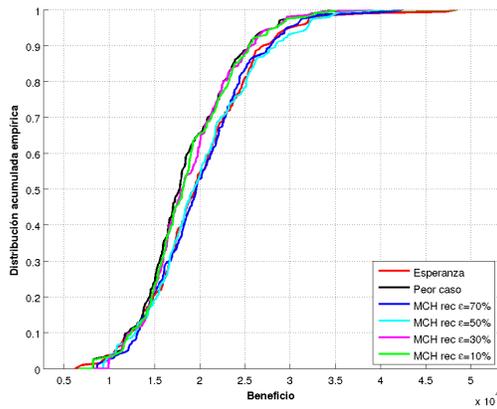


(d) Histograma, en muestra de evaluación

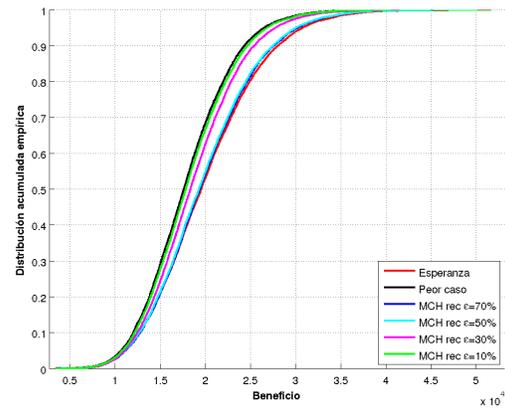
Figura 4.8: Distribución empírica de beneficios, modelo minimización de CVaR con recourse, para planes estadísticamente óptimos obtenidos resolviendo  $M = 30$  iteraciones del problema SAA con  $N = 200$  muestras.

Los resultados aquí expuestos corresponden a una planificación con una muestra de tamaño  $N = 200$ . El siguiente tamaño probado,  $N = 400$ , no se ejecutó satisfactoriamente, al igual que en el caso de estos modelos sin recourse y por los mismos motivos. Los tiempos de ejecución en las pruebas con  $N = 200$  muestras fueron siempre menores a una hora. La diferencia de tiempo con las pruebas sin recourse se debe a que éstas últimas se corrieron en el *cluster* Mansci, de la Universidad Adolfo Ibáñez, en cambio las pruebas con recourse fueron computadas en el *cluster* Vulcan, en el DII, Universidad de Chile. También, igual al caso sin recourse, en estas pruebas tomamos siempre las mismas  $M = 30$  colecciones de tamaño  $N = 200$  para hacer todas las pruebas, para los tres modelos y para varios valores de  $\epsilon$ .

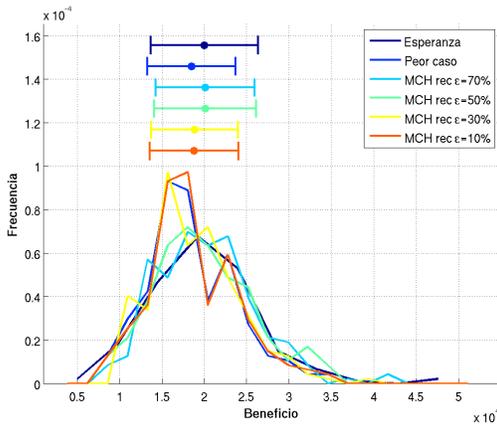
En primera instancia observamos que usar la variante con recourse da beneficios no-



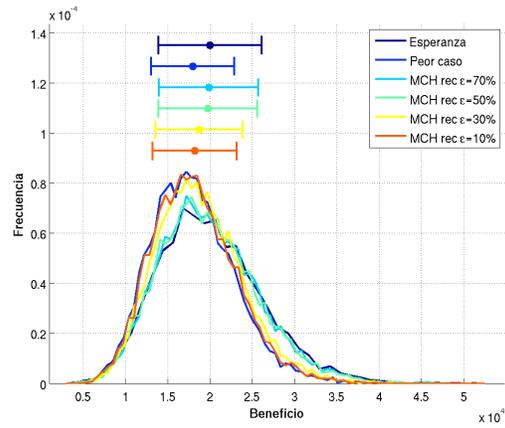
(a) Dist. acumulada, en muestra de planificación



(b) Dist. acumulada, en muestra de evaluación



(c) Histograma, en muestra de planificación

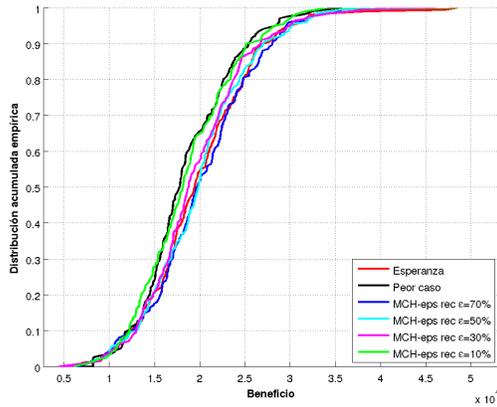


(d) Histograma, en muestra de evaluación

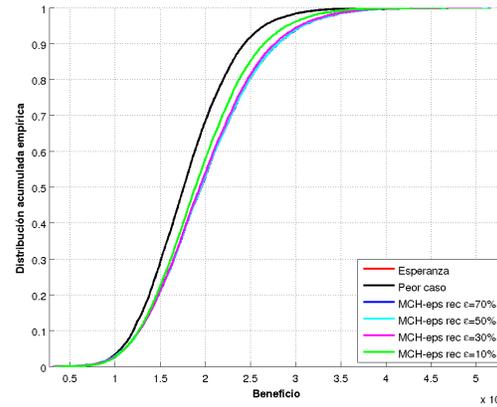
Figura 4.9: Distribución empírica de beneficios, modelo MCH con recourse, para planes estadísticamente óptimos obtenidos resolviendo  $M = 30$  iteraciones del problema SAA con  $N = 200$  muestras.

