

UNIVERSIDAD DEL VALLE DE GUATEMALA

Facultad de Ciencias y Humanidades
Departamento de Matemática



Modelo de Black Litterman

Trabajo de graduación presentado por
Silvia María Ruiz Colmenares
para optar al grado académico
de Licenciada en Matemática

Guatemala
2017

Modelo de Black Litterman

UNIVERSIDAD DEL VALLE DE GUATEMALA

Facultad de Ciencias y Humanidades
Departamento de Matemática

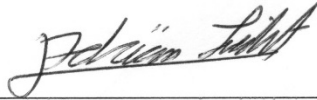


Modelo de Black Litterman

Trabajo de graduación presentado por
Silvia María Ruiz Colmenares
para optar al grado académico
de Licenciada en Matemática

Guatemala
2017

Vo.Bo.




Lic. Adrián Licht

Tribunal Examinador:



Lic. Adrián Licht



Lic. Dorval Carías



MA. Nancy Zurita

Fecha de aprobación: Guatemala, 21 de abril de 2017

Prefacio

A mi papá, Enrique Ruiz, que siempre me acompaña en mi corazón.

A mi mamá, Ana Silvia Colmenares, ejemplo de fortaleza y bondad.

A mis hermanas, Lucy y Cristy, por su amor y compañía.

A mi asesor, Adrián Licht, por toda su ayuda dentro y fuera del proceso de elaboración de este trabajo.

En mi opinión, el éxito de un individuo está fuertemente relacionado con tres aspectos: su pasión por lo que hace, su capacidad para superar la adversidad y su posición de privilegio. Siendo la última la más olvidada. La mayoría de oportunidades que tenemos, tanto académicas, económicas, sociales y hasta de género, no son merecidas, más bien es una cuestión de probabilidad y suerte. Es importante que como profesionales entendamos que muchos de nuestros logros no hubieran sido posibles sin esta posición de privilegio, claro también conllevan esfuerzo y sacrificio propio. Estar consciente de nuestra posición de privilegio nos hace tolerantes, nos hace humildes y nos motiva a generar oportunidades para otra gente que por cuestiones del azar no tiene esta posición. Como mensaje para todos los estudiantes y otras personas que les pueda servir esta tesis: aprendan a reconocer su posición de privilegio. Que esto los motive, como me motiva a mí, a ser mejores, a ser agradecidos y a crear oportunidades más justas y humanas en este país. Citando a mi papá: "Si se desea establecer una cultura de permanencia, debemos encontrar el coraje moral para enfrentar los desafíos de nuestros tiempos. De otra manera, perecemos".

Índice de contenido

RESUMEN	viii
I INTRODUCCIÓN	1
II Revisión de temas	2
A Teoría de Probabilidades	2
B Optimización Convexa	7
C Teoría de Muestreo	22
III Modelo de Black Litterman	25
A Notación	25
IV Enfoque de Bayes	26
A Inferencia Bayesiana	26
B Supuestos del modelo	28
C Visiones certeras acerca de los retornos esperados	29
D Visiones inciertas acerca de los retornos esperados	30
E Observaciones del enfoque	33
V Enfoque de Optimización Inversa	35
A Problemas de optimización convexa	35
B Teoremas de Black Litterman	41
C Observaciones del enfoque	47
VI Teoría de Muestreo	48
A Portafolio de Equilibrio	48
B Visiones del inversionista	50
C Combinación del Equilibrio de Mercado con las visiones del inversionista	51
D Observaciones del enfoque	54
VII Aplicación	55
A Ejemplo 1	55
1 Entradas del Programa	56
2 Procedimiento	57

3	Resultados	58
4	Discusión de resultados	60
B	Ejemplo 2	61
1	Entradas del Programa	61
2	Resultados	62
3	Discusión de resultados	63
C	Ejemplo 3	64
1	Entradas	64
2	Procedimiento	65
3	Resultados	65
4	Discusión de resultados	68
VII Conclusiones		69
IX Bibliografía		70
X Apéndice		71
A	Apéndice para Inferencia Bayesiana	71
B	Apéndice para Optimización Convexa	73

RESUMEN

Este trabajo de graduación se realizó con el objetivo de compilar tres diferentes enfoques desde los cuales se pueden derivar el modelo de optimización de portafolios Black Litterman. Estos enfoques fueron Inferencia Bayesiana, Optimización Convexa y Teoría de Muestreo. En la sección III se desarrolló la teoría necesaria para entender cada uno de los desarrollos en base a: Gelman (2004), Boyd y Vandenberghe (2004) y Kreyszig (1970) respectivamente. En las secciones V, VI, VII se buscó reproducir los resultados más importantes, llenando los detalles de las pruebas y agregando las demostraciones que hacían falta de Christodoulakes (2002), Bertsimas, Gupta y Paschalidis (2012) y Mankert, Seiler, (2011) respectivamente. Se hizo énfasis principalmente en el enfoque de Optimización Convexa, ya la mayoría de mejoras y aplicaciones actuales del modelo se basan en este desarrollo. Por último, en la sección VIII se realizaron dos ejemplos prácticos basados en información obtenida de Idzorek (2005) y He, Litterman (1999) para mostrar el uso del modelo.

I INTRODUCCIÓN

La teoría moderna de portafolios surgió a partir de las ideas de Markowitz en 1952. Esta teoría es un desarrollo matemático que busca construir portafolios de inversión, los cuales tengan un retorno esperado máximo sujeto a un nivel de tolerancia al riesgo, definido por el inversionista. A pesar del valor teórico que tienen los portafolios óptimos de media varianza, está ampliamente documentado que en la práctica estos tienden a ser sumamente concentrados, poco intuitivos, muy sensibles a las entradas y a maximizar el error de estimación. Esto debido a que son muy sensibles a los cambios en la matriz de varianza-correlación. El modelo de inversión Black Litterman busca solucionar estos problemas y construir portafolios de media-varianza eficientes y estables. Este modelo de localización de activos financieros fue planteado por Fischer Black y Robert Litterman en 1990. El modelo parte de la Teoría del Equilibrio General para estimar los retornos de inversiones inciertas. Luego, emplea la metodología Bayesiana para unir estas estimaciones de equilibrio con las visiones personales del inversionista acerca de los activos. (Idzorek, 2005)

Los modelos de optimización de portafolios no deben utilizarse como una caja negra. Un inversionista debe estar consciente del significado de cada uno de los parámetros del modelo y entender los alcances y limitaciones del mismo. Además, es importante considerar que el inversionista puede tener opiniones o información propia distintas a las del mercado. Por esta razón, se seleccionó el modelo de Black Litterman, ya que este utiliza las opiniones del mercado y las une con las posturas personales del inversionista. El objetivo de estudiar este modelo bajo tres perspectivas es para que el lector tenga un completo entendimiento de las premisas, los alcances, las limitaciones y los parámetros del mismo. La teoría del capítulo V mostrará como la Inferencia Bayesiana permite que la idea detrás del modelo sea bastante intuitiva y en el capítulo VIII se harán dos ejemplos prácticos usando este enfoque. Como se verá en el capítulo VII, algunos parámetros del modelo son más fácil de entender bajo el enfoque de Teoría de Muestreo. Se le dio mayor énfasis al enfoque a la teoría del capítulo VI, ya que la Optimización Convexa es el enfoque más reciente y con un creciente número de aplicaciones a finanzas. Se buscó desarrollar los conocimientos necesarios para que el lector pueda formular y resolver problemas de optimización convexa. Actualmente se están desarrollando métodos con Optimización Robusta para generalizar el modelo de Black Litterman de manera de combatir algunas de sus limitaciones.

II Revisión de temas

A Teoría de Probabilidades

Definición A.1. *Sigma Álgebra*

Sea X un conjunto, un σ -álgebra (\mathcal{A}) sobre el conjunto X se define como la familia de subconjuntos de X que cumplen las siguientes propiedades:

1. $X \in \mathcal{A}$
2. $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
3. $(A_j)_{j \in \mathbb{N}} \Rightarrow \bigcup_{j \in \mathbb{N}} A_j \in \mathcal{A}$

Teorema A.1. *Propiedad Sigma Álgebra*

Sea \mathcal{A} un σ -álgebra, entonces:

1. $\emptyset \in \mathcal{A}$
2. $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$
3. $(A_j)_{j \in \mathbb{N}} \Rightarrow \bigcap_{j \in \mathbb{N}} A_j \in \mathcal{A}$

Demostración.

1. Sabemos que $X^c = \emptyset$. Por 1.a, $X \in \mathcal{A}$. Por 1.b, $X^c = \emptyset \in \mathcal{A}$.
2. Sea $(A_j)_{j \in \mathbb{N}}$ tal que $A_1 = A, A_2 = B, A_3 = A_4 = \dots = \emptyset$. Por 1.c, $A \cup B \cup \emptyset \cup \dots \cup \emptyset = A \cup B \in \mathcal{A}$
3. Si $A_j \in \mathcal{A}$ por 1.b $\Rightarrow A_j^c \in \mathcal{A}$. Por 1.c, $\bigcup_{j \in \mathbb{N}} A_j^c \in \mathcal{A}$. Por 1.b, $(\bigcup_{j \in \mathbb{N}} A_j^c)^c = \bigcap_{j \in \mathbb{N}} A_j \in \mathcal{A}$

Definición A.2. *Medida*

Una medida, denotada por μ , se define como un mapa $\mu : \mathcal{S} \rightarrow [0, \infty]$ definido sobre un sigma álgebra \mathcal{A} que satisface:

1. $\mu(\emptyset) = 0$
2. Para una familia contable de conjuntos $(S_j)_{j \in \mathbb{N}} \subset \mathcal{S}$ disjuntos a pares, $\mu(\bigcup_{j \in \mathbb{N}} S_j) = \sum_{j \in \mathbb{N}} \mu(S_j)$

Definición A.3. *Espacio Medible*

Sea S un conjunto no vacío y \mathcal{S} un sigma álgebra de los subconjuntos de S , entonces el par (S, \mathcal{S}) define un espacio medible. Sea $S_1 \in \mathcal{S}$, entonces se dice que S_1 es medible en \mathcal{S} .

Definición A.4. *Experimento Aleatorio*

Es un proceso mediante el cual se hace una observación. Se caracteriza por llevarse a cabo bajo un conjunto bien definido de reglas, ser replicable y porque el resultado de cada repetición no se puede predecir de manera exclusiva.

Definición A.5. *Espacio Muestral*

Un espacio muestral, denotado por Ω , se define como el conjunto de todos los posibles resultados de un experimento aleatorio. Este puede ser finito o infinito.

Definición A.6. *Evento*

Un evento se define como cualquier subconjunto A del espacio muestral Ω .

Definición A.7. *Función de Probabilidad*

Sea Ω un espacio muestral y \mathcal{A} un σ -álgebra. Una función $P : \Omega \Rightarrow [0, 1]$ se define como función de probabilidad si se cumple:

1. $P(A) \geq 0 \forall A \in \mathcal{A}$
2. $P(\Omega) = 1$
3. Si $\forall n$ A_n son disjuntos a pares ($i \neq j \Rightarrow A_i \cap A_j = \emptyset$), entonces $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$

Definición A.8. *Probabilidad Condicional*

Sean (Ω, \mathcal{A}, P) un espacio de probabilidad y $A, B \in \mathcal{A}$ eventos tal que $P(B) > 0$, entonces la probabilidad condicional de A dado B se define como:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (\text{II.1})$$

Teorema A.2. *Regla de Bayes*

Sean (Ω, \mathcal{A}, P) un espacio de probabilidad y $A, B \in \mathcal{A}$ eventos tal que $P(B) > 0$, entonces:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (\text{II.2})$$

Demostración.

Por definición de probabilidad condicional se sabe que $P(A \cap B) = P(A|B)P(B)$ y $P(A \cap B) = P(B|A)P(A)$. Por tanto, $P(A|B)P(B) = P(B|A)P(A) \Rightarrow P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

Teorema A.3. *Regla de Bayes para más eventos*

Sean (Ω, \mathcal{A}, P) un espacio de probabilidad y $A, B_1, \dots, B_k \in \mathcal{A}$ eventos tal que $P(A) > 0$, entonces:

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^k P(A|B_i)P(B_i)} \quad (\text{II.3})$$

Teorema A.4. *Regla de Bayes para variables aleatorias continuas*

Sean X y Y variables aleatorias continuas, entonces:

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)} = \frac{f_{Y|X}(y|x)f_X(x)}{\int f_{Y|X}(y|x)f_X(x)dx} \quad (\text{II.4})$$

Definición A.9. *Variable Aleatoria*

Sea Ω un espacio muestral, una variable aleatoria se define como una función $X : \Omega \rightarrow R \ni X(w_j) = x_j, \forall w_j \in \Omega$, donde $x_j \in R$. Si la imagen de X es finita o contablemente infinita, entonces X se denomina una variable aleatoria discreta. Por el contrario, si la imagen es infinita, entonces X se denomina una variable aleatoria continua.

Definición A.10. *Función de Distribución*

Sea X una variable aleatoria, la función de distribución de X se define como $F(x) = P(X \leq x)$ para $-\infty < x < \infty$. Además, $F(x)$ cumple con las siguientes propiedades:

1. $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$
2. $F(\infty) = \lim_{x \rightarrow \infty} F(x) = 1$
3. $F(x)$ es una función no decreciente

Definición A.11. *Función de Densidad*

Sea X una variable aleatoria continua con función distribución $F(x) = P(X \leq x)$, esta se puede representar por una integral de la forma:

$$F(x) = \int_{-\infty}^x f(v)dv \quad (\text{II.5})$$

Donde $f(v)$ se define como función densidad de probabilidad y es una función continua, excepto en una cantidad finita de valores de v , la cual posee las siguientes características:

1. $F'(x) = f(x) \forall x$ para el cual $f(x)$ es continua.
2. $\int_{-\infty}^{\infty} f(v)dv = 1$
3. $P(a < X \leq b) = F(b) - F(a) = \int_a^b f(v)dv$
4. $f(x) \geq 0 \forall x$

Definición A.12. *Función de Densidad Conjunta*

Sean X, Y variables aleatorias, si existe una función no negativa $f_{XY} : \mathbb{R}^2 \rightarrow \mathbb{R}$ tal que para cada conjunto $A \in \mathbb{R}^2$ se tiene que:

$$P((X, Y) \in A) = \int \int_A f_{XY}(x, y) dx dy \quad (\text{II.6})$$

Entonces esta función $f_{XY}(x, y)$ se define como función densidad de probabilidad conjunta de X y Y

Definición A.13. *Distribución conjunta*

Dadas dos o más variables aleatorias X, Y, \dots en un espacio de probabilidad (Ω, \mathcal{A}, P) , la distribución de probabilidad conjunta para X, Y, \dots se define como la probabilidad que cada X, Y, \dots este dentro de un rango particular para esa variable. Si son dos variables aleatorias, se denomina distribución bivariada. Si son más de dos variables aleatorias, se denomina distribución multivariada.

1. Caso discreto:

Sean X, Y dos variables aleatorias discretas, la función de masa de probabilidad conjunta de estas se define como:

$$P(X = x \cap Y = y) = P(Y = y|X = x)P(X = x) = P(X = x|Y = y)P(Y = y) \quad (\text{II.7})$$

Donde:

$$\sum_i \sum_j P(X = x_i \cap Y = y_j) = 1 \quad (\text{II.8})$$

2. Caso continuo:

Sean X, Y dos variables aleatorias continuas, la función de densidad de probabilidad conjunta de estas se define como:

$$f_{X,Y}(x, y) = f_{Y|X}(y|x)f_X(x) = f_{X|Y}(x|y)f_Y(y) \quad (\text{II.9})$$

Donde:

$$\int_x \int_y f_{X,Y}(x, y) dy dx = 1 \quad (\text{II.10})$$

Definición A.14. *Distribución Marginal*

Sean X, Y dos variables aleatorias cuya función de distribución conjunta es conocida, entonces la distribución marginal de X se define como la distribución de X cuando el valor de Y es desconocido.

1. Caso discreto:

Si X, Y son variables aleatorias discretas, la función masa de probabilidad marginal se define como:

$$P(X = x) = \sum_y P(X = x, Y = y) = \sum_y P(X = x|Y = y)P(Y = y) \quad (\text{II.11})$$

2. Caso continuo:

Si X, Y son variables aleatorias continuas, la función densidad de probabilidad se define como:

$$f_X(x) = \int_y f_{X,Y}(x, y)dy = \int_y f_{X|Y}(x|y)f_Y(y)dy \quad (\text{II.12})$$

Definición A.15. *Distribución de Probabilidad Normal*

Sea X una variable aleatoria continua. Se dice que X tiene una distribución normal de probabilidad si y sólo si para $\sigma > 0$ y $-\infty < \mu < \infty$ la función densidad de probabilidad de X se define como:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty \quad (\text{II.13})$$

Donde:

- μ es la media o valor esperado de la distribución ($E[X]$)
- σ^2 es la varianza ($V[X]$)

Definición A.16. *Distribución de probabilidad normal multivariada*

Sean X_1, \dots, X_n variables aleatorias y $\mathbf{X} = [X_1 \dots X_n]^T$ un vector. La distribución de \mathbf{X} se define como distribución normal multivariada con media $\boldsymbol{\mu} \in R^n$ y matriz de covarianza $\Sigma \in S_{++}^n$ si su función densidad de probabilidad está dada por:

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^N |\Sigma|}} e^{-\frac{(\mathbf{x}-\boldsymbol{\mu})\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2}} \quad (\text{II.14})$$

B Optimización Convexa

Definición B.1. Problema de Optimización

Un problema de optimización se define como el problema de encontrar la mejor solución entre las soluciones factibles y se lleva a cabo maximizando o minimizando una función relativa a algún conjunto.

Definición B.2. Optimización Directa

Un problema de optimización se define como directo, si se conocen los datos \mathcal{P} y a partir de ellos se busca una solución óptima \mathbf{x}^* .

Definición B.3. Definición Optimización Inversa

Un problema de optimización se define como inverso si se propone un \mathbf{x}^* candidato y se busca encontrar los datos \mathcal{P} que hagan a este \mathbf{x}^* óptimo.

Definición B.4. Conjunto Convexo

Sea A un conjunto, A se define como un conjunto convexo ssi $\forall x_1, x_2 \in A, 0 \leq \theta \leq 1 \Rightarrow \theta x_1 + (1 - \theta)x_2 \in A$

Teorema B.1. Propiedad Conjunto Convexo

La intersección de cualquier colección de conjuntos convexos es un conjunto convexo.

Demostración.

Sea \mathcal{A} una familia de conjuntos convexos y sean $x_1, x_2 \in \cap \mathcal{A} \Rightarrow x_1, x_2 \in A_i, \forall A_i \in \mathcal{A}$. Por definición de conjunto convexo, $\forall x_1, x_2 \in A_i$ y $\theta \in [0, 1]$ se cumple $\theta x_1 + (1 - \theta)x_2 \in A_i$. Por tanto, como se cumple $\forall A_i \in \mathcal{A}, \therefore \theta x_1 + (1 - \theta)x_2 \in \cap \mathcal{A}$.

Definición B.5. Función Convexa

Sea S un conjunto no vacío convexo en R^n . La función $f : S \rightarrow R$ se define como convexa sobre S si:

$$f(\lambda \mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2) \leq \lambda f(\mathbf{x}_1) + (1 - \lambda)f(\mathbf{x}_2) \quad (\text{II.15})$$

Para cualesquiera $\mathbf{x}_1, \mathbf{x}_2 \in S$ y cualquier $\lambda \in (0, 1)$. Si la desigualdad es estricta, entonces la función se define como estrictamente convexa.