tablemente mejores que la variante sin recourse. En la Tabla 4.1 se comparan las variantes con y sin recourse de todos los enfoques con nivel de riesgo  $\epsilon = 10\%$ , salvo VaR, mostrando estimaciones de pérdida esperada y CVaR 5%. Estas dos estimaciones se hacen evaluando cada plan estadísticamente óptimo en la muestra de evaluación. Como se argumentó al inicio de este capítulo, el gran tamaño de esta muestra,  $N' = 10000$ , nos permite asumir que estas son buenas estimaciones. Vemos que el beneficio esperado prácticamente se triplica, y el riesgo –cuantificado en el CVaR 5%– disminuye casi en 900%.

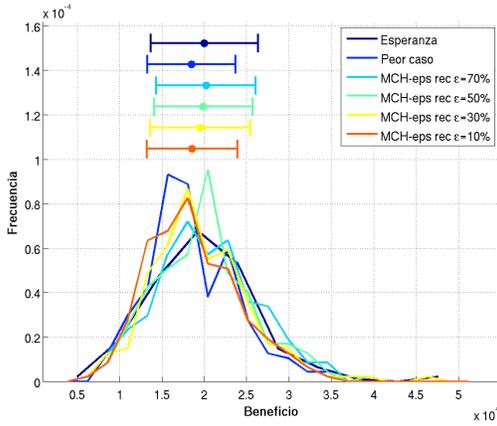
Se observa en los tres modelos que, en general, variar el nivel de riesgo  $\epsilon$  no produce una diferencia drástica en la distribución de los beneficios. En el caso de la muestra de planificación, al pasar de la variante sin recourse a con recourse se ha difuminado el efecto de que a menor nivel de riesgo hay más aversión el riesgo. Por ejemplo, en la Figura



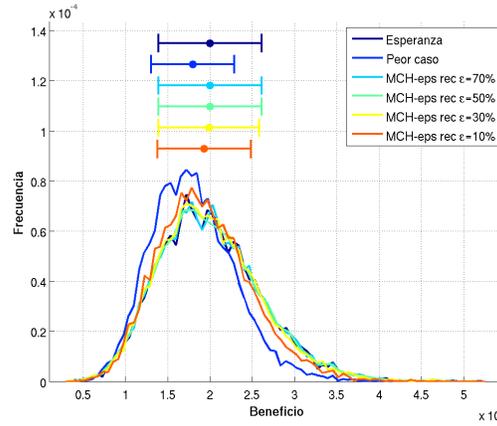
(a) Dist. acumulada, en muestra de planificación



(b) Dist. acumulada, en muestra de evaluación



(c) Histograma, en muestra de planificación



(d) Histograma, en muestra de evaluación

Figura 4.10: Distribución empírica de beneficios, modelo MCH- $\epsilon$  con recourse, para planes estadísticamente óptimos obtenidos resolviendo  $M = 30$  iteraciones del problema SAA con  $N = 200$  muestras.

	Pérdida esperada		CVaR 5%	
	Sin rec.	Con rec.	Sin rec.	Con rec.
Minimizar esperanza	-7971	-20028	-1054	-9340
CVaR, $\epsilon = 10\%$	-6717	-18930	-1193	-9606
MCH, $\epsilon = 10\%$	-6576	-18199	-1243	-9362
MCH- $\epsilon$ , $\epsilon = 10\%$	-6990	-19333	-1225	-9592
Minimizar peor caso	-5696	-17968	-942	-9239

Cuadro 4.1: Comparación de variantes con y sin recourse

4.10(a) se ve que tomar  $\epsilon = 70\%$  es incluso más riesgoso que el modelo de minimizar esperanza; en la Figura 4.8(a) las curvas de dos niveles de riesgo distintos se cruzan varias

veces; y en 4.9(a) la diferenciación entre distintos  $\epsilon$  es muy leve.

En cambio, al pasar a la muestra de evaluación vemos que la diferenciación por  $\epsilon$  es leve pero coherente con el nivel de riesgo. Por ejemplo, en 4.9(b) las curvas para distintos  $\epsilon$  son distinguibles, y a mayor  $\epsilon$  la curva se acerca uniformemente a la de minimizar esperanza; y en 4.10(b) hay niveles de riesgo distintos con curvas iguales, pero cuando las curvas son distintas éstas no se cruzan, a lo más se superponen en los sectores de beneficios bajos y altos.

Por otro lado, se observa en la muestra de evaluación que, para los tres modelos, disminuir el nivel de riesgo no produce que se de menos peso o probabilidad a los beneficios bajos. En ese sentido, pareciera ser que minimizar pérdida esperada da resultados más robustos que los obtenidos usando los modelos aversos al riesgo, pues en la muestra de evaluación no sólo pareciera dar menos peso a los beneficios bajos, sino que además daría más peso a los beneficios altos. Esto es paradójico, pues los tres modelos en cuestión se basan en reducir el CVaR, y esto, en cierta manera, considera reducir la probabilidad de los peores beneficios. Una posible explicación para esto es que el tamaño de muestra para la planificación,  $N = 200$ , no es suficientemente grande, y el argumento es el mismo que se dio para el caso sin recourse.