Definición B.6. Función Cóncava

Sea S un conjunto no vacío convexo en R^n . La función $f : S \rightarrow R$ se define como cóncava sobre S si $-f$ es convexa. En otras palabras, f se define como cóncava si

$$f(\lambda \mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2) \geq \lambda f(\mathbf{x}_1) + (1 - \lambda)f(\mathbf{x}_2) \quad (\text{II.16})$$

Para cualesquiera $\mathbf{x}_1, \mathbf{x}_2 \in S$ y cualquier $\lambda \in (0, 1)$. Si la desigualdad es estricta, entonces la función se denomina estrictamente cóncava.

Definición B.7. *Función Afín*

Sea $f : R^m \rightarrow R^n$ una función, f se define como afín si existen M (una matriz $n \times m$) y $b \in R^n$ tales que $\forall \mathbf{x} \in R^m$:

$$f(\mathbf{x}) = M\mathbf{x} + \mathbf{b} \quad (\text{II.17})$$

Definición B.8. *Aproximación de Taylor de Primer Orden*

Sea $f : R^n \rightarrow R$ una función infinitamente diferenciable en un intervalo abierto alrededor de $\mathbf{x} = \mathbf{a}$, entonces la aproximación de Taylor de primer orden se define como:

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \nabla f(\mathbf{a})^T (\mathbf{x} - \mathbf{a}) \quad (\text{II.18})$$

Definición B.9. *Aproximación de Taylor de Segundo Orden*

Sea $f : R^n \rightarrow R$ una función infinitamente diferenciable en un intervalo abierto alrededor de $\mathbf{x} = \mathbf{a}$, entonces la aproximación de Taylor de segundo orden está definida por:

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \nabla f(\mathbf{a})^T (\mathbf{x} - \mathbf{a}) + \frac{1}{2} (\mathbf{x} - \mathbf{a})^T \nabla^2 f(\mathbf{a}) (\mathbf{x} - \mathbf{a}) \quad (\text{II.19})$$

Definición B.10. *Optimización Convexa*

El término optimización convexa se define como maximizar o minimizar funciones convexas sobre conjuntos convexas.

Definición B.11. *Problema de optimización convexa*

Un problema de optimización convexa se define como un problema de optimización que busca minimizar:

$$f_0(x) \quad (\text{II.20})$$

Sujeto a:

$$\begin{aligned} f_i(\mathbf{x}) &\leq 0, i = 1, \dots, m \\ h_i(\mathbf{x}) &= 0, i = 1, \dots, p \end{aligned} \quad (\text{II.21})$$

Donde:

- $\mathbf{x} \in R^n$ es la variable de optimización.
- $f_0 : R^n \rightarrow R$ es la función objetivo, la cual es convexa.

- $f_i(x) \leq 0$ ($f_i : R^n \rightarrow R$) son las desigualdades de restricción, las cuales son convexas.
- $h_i(x) = 0$ son las restricciones de igualdad, las cuales son afines.
- $\mathcal{D} = \bigcap_{i=0}^m \text{dom}(f_i) \cap \bigcap_{i=1}^p \text{dom}(h_i)$ es el dominio del problema de optimización, el cual es un conjunto convexo (ya que es la intersección finita de conjuntos convexas).
- $x \in \mathcal{D}$ es factible si cumple con las restricciones de desigualdad e igualdad. El conjunto de puntos factibles se denomina conjunto o región factible.
- el punto $p^* = \inf\{f_0(\mathbf{x}) \mid f_i(\mathbf{x}), i = 1, \dots, m, h_i(\mathbf{x}) = 0, i = 1, \dots, p\}$ es el valor óptimo y $p^* = \infty$ si el problema no es factible.
- Si existen puntos factibles \mathbf{x}_k con $f_0(\mathbf{x}_k) \rightarrow -\infty$ cuando $k \rightarrow \infty$, entonces $p^* = -\infty$ y el problema es ilimitado por abajo.

Teorema B.2. *Propiedad*

En un problema de optimización convexa cualquier óptimo local es un óptimo global.

Demostración.

Sea \mathbf{x} un óptimo local para un problema de optimización convexa, entonces \mathbf{x} es factible y cumple:

$$f_0(\mathbf{x}) = \inf\{f_0(\mathbf{z}) \mid \mathbf{z} \text{ factible}, \|\mathbf{z} - \mathbf{x}\|_2 \leq R\}, \quad R > 0 \quad (\text{II.22})$$

Suponga que $\exists \mathbf{y}$ factible tal que $f_0(\mathbf{y}) < f_0(\mathbf{x})$. Note que $\|\mathbf{y} - \mathbf{x}\|_2 > R$, ya que de otra manera, por (II.22) $f_0(\mathbf{x}) \leq f_0(\mathbf{y})$. Ahora considere el punto \mathbf{z} dado por $\mathbf{z} = (1 - \lambda)\mathbf{x} + \lambda\mathbf{y}$, donde $\lambda = \frac{R}{2\|\mathbf{y} - \mathbf{x}\|_2}$. Por tanto,

$$\begin{aligned} \|\mathbf{z} - \mathbf{x}\|_2 &= \left\| \frac{2\|\mathbf{y} - \mathbf{x}\|_2 - R}{2\|\mathbf{y} - \mathbf{x}\|_2} \mathbf{x} + \frac{2\|\mathbf{y} - \mathbf{x}\|_2}{2\|\mathbf{y} - \mathbf{x}\|_2} \mathbf{y} - \mathbf{x} \right\|_2 \\ &= \left\| \frac{R(\mathbf{y} - \mathbf{x})}{2\|\mathbf{y} - \mathbf{x}\|_2} \right\|_2 = \frac{R}{2\|\mathbf{y} - \mathbf{x}\|_2} \|\mathbf{y} - \mathbf{x}\|_2 = \frac{R}{2} < R \end{aligned} \quad (\text{II.23})$$

Por tanto, \mathbf{z} es factible. Por convexidad de f_0 se tiene que:

$$f_0(\mathbf{z}) = f_0((1 - \lambda)\mathbf{x} + \lambda\mathbf{y}) \leq (1 - \lambda)f_0(\mathbf{x}) + \lambda f_0(\mathbf{y}) < f_0(\mathbf{x}) - \lambda f_0(\mathbf{x}) + \lambda f_0(\mathbf{y}) \quad (\text{II.24})$$

Por tanto $\exists \mathbf{z}$ tal que $\|\mathbf{z} - \mathbf{x}\|_2 < R$ y $f_0(\mathbf{z}) < f_0(\mathbf{x})$ ($\rightarrow \leftarrow$) $\Rightarrow \mathbf{x}$ es un óptimo global.

Teorema B.3. *Condiciones de Primer Orden*

Suponga que $f : R^n \rightarrow R$ es una función diferenciable (el gradiente ∇f existe en cada punto del dominio de f). Entonces, f es convexa si y solo si el dominio de f es convexo y

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y} \quad (\text{II.25})$$

Demostración.

1. Primero se considera el caso de $n = 1$.

Se mostrará que una función $f : R \rightarrow R$ es convexa $\Leftrightarrow f(y) = f'(x)(y-x) \quad \forall x, y \in \text{dom}(f)$.

(\Rightarrow) Suponga que f es una función convexa y que $x, y \in \text{dom}(f)$. Como el dominio de f es convexo $\forall t \in [0, 1], x + t(y-x) \in \text{dom}(f)$. Por convexidad de f ,

$$\begin{aligned} f(x + t(y-x)) &\leq (1-t)f(x) + tf(y) \\ f(x + t(y-x)) &\leq f(x) - tf(x) + tf(y) \\ f(y) &\geq f(x) + \frac{f(x + t(y-x)) - f(x)}{t} \\ f(y) &\geq f(x) + \frac{f(x + t(y-x)) - f(x)}{t(y-x)}(y-x) \end{aligned} \tag{II.26}$$

Haciendo $t \rightarrow 0$ en (II.26) se obtiene:

$$f(y) \geq f(x) + f'(x)(y-x) \tag{II.27}$$

(\Leftarrow) Suponga que la función f satisface (II.27) $\forall x, y \in \text{dom}(f)$. Escoja $x \neq y, \lambda \in [0, 1]$ y sea $z = \lambda x + (1-\lambda)y$. Evaluando x, y en (II.27) se obtiene:

$$f(x) \geq f(z) + f'(z)(x-z) \tag{II.28}$$

$$f(y) \geq f(z) + f'(z)(y-z) \tag{II.29}$$

Multiplicando (II.28) por λ y (II.29) por $(1-\lambda)$ se obtiene:

$$\lambda f(x) \geq \lambda f(z) + \lambda f'(z)x - \lambda f'(z)z \tag{II.30}$$

$$(1-\lambda)f(y) \geq f(z) - \lambda f(z) + f'(z)y - f'(z)z - \lambda f'(z)y + \lambda f'(z)z \tag{II.31}$$

Sumando (II.30) con (II.31) se obtiene:

$$\begin{aligned} \lambda f(x) + (1-\lambda)f(y) &\geq f(z) + f'(z)(\lambda x - \lambda z + \lambda z - \lambda y + y - z) \\ \lambda f(x) + (1-\lambda)f(y) &\geq f(z) + f'(z)(\lambda x + (1-\lambda)y - z) \\ \lambda f(x) + (1-\lambda)f(y) &\geq f(z) + f'(z)(z-z) \\ \lambda f(x) + (1-\lambda)f(y) &\geq f(z) \end{aligned} \tag{II.32}$$

Por tanto, f es convexa.

2. Ahora se considera el caso general $f : R^n \rightarrow R$. Sean $\mathbf{x}, \mathbf{y} \in R^n$. Considere que f está

restringida a la línea que pasa por ambos puntos, en otras palabras f está restringida a la función $g(t) = f(t\mathbf{y} + (1-t)\mathbf{x})$. Note que $g'(t) = \nabla f(t\mathbf{y} + (1-t)\mathbf{x})^T(\mathbf{y} - \mathbf{x})$.

(\Rightarrow) Suponga que f es convexa, entonces g es convexa. Por (II.27) se sabe que:

$$g(1) \geq g(0) + g'(0)(1-0) \quad (\text{II.33})$$

Además,

$$\begin{aligned} g(1) &= f(\mathbf{y}) \\ g(0) &= f(\mathbf{x}) \\ g'(0) &= \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \end{aligned} \quad (\text{II.34})$$

Sustituyendo (II.34) en (II.33) se obtiene:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \quad (\text{II.35})$$

(\Leftarrow) Ahora suponga que f cumple (II.35) $\forall \mathbf{x}, \mathbf{y}$. Por tanto, si $[t\mathbf{y} + (1-t)\mathbf{x}], [\bar{t}\mathbf{y} + (1-\bar{t})\mathbf{x}] \in \text{dom}(f)$, entonces:

$$f(t\mathbf{y} + (1-t)\mathbf{x}) \geq f(\bar{t}\mathbf{y} + (1-\bar{t})\mathbf{x}) + \nabla f(\bar{t}\mathbf{y} + (1-\bar{t})\mathbf{x})(\mathbf{y}(t-\bar{t}) - \mathbf{x}(t-\bar{t})) \quad (\text{II.36})$$

Por definición de g (II.36) se modifica a:

$$g(t) \geq g(\bar{t}) + g'(\bar{t})(t - \bar{t}) \quad (\text{II.37})$$

Por tanto, g es convexa $\Rightarrow f$ es convexa.

Nota. La desigualdad (II.25) establece que para una función convexa, la aproximación de Taylor de primer orden (II.18) es un subestimador global de la función. Por otro lado, si la aproximación de primer orden de Taylor es siempre un subestimador global, entonces la función es convexa.

Nota. Es importante notar que la desigualdad (II.25) permite derivar información global (un subestimador global de la función) de información local (su valor y su derivada en algún punto). Por ejemplo si $\nabla f(\mathbf{x}) = 0$, entonces $\forall \mathbf{y} \in \text{dom}(f)$, $f(\mathbf{y}) \geq f(\mathbf{x})$. Por tanto, \mathbf{x} es un mínimo global de la función f .

Teorema B.4.

Sea $f : R^n \rightarrow R$ una función diferenciable y convexa, sea $C \subset R^n$ un conjunto no vacío cerrado

y convexo. Considere el problema:

$$\begin{aligned} & \text{minimizar} && f(\mathbf{x}) \\ & \text{sujeto a} && \mathbf{x} \in C \end{aligned} \tag{II.38}$$

Un vector \mathbf{x}^* es óptimo para (II.38) $\Leftrightarrow \mathbf{x}^* \in C$ y

$$\nabla f(\mathbf{x}^*)^T(\mathbf{z} - \mathbf{x}^*) \geq 0 \quad \forall \mathbf{z} \in C \tag{II.39}$$

Demostración.

(\Rightarrow) Sea \mathbf{x}^* el punto óptimo y suponga que $\exists \hat{\mathbf{z}} \in C$ tal que $\nabla f(\mathbf{x}^*)^T(\hat{\mathbf{z}} - \mathbf{x}^*) < 0$. Sea $\alpha \in (0, 1)$ Como $\hat{\mathbf{z}}, \mathbf{x}^* \in C \Rightarrow \alpha \hat{\mathbf{z}} + (1 - \alpha)\mathbf{x}^* = \mathbf{x}^* + \alpha(\hat{\mathbf{z}} - \mathbf{x}^*) \in C$. Utilizando la aproximación de primer orden de Taylor (II.18), en este punto se obtiene:

$$f(\mathbf{x}^* + \alpha(\hat{\mathbf{z}} - \mathbf{x}^*)) = f(\mathbf{x}^*) + \alpha \nabla f(\mathbf{x}^*)^T(\hat{\mathbf{z}} - \mathbf{x}^*)$$

Como $\nabla f(\mathbf{x}^*)^T(\hat{\mathbf{z}} - \mathbf{x}^*) < 0$, entonces

$$f(\mathbf{x}^* + \alpha(\hat{\mathbf{z}} - \mathbf{x}^*)) < f(\mathbf{x}^*)$$

Lo cual es una contradicción, ya que \mathbf{x}^* es el óptimo. Por tanto, $\nabla f(\mathbf{x}^*)^T(\hat{\mathbf{z}} - \mathbf{x}^*) \geq 0 \quad \forall \mathbf{z} \in C$
 (\Leftarrow) Suponga que $\mathbf{x}^* \in C$ y $\nabla f(\mathbf{x}^*)^T(\mathbf{z} - \mathbf{x}^*) \geq 0 \quad \forall \mathbf{z} \in C$. Por las condiciones de primer orden (II.25) se tiene que $f(\mathbf{z}) - f(\mathbf{x}^*) \geq \nabla f(\mathbf{x}^*)^T(\mathbf{z} - \mathbf{x}^*) \quad \forall \mathbf{z} \in C$. Por tanto, $f(\mathbf{z}) - f(\mathbf{x}^*) \geq 0 \quad \forall \mathbf{z} \in C \Rightarrow f(\mathbf{z}) \geq f(\mathbf{x}^*) \quad \forall \mathbf{z} \in C \Rightarrow \mathbf{x}^*$ es el valor óptimo

Definición B.12. *Matriz Definida Positiva*

Sea M una matriz real, simétrica y de dimensión $n \times n$, entonces M está definida positiva si todos sus eigenvalores son positivos. Equivalentemente, M está definida positiva si $\forall \mathbf{x} \in R^n \quad \mathbf{x}^T M \mathbf{x} > 0$. Esto se denota por $M \succ 0$.

Definición B.13. *Matriz Semidefinida Positiva*

Sea M una matriz real, simétrica y de dimensión $n \times n$, si todos los eigenvalores de M son no negativos o $\forall \mathbf{x} \in R^n \quad \mathbf{x}^T M \mathbf{x} \geq 0$, entonces M es una matriz semidefinida positiva y se denota por $M \succeq 0$.

Teorema B.5. *Condiciones de Segundo Orden*

Sea f una función dos veces diferenciable con dominio abierto. (Esto también se puede expresar como que su Hessiano ∇^2 existe en cada punto de su dominio). Entonces, f es convexa si y solo si $dom(f)$ es convexo y su Hessiano está positivamente semi-definido (esto quiere decir $\nabla^2 f(\mathbf{x}) \succeq 0$).

Demostración.

(\Rightarrow) Se demuestra por contraposición. Suponga que el Hessiano de f no está positivamente definido en un $\mathbf{x} \in \text{dom}(f)$. Por continuidad del Hessiano esto significa que $\exists \mathbf{y} \in \text{dom}(f)$ tal que:

$$(\mathbf{y} - \mathbf{x})^T \nabla^2 f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x}) < 0 \quad (\text{II.40})$$

Utilizando la Aproximación de Taylor de segundo orden (II.19) se tiene que:

$$(\mathbf{y} - \mathbf{x})^T \nabla^2 f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x}) = 2f(\mathbf{y}) - 2f(\mathbf{x}) - 2\nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \quad (\text{II.41})$$

Sustituyendo (II.41) en (II.40) se obtiene:

$$\begin{aligned} f(\mathbf{y}) - f(\mathbf{x}) - \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) &< 0 \\ f(\mathbf{y}) &< f(\mathbf{x}) - \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \end{aligned} \quad (\text{II.42})$$

Por las condiciones de primer orden (II.25) $\Rightarrow f$ no es convexa. Por contraposición, f convexa $\Rightarrow \nabla^2 f \succeq 0$.

(\Leftarrow) Suponga que el Hessiano de f está semi-definido positivamente, entonces por definición:

$$\mathbf{x}^T [\nabla^2 f(\mathbf{x})] \mathbf{x} \geq 0 \quad \forall \mathbf{x} \in \text{dom}(f) \quad (\text{II.43})$$

Ahora, por la Aproximación de Taylor de segundo orden (II.19) se tiene que:

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) + \frac{1}{2}(\mathbf{y} - \mathbf{x})^T \nabla^2 f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x}) \quad \forall t \in [0, 1] \quad (\text{II.44})$$

Por (II.43) se sabe que $\frac{1}{2}(\mathbf{y} - \mathbf{x})^T \nabla^2 f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x}) \geq 0$. Por tanto, (II.44) se puede modificar a:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \quad (\text{II.45})$$

Por las condiciones de primer orden (II.25) $\Rightarrow f$ es convexa.

Nota.

Si el Hessiano de f está definido positivamente ($\nabla^2 f \succ 0$), entonces f es estrictamente convexa.

Teorema B.6.

Para $M \in R^{n \times n}$ los siguientes enunciados son equivalentes:

1. $M \succeq 0$
2. Existe una matriz $Q \in R^{n \times n}$ con $M = Q^T Q$

Definición B.14. *Norma*

Sea V un espacio vectorial y F un campo, entonces una norma sobre V es una función $\|\cdot\| : V \rightarrow R$ que cumple que para todo $c \in F$ y $\mathbf{x}, \mathbf{y} \in V$:

1. $\|c\mathbf{x}\| = |c| \|\mathbf{x}\|$ (homogeneidad absoluta)
2. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (desigualdad triangular)
3. Si $\|\mathbf{x}\| = 0$ entonces \mathbf{x} es el vector cero

Teorema B.7. *Proposición*

Cualquier norma es convexa.

Demostración.

Sea $\|\cdot\|$ una norma sobre un espacio vectorial V y sean $\mathbf{x}, \mathbf{y} \in V$ y $\lambda \in (0, 1)$, entonces por desigualdad triangular y homogeneidad absoluta:

$$\begin{aligned} \|\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}\| &\leq \|\lambda\mathbf{x}\| + \|(1 - \lambda)\mathbf{y}\| = |\lambda| \|\mathbf{x}\| + |1 - \lambda| \|\mathbf{y}\| = \lambda \|\mathbf{x}\| + (1 - \lambda) \|\mathbf{y}\| \\ \therefore \|\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}\| &\leq \lambda \|\mathbf{x}\| + (1 - \lambda) \|\mathbf{y}\| \end{aligned} \quad (\text{II.46})$$

Por tanto, $\|\cdot\|$ es convexa.

Lema B.8.

Una función $f : R^n \rightarrow R$ afín es convexa.

Demostración.

Se utilizarán las condiciones de primer orden. Sea $f(\mathbf{x}) = \mathbf{a}^T \mathbf{x} + b$ una función afín, entonces:

$$\begin{aligned} f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{z} - \mathbf{x}) &= \mathbf{a}^T \mathbf{x} + b + \mathbf{a}^T \mathbf{z} - \mathbf{a}^T \mathbf{x} = \mathbf{a}^T \mathbf{z} + b = f(\mathbf{z}) \\ \therefore f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{z} - \mathbf{x}) &= f(\mathbf{z}) \end{aligned} \quad (\text{II.47})$$

Por tanto, como f es una función diferenciable que cumple las condiciones de primer orden (II.25) \Rightarrow es una función convexa.

Definición B.15. *Hiperplano*

Un hiperplano se define como un conjunto $c = \{\mathbf{x} \in R^n : \mathbf{p}\mathbf{x} = \alpha\}$ donde $\mathbf{p} \in R^n$ y $\alpha \in R$.

Definición B.16. *Función semi-continua por abajo*

Una función f se define como semi-continua por abajo en \mathbf{x}_0 cuando para cada $\epsilon > 0$ existe una vecindad U de \mathbf{x}_0 tal que $f(\mathbf{x}) \geq f(\mathbf{x}_0) - \epsilon \forall \mathbf{x} \in U$.

Definición B.17. *Función coerciva*

Una función $f : R^n \rightarrow R$ se define coerciva si

$$\lim_{\|\mathbf{x}\| \rightarrow \infty} f(\mathbf{x}) = \infty$$

Esto quiere decir que para cualquier constante $M > 0$ existe una constante $R_M > 0$ tal que $\|f(\mathbf{x})\| > M$ cuando $\|\mathbf{x}\| > R_M$

Teorema B.9. *Existencia de la solución*

Sea $X \subset R^n$ un conjunto no vacío y $f : X \rightarrow R$ una función semi-continua por abajo sobre X . Si alguna de las siguientes condiciones es satisfecha:

1. La función f es coerciva sobre X y X es cerrado.
2. Para algún $\delta \in R$ el conjunto $\{\mathbf{x} \in X | f(\mathbf{x}) \leq \delta\}$ es no vacío y compacto.
3. El conjunto X es compacto.

Entonces $\inf_{\mathbf{x} \in X} f(\mathbf{x})$ es finito y existe un vector $\mathbf{x}^* \in X$ tal que:

$$f(\mathbf{x}^*) = \inf_{\mathbf{x} \in X} f(\mathbf{x}) \tag{II.48}$$

Teorema B.10. *Unicidad de la solución*

Sea $C \subset R^n$ un conjunto no vacío cerrado y sea f una función estrictamente convexa sobre C . Si el problema de minimizar f sobre C tiene una solución, entonces la solución es única.

Definición B.18. *Proyección*

Sea $C \subset R^n$ un conjunto no vacío convexo y $\hat{\mathbf{x}}$ un vector arbitrario. El problema de proyección se define como el problema de determinar el punto $\mathbf{x}^* \in C$ que sea más cercano a $\hat{\mathbf{x}}$ entre todos los $\mathbf{x} \in C$. El problema consiste en

$$\begin{aligned} &\text{minimizar} && \|\mathbf{x} - \hat{\mathbf{x}}\|^2 \\ &\text{sujeto a} && \mathbf{x} \in C \end{aligned} \tag{II.49}$$

Teorema B.11. *Teorema de Proyección*

Sea $C \subset R^n$ un conjunto no vacío cerrado y $\hat{\mathbf{x}} \in R^n$ un vector arbitrario. Entonces:

1. La proyección dada en (II.49) tiene solución única

2. El vector $\mathbf{x}^* \in C$ es la solución del problema de proyección \Leftrightarrow

$$(\mathbf{x}^* - \hat{\mathbf{x}})^T(\mathbf{x} - \mathbf{x}^*) \geq 0 \quad \forall \mathbf{x} \in C \quad (\text{II.50})$$

Demostración.

1. Como la función $f(\mathbf{x}) = \|\mathbf{x} - \hat{\mathbf{x}}\|$ es coerciva en R^n , es coerciva en C . Además, como C es cerrado, por el Teorema (B.9) apartado 1, el conjunto óptimo X^* para el problema de proyección (II.49) es no vacío. Además, el Hessiano de f está dado por $\nabla^2 f(\mathbf{x}) = 2I$. Como la identidad está definida positivamente, las condiciones de segundo orden (B.5) implican que f es estrictamente convexa y por el Teorema (B.10) la solución es única.

2. Por el Teorema (B.4) se sabe que $\mathbf{x}^* \in C$ es solución del problema de proyección si y solo si:

$$\nabla f(\mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*) \geq 0 \quad \forall \mathbf{x} \in C \quad (\text{II.51})$$

Como $\nabla f(\mathbf{x}^*) = 2(\mathbf{x}^* - \hat{\mathbf{x}})$, entonces se puede modificar (II.51) y se obtiene que $\mathbf{x}^* \in C$ es solución del problema de proyección si y solo si:

$$\nabla(\mathbf{x}^* - \hat{\mathbf{x}})^T(\mathbf{x} - \mathbf{x}^*) \geq 0 \quad \forall \mathbf{x} \in C \quad (\text{II.52})$$

Teorema B.12. *Teorema de Bolzano-Weierstrass*

Cualquier secuencia de R^n limitada tiene una subsecuencia convergente.

Teorema B.13.

Si una secuencia converge a un punto límite, entonces cada subsecuencia converge al mismo límite.

Teorema B.14. *Hiperplano de soporte*

Sea $C \subset R^n$ un conjunto no vacío convexo. Sea \mathbf{x}_0 tal que $\mathbf{x}_0 \in \text{borde}(C)$ o $\mathbf{x}_0 \notin C$. Entonces existe un hiperplano que pasa por \mathbf{x}_0 y que contiene al conjunto C en uno de sus semi-espacios. Esto quiere decir que existe un vector $\mathbf{a} \in R^n$, $\mathbf{a} \neq \mathbf{0}$ tal que:

$$\sup_{\mathbf{z} \in C} \mathbf{a}^T \mathbf{z} \leq \mathbf{a}^T \mathbf{x}_0 \quad (\text{II.53})$$

Demostración.

Sea $\{\mathbf{x}_k\} \not\subseteq C$ tal que $\mathbf{x}_k \rightarrow \mathbf{x}_0$. Sea \mathbf{z}_k^* la proyección de \mathbf{x}_k en el conjunto C para cada k . Ahora considere la secuencia $\mathbf{a}_k = \frac{\mathbf{x}_k - \mathbf{z}_k^*}{\|\mathbf{x}_k - \mathbf{z}_k^*\|}$ para $k \geq 1$. Como esta secuencia es limitada,

tiene un punto límite \mathbf{a} . Además, en virtud del Teorema de Bolzano Weierstrass (B.12), tiene una subsecuencia $\{\mathbf{a}_k\}_k$ convergente. Por el Teorema (B.13), esta subsecuencia converge a \mathbf{a} . Como \mathbf{z}_k^* es la proyección de \mathbf{x}_k para cada k , por (II.50) se tiene que:

$$(\mathbf{z}_k^* - \mathbf{x}_k)^T (\mathbf{z} - \mathbf{z}_k^*) \geq 0 \quad \forall \mathbf{z} \in C \quad (\text{II.54})$$

(II.54) se puede modificar a:

$$\begin{aligned} (\mathbf{z}_k^* - \mathbf{x}_k)^T \mathbf{z} &\geq (\mathbf{z}_k^* - \mathbf{x}_k)^T \mathbf{z}_k^* \\ (\mathbf{x}_k - \mathbf{z}_k^*)^T \mathbf{z} &\leq (\mathbf{x}_k - \mathbf{z}_k^*)^T \mathbf{z}_k^* \\ \frac{(\mathbf{x}_k - \mathbf{z}_k^*)^T}{\|\mathbf{x}_k - \mathbf{z}_k^*\|} \mathbf{z} &\leq \frac{(\mathbf{x}_k - \mathbf{z}_k^*)^T}{\|\mathbf{x}_k - \mathbf{z}_k^*\|} \mathbf{z}_k^* \end{aligned} \quad (\text{II.55})$$

Por definición de \mathbf{a}_k (II.55) se modifica a:

$$\mathbf{a}_k^T \mathbf{z} \leq \mathbf{a}_k^T \mathbf{z}_k^* \quad \forall \mathbf{z} \in C \quad (\text{II.56})$$

Como $\mathbf{a}_k^T \mathbf{z}_k^* = \mathbf{a}_k^T (\mathbf{z}_k^* - \mathbf{x}_k) + \mathbf{a}_k^T \mathbf{x}_k < \mathbf{a}_k^T \mathbf{x}_k$, (II.56) se modifica a:

$$\mathbf{a}_k^T \mathbf{z} < \mathbf{a}_k^T \mathbf{x}_k \quad \forall \mathbf{z} \in C \quad (\text{II.57})$$

Como $\mathbf{x}_k \rightarrow \mathbf{x}_0$ y $\mathbf{a}_k \rightarrow \mathbf{a}$ se obtiene que:

$$\mathbf{a}^T \mathbf{z} \leq \mathbf{a}^T \mathbf{x}_0 \quad \forall \mathbf{z} \in C \quad (\text{II.58})$$

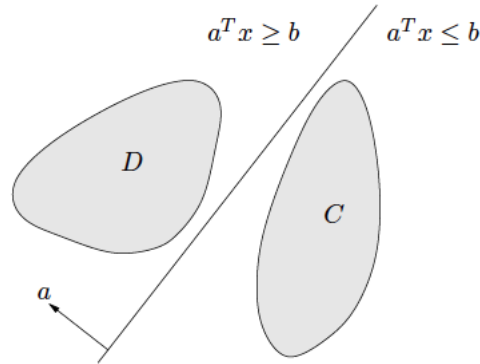
Teorema B.15. *Teorema del hiperplano de separación*

Sean C, D dos conjuntos convexos no vacíos y disjuntos ($C \cap D = \emptyset$). Entonces existen $\mathbf{a} \neq 0, b$ tales que $\mathbf{a}^T \mathbf{x} \leq b \quad \forall \mathbf{x} \in C$ y $\mathbf{a}^T \mathbf{z} \geq b \quad \forall \mathbf{z} \in D$. Esto quiere decir que

$$\sup_{\mathbf{x} \in C} \mathbf{a}^T \mathbf{x} \leq \inf_{\mathbf{z} \in D} \mathbf{a}^T \mathbf{z} \quad (\text{II.59})$$

El hiperplano $\{\mathbf{y} | \mathbf{a}^T \mathbf{y} = b\}$ se denomina el hiperplano de separación para los conjuntos C, D .

Figura II.1: El Hiperplano de separación (Boyd, 2004)

**Demostración.**

Considere el conjunto $A = C - D$. Este conjunto es no vacío. Sea $\mathbf{x}_0 = \mathbf{0}$. Como $C \cap D = \emptyset$ por el Teorema de Hiperplano de Soporte (B.14) se sabe que existe un $\mathbf{a} \in R^n$ tal que:

$$\mathbf{a}^T \mathbf{y} \leq 0 \quad \forall \mathbf{y} \in A \quad (\text{II.60})$$

Como $A = C - D$, (II.60) se modifica a:

$$\mathbf{a}^T \mathbf{x} \leq \mathbf{a}^T \mathbf{z} \quad \forall \mathbf{x} \in C, \forall \mathbf{z} \in D \quad (\text{II.61})$$

Tomando el supremo sobre C y el ínfimo sobre D de (II.61) se obtiene:

$$\sup_{\mathbf{x} \in C} \mathbf{a}^T \mathbf{x} \leq \inf_{\mathbf{z} \in D} \mathbf{a}^T \mathbf{z} \quad (\text{II.62})$$

Teorema B.16. *Complemento de Schur:*

Sea A una matriz $n \times n$ invertible y \mathbf{b} un vector de dimensión n . Suponga que A se puede escribir de la siguiente manera:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

Donde A_{11} , A_{12} , A_{21} y A_{22} son matrices $p \times p$, $p \times q$, $q \times p$ y $q \times q$ invertibles y $p + q = n$. Entonces, el complemento de Schur del bloque A_{22} está dado por:

$$S = A_{11} - A_{12}A_{22}^{-1}A_{21} \quad (\text{II.63})$$

Equivalentemente, el complemento de Schur del bloque A_{12} está dado por:

$$S = A_{21} - A_{12}A_{11}^{-1}A_{12} \quad (\text{II.64})$$

Este es un método para resolver $A\mathbf{x} = \mathbf{b}$. Consiste en eliminar un subconjunto de las variables y resolver un sistema más pequeño de ecuaciones.

Demostración.

Sea $\mathbf{x} \in R^n$ y suponga que se particiona de la siguiente manera $\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}$ donde $\mathbf{x}_1 \in R^{n_1}$ y $\mathbf{x}_2 \in R^{n_2}$. Entonces el problema se puede reescribir de la siguiente manera:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} \quad (\text{II.65})$$

Donde $A_{11} \in R^{n_1 \times n_1}$, $A_{22} \in R^{n_2 \times n_2}$.

$$A_{11}\mathbf{x}_1 + A_{12}\mathbf{x}_2 = \mathbf{b}_1 \quad (\text{II.66})$$

$$A_{21}\mathbf{x}_1 + A_{22}\mathbf{x}_2 = \mathbf{b}_2 \quad (\text{II.67})$$

Suponiendo que A_{22} es invertible, se puede expresar \mathbf{x}_2 de la siguiente manera:

$$\mathbf{x}_2 = A_{22}^{-1}(\mathbf{b}_2 - A_{21}\mathbf{x}_1) \quad (\text{II.68})$$

Sustituyendo (II.68) en (II.66) se obtiene:

$$\begin{aligned} A_{11}\mathbf{x}_1 + A_{12}A_{22}^{-1}(\mathbf{b}_2 - A_{21}\mathbf{x}_1) &= \mathbf{b}_1 \\ \mathbf{x}_1 &= (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}(\mathbf{b}_1 - A_{12}A_{22}^{-1}\mathbf{b}_2) \end{aligned} \quad (\text{II.69})$$

El complemento de Schur para el bloque A_{22} está dado por:

$$S = A_{11} - A_{12}A_{22}^{-1}A_{21} \quad (\text{II.70})$$

Expresando \mathbf{x}_2 en términos de \mathbf{x}_1

$$\begin{aligned} \mathbf{x}_2 &= A_{22}^{-1}(\mathbf{b}_2 - A_{21}\mathbf{x}_1) \\ \mathbf{x}_2 &= A_{22}^{-1}(\mathbf{b}_2 - A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}(\mathbf{b}_1 - A_{12}A_{22}^{-1}\mathbf{b}_2)) \end{aligned} \quad (\text{II.71})$$

Equivalentemente si se despeja para la segunda variable el complemento de Schur para el bloque

A_{12} está dado por:

$$S = A_{21} - A_{12}A_{11}^{-1}A_{12} \quad (\text{II.72})$$

Definición B.19. *Gráfica*

La gráfica de una función $f : R^n \rightarrow R$ se define como $\{(\mathbf{x}, f(\mathbf{x})) | \mathbf{x} \in \text{dom}(f)\} \subseteq R^{n+1}$.

Definición B.20. *Epigrafo*

El epigrafo de una función $f : R^n \rightarrow R$ se define como el conjunto de puntos situados en o sobre la gráfica $\text{epi}(f) = \{(\mathbf{x}, t) | \mathbf{x} \in \text{dom}(f), f(\mathbf{x}) \leq t\} \subseteq R^{n+1}$

Teorema B.17. *Propiedad*

Una función es convexa \Leftrightarrow su epigrafo es un conjunto convexo.

Demostración.

(\Rightarrow) Sea f una función convexa y sean $(\mathbf{x}, a), (\mathbf{y}, b) \in \text{epi}(f)$. Como f es convexa y por definición de epigrafo:

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y}) \leq \lambda a + (1-\lambda)b \quad \forall \lambda \in [0, 1]$$

Por tanto, $\lambda(\mathbf{x}, a) + (1-\lambda)(\mathbf{y}, b) \in \text{epi}(f) \Rightarrow \text{epi}(f)$ es convexo.

(\Leftarrow) Suponga que el epigrafo de f es convexo y sean $\mathbf{x}, \mathbf{y} \in \text{dom}(f)$, tal que $f(\mathbf{x}) \leq a$ y $f(\mathbf{y}) \leq b$. Como el epigrafo es convexo se sabe que $\lambda(\mathbf{x}, a) + (1-\lambda)(\mathbf{y}, b) \in \text{epi}(f)$, por lo que:

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda a + (1-\lambda)b$$

Haciendo $a = f(\mathbf{x})$ y $b = f(\mathbf{y})$ se obtiene:

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y})$$

$\Rightarrow f$ es convexo

Teorema B.18. *Propiedad*

Suponga que $A \succeq 0$ y considere el problema de minimización (donde \mathbf{u} es la variable):

$$\mathbf{u}^T A \mathbf{u} + 2\mathbf{v}^T B^T \mathbf{u} + \mathbf{v}^T C \mathbf{v} \quad (\text{II.73})$$

La solución está dada por:

$$\mathbf{u} = -A^{-1}B\mathbf{v} \quad (\text{II.74})$$

El valor optimal es:

$$\inf_{\mathbf{u}} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix}^T \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \mathbf{v}^T S \mathbf{v} \quad (\text{II.75})$$

Donde S es el complemento de Schur y $X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$. A partir de (II.75) se obtienen dos corolarios acerca de las siguientes caracterizaciones de definida o semidefinida positivamente para la matriz X .

Corolario B.18.1.

$X \succ 0$ si y solo si $A \succ 0$ y $S \succ 0$

Corolario B.18.2.

Si $A \succ 0$, entonces $X \succeq 0$ si y solo si $S \succeq 0$

Teorema B.19. Propiedad

La función $f : R^n \times S^n \rightarrow R$ definida por $f(\mathbf{x}, M) = \mathbf{x}^T M^{-1} \mathbf{x}$ es convexa.

Demostración.

Por definición de epigrafo sabemos que:

$$\text{epi}(f) = \{(\mathbf{x}, M, t) | M \succeq 0, \mathbf{x}^T M^{-1} \mathbf{x} \leq t\} \quad (\text{II.76})$$

Sea

$$X = \begin{bmatrix} M & \mathbf{x}^T \\ \mathbf{x} & t \end{bmatrix}$$

Utilizando (II.72) se obtiene que $S = t - \mathbf{x}^T M^{-1} \mathbf{x}$. Por hipótesis se sabe que $\mathbf{x}^T M^{-1} \mathbf{x} \leq t \Rightarrow t - \mathbf{x}^T M^{-1} \mathbf{x} \geq 0$. Por tanto $S \succeq 0$. Además, por hipótesis $M \succ 0$. Utilizando el corolario (B.18.2) se obtiene $X \succeq 0$. Por tanto, (II.76) se modifica a:

$$\text{epi}(f) = \{(\mathbf{x}, M, t) | X = \begin{bmatrix} M & \mathbf{x}^T \\ \mathbf{x} & t \end{bmatrix} \succeq 0, M \succ 0\} \quad (\text{II.77})$$

\Rightarrow el epigrafo de f es convexo. Por (B.17), f es convexa.