En síntesis, no obstante el enorme aumento en los beneficios que resulta de usar la variante con recourse, los modelos CVaR, MCH y MCH- $\epsilon$  no producen resultados significativamente distintos, y más aún, no es posible concluir que alguno de estos modelos d resultados menos riesgosos que los que da minimizar pérdida esperada.

### 4.3.2. Convergencia

Nos dedicamos ahora a mostrar y estudiar los resultados del análisis estadístico y de convergencia para los cuatro modelos, en sus variantes con y sin recourse. La idea fundamental es examinar cómo cambian, para una prueba y un nivel de riesgo dado, estimadores de interés al tomar un tamaño  $N$  de muestra mayor, y también estudiar cómo convergen estos modelos.

Para estudiar la convergencia de cada modelo nos concentramos en la cota superior estadística del gap de optimalidad. Recordemos que, en notación del Capítulo 3 y usando como ejemplo el modelo de CVaR, dado un plan  $x$  candidato, su gap de optimalidad es la distancia entre  $v^*$ , el verdadero mínimo CVaR, y el CVaR del plan  $x$ . Nos concentramos también en estudiar cómo varía la pérdida esperada de cada plan solución, y estudiar también la variación del CVaR 5%, entendiendo que éste último es un cuantificador razonable del riesgo de cada plan. Aproximaremos estos dos valores por sus estimadores en la muestra de evaluación, ya que, como hemos mencionado previamente, su gran tamaño  $-N' = 10000-$  nos permite asumir que éstas serán buenas aproximaciones.

En las siguientes siete tablas mostramos los resultados, para las variantes con y sin recourse, de todos los modelos salvo para VaR sin recourse, pues no pudimos resolver este modelo para  $N = 100$  y menos aún  $N = 200$ . Comparamos distintos niveles de riesgo  $\epsilon$

y además incluimos los resultados para los enfoques de minimizar pérdida esperada (mín E) y también minimizar peor pérdida (mín PC). Para este último enfoque no mostramos una cota superior del gap, pues la metodología SAA no permite estimar este valor.

Además, llamamos “Coef. R-P”, de *coeficiente riesgo-pérdida*, al coeficiente porcentual entre el CVaR 5 % y la pérdida esperada, para el plan estadísticamente óptimo. Dado que obtenemos siempre que el CVaR 5 % y la pérdida esperada son negativos, entonces este coeficiente será siempre positivo y representará la *certeza* del plan estadísticamente óptimo. Esto pues este coeficiente representa en cierta manera un riesgo normalizado, y como adicionalmente siempre los dos términos involucrados serán negativos, representando beneficios, entonces éste es un índice que cuánta certeza hay en el resultado final de un plan.

Cota gap	mín PC	$\epsilon = 1\%$	$\epsilon = 10\%$	$\epsilon = 30\%$	mín E	
						CVaR 5 %
						Pérdida E
						Coef. R-P [%]
$N = 50$	— -650 -5342 12.2	7398 -669 -5385 12.4	5413 -844 -5707 14.8	4492 -702 -6316 11.1	2725 -393 -7142 5.5	
$N = 100$	— -989 -6063 16.3	5867 -910 -5921 15.4	4080 -941 -6517 14.4	2801 -1156 -6649 17.4	1483 -789 -7668 10.3	
$N = 200$	— -942 -5696 16.5	4631 -1125 -6146 18.3	2830 -1193 -6717 17.8	1835 -1268 -7262 17.5	837 -1054 -7971 13.2	

Cuadro 4.2: Resumen de estadísticos para modelo minimización de CVaR

En líneas generales se observan las siguientes tendencias.

- Gap: Disminuye al aumentar  $N$  y también al aumentar  $\epsilon$ . Esto para todos los modelos salvo VaR con recourse, y salvo en MCH y MCH con recourse puse no hay una estimación del gap.
- CVaR 5 %: No se observa una tendencia monótona en  $\epsilon$ . Existe una tendencia clara a disminuir al aumentar  $N$  en modelos CVaR, con y sin recourse, MCH y MCH- $\epsilon$  sin recourse, y minimizar esperanza con y sin recourse. Y la tendencia es más bien difusa en los modelos MCH y MCH- $\epsilon$  con recourse.
- Pérdida esperada: aumenta al aumentar  $\epsilon$ . Aumenta con  $N$  en todos los modelos salvo peor caso con y sin recourse, y VaR con recourse. Sin embargo en los modelos MCH y MCH- $\epsilon$  con recourse esta tendencia es menos clara que en los otros.

CVaR 5 % Pérdida E Coef. R-P [%]	mín PC	$\epsilon = 1 \%$	$\epsilon = 10 \%$	$\epsilon = 30 \%$	mín E						
						$N = 50$	-650	-673	-710	-1012	-393
							-5342	-5400	-5521	-6673	-7142
12.2	12.5	12.9	15.2	5.5							
$N = 100$	-989	-947	-1057	-1342	-789						
	-6063	-5986	-6205	-6528	-7668						
	16.3	15.8	17.0	20.6	10.3						
$N = 200$	-942	-1214	-1243	-1324	-1054						
	-5696	-6199	-6576	-6592	-7971						
	16.5	19.6	18.9	20.1	13.2						