C Teoría de Muestreo

Definición C.1. *Parámetro*

Un parámetro se define como una característica poblacional que aparece en las funciones de distribución. Por ejemplo, μ, σ en la distribución normal.

Definición C.2. *Estimación puntual*

El empleo de un único valor extraído de la muestra para estimar un parámetro desconocido de la población se define como estimación puntual. Por ejemplo, la estimación puntual de la media poblacional μ , en base a una muestra de tamaño n , se hace a través de la media muestral:

$$\hat{\mu} = \bar{x} = \frac{1}{n}(\mathbf{x}_1 + \dots + \mathbf{x}_n) \quad (\text{II.78})$$

De igual manera, la estimación puntual de la varianza poblacional σ^2 se hace a través de la varianza muestral:

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1} \sum_{j=1}^n (\mathbf{x}_j - \bar{x})^2 \quad (\text{II.79})$$

Definición C.3. *Estimador y estimación*

Considere una distribución que involucra un solo parámetro desconocido θ y suponga que se cuenta con una fórmula para obtener una estimación $\hat{\theta}$ de θ dada una muestra $\mathbf{x}_1, \dots, \mathbf{x}_n$ (donde estos son n valores observados de la misma variable aleatoria X). Como $\hat{\theta}$ depende de los valores muestral, se puede considerar como una función de esos valores:

$$\hat{\theta} = g(\mathbf{x}_1, \dots, \mathbf{x}_n) \quad (\text{II.80})$$

Para valores dados $\mathbf{x}_1, \dots, \mathbf{x}_n$ corresponde un cierto valor numérico de g . Este valor se define como estimación del parámetro θ . Ahora si se considera la muestra $\mathbf{x}_1, \dots, \mathbf{x}_n$ como observaciones individuales de n variables aleatorias X_1, \dots, X_n , las cuales tienen la misma distribución X y son independientes (porque los valores muestrales se asumen como independientes). Entonces se puede considerar $\hat{\theta} = g(\mathbf{x}_1, \dots, \mathbf{x}_n)$ como una observación individual de la variable aleatoria:

$$\hat{\Theta} = g(X_1, \dots, X_n) \quad (\text{II.81})$$

Esta variable aleatoria $\hat{\Theta}$ se define como estimador para el parámetro θ . Como esta es una función no constante de las variables aleatorias X_1, \dots, X_n se denomina un estadístico.

Definición C.4. *Estimador y estimación no-sesgada*

Un estimador $g(X_1, \dots, X_n)$ para un parámetro θ se define no-sesgado si:

$$E[g(X_1, \dots, X_n)] = \theta \quad (\text{II.82})$$

Un valor correspondiente de $g(\mathbf{x}_1, \dots, \mathbf{x}_n)$ obtenido de la muestra $\mathbf{x}_1, \dots, \mathbf{x}_n$ se define como una estimación no-sesgada.

Definición C.5. *Varianza de un estimador*

La varianza de un estimador $\hat{\Theta} = g(X_1, \dots, X_m)$ para un parámetro θ se define como:

$$E\left([\hat{\Theta} - \theta]^2\right) \quad (\text{II.83})$$

Definición C.6. *Estimador eficiente*

Un estimador no sesgado $\hat{\Theta} = g(X_1, \dots, X_m)$ para un parámetro θ se define como eficiente si tiene una varianza finita y si no existe otro estimador $\hat{\Theta}_* = g^*(X_1, \dots, X_n)$ para θ , cuya varianza sea más pequeña que la de $\hat{\Theta}$.

Definición C.7. *Función Verosimilitud*

Sea X una variable aleatoria discreta o continua, cuya función de masa o densidad de probabilidad f , respectivamente, depende únicamente del parámetro θ . Suponga que se realiza un experimento aleatorio n veces y se obtiene la siguiente muestra:

$$\mathbf{x}_1, \dots, \mathbf{x}_n \quad (\text{II.84})$$

Suponga que los n elementos de la muestra son independientes. En el caso discreto, la probabilidad de que una muestra consista precisamente de esos n valores (II.84) está dada por el producto:

$$l = f(\mathbf{x}_1)f(\mathbf{x}_2)\dots f(\mathbf{x}_n) \quad (\text{II.85})$$

En el caso continuo se considera la probabilidad de que la muestra consista de valores dentro de los siguientes intervalos:

$$\mathbf{x}_1 \leq \mathbf{x} \leq \mathbf{x}_1 + \Delta\mathbf{x}, \dots, \mathbf{x}_n \leq \mathbf{x} \leq \mathbf{x}_n + \Delta\mathbf{x} \quad (\text{II.86})$$

Esta probabilidad está dada por:

$$f(\mathbf{x}_1)\Delta\mathbf{x}\dots f(\mathbf{x}_n)\Delta\mathbf{x} = l(\Delta\mathbf{x})^n \quad (\text{II.87})$$

Los valores $f(\mathbf{x}_1), \dots, f(\mathbf{x}_n)$ dependen del parámetro θ . Por lo que la función l depende de $\mathbf{x}_1, \dots, \mathbf{x}_n$ y θ . Si los valores $\mathbf{x}_1, \dots, \mathbf{x}_n$ son dados y fijos, entonces l es una función de θ , definida como función de verosimilitud.

Definición C.8. *Método de Máxima Verosimilitud*

El método de máxima verosimilitud se utiliza para obtener estimaciones de los parámetros. Este consiste en escoger la aproximación para el valor desconocido de θ , para la cual el valor de l sea lo mayor posible. Si l es una función diferenciable de θ , entonces se iguala su primera derivada respecto a θ a 0:

$$\frac{\partial l}{\partial \theta} = 0 \quad (\text{II.88})$$

Una solución para (II.88) que dependa de (II.84) se define como una estimación de máxima verosimilitud para el parámetro θ . Si en la solución se reemplazan (II.84) por las variables aleatorias X_1, \dots, X_n , las cuales tienen la distribución de X , se obtiene una variable aleatoria definida como estimador de máxima verosimilitud para el parámetro θ . Si la distribución posee varios parámetros $\theta_1, \dots, \theta_k$, entonces en lugar de la ecuación (II.88) se tienen k ecuaciones de la misma forma:

$$\frac{\partial l}{\partial_1 \theta} = 0, \dots, \frac{\partial l}{\partial_k \theta} = 0 \quad (\text{II.89})$$

Nota. Como $f(\mathbf{x})$ es una función no-negativa, el valor máximo de l es positivo. Como el logaritmo natural es una función monótona creciente, $\ln(l)$ alcanza su máximo precisamente en el mismo punto que l tiene su máximo. Para simplificar los cálculos normalmente se opta por trabajar con el logaritmo natural de la función de verosimilitud. Por tanto, (II.88) y (II.89) se modifican a:

$$\frac{\partial \ln(l)}{\partial \theta} = 0 \quad (\text{II.90})$$

$$\frac{\partial \ln(l)}{\partial_1 \theta} = 0, \dots, \frac{\partial \ln(l)}{\partial_k \theta} = 0 \quad (\text{II.91})$$

III Modelo de Black Litterman

El modelo de Black Litterman es una estrategia de optimización de portafolios de inversión, la cual permite combinar las visiones subjetivas del inversionista con un punto de referencia neutral del mercado. Este modelo es muy intuitivo. Parte de los retornos en equilibrio observados por el mercado y se desvía según las expectativas y grado de confianza del inversionista. Los portafolios optimizados con este método resuelven varios de los problemas de los portafolios de Markowitz, los cuales muchas veces son poco intuitivos, muy concentrados, poco diversificados, bastante sensibles a los parámetros de entrada e inestables. (Idzorek, 2005)

Antes de iniciar el desarrollo del modelo desde las tres perspectivas anteriormente mencionadas, se presenta la notación que se utilizará en todos los enfoques.

A Notación

- $\boldsymbol{\Pi}$ es el vector $(n \times 1)$ de los retornos en exceso implícitos de equilibrio
- λ es el coeficiente de aversión al riesgo
- \mathbf{r} es el vector $(n \times 1)$ de retornos en exceso de los n activos riesgosos
- r_f es el escalar que representa el retorno en exceso del activo libre de riesgo
- Σ es la matriz $(n \times n)$ de covarianza de los retornos en exceso
- \mathbf{x}_m es el vector $(n \times 1)$ de los pesos de capitalización
- w_i es el nivel de confianza asignado a la visión i . Es la desviación estándar del retorno esperado
- Ω es la matriz $(k \times k)$ diagonal de covarianza de los errores de las visiones (su diagonal consiste de w_1^2, \dots, w_k^2)
- τ es un escalar
- P es una matriz $(k \times n)$ que representa parte de las visiones. Cada fila de la matriz contiene los pesos de los activos en una visión. El número máximo de filas es el número máximo de activos en el portafolio
- \mathbf{q} es un vector $(k \times 1)$ que representa los retornos esperados en cada visión
- $E(\mathbf{r})$ es el vector $(n \times 1)$ posterior de los retornos esperados modificado por Black-Litterman

IV Enfoque de Bayes

A Inferencia Bayesiana

La teoría acerca de la Inferencia Bayesiana se desarrollará con base en Gelman (2014). El objetivo de la Inferencia Bayesiana es incorporar conocimiento *a priori* en probabilidades estadísticas. La cantidad desconocida θ se considera una variable aleatoria y se supone una estimación inicial de su distribución. Luego, cuando haya más información o datos disponibles la distribución de θ se actualiza utilizando la regla de Bayes (II.2). Se utilizará la siguiente notación:

- θ es el vector de cantidades no observables o parámetros poblacionales de interés
- \mathbf{y} es el vector de datos observables
- $\tilde{\mathbf{y}}$ denota el vector de las cantidades desconocidas pero potencialmente observables
- $f(\cdot|\cdot)$ es la función densidad de probabilidad condicional entre los dos argumentos
- $f(\cdot)$ denota la distribución marginal
- $P(\cdot)$ es la probabilidad de un evento
- $N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ es la distribución normal para variables aleatorias
- $N(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ es la distribución normal para funciones densidad

Para poder realizar afirmaciones sobre θ dado el valor conocido de \mathbf{y} , se inicia con la distribución de probabilidad conjunta de θ y \mathbf{y} . La función masa o densidad de probabilidad conjunta se puede escribir como producto de dos funciones densidad: la distribución a priori $f(\boldsymbol{\theta})$ y la distribución de muestreo $f(\mathbf{y}|\boldsymbol{\theta})$.

$$f(\boldsymbol{\theta}, \mathbf{y}) = f(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta}) \quad (\text{IV.1})$$

La función densidad de probabilidad posterior se obtiene al aplicar la regla de Bayes (II.2).

$$f(\boldsymbol{\theta}|\mathbf{y}) = \frac{f(\boldsymbol{\theta}, \mathbf{y})}{f(\mathbf{y})} = \frac{f(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})}{f(\mathbf{y})} \quad (\text{IV.2})$$

Donde, en el caso discreto $f(\mathbf{y}) = \sum_{\boldsymbol{\theta}} f(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})$ sobre todos los valores de $\boldsymbol{\theta}$ o en el caso continuo $f(\mathbf{y}) = \int f(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})d\boldsymbol{\theta}$. Debido a que el factor $f(\mathbf{y})$ no depende de $\boldsymbol{\theta}$, al fijar \mathbf{y} se

puede considerar como una constante. Por tanto, se obtiene la función de densidad posterior no normalizada:

$$f(\boldsymbol{\theta}|\mathbf{y}) \propto f(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta}) \quad (\text{IV.3})$$

Definición A.1. *Distribución a priori*

Las inferencias predictivas son inferencias que se hacen acerca de una cantidad desconocida observable. Antes que los datos \mathbf{y} sean considerados, la distribución de los datos desconocidos, pero observables, \mathbf{y} está dada por (II.12):

$$f(\mathbf{y}) = \int f(\mathbf{y}, \boldsymbol{\theta})d\boldsymbol{\theta} = \int f(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})d\boldsymbol{\theta} \quad (\text{IV.4})$$

Usualmente esta se denomina distribución marginal de \mathbf{y} o distribución predictiva *a priori*. Se llama *a priori* porque no depende de observaciones anteriores del proceso y predictiva porque es la distribución para una cantidad que es observable. Esta distribución refleja el conocimiento o visiones previas del experimentador acerca de la hipótesis antes que los datos se consideren. La probabilidad *a priori* es el término más subjetivo de la probabilidad Bayesiana.

Definición A.2. *Distribución posterior*

Después que los datos \mathbf{y} han sido observados, se puede predecir un $\tilde{\mathbf{y}}$ desconocido observable del mismo proceso. La distribución de $\tilde{\mathbf{y}}$ se denomina distribución predictiva posterior. Posterior porque es condicional al valor observado \mathbf{y} y predictiva porque es una predicción para un $\tilde{\mathbf{y}}$ observable. Esta refleja la validez de la hipótesis después de considerar los datos.

$$f(\tilde{\mathbf{y}}|\mathbf{y}) = \int f(\tilde{\mathbf{y}}, \boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} \quad (\text{IV.5})$$

Por definición de función densidad de probabilidad conjunta (II.9), (IV.5) se modifica a:

$$f(\tilde{\mathbf{y}}|\mathbf{y}) = \int f(\tilde{\mathbf{y}}|\boldsymbol{\theta}, \mathbf{y})f(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} \quad (\text{IV.6})$$

Debido al supuesto de que la y y \tilde{y} son condicionalmente independientes dada θ , (IV.6) se modifica a:

$$f(\tilde{\mathbf{y}}|\mathbf{y}) = \int f(\tilde{\mathbf{y}}|\boldsymbol{\theta})f(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} \quad (\text{IV.7})$$

Definición A.3. *Función de Verosimilitud*

El usar la regla de Bayes con un modelo de probabilidad particular significa que los datos \mathbf{y} afectan la inferencia posterior únicamente a través de $f(\mathbf{y}|\boldsymbol{\theta})$. Esta se puede considerar como una función de $\boldsymbol{\theta}$ cuando se fija \mathbf{y} y se denomina función de verosimilitud. Esta función evalúa la probabilidad de los datos observados dada la hipótesis. En este sentido, la Inferencia Bayesiana

obedece el principio de verosimilitud, el cual dice que para una muestra dada, cualesquiera dos modelos de probabilidad $f(\mathbf{y}|\boldsymbol{\theta})$ que tengan la misma función de verosimilitud producen la misma inferencia para $\boldsymbol{\theta}$.

Nota.

El enfoque de modelación Bayesiana se puede resumir en:

$$\text{posterior} \sim \text{verosimilitud} \times \text{priori} \quad (\text{IV.8})$$

B Supuestos del modelo

Es necesario establecer algunos supuestos para poder construir un modelo de Inferencia de Bayes de la forma:

$$P(E(\mathbf{r})|\mathbf{\Pi}) = \frac{P(\mathbf{\Pi}|E(\mathbf{r}))P(E(\mathbf{r}))}{P(\mathbf{\Pi})} \quad (\text{IV.9})$$

Supuesto IV.1.

Se supondrá que las visiones *a priori* $P(E(\mathbf{r}))$ serán k restricciones lineales sobre el vector (de dimensión n) de retornos esperados $E(\mathbf{r})$. Por tanto, estas se expresarán como una matriz $P(k \times n)$. De manera que:

$$PE(\mathbf{r}) = \mathbf{q} + \mathbf{v} \quad (\text{IV.10})$$

Supuesto IV.2.

El término \mathbf{v} en (IV.10) es un vector de error que representa la incertidumbre de las visiones del inversionista. Además, $\mathbf{v} \sim N(0, \Omega)$. Donde Ω es una matriz diagonal de covarianza, lo cual significa que las visiones del inversionista son independientes entre si. Por tanto,

$$PE(\mathbf{r}) \sim N(\mathbf{q}, \Omega) \quad (\text{IV.11})$$

Supuesto IV.3.

Se supondrá que los retornos en equilibrio condicionales a las visiones *a priori* del inversionista tienen una distribución normal de la siguiente forma:

$$\mathbf{\Pi}|E(\mathbf{r}) \sim N(E(\mathbf{r}), \tau\Sigma) \quad (\text{IV.12})$$

El hecho que $E[\mathbf{\Pi}] = E(\mathbf{r})$ refleja el supuesto del modelo CAPM de las visiones homogéneas de todos los inversionistas.

C Visiones certeras acerca de los retornos esperados

La certeza acerca de las visiones de los inversionistas corresponde a una desviación estándar de cero. Por tanto, las visiones se expresan como una relación exacta, la cual formará una restricción para el problema de optimización. En particular se tiene que

$$\min_{E(\mathbf{r})} (E(\mathbf{r}) - \mathbf{\Pi})^T \tau \Sigma (E(\mathbf{r}) - \mathbf{\Pi}) \quad (\text{IV.13})$$

Sujeto a la condición

$$PE(\mathbf{r}) = \mathbf{q} \quad (\text{IV.14})$$

Teorema C.1. *Proposición 1*

El predictor óptimo de $E(\mathbf{r})$ que minimiza su varianza al rededor de los retornos en equilibrio $\mathbf{\Pi}$ y satisface k restricciones lineales está dado por

$$\widehat{E(\mathbf{r})} = \mathbf{\Pi} + \Sigma^{-1} P^T (P \Sigma^{-1} P^T)^{-1} (\mathbf{q} - P \mathbf{\Pi}) \quad (\text{IV.15})$$

Demostración.