Cuadro 4.3: Resumen de estadísticos para modelo MCH

Cota gap CVaR 5 % Pérdida E Coef. R-P [%]	mín PC	$\epsilon = 1 \%$	$\epsilon = 10 \%$	$\epsilon = 30 \%$	mín E						
						$N = 50$	—	7350	4958	3558	2725
							-650	-647	-939	-864	-393
							-5342	-5340	-5909	-6761	-7142
12.2	12.1	15.9	12.8	5.5							
$N = 100$	—	5816	3664	2131	1483						
	-989	-947	-1053	-1290	-789						
	-6063	-5986	-6480	-7134	-7668						
	16.3	15.8	16.3	18.1	10.3						
$N = 200$	—	4583	2524	1325	837						
	-942	-1153	-1225	-1383	-1054						
	-5696	-6228	-6990	-7586	-7971						
	16.5	18.5	17.5	18.2	13.2						

Cuadro 4.4: Resumen de estadísticos para modelo MCH- $\epsilon$ 

- Coeficiente entre CVaR 5 % y pérdida esperada: En todo los modelos, salvo MCH- $\epsilon$  con recourse, este coeficiente tiende a crecer al aumentar el tamaño  $N$  de muestra.

En la sección §4.3.1 ya mencionamos los graves problemas que presentaba la tratabilidad computacional de este enfoque. Agregamos a esto los resultados aquí presentados que sugieren que el modelo no guarda relación, en ningún plano, con los enfoques de minimización de pérdida esperada y minimización de peor pérdida. Más aún, los resultados muestran que, dado  $\epsilon$ , el aumento de  $N$  en la práctica no disminuye el gap de optimalidad y tampoco se traduce en disminución de indicadores de interés, por ejemplo pérdida esperada o CVaR 5 %. Esta evidencia, en suma, contribuye a desechar el modelo de mini-

Cota gap CVaR 5 % Pérdida E Coef. R-P [%]	mín PC	$\epsilon = 1\%$	
		mín E	
$N = 50$	—	3387	480
	-9218	-9259	-9283
	-18826	-18973	-19968
	49.0	48.8	46.5
$N = 100$	—	3477	538
	-9178	-9183	-9334
	-17923	-18529	-20007
	51.2	49.6	46.7
$N = 200$	—	2350	355
	-9239	-9447	-9340
	-17968	-18659	-20028
	51.4	50.6	46.6

Cuadro 4.5: Resumen de estadísticos para modelo minimización de VaR con recourse

Cota gap CVaR 5 % Pérdida E Coef. R-P [%]	mín PC	$\epsilon$			mín E
		$\epsilon = 1\%$	$\epsilon = 10\%$	$\epsilon = 30\%$	
$N = 50$	—	4618	2098	1352	480
	-9218	-9034	-9238	-9480	-9283
	-18826	-17956	-18901	-19495	-19968
	49.0	50.3	48.9	48.6	46.5
$N = 100$	—	3678	1657	1059	538
	-9178	-9172	-9461	-9557	-9334
	-17923	-17916	-19192	-19614	-20007
	51.2	51.2	49.3	48.7	46.7
$N = 200$	—	2070	890	642	355
	-9239	-9372	-9606	-9539	-9340
	-17968	-18224	-18930	-19685	-20028
	51.4	51.4	50.7	48.5	46.6

Cuadro 4.6: Resumen de estadísticos para modelo minimización de CVaR con recourse

mización de VaR.

Para los modelos CVaR, MCH y MCH- $\epsilon$  el comportamiento es predecible y razonable: aumentar el tamaño de muestra aumenta la certeza del plan, el beneficio esperado y disminuye el CVaR 5 %. También, al tomar menores  $\epsilon$  estamos concentrándonos en una porción más extrema de la distribución de beneficios, por lo que se hace necesario tener un

CVaR 5 % Pérdida E Coef. R-P [%]	mín PC	$\epsilon = 1 \%$	$\epsilon = 10 \%$	$\epsilon = 30 \%$	mín E						
						$N = 50$	-9218	-9218	-9304	-9409	-9283
							-18826	-18826	-19004	-19061	-19968
49.0	49.0	49.0	49.4	46.5							
$N = 100$	-9178	-9172	-9351	-9533	-9334						
	-17923	-17916	-18221	-19065	-20007						
	51.2	51.2	51.3	50.0	46.7						
$N = 200$	-9239	-9239	-9362	-9604	-9340						
	-17968	-17968	-18199	-18728	-20028						
	51.4	51.4	51.4	51.3	46.6						

Cuadro 4.7: Resumen de estadísticos para modelo MCH con recourse

Cota gap CVaR 5 % Pérdida E Coef. R-P [%]	mín PC	$\epsilon = 1 \%$	$\epsilon = 10 \%$	$\epsilon = 30 \%$	mín E						
						$N = 50$	—	4560	1784	1035	480
							-9218	-9046	-9481	-9527	-9283
							-18826	-17968	-18539	-19578	-19968
$N = 100$	—	3641	1505	819	538						
	-9178	-9172	-9468	-9334	-9334						
	-17923	-17916	-19252	-19955	-20007						
$N = 200$	—	2047	790	510	355						
	-9239	-9372	-9592	-9427	-9340						
	-17968	-18224	-19333	-19899	-20028						
	51.4	51.4	49.6	47.4	46.6						

Cuadro 4.8: Resumen de estadísticos para modelo MCH- $\epsilon$  con recourse

mayor tamaño de muestra  $N$ .

Consideramos importante que en los modelos sin recourse las cotas obtenidas son muy gruesas y por tanto poco útiles. Por ejemplo, para el modelo CVaR con nivel de riesgo  $\epsilon = 10 \%$ , el análisis estadístico para  $N = 200$  nos da el intervalo  $[-4740,5, -1864,2]$ , con confianza 95 %, para el mínimo CVaR 10 %. Por otro lado, el CVaR 10 % en la muestra de evaluación es estimado por  $-1925,1$  y la pérdida esperada por  $-6717,5$ .