Este es un problema de optimización con restricciones lineales, por lo que se resolverá utilizando multiplicadores de Lagrange. Defínase la función que se desea maximizar $f(E(\mathbf{r}))$ y la condición de la siguiente manera:

$$f(E(\mathbf{r})) = (E(\mathbf{r}) - \mathbf{\Pi})^T \tau \Sigma (E(\mathbf{r}) - \mathbf{\Pi}) \quad (\text{IV.16})$$

$$c(E(\mathbf{r})) = PE(\mathbf{r}) - \mathbf{q} = 0 \quad (\text{IV.17})$$

Entonces utilizando multiplicadores de Lagrange se tiene que

$$\nabla f(E(\mathbf{r})) - \lambda \nabla c(E(\mathbf{r})) = 0 \quad (\text{IV.18})$$

Por tanto

$$\begin{aligned} \frac{\partial}{\partial E(\mathbf{r})} [(E(\mathbf{r}) - \mathbf{\Pi})^T \tau \Sigma (E(\mathbf{r}) - \mathbf{\Pi}) - \lambda (PE(\mathbf{r}) - \mathbf{q})] &= 0 \\ [E(\mathbf{r})^T \tau \Sigma E(\mathbf{r}) - E(\mathbf{r})^T \tau \Sigma \mathbf{\Pi} - \mathbf{\Pi}^T \tau \Sigma E(\mathbf{r}) + \mathbf{\Pi}^T \tau \Sigma \mathbf{\Pi} - \lambda PE(\mathbf{r}) + \lambda \mathbf{q}] &= 0 \\ \tau \Sigma^T E(\mathbf{r}) + \tau \Sigma E(\mathbf{r}) - 2\tau \Sigma \mathbf{\Pi} - P^T \lambda &= 0 \end{aligned} \quad (\text{IV.19})$$

Despejando $E(\mathbf{r})$ en (IV.19) se obtiene:

$$E(\mathbf{r}) = \mathbf{\Pi} + \frac{\lambda}{2\tau} \Sigma^{-1} P^T \quad (\text{IV.20})$$

Ahora, sustituyendo (IV.20) en (IV.17) se obtiene:

$$\begin{aligned}
PE(\mathbf{r}) - \mathbf{q} &= 0 \\
P\mathbf{\Pi} + \frac{\lambda}{2\tau}P\Sigma^{-1}P^T - \mathbf{q} &= 0 \\
\lambda P\Sigma^{-1}P^T &= 2\tau(\mathbf{q} - P\mathbf{\Pi}) \\
\Rightarrow \lambda &= (P\Sigma^{-1}P^T)^{-1}2\tau(\mathbf{q} - P\mathbf{\Pi})
\end{aligned} \tag{IV.21}$$

Sustituyendo (IV.21) en (IV.20) se obtiene el valor óptimo de $E(\mathbf{r})$:

$$\begin{aligned}
E(\mathbf{r}) &= \mathbf{\Pi} + \frac{(P\Sigma^{-1}P^T)^{-1}2\tau(\mathbf{q} - P\mathbf{\Pi})\Sigma^{-1}P^T}{2\tau} \\
E(\mathbf{r}) &= \mathbf{\Pi} + \Sigma^{-1}P^T(P\Sigma^{-1}P^T)^{-1}(\mathbf{q} - P\mathbf{\Pi})
\end{aligned} \tag{IV.22}$$

Corolario C.1.1.

Si no existen visiones previas por parte del inversionista, entonces $E(\mathbf{r})$ converge al retorno dado por los datos de equilibrio $\mathbf{\Pi}$.

Demostración.

Haciendo $P = 0$ en (IV.22) obtenemos el resultado deseado.

D Visiones inciertas acerca de los retornos esperados

Cuando el inversionista forma sus propias visiones con un grado de incertidumbre, esto se refleja en los valores distintos a cero de los elementos de la diagonal de la matriz Ω .

Teorema D.1. *Proposición 2*

La función de densidad de probabilidad posterior $f(\mathbf{\Pi}|E(\mathbf{r}))$ es multivariada normal con media:

$$\boldsymbol{\mu} = [(\tau\Sigma)^{-1} + P^T\Omega^{-1}P]^{-1} [(\tau\Sigma)^{-1}\mathbf{\Pi} + P^T\Omega^{-1}\mathbf{q}] \tag{IV.23}$$

y varianza:

$$\boldsymbol{\sigma}^2 = [(\tau\Sigma)^{-1} + P^T\Omega^{-1}P]^{-1} \tag{IV.24}$$

Demostración.

Por el supuesto (IV.11) se sabe que $PE(\mathbf{r})$ posee una distribución normal con media \mathbf{q} y varianza Ω . Por tanto, su función de densidad de probabilidad multivariada está dada por:

$$f(PE(\mathbf{r})) = \frac{1}{\sqrt{2\pi|\Omega|}} e^{-\frac{1}{2}(PE(\mathbf{r})-\mathbf{q})^T\Omega^{-1}(PE(\mathbf{r})-\mathbf{q})} \tag{IV.25}$$

Por el supuesto (IV.12) $\mathbf{\Pi}|E(\mathbf{r})$ tiene una distribución normal con media $E(\mathbf{r})$ y varianza $\tau\Sigma$. Por tanto, su función de densidad de probabilidad multivariada está dada por:

$$f(\mathbf{\Pi}|E(\mathbf{r}), \tau\Sigma) = \frac{1}{\sqrt{2\pi|\tau\Sigma|}} e^{-\frac{1}{2}(\mathbf{\Pi}-E(\mathbf{r}))^T(\tau\Sigma)^{-1}(\mathbf{\Pi}-E(\mathbf{r}))} \quad (\text{IV.26})$$

Notando que $P(\mathbf{\Pi})$ es una constante y sustituyendo (IV.25), (IV.26) en (IV.9) se obtiene:

$$f(E(\mathbf{r})|\mathbf{\Pi}) = \frac{1}{P(\mathbf{\Pi})} \exp\left(-\frac{1}{2}(\mathbf{\Pi}-E(\mathbf{r}))^T(\tau\Sigma)^{-1}(\mathbf{\Pi}-E(\mathbf{r})) - \frac{1}{2}(PE(\mathbf{r})-\mathbf{q})^T\Omega^{-1}(PE(\mathbf{r})-\mathbf{q})\right) \quad (\text{IV.27})$$

Reescribiendo la ecuación anterior se obtiene:

$$f(E(\mathbf{r})|\mathbf{\Pi}) = \frac{1}{P(\mathbf{\Pi})} \exp\left(-\frac{1}{2}\mathbf{\Pi}^T(\tau\Sigma)^{-1}\mathbf{\Pi} + \mathbf{\Pi}(\tau\Sigma)^{-1}E(\mathbf{r}) - \frac{1}{2}E(\mathbf{r})^T(\tau\Sigma)^{-1}E(\mathbf{r}) - \frac{1}{2}E(\mathbf{r})^T P^T \Omega^{-1} P E(\mathbf{r}) + P^T \Omega^{-1} \mathbf{q} E(\mathbf{r}) - \frac{1}{2} \mathbf{q}^T \Omega^{-1} \mathbf{q}\right) \quad (\text{IV.28})$$

Ahora haciendo

$$\begin{aligned} A &= (\tau\Sigma)^{-1} + P^T \Omega^{-1} P \\ B &= (\tau\Sigma)^{-1} \mathbf{\Pi} + P^T \Omega^{-1} \mathbf{q} \\ C &= \mathbf{\Pi}^T (\tau\Sigma)^{-1} + \mathbf{q}^T \Omega^{-1} \mathbf{q} \end{aligned} \quad (\text{IV.29})$$

Sustituyendo (IV.29) en (IV.28) se obtiene:

$$f(E(\mathbf{r})|\mathbf{\Pi}) = \frac{1}{P(\mathbf{\Pi})} \exp\left(-\frac{1}{2}[E(\mathbf{r})^T A E(\mathbf{r}) - 2B^T E(\mathbf{r}) + C]\right) \quad (\text{IV.30})$$

Ahora, modificando algebraicamente (IV.30) se obtiene:

$$\begin{aligned} f(E(\mathbf{r})|\mathbf{\Pi}) &= \frac{1}{P(\mathbf{\Pi})} \exp\left(-\frac{1}{2}[E(\mathbf{r})^T A^T A A^{-1} E(\mathbf{r}) - 2B^T A^{-1} A E(\mathbf{r}) + C]\right) \\ &= \frac{1}{P(\mathbf{\Pi})} \exp\left(-\frac{1}{2}[E(\mathbf{r})^T A^T A A^{-1} E(\mathbf{r}) - 2B^T A^{-1} A E(\mathbf{r}) + B^T A^{-1} B - B^T A^{-1} B + C]\right) \\ &= \frac{1}{P(\mathbf{\Pi})} \exp\left(-\frac{1}{2}[(A E(\mathbf{r}) - B)^T A^{-1} (A E(\mathbf{r}) - B) - B^T A^{-1} B + C]\right) \\ &= \frac{1}{P(\mathbf{\Pi})} \exp\left(-\frac{1}{2}[C - B^T A^{-1} B]\right) \exp\left(-\frac{1}{2}[(A E(\mathbf{r}) - B)^T A^{-1} (A E(\mathbf{r}) - B)]\right) \end{aligned} \quad (\text{IV.31})$$

Sea $D = \frac{1}{P(\mathbf{\Pi})} \exp\left(-\frac{1}{2}[C - B^T A^{-1} B]\right)$, entonces la ecuación (IV.31) se modifica a:

$$f(E(\mathbf{r})|\mathbf{\Pi}) = D e\left(-\frac{1}{2}[(A E(\mathbf{r}) - B)^T A^{-1} (A E(\mathbf{r}) - B)]\right) \quad (\text{IV.32})$$

Manipulando (IV.32) algebraicamente se obtiene:

$$\begin{aligned}
f(E(\mathbf{r})|\mathbf{\Pi}) &= De^{(-\frac{1}{2}[(AA^{-1})(AE(\mathbf{r})-B)^T A^{-1}(AA^{-1})(AE(\mathbf{r})-B)])} \\
&= De^{(-\frac{1}{2}[(A^{-1}AE(\mathbf{r})-A^{-1}B)^T A(A^{-1}A)(A^{-1}AE(\mathbf{r})-A^{-1}B)])} \\
&= De^{(-\frac{1}{2}[(E(\mathbf{r})-A^{-1}B)^T A(E(\mathbf{r})-A^{-1}B)])} \\
&= De^{(-\frac{1}{2}[(E(\mathbf{r})-A^{-1}B)^T (A^{-1})^{-1}(E(\mathbf{r})-A^{-1}B)])}
\end{aligned} \tag{IV.33}$$

De (IV.33) se obtienen la media y varianza deseadas:

$$\boldsymbol{\mu} = A^{-1}B = [(\tau\Sigma)^{-1}P^T\Omega^{-1}P]^{-1}[(\tau\Sigma)^{-1}\mathbf{\Pi} + P^T\Omega^{-1}\mathbf{q}] \tag{IV.34}$$

$$\boldsymbol{\sigma}^2 = A^{-1} = [(\tau\Sigma)^{-1}P^T\Omega^{-1}P]^{-1} \tag{IV.35}$$

Nota.

La expresión obtenida para la media $\boldsymbol{\mu}$ (IV.33) se puede reescribir como (ver Apéndice):

$$\boldsymbol{\mu} = \mathbf{\Pi} + (\tau\Sigma)P^T[P(\tau\Sigma)P^T + \Omega]^{-1}(\mathbf{q} - P\mathbf{\Pi}) \tag{IV.36}$$

Corolario D.1.1.

A medida que la incertidumbre sobre las visiones del inversionista reducen (el inversionista se siente más confiado con sus visiones), el retorno esperado se acerca al caso determinístico, en el que

$$E[\mathbf{r}] = \mathbf{\Pi} + (\tau\Sigma)P^T(P\Sigma^{-1}P^T)^{-1}(\mathbf{q} - P\mathbf{\Pi}) \tag{IV.37}$$

Demostración.

Si la incertidumbre en las visiones disminuyen, $\Omega \rightarrow 0$, por lo que la fórmula alternativa de Black-Litterman (IV.36) se modifica a:

$$E[\mathbf{r}] = \mathbf{\Pi} + (\tau\Sigma)P^T(P(\tau\Sigma)P^T)^{-1}(\mathbf{q} - P\mathbf{\Pi}) \tag{IV.38}$$

Nota. Análisis de la proposición 2

Se probó que $E(\mathbf{r})$ tiene la siguiente media posterior:

$$\boldsymbol{\mu} = [(\tau\Sigma)^{-1} + P^T\Omega^{-1}P]^{-1} [(\tau\Sigma)^{-1}\mathbf{\Pi} + P^T\Omega^{-1}\mathbf{q}] \tag{IV.39}$$

Reescribiendo (IV.39) se obtiene:

$$\begin{aligned}\boldsymbol{\mu} &= [(\tau\Sigma)^{-1} + P^T\Omega^{-1}P]^{-1} [(\tau\Sigma)^{-1}\boldsymbol{\Pi} + P^T\Omega^{-1}(PP^{-1})((P^T)^{-1}P^T)\mathbf{q}] \\ &= [(\tau\Sigma)^{-1} + P^T\Omega^{-1}P]^{-1} [(\tau\Sigma)^{-1}\boldsymbol{\Pi} + (P^T\Omega^{-1}P)(P^T P)^{-1}P^T\mathbf{q}]\end{aligned}\quad (\text{IV.40})$$

Además, se sabe que:

$$\begin{aligned}PE(\mathbf{r}) &= \mathbf{q} + \mathbf{v} \\ \mathbf{q} &= PE(\mathbf{r}) - \mathbf{v}\end{aligned}\quad (\text{IV.41})$$

Por tanto, (IV.41) se puede considerar como una regresión de \mathbf{q} respecto a P , donde $E(\mathbf{r})$ es el vector de coeficientes que deben ser estimados. Utilizando el método de mínimos cuadrados la estimación de los retornos esperados según las visiones de los inversionistas es:

$$\widehat{E(\mathbf{r})} = (P^T P)^{-1} P^T \mathbf{q} \quad (\text{IV.42})$$

Por tanto, la ecuación (IV.40) se puede reescribir como:

$$\boldsymbol{\mu} = [(\tau\Sigma)^{-1} + P^T\Omega^{-1}P]^{-1} [(\tau\Sigma)^{-1}\boldsymbol{\Pi} + P^T\Omega^{-1}(PP^{-1})((P^T)^{-1}P^T)\widehat{E(\mathbf{r})}] \quad (\text{IV.43})$$

En la ecuación (IV.43) se puede notar como se combinan las visiones subjetivas del inversionista con los datos en equilibrio. El término en el segundo paréntesis es un promedio ponderado de los datos en equilibrio $\boldsymbol{\Pi}$ y del estimado de mínimos cuadrados $\widehat{E(\mathbf{r})}$. Los pesos (vectores de peso) son $(\tau\Sigma)^{-1}$ y $(P^T\Omega^{-1}P)$ respectivamente.

Si la distribución de los retornos esperados es cercana a los datos de equilibrio $\boldsymbol{\Pi}$, entonces $(\tau\Sigma)$ será pequeño y $(\tau\Sigma)^{-1}$ será grande y se pondrá más peso a $\boldsymbol{\Pi}$. Si el inversionista está seguro acerca de sus visiones, entonces Ω será pequeño, lo que resultará en un $P^T\Omega^{-1}P$ grande, el cual se pondrá más peso en las visiones de mínimos cuadrados $\widehat{E(\mathbf{r})}$.

E Observaciones del enfoque

1. El desarrollo del modelo de Black Litterman utilizando Inferencia Bayesiana es el más estudiado, ya que es muy intuitivo y accesible. Por tanto, existe mucha literatura al respecto.
2. Además de ser muy intuitivo, este enfoque permite una fácil implementación. Se pueden programar algoritmos numéricos rápidos y eficientes para obtener los retornos esperados y pesos recomendados.
3. En general la Inferencia de Bayes es bastante aplicable para el análisis de portafolios de inversión similares al modelo de Black Litterman. Primero porque puede emplear información *a priori* útil sobre cantidades de interés (por ejemplo eventos macro-económicos, teorías de

valuación de activos y otras cantidades relevantes para la dinámica de los retornos esperados). Segundo porque puede estimar riesgo e incertidumbre del modelo. Por último, porque se presta a el uso algoritmos numéricos rápidos e intuitivos para simular cantidades económicas complejas.

4. Para el lector interesado en otros modelos de optimización de portafolios que pueden ser deducidos a partir de Inferencia Bayesiana, además de Black Litterman, puede consultar (Avramov y Zhou, 2009).

V Enfoque de Optimización Inversa

El objetivo de esta sección es caracterizar el método de Black Litterman como solución de un problema particular de optimización convexa (B). De esta manera se elimina la necesidad de un modelo estadístico y varias de las limitaciones se solucionan. La teoría de optimización convexa se desarrollará en base a Boyd y Vandenberghe (2004).

Suponga que se tiene el siguiente problema de optimización:

$$\min_{\mathbf{x} \in \mathcal{D}} f(\mathbf{x}, \mathcal{P}) \tag{V.1}$$

Donde:

- $f(\mathbf{x}, \mathcal{P})$ es la función objetivo
- $\mathbf{x} \in R^n$ son las variables de decisión
- $\mathcal{P} \in R^m$ son los datos
- \mathcal{D} es la región factible

La optimización convexa es de caracter inverso. Por ejemplo, en el problema de Markowitz, los datos \mathcal{P} corresponden a $(\boldsymbol{\mu}, \Sigma, r_f, L)$ y el portafolio óptimo será \mathbf{x}^* . En la optimización directa se busca encontrar una solución óptima \mathbf{x}^* en base a los datos $(\boldsymbol{\mu}, \Sigma, r_f, L)$. En la optimización inversa se inicia con un portafolio candidato \mathbf{x}^* y se buscan los datos $(\boldsymbol{\mu}, \Sigma, r_f, L)$ que lo hagan óptimo.

A Problemas de optimización convexa

Definición A.1. *Principio de Dualidad en Optimización*

En optimización matemática, el principio de dualidad establece que los problemas de optimización se pueden ver desde dos perspectivas, el problema primal y el dual. La solución del dual proporciona una cota inferior para la solución del problema primal. En general, los valores óptimos de los problemas primal y dual no son iguales. Su diferencia se denomina "brecha dual". Para los problemas de optimización convexa, la brecha dual es cero bajo las condiciones de regularidad Karush-Kuhn-Tucker.

Definición A.2. *Función Lagrangiana*

Para un problema de optimización en la forma estándar:

$$\begin{aligned} & \text{minimizar } f_0(\mathbf{x}) \\ & \text{sujeto a } f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m \\ & \quad \quad \quad h_i(\mathbf{x}) = 0, \quad i = 1, \dots, m \end{aligned} \tag{V.2}$$

Se supone que el dominio de este problema $\mathcal{D} = [\cap_{i=0}^m \text{dom}(f_i)] \cap [\cap_{i=1}^p \text{dom}(h_i)]$ es no vacío. Denotamos el valor óptimo de (V.2) por \mathbf{p}^* . No se asume que el problema (V.2) sea convexo. Lo que se busca con una función Lagrangiana es tomar las restricciones de (V.2) en consideración al aumentar la función objetivo como una suma ponderada de las funciones restricción. La función Lagrangiana asociada a (V.2) $L : R^n \times R^m \times R^p \rightarrow R$ se define como:

$$L(\mathbf{x}, \lambda, v) = f_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^p v_i h_i(\mathbf{x}) \tag{V.3}$$

Donde

- el dominio de L es $\text{dom}(L) = \mathcal{D} \times R^m \times R^p$.
- v_i se denomina el multiplicador de Lagrange asociado a la i -ésima restricción de igualdad $h_i(\mathbf{x}) = 0$
- λ_i se denomina el multiplicador de Lagrange asociado a la i -ésima restricción de desigualdad: $f_i(\mathbf{x}) \leq 0$
- Los vectores λ, v se denominan las variables duales o los vectores multiplicadores de Lagrange asociados a (V.2)

Definición A.3. *Función de Lagrange dual*

Esta es una función $g : R^m \times R^p \rightarrow R$ que se define como el valor mínimo de la función Lagrangiana sobre \mathbf{x} para $\lambda \in R^m, v \in R^p$

$$g(\lambda, v) = \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \lambda, v) = \inf_{\mathbf{x} \in \mathcal{D}} \left(f_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^p v_i h_i(\mathbf{x}) \right) \tag{V.4}$$

Como la función dual es el ínfimo punto a punto de una familia de función afines de (λ, v) es cóncava, aún y cuando el problema (V.2) no sea convexo.

Definición A.4. *Función conjugada*

Sea $f : R^n \rightarrow R$ una función, entonces la función conjugada de f se denota por f^* y se define como:

$$f^*(\mathbf{y}) = \sup_{\mathbf{x}} (\mathbf{y}^T \mathbf{x} - f(\mathbf{x})), \quad \mathbf{x} \in \text{dom}(f) \quad (\text{V.5})$$

Teorema A.1. *Cotas inferiores en valores óptimos*

La función dual proporciona cotas inferiores para el valor óptimo p^* del problema (V.2). Para cualquier $\lambda \succeq 0$ y v se tiene que:

$$g(\lambda, v) \leq p^* \quad (\text{V.6})$$

Demostración.