Finalmente, es destacable los múltiples beneficios que trae usar la variante con recourse de los modelos CVaR, MCH o MCH- $\epsilon$ . Además del claro aumento en los beneficios, también evidenciamos una reducción notoria en la cota del gap de optimalidad. Por

ejemplo, para el modelo MCH- $\epsilon$  con  $\epsilon = 10\%$  y usando  $N = 200$  para la aproximación SAA, la cota del gap se reduce de 2524 a 790 al pasar de la variante sin recourse a con recourse. También, el coeficiente entre el CVaR 5% y la pérdida esperada, que dijimos sería un índice de la *certeza* del plan estadísticamente óptimo, en las variantes con recourse siempre es de alrededor del 50%, en circunstancias en que para los modelos sin recourse está siempre entre 15 y 20%.

Como hemos obtenido de estos resultados, es un problema importante a considerar la convergencia de la aproximación SAA en cada modelo. A continuación exploramos algunos factores propios de nuestro problema que podrían afectar esta convergencia. Particularmente, veremos cuánto afecta que el problema use mayormente variables binarias, y también estudiaremos la convergencia cuando conocemos la verdadera distribución de probabilidad del vector  $\rho$ .

### 4.3.3. Estudio de algunos factores que afectan convergencia

#### Relajación de variables enteras

Una característica del problema de planificación minera es que usa una gran cantidad de variables binarias. Recordamos que, para cada bloque  $b \in \mathcal{B}$ , usamos la variable binaria  $x_b^e \in \{0, 1\}$  para representar si un bloque es extraído o no, y  $x_b^p \in \{0, 1\}$  para representar si un bloque es procesado o no. Queremos examinar si el uso de esta gran cantidad de variables binarias afecta la convergencia de las aproximaciones SAA.

Nos concentramos en los modelos CVaR y MCH- $\epsilon$  sin recourse, pues son los modelos, fuera del modelo VaR, donde es atingente hablar de convergencia. Para realizar este estudio repetimos las pruebas, para varios niveles de riesgo, siguiendo el procedimiento general de repeticiones y análisis estadístico de soluciones. Usamos los mismos parámetros de las secciones anteriores, e.g.  $M = 30$ ,  $N' = 10000$ ,  $(1 - \beta) = 95\%$ , etcétera; y la única diferencia es que ahora usaremos la relajación continua para las variables binarias, i.e. pediremos simplemente  $x_b^e, x_b^p \in [0, 1]$  para todo bloque  $b \in \mathcal{B}$ . Remarcamos que con esto son continuas todas las variables en las formulaciones de los problemas.

En la Tabla 4.9 mostramos los resultados para el modelo minimización de CVaR sin recourse, para  $\epsilon = 1\%$  y  $\epsilon = 10\%$ . Aquí “Cota gap” es la cota superior estadística, y “Coef. BE” es el coeficiente entre la cota del gap y el beneficio esperado. Vemos que no hay una diferencia significativa entre los dos resultados. Esta situación se repite para el modelo MCH- $\epsilon$  y en varios niveles de riesgo  $\epsilon$ . Esto sugiere descartar que la integralidad en variables de decisión en nuestro problema esté afectando la convergencia de la aproximación SAA.

Cota gap Coef. C-BE	$\epsilon = 1\%$		$\epsilon = 10\%$	
	LP	IP	LP	IP
$N = 50$	7312 1.27	7398 1.37	5470 0.89	5413 0.95
$N = 100$	5806 0.91	5867 0.99	3911 0.58	4080 0.63
$N = 200$	4715 0.72	4631 0.75	2874 0.44	2830 0.42
$N = 400$	3827 0.62	— —	1875 0.28	— —

Cuadro 4.9: Comparación entre versiones IP y LP para modelo CVaR sin recourse

### Distribución de probabilidad para vector $\rho$

Al presentar el análisis teórico asociado a la convergencia de cada modelo siempre establecimos que, si el vector  $\rho$  seguía una distribución estocástica que cumplía ciertas propiedades, entonces efectivamente la aproximación SAA convergería al problema verdadero. Nos dedicamos ahora a estudiar el efecto de la distribución estocástica para  $\rho$ . O más específicamente, nos concentraremos en descartar que los pobres resultados de convergencia obtenidos en los experimentos anteriores –especialmente para las variantes sin recourse– se deban a que las muestras con las que estamos trabajando no estén describiendo alguna distribución.

Para hacer esta comparación realizaremos de nuevo los pruebas, haciendo análisis estadístico y ocupando exactamente los mismos parámetros que en las veces anteriores, salvo que ahora tomaremos otra muestra de vectores de leyes  $\rho$ . Recordemos que hasta ahora estábamos tomando todas las muestras desde una gran colección de 20000 vectores  $\rho$ , obtenida con el algoritmo de [EL06]. Lo que hacemos ahora es tomar todas las muestras a partir de otra colección, también de tamaño 20000, obtenida de la siguiente manera.

1. Para cada bloque  $b \in \mathcal{B}$ , consideramos las 20000 muestras de leyes de él,  $\rho_b^1, \dots, \rho_b^{20000}$ , y obtenemos su ley promedio y desviación estándar:

$$\bar{\rho}_b := \frac{1}{20000} \sum_{k=1}^{20000} \rho_b^k, \quad \sigma_b := \sqrt{\frac{1}{20000} \sum_{k=1}^{20000} (\rho_b^k - \bar{\rho}_b)^2}.$$

2. Obtenemos 20000 muestras para  $\rho_b$  muestreando independientemente una distribución log-normal con media  $\bar{\rho}_b$  y varianza  $\sigma_b^2$ .
3. Realizando este procedimiento para cada bloque  $b \in \mathcal{B}$  obtenemos 20000 nuevas muestras del vector  $\rho$ .