Suponga que $\bar{\mathbf{x}}$ es un punto factible para el problema (V.2), entonces por las restricciones del problema se tiene que $f_i(\bar{\mathbf{x}}) \leq 0$ y $h_i(\bar{\mathbf{x}}) = 0$ y $\lambda \succeq 0$. Sumando estas condiciones se obtiene:

$$\sum_{i=1}^m \lambda_i f_i(\bar{\mathbf{x}}) + \sum_{i=1}^p v_i h_i(\bar{\mathbf{x}}) \leq 0 \quad (\text{V.7})$$

Sumando $f_0(\bar{\mathbf{x}})$ de ambos lados de (V.7) se obtiene:

$$L(\bar{\mathbf{x}}) = f_0(\bar{\mathbf{x}}) + \sum_{i=1}^m \lambda_i f_i(\bar{\mathbf{x}}) + \sum_{i=1}^p v_i h_i(\bar{\mathbf{x}}) \leq f_0(\bar{\mathbf{x}}) \quad (\text{V.8})$$

Al aplicar ínfimo a (V.8) se obtiene:

$$\begin{aligned} g(\lambda, v) &= \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \lambda, v) \leq L(\bar{\mathbf{x}}, \lambda, v) \leq f_0(\bar{\mathbf{x}}) \\ &g(\lambda, v) \leq f_0(\bar{\mathbf{x}}) \end{aligned} \quad (\text{V.9})$$

La función dual da una cota inferior no trivial en p^* solo cuando $\lambda \succeq 0$ y $(\lambda, v) \in \text{dom}(g)$.

Definición A.5. *El problema dual de Lagrange*

Para cada par (λ, v) con $\lambda \succeq 0$ la función dual de Lagrange da una cota inferior para el valor óptimo p^* del problema de optimización (V.2). Ahora se quiere obtener la mayor cota inferior (ínfimo). Por tanto, se obtiene el siguiente problema de optimización:

$$\begin{aligned} &\text{maximizar } g(\lambda, v) \\ &\text{sujeto a } \lambda \succeq 0 \end{aligned} \quad (\text{V.10})$$

Al problema de optimización (V.10) se le denomina a problema dual de Lagrange asociado al problema primal (V.2). El par (λ, v) se denomina dual factible, si es factible para el problema

dual (V.10). Si el par (λ^*, v^*) es el punto óptimo del problema de optimización (V.10), entonces se denomina óptimo dual. El problema dual de Lagrange (V.10) es un problema de optimización convexa, ya que la función objetivo a maximizar es cóncava y la restricción es convexa. Esto no depende de si el problema primal (V.2) es convexo o no.

Definición A.6. *Dualidad débil*

El valor óptimo del problema dual de Lagrange se denota por d^* y se define como la mayor cota inferior de p^* que se puede obtener de la función dual de Lagrange. Por tanto, se tiene la propiedad de dualidad débil, la cual se cumple aún y cuando el problema no sea convexo:

$$d^* \leq p^* \tag{V.11}$$

Definición A.7. *Brecha óptima de dualidad*

La brecha óptima dual es la diferencia entre el valor óptimo del problema primal y la mayor cota inferior que se obtiene de la función dual de Lagrange. Esta es siempre no negativa.

$$p^* - d^* \tag{V.12}$$

Definición A.8. *Dualidad fuerte y cualificación de restricción de Slater*

Si se cumple la igualdad en (V.11), la brecha óptima dual es cero y se cumple la dualidad fuerte. La dualidad fuerte en general no se cumple. Si el problema primal (V.2) es convexo, usualmente se tiene dualidad fuerte. Hay resultados que establecen condiciones, bajo las cuales se cumple la dualidad fuerte. Estas condiciones se llaman condiciones de regularidad (o cualificación de las restricciones). Una de ellas es la condición de Slater.

Teorema A.2. *Condición de Slater*

La condición de Slater dice que existe un $x \in \text{relint } D$ ¹ tal que

$$f_i(\mathbf{x}) < 0, \quad i = 1, \dots, m \quad A\mathbf{x} = b \tag{V.13}$$

Este punto se denomina estrictamente factible.

Teorema A.3. *Teorema de Slater*

El teorema de Slater dice que si el problema es convexo y se cumple la condición de Slater, se cumple la dualidad fuerte.

¹*relint(D)* se define como el interior de D dentro del conjunto de todas las combinaciones afines de D

Demostración.

Para simplificar la prueba, se agregaran algunos supuestos adicionales. Primero que \mathcal{D} tiene interior no vacío (en otras palabras que $\text{relint}(\mathcal{D}) = \text{int}(\mathcal{D})$). Segundo, que $\text{rank}(A) = p$. Por último, que p^* es finito. Ahora, se definen los conjuntos \mathcal{A}, \mathcal{B} de la siguiente manera:

$$\begin{aligned} f_i(\mathbf{x}) &\leq u_i, \quad i = 1, \dots, m \\ \mathcal{A} &= \{(u, v, t) | \exists \mathbf{x} \in \mathcal{D}, \quad h_i(\mathbf{x}) = v_i, \quad i = 1, \dots, p \} \\ f_0(\mathbf{x}) &\leq t \end{aligned} \quad (\text{V.14})$$

$$B = \{(0, 0, s) \in R^m \times R^p \times R | s < p^*\} \quad (\text{V.15})$$

Primero se probará que los conjuntos \mathcal{A}, \mathcal{B} son disjuntos. Para esto suponga que no son disjuntos, entonces existe $(\mathbf{u}, \mathbf{v}, t) \in \mathcal{A} \cap \mathcal{B}$. Como $(\mathbf{u}, \mathbf{v}, t) \in \mathcal{B} \Rightarrow \mathbf{u} = \mathbf{0}, \mathbf{v} = \mathbf{0} \ \& \ t < p^*$. Además, como $(\mathbf{u}, \mathbf{v}, t) \in \mathcal{A} \Rightarrow \exists \mathbf{x} \in \mathcal{D}$ tal que $f_i(\mathbf{x}) \leq 0, A\mathbf{x} - \mathbf{b} = 0$ y $f_0(\mathbf{x}) \leq t$. Pero como $t < p^*$ esto implica que $f_0(\mathbf{x}) < p^*$ ($\rightarrow \leftarrow$) ya que p^* es el valor óptimo del problema primario $\Rightarrow \therefore \mathcal{A} \cap \mathcal{B} = \emptyset$.

Ahora, por el teorema del hiperplano de separación (B.15) $\exists(\bar{\boldsymbol{\lambda}}, \bar{\mathbf{v}}, \mu) \neq \mathbf{0}$ y $\alpha \in R$ tal que:

$$(\mathbf{u}, \mathbf{v}, t) \in \mathcal{A} \Rightarrow \bar{\boldsymbol{\lambda}}^T \mathbf{u} + \bar{\mathbf{v}}^T \mathbf{v} + \mu t \geq \alpha \quad (\text{V.16})$$

$$(\mathbf{u}, \mathbf{v}, t) \in \mathcal{B} \Rightarrow \bar{\boldsymbol{\lambda}}^T \mathbf{u} + \bar{\mathbf{v}}^T \mathbf{v} + \mu t \leq \alpha \quad (\text{V.17})$$

De (V.16) se concluye que $\bar{\boldsymbol{\lambda}} \succeq 0$ y que $\mu \geq 0$. Por otro lado, de (V.17) se concluye que $\mu t \leq \alpha \ \forall t < p^*$. Por tanto, $\mu p^* \leq \alpha$. Ahora, por (V.16):

$$\sum_{i=1}^m \bar{\lambda}_i f_i(\mathbf{x}) + \bar{\mathbf{v}}^T (A\mathbf{x} - \mathbf{b}) + \mu f_0(\mathbf{x}) \geq \alpha \geq \mu p^* \quad (\text{V.18})$$

Suponiendo que $\mu > 0$ (V.18) se modifica a:

$$\sum_{i=1}^m \frac{\bar{\lambda}_i}{\mu} f_i(\mathbf{x}) + \frac{\bar{\mathbf{v}}^T}{\mu} (A\mathbf{x} - \mathbf{b}) + f_0(\mathbf{x}) = L\left(\mathbf{x}, \frac{\bar{\boldsymbol{\lambda}}}{\mu}, \frac{\bar{\mathbf{v}}}{\mu}\right) \geq p^* \quad (\text{V.19})$$

Haciendo las sustituciones $\boldsymbol{\lambda} = \frac{\bar{\boldsymbol{\lambda}}}{\mu}, \mathbf{v} = \frac{\bar{\mathbf{v}}}{\mu}$ y al minimizar sobre \mathbf{x} se obtiene que $g(\boldsymbol{\lambda}, \mathbf{v}) \geq p^*$. Además, por dualidad débil se sabe que $g(\boldsymbol{\lambda}, \mathbf{v}) \leq p^*$. Por tanto, $g(\boldsymbol{\lambda}, \mathbf{v}) = p^*$, por lo que se cumple la dualidad fuerte y p^* es valor mínimo del el mínimo del problema de optimización (V.2).

Ahora si se considera el caso en el que $\mu = 0$, entonces $\forall \mathbf{x} \in \mathcal{D}$ la ecuación (V.18) se modifica a:

$$\sum_{i=1}^m \bar{\lambda}_i f_i(\mathbf{x}) + \bar{\mathbf{v}}^T (A\mathbf{x} - \mathbf{b}) \geq 0 \quad (\text{V.20})$$

Aplicando (V.20) al punto $\bar{\mathbf{x}}$ que satisface la condición de Slater ($f_i(\bar{\mathbf{x}}) < 0$ y $A\bar{\mathbf{x}} - \mathbf{b} = 0$) se obtiene:

$$\bar{\lambda}_i f_i(\bar{\mathbf{x}}) \geq 0 \quad (\text{V.21})$$

Por la condición de Slater se sabe que $f_i(\bar{\mathbf{x}}) < 0$ y pero $\bar{\lambda}_i \geq 0$. Por tanto se concluye que $\bar{\lambda} = \mathbf{0}$. Como $(\bar{\lambda}, \bar{\mathbf{v}}, \mu) \neq \mathbf{0}$, $\bar{\lambda} = 0$ y $\mu = 0$, se concluye que $\bar{\mathbf{v}} \neq 0$. Entonces la ecuación (V.20) implica que para todo $\mathbf{x} \in \mathcal{D}$, $\bar{\mathbf{v}}^T(A\mathbf{x} - \mathbf{b}) \geq 0$. Pero $\bar{\mathbf{x}}$ satisface $\bar{\mathbf{v}}^T(A\bar{\mathbf{x}} - \mathbf{b}) = 0$ y como $\bar{\mathbf{x}} \in \text{int}(\mathcal{D})$ existen puntos de \mathcal{D} con $\bar{\mathbf{v}}^T(A\mathbf{x} - \mathbf{b}) < 0$ a menos que $A^T\bar{\mathbf{v}} = 0$. Pero esto contradice el supuesto que $\text{rank}(A) = p$. Por tanto $\mu > 0$.

Definición A.9. *Condiciones KKT para un problema no convexo*

Sea \mathbf{x}^* un punto óptimo primal y (λ^*, v^*) un punto dual óptimo con desface de dualidad cero. Como \mathbf{x}^* minimiza $L(\mathbf{x}, \lambda^*, v^*)$ sobre x , su gradiente debe ser cero en \mathbf{x}^* :

$$\nabla f_0(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(\mathbf{x}^*) + \sum_{i=1}^p v_i^* \nabla h_i(\mathbf{x}^*) = 0 \quad (\text{V.22})$$

Por tanto, se obtienen las condiciones de Karush-Kuhn-Tucker (KKT):

$$\begin{aligned} f_i(\mathbf{x}^*) &\leq 0, \quad i = 1, \dots, m \\ h_i(\mathbf{x}^*) &= 0, \quad i = 1, \dots, m \\ \lambda_i^* &\geq 0, \quad i = 1, \dots, m \\ \lambda_i^* f_i(\mathbf{x}^*) &= 0, \quad i = 1, \dots, m \\ \nabla f_0(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(\mathbf{x}^*) + \sum_{i=1}^p v_i^* \nabla h_i(\mathbf{x}^*) &= 0 \end{aligned} \quad (\text{V.23})$$

Para cualquier problema de optimización con función objetivo diferenciable y funciones restricción diferenciables para las cuales se cumple la dualidad fuerte, los puntos óptimos primal y dual satisfacen las condiciones KKT.

Definición A.10. *Condiciones KKT para un problema convexo*

Cuando el problema principal es convexo, las condiciones KKT son suficientes para que los puntos sean primal y dual óptimos. En otras palabras, si f_i son convexas y las h_i son afín, y si los

puntos $\bar{\mathbf{x}}, \bar{\lambda}, \bar{v}$ satisfacen las condiciones KKT:

$$\begin{aligned}
f_i(\bar{\mathbf{x}}) &\leq 0, \quad i = 1, \dots, m \\
h_i(\bar{\mathbf{x}}) &= 0, \quad i = 1, \dots, m \\
\bar{\lambda}_i &\geq 0, \quad i = 1, \dots, m \\
\bar{\lambda}_i f_i(\bar{\mathbf{x}}) &= 0, \quad i = 1, \dots, m \\
\nabla f_0(\bar{\mathbf{x}}) + \sum_{i=1}^m \bar{\lambda} \nabla f_i(\bar{\mathbf{x}}) + \sum_{i=1}^p \bar{v}_i \nabla h_i(\bar{\mathbf{x}}) &= 0
\end{aligned} \tag{V.24}$$

Entonces $\bar{\mathbf{x}}$ y $(\bar{\lambda}, \bar{v})$ son los óptimos primal y dual con un desfase de dualidad de cero.

Nota. Condiciones KKT y condición de Slater

Si se satisface la condición de Slater, \mathbf{x} es óptimo $\Leftrightarrow \exists \lambda, v$ que satisfagan las condiciones KKT.

B Teoremas de Black Litterman

Teorema B.1. *Problema de Markowitz*

Suponga que el mercado consiste de n activos riesgosos y un activo libre de riesgo. Además, suponga que todos los inversionistas buscan maximizar el retorno esperado sujeto a un nivel límite de riesgo L . Entonces inversionistas resuelven el problema de Markowitz.²:

$$\begin{aligned}
\text{maximizar : } & \boldsymbol{\mu}^T \mathbf{x} + (1 - \mathbf{e}^T \mathbf{x}) r_f \\
\text{sujeto a : } & \sqrt{\mathbf{x}^T \Sigma \mathbf{x}} \leq L
\end{aligned} \tag{V.25}$$

Si $\mathbf{r} \in R^n$ es el vector de retornos de los activos riesgosos, $\boldsymbol{\mu} = E[\mathbf{r}]$

Definición B.1. *Equilibrio de Mercado*

Si todos los inversionistas resuelven el problema (V.25) para algún nivel de riesgo específico L , entonces existe un valor $\delta \in R_+$ tal que los retornos esperados $\boldsymbol{\mu}$ satisfacen:

$$\boldsymbol{\mu} = r_f \mathbf{e} + 2\delta \Sigma \mathbf{x}_m \tag{V.26}$$

Donde \mathbf{x}_m es el porcentaje de capitalización bursátil de cada activo. Además, note que:

$$\delta = \frac{\sqrt{(\boldsymbol{\mu} - r_f \mathbf{e})^T \Sigma^{-1} (\boldsymbol{\mu} - r_f \mathbf{e})}}{2\mathbf{x}_m^T \Sigma \mathbf{x}_m} \tag{V.27}$$

²Se utiliza la misma notación presentada anteriormente en el capítulo 2, agregando que L es el nivel (o límite) de riesgo aceptado por el inversionista y δ es el coeficiente de compensación de riesgo para el portafolio del mercado.

Lo cual significa que δ es el cociente del retorno y la varianza. Por lo que δ es una medida de compensación de riesgo para el portafolio del mercado. Además, existe un valor de L tal que \mathbf{x}_m es una solución para:

$$\delta = \frac{\sqrt{(\boldsymbol{\mu} - r_f \mathbf{e})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - r_f \mathbf{e})}}{2L} \quad (\text{V.28})$$

Estas condiciones forman el "Capital Asset Pricing Model Equilibrium".

Definición B.2. *Información Privada*

Como se mencionó en el capítulo anterior, los inversionistas pueden tener información privada o visiones diferentes a las de los demás inversionistas. Estas visiones se expresan en diferentes portafolios $\mathbf{p}_1, \dots, \mathbf{p}_n$ (columnas de la matriz P) para los cuales los inversionistas tienen una estimación de sus retornos esperados q_1, \dots, q_m .

$$P\boldsymbol{\mu} = \mathbf{q} \quad (\text{V.29})$$

Suponiendo que las fuentes de información son confiables el modelo de Black-Litterman sugiere el siguiente modelo estadístico:

$$\hat{\mathbf{r}} = r_f \mathbf{e} + 2\hat{\delta}\hat{\boldsymbol{\Sigma}}\mathbf{x}_m \quad (\text{V.30})$$

$$\boldsymbol{\mu} = \hat{\mathbf{r}} + \mathbf{e}^r \quad (\text{V.31})$$

Donde $\hat{\delta}$, $\hat{\boldsymbol{\Sigma}}$ son estimaciones de δ , $\boldsymbol{\Sigma}$ y $(\mathbf{e}^r, \mathbf{e}^v)$ es un vector aleatorio de ruido. Se supone que este vector tiene una distribución normal multivariada $N(\mathbf{0}, \boldsymbol{\Omega})$. Donde $\boldsymbol{\Omega}$ es una matriz de covarianza.

Teorema B.2.

Considere el problema de Markowitz (V.25) Suponga que se tiene un portafolio candidato \mathbf{x}^* y que a priori se sabe que $\boldsymbol{\Sigma} \succeq 0$ y $L > 0$. Entonces $(\boldsymbol{\mu}, \boldsymbol{\Sigma}, r_f, L)$ resuelven el problema inverso de (V.25) si y solo si existe un δ tal que:

$$\begin{aligned} \mathbf{x}^{*T} \boldsymbol{\Sigma} \mathbf{x}^* &\leq L^2 \\ \delta(L^2 - \mathbf{x}^{*T} \boldsymbol{\Sigma} \mathbf{x}^*) &= 0 \\ \boldsymbol{\mu} - r_f \mathbf{e} - 2\delta \boldsymbol{\Sigma} \mathbf{x}^* &= 0 \\ \delta \geq 0, L > 0, \boldsymbol{\Sigma} \succeq 0 \end{aligned} \quad (\text{V.32})$$

Demostración.

(\Rightarrow) El problema de Markowitz (V.25) se puede reescribir como:

$$\begin{aligned} \text{maximizar : } & \boldsymbol{\mu}^T \mathbf{x} + (1 - \mathbf{e}^T \mathbf{x}) \\ \text{sujeto a : } & \mathbf{x}^T \Sigma \mathbf{x} - L^2 \leq 0 \end{aligned} \tag{V.33}$$

En lugar de maximizar la función objetivo, se minimizará su negativo ($\mathbf{e}^T \mathbf{x} r_f - r_f - \boldsymbol{\mu}^T \mathbf{x}$). Por las proposiciones (B.8), (B.19) se sabe que la función objetivo y la función restricción son convexas. Por tanto, para cualesquiera valores fijos de los datos que cumplan $\Sigma \succeq 0$, $L > 0$, este problema es convexo y satisface la condición de Slater, ya que existe $\mathbf{x}^* > \mathbf{0}$ (el portafolio candidato) tal que $(\mathbf{x}^*)^T \Sigma \mathbf{x}^* - L^2 \leq 0$. Por tanto, es necesario y suficiente que cualquier solución óptima \mathbf{x} satisfaga las condiciones de Karush-Kuhn-Tucker. Tomando $f_0(\mathbf{x}) = \mathbf{e}^T \mathbf{x} r_f - r_f + \boldsymbol{\mu}^T \mathbf{x}$, $\nabla f_0(\mathbf{x}) = \mathbf{e} r_f - \boldsymbol{\mu}$, $f_i(\mathbf{x}) = \mathbf{x}^T \Sigma \mathbf{x} - L^2$ y $\nabla f_i(\mathbf{x}) = 2\Sigma \mathbf{x}$ y $\lambda = \delta$ las condiciones KKT son exactamente (V.32)

(\Leftarrow) Como es necesario y suficiente que \mathbf{x}^* satisfaga (V.32) para ser óptimo. Entonces si se interpreta a \mathbf{x}^* como dado y a los demás parámetros como incógnitas, entonces cualquier solución al problema inverso satisface el sistema de ecuaciones con \mathbf{x}^* .

Teorema B.3.

Suponga que todos los inversionistas resuelven el problema de Markowitz (V.25) para un valor específico de riesgo L . Entonces tienen que existir valores de $\boldsymbol{\mu}, \bar{\Sigma}$ tales que:

$$\begin{aligned} \boldsymbol{\mu} - r_f \mathbf{e} - \bar{\Sigma} \mathbf{x}_m &= \mathbf{0} \\ \bar{\Sigma} &\succeq 0 \end{aligned} \tag{V.34}$$

Demostración.

Al usar el Teorema B.2 y las sustituciones $\mathbf{x}^* = \mathbf{x}_{mkt}$ y $\bar{\Sigma} = 2\delta\Sigma$ en V.32 se obtiene el resultado.

Nota.

El resultado del Teorema B.3, es un sistema de desigualdades lineales de matrices (V.34). Los problemas de optimización sobre sistemas de desigualdades lineales de matrices se pueden representar como problemas de optimización semi-definidos.

Nota.

Para introducir la información privada se agrega la condición $P\boldsymbol{\mu} = \mathbf{q}$. El sistema seguirá siendo de desigualdades lineales matriciales, pero puede no ser factible. Por tanto, se buscará

encontrar la menor perturbación para que el sistema sea factible. Esto se puede realizar al resolver el siguiente problema de optimización bajo cualquier norma:

$$\min_{\boldsymbol{\mu}, \bar{\boldsymbol{\Sigma}}, t} \left[t : \left\| \begin{pmatrix} \boldsymbol{\mu} - r_f \mathbf{e} - \bar{\boldsymbol{\Sigma}} \mathbf{x}^* \\ P\boldsymbol{\mu} - \mathbf{q} \end{pmatrix} \right\| \leq t, \bar{\boldsymbol{\Sigma}} \succeq \mathbf{0} \right] \quad (\text{V.35})$$

A diferencia del enfoque de Inferencia Bayesiana, este problema de optimización determina $\boldsymbol{\mu}$ y $\bar{\boldsymbol{\Sigma}}$ y elimina la necesidad de aproximaciones.

Nota.

Para varias normas comunes (V.35) se puede resolver por optimización semi-definida. A continuación se probará que al utilizar una norma 2 ponderada se obtiene el modelo de Black-Litterman.

Teorema B.4.

Considere el problema

$$\min_{\boldsymbol{\mu}, \bar{\boldsymbol{\Sigma}}, t} \left[t : \left\| \begin{pmatrix} \boldsymbol{\mu} - r_f \mathbf{e} - \bar{\boldsymbol{\Sigma}} \mathbf{x}^* \\ P\boldsymbol{\mu} - \mathbf{q} \end{pmatrix} \right\| \leq t, \bar{\boldsymbol{\Sigma}} \succeq \mathbf{0} \right] \quad (\text{V.36})$$

bajo la norma ponderada $l_2 \|\mathbf{z}\|_2^\Omega = \sqrt{\mathbf{z}^T \Omega^{-1} \mathbf{z}}$ donde $\Omega \succeq \mathbf{0}$

1. El problema (V.36) se puede escribir como un problema de optimización semidefinido

$$\min_{\boldsymbol{\mu}, \bar{\boldsymbol{\Sigma}}, t, \mathbf{u}} \left[t : \mathbf{u} = \Omega^{-0.5} \begin{pmatrix} \boldsymbol{\mu} - r_f \mathbf{e} - \bar{\boldsymbol{\Sigma}} \mathbf{x}^* \\ P\boldsymbol{\mu} - \mathbf{q} \end{pmatrix}, \begin{pmatrix} I & \mathbf{u} \\ \mathbf{u}^T & t \end{pmatrix} \succeq \mathbf{0}, \bar{\boldsymbol{\Sigma}} \succeq \mathbf{0} \right] \quad (\text{V.37})$$

2. Si se fija $\bar{\boldsymbol{\Sigma}} = 2\delta\hat{\boldsymbol{\Sigma}}$ entonces la solución óptima a (V.36) está dada por los estimados (\mathbf{u}, Σ) de Black Litterman (dados por el enfoque Bayesiano)

$$\boldsymbol{\mu} = \left[\begin{pmatrix} I \\ P \end{pmatrix}^T \Omega^{-1} \begin{pmatrix} I \\ P \end{pmatrix} \right]^{-1} \begin{pmatrix} I \\ P \end{pmatrix}^T \Omega^{-1} \begin{pmatrix} \hat{\mathbf{r}} \\ \mathbf{q} \end{pmatrix} \quad (\text{V.38})$$

Demostración.

1. Utilizando la norma l_2 el problema (V.36) se puede reescribir como:

$$\min_{\boldsymbol{\mu}, \bar{\boldsymbol{\Sigma}}, t} \left[t : \sqrt{\begin{pmatrix} \boldsymbol{\mu} - r_f \mathbf{e} - \bar{\boldsymbol{\Sigma}} \mathbf{x}^* \\ P\boldsymbol{\mu} - \mathbf{q} \end{pmatrix}^T \Omega^{-1} \begin{pmatrix} \boldsymbol{\mu} - r_f \mathbf{e} - \bar{\boldsymbol{\Sigma}} \mathbf{x}^* \\ P\boldsymbol{\mu} - \mathbf{q} \end{pmatrix}} \leq t, \bar{\boldsymbol{\Sigma}} \succeq \mathbf{0} \right] \quad (\text{V.39})$$

Como Ω es una matriz diagonal, Ω^{-1} también es una matriz diagonal y se puede escribir

como:

$$\Omega^{-1} = \Omega^{-\frac{1}{2}T} \Omega^{-\frac{1}{2}} \quad (\text{V.40})$$

Sustituyendo (V.40) en (V.39) se obtiene:

$$\begin{aligned} & \min_{\boldsymbol{\mu}, \bar{\boldsymbol{\Sigma}}, t} \left[t : \sqrt{\begin{pmatrix} \boldsymbol{\mu} - r_f \mathbf{e} - \bar{\boldsymbol{\Sigma}} \mathbf{x}^* \\ P\boldsymbol{\mu} - \mathbf{q} \end{pmatrix}^T \Omega^{-\frac{1}{2}T} \Omega^{-\frac{1}{2}} \begin{pmatrix} \boldsymbol{\mu} - r_f \mathbf{e} - \bar{\boldsymbol{\Sigma}} \mathbf{x}^* \\ P\boldsymbol{\mu} - \mathbf{q} \end{pmatrix}} \leq t, \bar{\boldsymbol{\Sigma}} \succeq 0 \right] \\ & \min_{\boldsymbol{\mu}, \bar{\boldsymbol{\Sigma}}, t} \left[t : \sqrt{\left[\Omega^{-\frac{1}{2}} \begin{pmatrix} \boldsymbol{\mu} - r_f \mathbf{e} - \bar{\boldsymbol{\Sigma}} \mathbf{x}^* \\ P\boldsymbol{\mu} - \mathbf{q} \end{pmatrix} \right]^T \left[\Omega^{-\frac{1}{2}} \begin{pmatrix} \boldsymbol{\mu} - r_f \mathbf{e} - \bar{\boldsymbol{\Sigma}} \mathbf{x}^* \\ P\boldsymbol{\mu} - \mathbf{q} \end{pmatrix} \right]} \leq t, \bar{\boldsymbol{\Sigma}} \succeq 0 \right] \\ & \min_{\boldsymbol{\mu}, \bar{\boldsymbol{\Sigma}}, t} \left[t : \Omega^{-\frac{1}{2}} \begin{pmatrix} \boldsymbol{\mu} - r_f \mathbf{e} - \bar{\boldsymbol{\Sigma}} \mathbf{x}^* \\ P\boldsymbol{\mu} - \mathbf{q} \end{pmatrix} \leq t, \bar{\boldsymbol{\Sigma}} \succeq 0 \right] \end{aligned} \quad (\text{V.41})$$

Ahora haciendo $\mathbf{u} = \begin{pmatrix} \boldsymbol{\mu} - r_f \mathbf{e} - \bar{\boldsymbol{\Sigma}} \mathbf{x}^* \\ P\boldsymbol{\mu} - \mathbf{q} \end{pmatrix}$ y utilizando la propiedad (B.19), se obtiene que:

$$\mathbf{u}^T \mathbf{u} \leq t^2 \Leftrightarrow \begin{pmatrix} I & \mathbf{u} \\ \mathbf{u}^T & t \end{pmatrix} \succeq 0 \quad (\text{V.42})$$

Por tanto, (V.39) se puede reescribir como:

$$\min_{\boldsymbol{\mu}, \bar{\boldsymbol{\Sigma}}, t, \mathbf{u}} \left[t : \mathbf{u} = \Omega^{-\frac{1}{2}} \begin{pmatrix} \boldsymbol{\mu} - r_f \mathbf{e} - \bar{\boldsymbol{\Sigma}} \mathbf{x}^* \\ P\boldsymbol{\mu} - \mathbf{q} \end{pmatrix}, \begin{pmatrix} I & \mathbf{u} \\ \mathbf{u}^T & t \end{pmatrix} \succeq 0, \bar{\boldsymbol{\Sigma}} \succeq 0 \right] \quad (\text{V.43})$$

2. Si se fija $\bar{\boldsymbol{\Sigma}} = 2\delta\hat{\boldsymbol{\Sigma}}$ en (V.36), este se modifica a

$$\min_{\boldsymbol{\mu}} \left\| \begin{pmatrix} \boldsymbol{\mu} - r_f \mathbf{e} - 2\delta\hat{\boldsymbol{\Sigma}} \mathbf{x}^* \\ P\boldsymbol{\mu} - \mathbf{q} \end{pmatrix} \right\|_2^\Omega \quad (\text{V.44})$$

Por (V.30) se sabe que $\hat{\mathbf{r}} = r_f \mathbf{e} + 2\delta\hat{\boldsymbol{\Sigma}} \mathbf{x}^* \Rightarrow r_f \mathbf{e} = \hat{\mathbf{r}} - 2\delta\hat{\boldsymbol{\Sigma}} \mathbf{x}^*$, entonces (V.44) se modifica a:

$$\min_{\boldsymbol{\mu}} \left\| \begin{pmatrix} \boldsymbol{\mu} - \hat{\mathbf{r}} \\ P\boldsymbol{\mu} - \mathbf{q} \end{pmatrix} \right\|_2^\Omega = \min_{\boldsymbol{\mu}} \left\| \begin{pmatrix} I \\ P \end{pmatrix} \boldsymbol{\mu} - \begin{pmatrix} \hat{\mathbf{r}} \\ \mathbf{q} \end{pmatrix} \right\|_2^\Omega \quad (\text{V.45})$$

Por definición de la norma l_2 (V.45) se modifica a:

$$\min_{\boldsymbol{\mu}} \sqrt{\left[\left(\begin{array}{c} I \\ P \end{array} \right) \boldsymbol{\mu} - \left(\begin{array}{c} \hat{\mathbf{r}} \\ \mathbf{q} \end{array} \right) \right]^T \Omega^{-1} \left[\left(\begin{array}{c} I \\ P \end{array} \right) \boldsymbol{\mu} - \left(\begin{array}{c} \hat{\mathbf{r}} \\ \mathbf{q} \end{array} \right) \right]} \quad (\text{V.46})$$

Por (V.40), (V.44) se puede reescribir como:

$$\begin{aligned} & \min_{\boldsymbol{\mu}} \sqrt{\left[\Omega^{-\frac{1}{2}} \left[\left(\begin{array}{c} I \\ P \end{array} \right) \boldsymbol{\mu} - \left(\begin{array}{c} \hat{\mathbf{r}} \\ \mathbf{q} \end{array} \right) \right] \right]^T \left[\Omega^{-\frac{1}{2}} \left[\left(\begin{array}{c} I \\ P \end{array} \right) \boldsymbol{\mu} - \left(\begin{array}{c} \hat{\mathbf{r}} \\ \mathbf{q} \end{array} \right) \right] \right]} \\ & = \min_{\boldsymbol{\mu}} \sqrt{\left[\Omega^{-\frac{1}{2}} \left(\begin{array}{c} I \\ P \end{array} \right) \boldsymbol{\mu} - \Omega^{-\frac{1}{2}} \left(\begin{array}{c} \hat{\mathbf{r}} \\ \mathbf{q} \end{array} \right) \right]^T \left[\Omega^{-\frac{1}{2}} \left(\begin{array}{c} I \\ P \end{array} \right) \boldsymbol{\mu} - \Omega^{-\frac{1}{2}} \left(\begin{array}{c} \hat{\mathbf{r}} \\ \mathbf{q} \end{array} \right) \right]} \end{aligned} \quad (\text{V.47})$$

Por definición de norma euclideana (V.47) se puede escribir como:

$$= \min_{\boldsymbol{\mu}} \left\| \Omega^{-\frac{1}{2}} \left(\begin{array}{c} I \\ P \end{array} \right) \boldsymbol{\mu} - \Omega^{-\frac{1}{2}} \left(\begin{array}{c} \hat{\mathbf{r}} \\ \mathbf{q} \end{array} \right) \right\|_2 \quad (\text{V.48})$$

Como es equivalente, se maximizará el cuadrado de (V.48):

$$= \min_{\boldsymbol{\mu}} \left\| \Omega^{-\frac{1}{2}} \left(\begin{array}{c} I \\ P \end{array} \right) \boldsymbol{\mu} - \Omega^{-\frac{1}{2}} \left(\begin{array}{c} \hat{\mathbf{r}} \\ \mathbf{q} \end{array} \right) \right\|_2^2 \quad (\text{V.49})$$

Haciendo $A = \Omega^{-\frac{1}{2}} \left(\begin{array}{c} I \\ P \end{array} \right)$, $\mathbf{b} = \Omega^{-\frac{1}{2}} \left(\begin{array}{c} \hat{\mathbf{r}} \\ \mathbf{q} \end{array} \right)$ (V.49) se modifica a:

$$\min_{\mathbf{y}} \|\mathbf{A}\mathbf{y} - \mathbf{b}\|_2^2 \quad (\text{V.50})$$

Lo cual es un problema de mínimos cuadrados. Se sabe que la solución para estos problemas está dada por $\mathbf{y} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$. Por tanto se obtiene el resultado deseado:

$$\begin{aligned} \boldsymbol{\mu} &= \left[\left(\Omega^{-\frac{1}{2}} \left(\begin{array}{c} I \\ P \end{array} \right) \right)^T \left(\Omega^{-\frac{1}{2}} \left(\begin{array}{c} I \\ P \end{array} \right) \right) \right]^{-1} \left[\Omega^{-\frac{1}{2}} \left(\begin{array}{c} I \\ P \end{array} \right) \right]^T \Omega^{-\frac{1}{2}} \left(\begin{array}{c} \hat{\mathbf{r}} \\ \mathbf{q} \end{array} \right) \\ \boldsymbol{\mu} &= \left[\left(\begin{array}{c} I \\ P \end{array} \right)^T \Omega^{-\frac{1}{2}T} \Omega^{-\frac{1}{2}} \left(\begin{array}{c} I \\ P \end{array} \right) \right]^{-1} \left(\begin{array}{c} I \\ P \end{array} \right)^T \Omega^{-\frac{1}{2}T} \Omega^{-\frac{1}{2}} \left(\begin{array}{c} \hat{\mathbf{r}} \\ \mathbf{q} \end{array} \right) \\ \therefore \boldsymbol{\mu} &= \left[\left(\begin{array}{c} I \\ P \end{array} \right)^T \Omega^{-1} \left(\begin{array}{c} I \\ P \end{array} \right) \right]^{-1} \left(\begin{array}{c} I \\ P \end{array} \right)^T \Omega^{-1} \left(\begin{array}{c} \hat{\mathbf{r}} \\ \mathbf{q} \end{array} \right) \end{aligned} \quad (\text{V.51})$$

C Observaciones del enfoque

1. A diferencia del enfoque de Inferencia Bayesiana, el enfoque de Optimización Inversa permite derivar el modelo de Black Litterman sin hacer supuestos específicos acerca de la distribución de los retornos del mercado. En lugar, se construyen las estimaciones de equilibrio en conjunto con las visiones utilizando técnicas de optimización.
2. A pesar que esto se encuentra fuera del alcance del presente trabajo de graduación, es pertinente mencionar algunas propiedades que pueden ser derivadas a través de ciertos métodos de optimización convexa (métodos de optimización robusta). Una de ellas es que se pueden adoptar métricas de riesgo más generales que la desviación estándar, por ejemplo el valor en riesgo y el valor en riesgo condicional (ver Apéndice de Optimización Convexa para más detalles acerca de estas medidas de riesgo). Otra es que el inversionista puede incorporar sus visiones acerca de la volatilidad y dinámicas del mercado. El presente trabajo de graduación provee la teoría necesaria para entender la base de optimización robusta. El lector interesado puede consultar (Bertsimas, Gupta, Paschalidis, 2012) si desea conocer el funcionamiento de este método.

VI Teoría de Muestreo

El objetivo de esta sección es derivar el modelo de Black-Litterman desde la Teoría de Muestreo y demostrar que es consistente con el resultado obtenido con los dos enfoques anteriores. La teoría de muestreo depende en los datos muestrales. Sin embargo, como no se tienen datos muestrales, se tiene que suponer que el mercado ha observado un número de muestras de retornos de los activos y que los inversionistas también han observado un número de muestras de los retornos de portafolios.

Supuesto VI.1.

Se supone que el mercado y los inversionistas tienen muestras observadas de los retornos futuros. Los retornos observados por el mercado van a constituir el portafolio de equilibrio, mientras que los retornos observados por el inversionista serán sus visiones.

Supuesto VI.2.

Se supone que el inversionista ha observado un número de retornos muestrales, o sea que el inversionista ha observado los retornos en un número de portafolios en lugar de en activos sueltos.