Elegimos una distribución log-normal pues es generalmente asumido en la industria que las leyes de los bloques siguen esta distribución. También, para obtener muestras de este tipo basta tomar, para cada bloque  $b$ ,  $N^1, \dots, N^{20000}$  muestras iid de una normal con media 0 y varianza 1, y  $e^{\bar{\rho}_b + \sigma_b N^1}, \dots, e^{\bar{\rho}_b + \sigma_b N^{20000}}$  será una muestra iid como la que necesitamos.

Cota gap Coef. C-BE	CVaR		MCH- $\epsilon$		MCH- $\epsilon$ rec.	
	Orig.	L-N	Orig.	L-N	Orig.	L-N
$N = 50$	5413	3121	4958	2841	1784	911
	0.95	0.43	0.84	0.38	0.10	0.06
$N = 100$	4080	2160	3664	1923	1505	462
	0.63	0.28	0.57	0.25	0.08	0.03
$N = 200$	2830	1628	2524	1473	790	300
	0.42	0.21	0.36	0.19	0.04	0.02

Cuadro 4.10: Comparación entre mina original y leyes log-normales, usando  $\epsilon = 10\%$

Usando esta nueva colección de muestras para  $\rho$  realizamos de nuevo las pruebas para los modelos CVaR y MCH- $\epsilon$ , con y sin recourse. En la Tabla 4.10 mostramos los resultados para CVaR, MCH- $\epsilon$  y MCH- $\epsilon$  con recourse, siempre con  $\epsilon = 10\%$ . “Cota gap” es la cota superior estadística para el gap, “Coef. C-BE”, de *coeficiente cota-beneficio esperado*, es la razón entre la cota superior del gap y el beneficio esperado, “Orig.” significa que son los resultados para las leyes originales, y “L-N” significa que son los resultados para las leyes log-normales.

Observamos que efectivamente la cota del gap estadístico disminuye. Se obtienen en general beneficios esperados (o menos pérdidas esperadas) menores con los datos log-normales, por lo que al dividir el gap por el beneficio esperado intentamos aislar el efecto que causa este cambio en las leyes. En este nuevo coeficiente, que representa una cota normalizada por los beneficios, al pasar de los datos originales a los datos log-normales este valor disminuye siempre a alrededor de la mitad, confirmando una disminución en el gap de convergencia.

Por otro lado, las cotas superiores para el gap sigue altas en los datos log-normales, representando una parte importante del beneficio esperado. Por ejemplo, en los modelos sin recourse el coeficiente cota-beneficio esperado generalmente es mayor a 20%. En el caso de los modelos con recourse este coeficiente es mucho menor, sin embargo ya en los datos originales estos valores siempre eran menores que 10%.

Este análisis nos sugiere que efectivamente ayuda a la convergencia de la aproximación SAA que los datos provengan de una distribución razonable, con varias propiedades teóricas. Sin embargo esto no permite disminuir al mínimo el gap de optimalidad de la aproximación, e.g. en CVaR 10% sin recourse la cota normalizada siempre era mayor a 20%. Esto evidencia que en la convergencia entran en juego otros tipos de factores.

# Capítulo 5

## Conclusiones y Trabajo Futuro

### 5.1. Conclusiones

Las conclusiones principales de este trabajo son las siguientes.

- Si bien el modelo de minimización de VaR es intuitivo y usa una herramienta, VaR, de enorme popularidad en áreas como finanzas y evaluación de proyectos, los resultados aquí mostrados sugieren descartar su uso en el problema de planificación minera bajo incertidumbre. En el ámbito teórico, este modelo es deficiente pues optimiza una medida de riesgo no convexa, y por tanto no coherente. Y en la práctica, nuestros experimentos computacionales evidencian un inadecuado, o incluso nulo, manejo de riesgo. Remarcamos que, si bien fue una desventaja que la formulación usada requirió muchos más recursos computacionales, e.g. tiempo y memoria, que los otros modelos, igualmente existen otras formulaciones más eficientes y de distintos niveles de complejidad. De todas maneras, el mal desempeño mostrado por el modelo en el manejo de riesgo sugiere que es innecesario estudiar la aplicación de las técnicas de mayor eficiencia al modelo.
- Los modelos de minimización de CVaR, MCH y MCH- $\epsilon$  muestran ser modelos atractivos para tratar el problema de planificación minera bajo riesgo aquí estudiado. Si bien estos modelos son abordados usando una muestra iid del parámetro incierto, teóricamente estas aproximaciones convergen, en el caso de los modelos de CVaR y MCH- $\epsilon$ , al problema verdadero bajo supuestos razonables. Además, estas aproximaciones poseen la ventaja de poder ser vistas indistintamente como problemas de optimización robusta o como problemas de optimización aversa al riesgo usando medidas de riesgo de distorsión. Por otro lado, en el ámbito práctico estos modelos también son atractivos pues, en primer lugar, no añaden complejidad al problema de planificación original, ya que son linealizables y usan pocas variables adicionales. Y en segundo lugar, los experimentos computacionales muestran un tratamiento del riesgo deseable, pues las planificaciones obtenidas efectivamente

son aversas al riesgo y este comportamiento es controlable con el parámetro de nivel de riesgo  $\epsilon$ .

- Las estimaciones estadísticas para la convergencia de los modelos de CVaR y MCH- $\epsilon$  en §4.3.2 evidencian una lenta convergencia de las aproximaciones SAA. Los experimentos realizados en §4.3.3 muestran que esto no se debe a que el problema de planificación minera use una gran cantidad de variables binarias. También permiten descartar que usando una distribución de probabilidad sencilla para el parámetro incierto  $\rho$  aumente considerablemente la velocidad de convergencia. Por otro lado, la discusión en §2.2.4 indica que la dimensión del parámetro  $\rho$ , i.e. el número de bloques sujetos a incertidumbre, es un parámetro sumamente relevante en cuán representativa de la distribución original puede ser una muestra de  $\rho$ . Esto sugiere que la cantidad de bloques con ley incierta, que en principio es un número grande, explica la lenta velocidad de convergencia de las aproximaciones SAA.
- Por último, la incorporación del paradigma multi-etapa en las variantes con recourse muestran una enorme oportunidad para el tratamiento de la incertidumbre en el problema de planificación. Esto, pues este paradigma permite, simultáneamente, mejorar de manera considerable los beneficios, disminuir la incerteza del resultado final de las planificaciones, y a la vez mejorar la convergencia de las aproximaciones SAA.

## 5.2. Trabajo futuro

Este trabajo genera las siguientes líneas y posibilidades de trabajo futuro.

- Investigar una implementación eficiente del modelo de minimización de CVaR en el problema de planificación minera, subentendiéndose que el éxito en esta tarea también implicaría poder implementar eficientemente los modelos MCH y MCH- $\epsilon$ . Se pueden aplicar métodos específicos para minimización eficiente de CVaR, o también implementar los métodos más novedosos para resolver los problemas de planificación minera a cielo abierto (ver [CEG<sup>+</sup>09] y [BZ10]). Avances en este ámbito permitirían robustecer otros modelos de planificación en minería a cielo abierto, por ejemplo el de planificación en varios períodos. Para más ejemplos de problemas en esta área recomendamos ver [CH03, §3.2].
- Estudiar la convergencia de una combinación lineal convexa entre peor pérdida y pérdida esperada. Vimos que, dada una muestra, el modelo MCH es precisamente esto pero en el espacio finito equiprobable que define la muestra. Conjeturamos que la aproximación SAA de la minimización de la peor pérdida efectivamente converge, sin embargo no es claro cómo se acopla esto a la optimización de la combinación entre peor caso y pérdida esperada.

- Vimos que el modelo MCH y las aproximaciones SAA de los modelos CVaR y MCH- $\epsilon$ , todos en su variante sin recourse, se podían ver como minimización de una medida de riesgo de distorsión definida en el espacio equiprobable  $(\Omega_N, \mathcal{P}, \mathcal{P}_N)$  que definía la muestra de  $\rho$ . Una pregunta interesante a resolver es si esta medida de riesgo sobre el espacio equiprobable es también una medida de riesgo, y de qué tipo, al tomarla en el espacio “verdadero”  $(\Xi, \mathcal{F}, \mathcal{P})$  de  $\rho$ .
- Investigar la convergencia de las aproximaciones SAA de los modelos CVaR, MCH y MCH- $\epsilon$  cuando la muestra no es iid. Esto daría pie a posibilidades como redefinir y manipular una muestra, obteniendo una nueva muestra de tamaño tratable para el modelo robusto de elección. Así por ejemplo, podríamos tomar una muestra iid de gran tamaño que sepamos represente fielmente la distribución estocástica, luego hacer una agrupación de los datos en base a algún criterio, y tomar un representante por grupo con probabilidad proporcional.
- Un problema que quedó pendiente en nuestro trabajo fue dar una formulación tratable para la contraparte con recourse directa (2.42) del modelo MCH. O, análogamente, demostrar que no existe una formulación que puede ser resuelta en tiempo polinomial para el problema (2.42).
- Por último, y quizás la línea de mayor impacto futuro en el área minera, estudiar la factibilidad de implementar en la práctica el esquema de decisión multi-etapa que usan los modelos con recourse aquí presentados. Como muestra nuestro trabajo, este paradigma permite aumentar notablemente los beneficios y reducir el riesgo de la operación.

# Bibliografía

- [ADEH99] P. Artzner, F. Delbaen, J.M. Eber, and D. Heath, *Coherent measures of risk*, *Mathematical finance* **9** (1999), no. 3, 203–228. 4, 11, 17
- [BB09] D. Bertsimas and D.B. Brown, *Constructing uncertainty sets for robust linear optimization*, *Operations research* **57** (2009), no. 6, 1483–1495. 4, 5, 19, 24, 25
- [BBFF10] A. Bley, N. Boland, C. Fricke, and G. Froyland, *A strengthened formulation and cutting planes for the open pit mine production scheduling problem*, *Computers & Operations Research* **37** (2010), no. 9, 1641–1647. 3
- [BS03] D. Bertsimas and M. Sim, *Robust discrete optimization and network flows*, *Mathematical Programming* **98** (2003), no. 1, 49–71. 5
- [BS04] ———, *The price of robustness*, *Operations research* (2004), 35–53. 5, 23
- [BT06] D. Bertsimas and A. Thiele, *A robust optimization approach to inventory theory*, *Operations Research* **54** (2006), no. 1, 150–168. 5
- [BTGGN04] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski, *Adjustable robust solutions of uncertain linear programs*, *Mathematical Programming* **99** (2004), no. 2, 351–376. 5, 13
- [BTN00] A. Ben-Tal and A. Nemirovski, *Robust solutions of linear programming problems contaminated with uncertain data*, *Mathematical Programming* **88** (2000), no. 3, 411–424. 4, 23
- [BTN08] ———, *Selected topics in robust convex optimization*, *Mathematical Programming* **112** (2008), no. 1, 125–158. 5
- [BZ10] D. Bienstock and M. Zuckerberg, *Solving lp relaxations of large-scale precedence constrained problems*, *Integer Programming and Combinatorial Optimization* (2010), 1–14. 3, 98
- [CCS58] A. Charnes, WW Cooper, and GH Symonds, *Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil*, *Management Science* **4** (1958), no. 3, 235–263. 3

- [CEG<sup>+</sup>09] R. Chicoisne, D. Espinoza, M. Goycoolea, E. Moreno, and E. Rubio, *A new algorithm for the open-pit mine scheduling problem*, Tech. report, Working Paper, Universidad Adolfo Ibañez and Universidad de Chile, 2009. 3, 98
- [CG10] MC Campi and S. Garatti, *A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality*, Journal of Optimization Theory and Applications (2010), 1–24. 4
- [CH03] L. Caccetta and S.P. Hill, *An application of branch and cut to open pit mine scheduling*, Journal of Global optimization **27** (2003), no. 2, 349–365. 3, 98
- [CSS07] X. Chen, M. Sim, and P. Sun, *A robust optimization perspective on stochastic programming.*, Operations research (2007), no. 6, 1058–1071. 5
- [DW88] J. Dupačová and R. Wets, *Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems*, The Annals of Statistics **16** (1988), no. 4, 1517–1549. 4
- [EL06] X. Emery and C. Lantuéjoul, *Tbsim: A computer program for conditional simulation of three-dimensional gaussian random fields via the turning bands method*, Computers & Geosciences **32** (2006), no. 10, 1615 – 1628. 3, 10, 75, 95
- [HC00] D.S. Hochbaum and A. Chen, *Performance analysis and best implementations of old and new algorithms for the open-pit mining problem*, Operations Research **48** (2000), no. 6, 894–914. 3
- [KSHdM02] A.J. Kleywegt, A. Shapiro, and T. Homem-de Mello, *The sample average approximation method for stochastic discrete optimization*, SIAM Journal on Optimization **12** (2002), no. 2, 479–502. 4
- [LAN10] J. Luedtke, S. Ahmed, and G.L. Nemhauser, *An integer programming approach for linear programs with probabilistic constraints*, Mathematical Programming **122** (2010), no. 2, 247–272. 3
- [LD07] A. Leite and R. Dimitrakopoulos, *Stochastic optimisation model for open pit mine planning: application and risk analysis at copper deposit*, Mining Technology: Transactions of the Institute of Mining and Metallurgy, Section A **116** (2007), no. 3, 109–118. 3
- [LSV11] A.E.B. Lim, J.G. Shanthikumar, and G.Y. Vahn, *Conditional value-at-risk in portfolio optimization: Coherent but fragile*, Operations Research Letters (2011). 4

- [LSW06] J. Linderoth, A. Shapiro, and S. Wright, *The empirical behavior of sampling methods for stochastic programming*, *Annals of Operations Research* **142** (2006), no. 1, 215–241. 4
- [Lue10] J. Luedtke, *An integer programming and decomposition approach to general chance-constrained mathematical programs*, *Integer Programming and Combinatorial Optimization* (2010), 271–284. 3
- [Mar52] H.M. Markowitz, *Portfolio selection, 1952*, *Journal of Finance* (1952), 77–91. 4
- [Mat02] J. Matousek, *Lectures on discrete geometry*, Springer-Verlag New York, Inc., 2002. 36
- [NPS09] K. Natarajan, D. Pachamanova, and M. Sim, *Constructing risk measures from uncertainty sets*, *Operations research* **57** (2009), no. 5, 1129–1141. 5
- [PAS09] B.K. Pagnoncelli, S. Ahmed, and A. Shapiro, *Sample average approximation method for chance constrained programming: theory and applications*, *Journal of optimization theory and applications* **142** (2009), no. 2, 399–416. 3
- [RU00] R.T. Rockafellar and S. Uryasev, *Optimization of conditional value-at-risk*, *Journal of risk* **2** (2000), 21–42. 4, 16
- [RU02] ———, *Conditional value-at-risk for general loss distributions*, *Journal of Banking & Finance* **26** (2002), no. 7, 1443–1471. 4, 16, 17
- [SDR09] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on stochastic programming: modeling and theory*, Society for Industrial Mathematics, 2009. 4, 17, 18, 19, 29, 38, 39, 40, 41, 42, 44, 45, 48, 78
- [Sio58] M. Sion, *On general minimax theorems.*, *Pacific Journal of Mathematics* **8** (1958), no. 1, 171–176. 33
- [Soy73] A.L. Soyster, *Convex programming with set-inclusive constraints and applications to inexact linear programming*, *Operations research* **21** (1973), no. 5, 1154–1157. 4
- [UT98] R. Underwood and B. Tolwinski, *A mathematical programming viewpoint for solving the ultimate pit problem*, *European Journal of Operational Research* **107** (1998), no. 1, 96–107. 3
- [VEM09] J.P. Vielma, D. Espinoza, and E. Moreno (eds.), *Risk control in ultimate pits using conditional simulations*, 34th International Symposium on Application of Computers and Operations Research in The Mineral Industry, 2009. 3, 10