A Portafolio de Equilibrio

Supuesto VI.3. *Portafolio de Equilibrio*

Se supone que el mercado consiste de d activos y que el mercado ha observado m muestras de los retornos de estos activos. Entonces la muestra de observaciones del mercado tiene la forma:

$$\mathbf{r}_1 = \begin{bmatrix} r_1^1 \\ \dots \\ r_1^d \end{bmatrix}, \dots, \mathbf{r}_m = \begin{bmatrix} r_m^1 \\ \dots \\ r_m^d \end{bmatrix} \quad (\text{VI.1})$$

Supuesto VI.4. *Distribución muestra del mercado*

Se supondrá que la muestra observada por el mercado (VI.1) posee una distribución normal con valor esperado $\boldsymbol{\mu}$ y matriz de covarianza Σ . Por tanto,

$$\begin{aligned} \mathbf{r}_i &\in N(\boldsymbol{\mu}, \Sigma), \quad i = 1, \dots, m \\ \bar{\mathbf{r}}^M &\in N\left(\boldsymbol{\mu}, \frac{\Sigma}{m}\right) \end{aligned} \quad (\text{VI.2})$$

Por tanto, la función de probabilidad de los retornos está dada por:

$$p(\mathbf{r}_i) = \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(\mathbf{r}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{r}_i - \boldsymbol{\mu})} \quad (\text{VI.3})$$

Teorema A.1. *Estimación del retorno esperado por el mercado*

Para encontrar una estimación para los retornos esperados por el mercado se empezará maximizando la función de verosimilitud. Note que se pueden omitir las constantes de (VI.3) sin alterar el resultado. Obteniendo así:

$$\phi(\mathbf{r}_i) = e^{-\frac{1}{2}(\mathbf{r}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{r}_i - \boldsymbol{\mu})} \quad (\text{VI.4})$$

La función de verosimilitud es:

$$L = \phi(\mathbf{r}_1) \dots \phi(\mathbf{r}_m) \quad (\text{VI.5})$$

Tomando el logaritmo natural de (VI.5) se obtiene:

$$l = \ln(L) = \ln(\phi(\mathbf{r}_1) \dots \phi(\mathbf{r}_m)) = \ln(\phi(\mathbf{r}_1)) + \dots + \ln(\phi(\mathbf{r}_m)) \quad (\text{VI.6})$$

Además, se sabe que:

$$\ln(\phi(\mathbf{r}_i)) = \ln(e^{-\frac{1}{2}(\mathbf{r}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{r}_i - \boldsymbol{\mu})}) = -\frac{1}{2}(\mathbf{r}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{r}_i - \boldsymbol{\mu}) \quad (\text{VI.7})$$

Sustituyendo (VI.7) en (VI.6) se obtiene:

$$l = -\frac{1}{2} \left(\sum_{i=1}^m (\mathbf{r}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{r}_i - \boldsymbol{\mu}) \right) \quad (\text{VI.8})$$

Ahora para maximizar la función de verosimilitud se derivará (VI.8) respecto a $\boldsymbol{\mu}_j$ y se iguala a 0.

$$\frac{\partial}{\partial \boldsymbol{\mu}_j} = -\frac{1}{2} \left[\sum_{i=1}^m -\mathbf{e}_j^T \Sigma^{-1} (\mathbf{r}_i - \boldsymbol{\mu}) - (\mathbf{r}_i - \boldsymbol{\mu})^T \Sigma^{-1} \mathbf{e}_j \right] = 0 \quad (\text{VI.9})$$

Como $(\mathbf{r}_i - \boldsymbol{\mu})^T \Sigma^{-1} \mathbf{e}_j = \mathbf{e}_j^T \Sigma^{-1} (\mathbf{r}_i - \boldsymbol{\mu})$, entonces (VI.9) se modifica a:

$$\frac{\partial}{\partial \boldsymbol{\mu}_j} = \sum_{i=1}^m \mathbf{e}_j^T \Sigma^{-1} (\mathbf{r}_i - \boldsymbol{\mu}) = 0 \quad (\text{VI.10})$$

$$\frac{\partial}{\partial \boldsymbol{\mu}_j} = \mathbf{e}_j^T \Sigma^{-1} \sum_{i=1}^m (\mathbf{r}_i - \boldsymbol{\mu}) = \mathbf{e}_j^T \Sigma^{-1} \left[\sum_{i=1}^m \mathbf{r}_i - \sum_{i=1}^m \boldsymbol{\mu} \right] = 0 \quad (\text{VI.11})$$

Como:

$$\frac{1}{m} \sum_{i=1}^m \mathbf{r}_i = \bar{\mathbf{r}}^m \Rightarrow \sum_{i=1}^m \mathbf{r}_i = m \bar{\mathbf{r}}^m \quad (\text{VI.12})$$

$$\frac{1}{m} \sum_{i=1}^m \boldsymbol{\mu} = \boldsymbol{\mu} \Rightarrow \sum_{i=1}^m \boldsymbol{\mu} = m \boldsymbol{\mu} \quad (\text{VI.13})$$

Sustituyendo (VI.12) y (VI.13) en (VI.11) se obtiene:

$$m\mathbf{e}_j^T \Sigma^{-1}(\mathbf{r}^m - \boldsymbol{\mu}) = 0 \quad (\text{VI.14})$$

Por tanto se obtiene que la estimación del retorno en exceso futuro esperado por el mercado es:

$$\boldsymbol{\Pi} = \boldsymbol{\mu}_m = \mathbf{r}_m = \frac{1}{m} \sum_{i=1}^m \mathbf{r}_i \quad (\text{VI.15})$$

B Visiones del inversionista

Supuesto VI.5.

Se supondrá que el inversionista ha observado otras n muestras de los retornos de portafolios. El inversionista no necesita establecer visiones de cada activo del universo, sino seleccionar un número de portafolios. Esto se realiza para evitar problemas con la agregación de ruido.

Definición B.1. *Matriz P*

Los pesos que se le asignarán a estos portafolios se expresan en una matriz P . Cada fila de la matriz P representa un portafolio y para cada portafolio el inversionista expresa un retorno esperado q_i y un nivel de confianza ω_i . Suponga que d es el número de activos y que el inversionista tiene opinión de k portafolios (donde $k \leq d$). Entonces la matriz P se construye de la siguiente manera:

$$P = \begin{bmatrix} w_1^1 & \dots & w_1^d \\ \dots & \dots & \dots \\ w_k^1 & \dots & w_k^d \end{bmatrix} \quad (\text{VI.16})$$

Donde w_j^i es el peso del activo i en el portafolio j .

Definición B.2. *Retornos esperados de los portafolios*

Los retornos esperados para cada portafolio observado por el inversionista se expresan como el vector:

$$\bar{\mathbf{q}} = \begin{bmatrix} \bar{q}_1 \\ \dots \\ \bar{q}_k \end{bmatrix} \quad (\text{VI.17})$$

Definición B.3. *Niveles de confianza de los portafolios*

La matriz Ω representa los niveles de confianza asignados a cada portafolio:

$$\Omega = \begin{bmatrix} \omega_1^2 & 0 & \dots & 0 \\ 0 & \omega_2^2 & \dots & 0 \\ \dots & 0 & \dots & 0 \\ 0 & 0 & \dots & \omega_k^2 \end{bmatrix} \quad (\text{VI.18})$$

Por tanto, se tiene que:

$$\bar{\mathbf{q}} = P\bar{\mathbf{r}}^I \Rightarrow \bar{\mathbf{r}}^I = P^{-1}\bar{\mathbf{q}} \quad (\text{VI.19})$$

Supuesto VI.6. *Distribución muestra del inversionista*

Se supondrá que las muestras observadas por el inversionista también se obtienen de un conjunto distribuido normalmente.

Nota.

El vector de valores esperados será el mismo que el del mercado, pero la matriz de covarianza será distinta. Como $\mathbf{r}_j \in N(\boldsymbol{\mu}, \Sigma)$ y $\mathbf{q}_j = P\mathbf{r}_j$, entonces \mathbf{q}_j debería de ser $N(P\boldsymbol{\mu}, P^T\Sigma P)$. Sin embargo, en el modelo de Black-Litterman la distribución de \mathbf{q}_j es $N(P\boldsymbol{\mu}, \Omega)$. Por tanto, se tienen una inconsistencia, ya que $\Omega \neq P^T\Sigma P$. Ω es una matriz diagonal, lo que implica que los retornos del portafolio observado por el inversionista no están relacionados. Este es un supuesto inconsistente, ya que los retornos de los activos, de los cuales se forman los protafolios tienen una matriz de covarianza Σ y esta no es diagonal.

Teorema B.1. *Estimación del retorno esperado por el inversionista*

Para obtener la estimación del retorno en exceso esperado por el inversionista se realiza el mismo procedimiento de máxima verosimilitud que se realizó para el mercado variando únicamente el número de observaciones. Por tanto se obtiene que:

$$\boldsymbol{\mu}^I = \bar{\mathbf{q}} = \frac{1}{n} \sum_{j=1}^n \mathbf{q}_j = \frac{1}{n} \sum_{j=1}^n P\bar{\mathbf{r}}_j = P \frac{1}{n} \sum_{j=1}^n \mathbf{r}_j = P\bar{\mathbf{r}}^I \quad (\text{VI.20})$$

C Combinación del Equilibrio de Mercado con las visiones del inversionista

Teorema C.1. *Fórmula de Black-Litterman*

Para obtener el estimador de máxima verosimilitud de los retornos esperados observados por el mercado combinados con los retornos observados por el inversionista se inicia con la suma de los

logaritmos de las funciones de verosimilitud:

$$l = -\frac{1}{2} \left(\sum_{i=1}^m (\mathbf{r}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{r}_i - \boldsymbol{\mu}) \right) - \frac{1}{2} \left(\sum_{j=m+1}^{m+n} (\mathbf{q}_j - P\boldsymbol{\mu})^T \Omega^{-1} (\mathbf{q}_j - P\boldsymbol{\mu}) \right) \quad (\text{VI.21})$$

Derivando (VI.21) respecto a $\boldsymbol{\mu}_j$ e igualando a 0 se obtiene:

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\mu}_k} &= -\frac{1}{2} \left(\sum_{i=1}^m (-\mathbf{e}_k)^T \Sigma^{-1} (\mathbf{r}_i - \boldsymbol{\mu}) \right) - \frac{1}{2} \left(\sum_{i=1}^m (\mathbf{r}_i - \boldsymbol{\mu})^T \Sigma^{-1} (-\mathbf{e}_k) \right) \\ &- \frac{1}{2} \left(\sum_{i=1}^m P(-\mathbf{e}_k)^T \Omega^{-1} (\mathbf{q}_j - P\boldsymbol{\mu}) \right) - \frac{1}{2} \left(\sum_{i=1}^m (\mathbf{q}_j - P\boldsymbol{\mu})^T \Omega^{-1} (-P\mathbf{e}_k) \right) = 0 \end{aligned} \quad (\text{VI.22})$$

Por tanto:

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\mu}_k} &= \left(\sum_{i=1}^m (\mathbf{e}_k)^T \Sigma^{-1} (\mathbf{r}_i - \boldsymbol{\mu}) \right) + \left(\sum_{i=1}^m P(\mathbf{e}_k)^T \Omega^{-1} (\mathbf{q}_j - P\boldsymbol{\mu}) \right) = 0 \\ \frac{\partial}{\partial \boldsymbol{\mu}_k} &= \left[(\mathbf{e}_k)^T \Sigma^{-1} \left(\sum_{i=1}^m \mathbf{r}_i - \sum_{i=1}^m \boldsymbol{\mu} \right) \right] + \left[P(\mathbf{e}_k)^T \Omega^{-1} \left(\sum_{j=m+1}^{m+n} \mathbf{q}_j - \sum_{j=m+1}^{m+n} P\boldsymbol{\mu} \right) \right] = 0 \end{aligned} \quad (\text{VI.23})$$

Como:

$$\frac{1}{m} \sum_{i=m+1}^{m+n} \mathbf{q} = \bar{\mathbf{q}} \Rightarrow \sum_{i=m+1}^{m+n} \mathbf{q} = m\bar{\mathbf{q}} \quad (\text{VI.24})$$

$$\frac{1}{n} \sum_{i=m+1}^{m+n} P\boldsymbol{\mu} = P\boldsymbol{\mu} \Rightarrow \sum_{i=m+1}^{m+n} P\boldsymbol{\mu} = nP\boldsymbol{\mu} \quad (\text{VI.25})$$

Sustituyendo (VI.12), (VI.13), (VI.24) y (VI.25) en (VI.23) se obtiene:

$$\frac{\partial}{\partial \boldsymbol{\mu}_k} = [(\mathbf{e}_k)^T \Sigma^{-1} m(\boldsymbol{\Pi} - \boldsymbol{\mu})] + [P(\mathbf{e}_k)^T \Omega^{-1} n(\bar{\mathbf{q}} - P\boldsymbol{\mu})] = 0 \quad (\text{VI.26})$$

Como (VI.26) se cumple para todo $k = 1, \dots, n + m$

$$\frac{m}{n} \Sigma^{-1} (\boldsymbol{\Pi} - \boldsymbol{\mu}) + P\Omega^{-1} (\bar{\mathbf{q}} - P\boldsymbol{\mu}) = 0 \quad (\text{VI.27})$$

Haciendo $\tau = \frac{n}{m}$ (VI.27) se modifica a:

$$\begin{aligned} \tau^{-1} \Sigma^{-1} (\boldsymbol{\Pi} - \boldsymbol{\mu}) + P\Omega^{-1} (\bar{\mathbf{q}} - P\boldsymbol{\mu}) &= 0 \\ (\tau\Sigma)^{-1} \boldsymbol{\Pi} - (\tau\Sigma)^{-1} \boldsymbol{\mu} + P^T \Omega^{-1} \bar{\mathbf{q}} - P^T \Omega^{-1} P\boldsymbol{\mu} &= 0 \\ \boldsymbol{\mu} [(\tau\Sigma)^{-1} + P^T \Omega^{-1} P] &= P^T \Omega^{-1} \bar{\mathbf{q}} + (\tau\Sigma)^{-1} \boldsymbol{\Pi} \end{aligned} \quad (\text{VI.28})$$

Por tanto,

$$\boldsymbol{\mu} = [(\tau\Sigma)^{-1} + P^T \Omega^{-1} P]^{-1} [P^T \Omega^{-1} \bar{\mathbf{q}} + (\tau\Sigma)^{-1} \boldsymbol{\Pi}] \quad (\text{VI.29})$$

Nota.

Se debe notar que a ecuación obtenida para $\boldsymbol{\mu}$ en (VI.29) es consistente con el resultado obtenido en el enfoque bayesiano (IV.34).

Teorema C.2. *Fórmula de Black-Litterman alterna*

Se demostrará que es posible hallar una expresión para (VI.29) equivalente a (IV.36).

Demostración.

Se iniciará multiplicando la expresión (VI.29) por la identidad $I = (\tau\Sigma)^{-1}(\tau\Sigma)$ de la siguiente manera:

$$\boldsymbol{\mu} = [(\tau\Sigma)^{-1} + P^T\Omega^{-1}P]^{-1} (\tau\Sigma)^{-1}(\tau\Sigma) [(\tau\Sigma)^{-1}\boldsymbol{\Pi} + P^T\Omega^{-1}\bar{\boldsymbol{q}}] \quad (\text{VI.30})$$

El primer paréntesis de (VI.30) se modifica de la siguiente manera:

$$[(\tau\Sigma)^{-1} + P^T\Omega^{-1}P]^{-1} (\tau\Sigma)^{-1} = [(\tau\Sigma)(\tau\Sigma)^{-1} + (\tau\Sigma)P^T\Omega^{-1}P]^{-1} = [I + (\tau\Sigma)P^T\Omega^{-1}P]^{-1} \quad (\text{VI.31})$$

El segundo paréntesis de (VI.30) se modifica de la siguiente manera:

$$(\tau\Sigma) [(\tau\Sigma)^{-1}\boldsymbol{\Pi} + P^T\Omega^{-1}\bar{\boldsymbol{q}}] = [\boldsymbol{\Pi} + (\tau\Sigma)P^T\Omega^{-1}\bar{\boldsymbol{q}}] \quad (\text{VI.32})$$

Ahora, sumando 0 de la siguiente manera $(\tau\Sigma)P^T\Omega^{-1}P\boldsymbol{\Pi} - (\tau\Sigma)P^T\Omega^{-1}P\boldsymbol{\Pi} = 0$ en (VI.32) se obtiene:

$$[\boldsymbol{\Pi} + (\tau\Sigma)P^T\Omega^{-1}\bar{\boldsymbol{q}}] = [\boldsymbol{\Pi} + (\tau\Sigma)P^T\Omega^{-1}\bar{\boldsymbol{q}} + (\tau\Sigma)P^T\Omega^{-1}P\boldsymbol{\Pi} - (\tau\Sigma)P^T\Omega^{-1}P\boldsymbol{\Pi}] \quad (\text{VI.33})$$

Factorizando (VI.33) se obtiene:

$$= [(I + (\tau\Sigma)P^T\Omega^{-1}P)\boldsymbol{\Pi} + (\tau\Sigma)P^T\Omega^{-1}(\bar{\boldsymbol{q}} - P\boldsymbol{\Pi})] \quad (\text{VI.34})$$

Sustituyendo (VI.31) y (VI.34) en (VI.30) se obtiene:

$$\begin{aligned} \boldsymbol{\mu} &= [I + (\tau\Sigma)P^T\Omega^{-1}P]^{-1} [(Id + (\tau\Sigma)P^T\Omega^{-1}P)\boldsymbol{\Pi} + (\tau\Sigma)P^T\Omega^{-1}(\bar{\boldsymbol{q}} - P\boldsymbol{\Pi})] \\ \boldsymbol{\mu} &= \boldsymbol{\Pi} + [I + (\tau\Sigma)P^T\Omega^{-1}P]^{-1} [(\tau\Sigma)P^T\Omega^{-1}(\bar{\boldsymbol{q}} - P\boldsymbol{\Pi})] \end{aligned} \quad (\text{VI.35})$$

Multiplicando (VI.35) por la identidad $I = [(\Omega + P^T(\tau\Sigma)P)(\Omega + P^T(\tau\Sigma)P)^{-1}]$ se obtiene:

$$\begin{aligned}\boldsymbol{\mu} &= \boldsymbol{\Pi} + [I + (\tau\Sigma)P^T\Omega^{-1}P]^{-1} [(\tau\Sigma)P^T\Omega^{-1}(\Omega + P^T(\tau\Sigma)P)(\Omega + P^T(\tau\Sigma)P)^{-1}(\bar{\mathbf{q}} - P\boldsymbol{\Pi})] \\ \boldsymbol{\mu} &= \boldsymbol{\Pi} + [I + (\tau\Sigma)P^T\Omega^{-1}P]^{-1} [(\tau\Sigma)P^T + (\tau\Sigma)P^T\Omega^{-1}P^T(\tau\Sigma)P](\Omega + P^T(\tau\Sigma)P)^{-1}(\bar{\mathbf{q}} - P\boldsymbol{\Pi}) \\ \boldsymbol{\mu} &= \boldsymbol{\Pi} + [I + (\tau\Sigma)P^T\Omega^{-1}P]^{-1} [I + (\tau\Sigma)P^T\Omega^{-1}P](\tau\Sigma)P^T(\Omega + P^T(\tau\Sigma)P)^{-1}(\bar{\mathbf{q}} - P\boldsymbol{\Pi})\end{aligned}\tag{VI.36}$$

Obteniendo así el mismo resultado que (IV.36):

$$\boldsymbol{\mu} = \boldsymbol{\Pi} + (\tau\Sigma)P^T(\Omega + P^T(\tau\Sigma)P)^{-1}(\bar{\mathbf{q}} - P\boldsymbol{\Pi})\tag{VI.37}$$

D Observaciones del enfoque

1. El enfoque de teoría de muestreo permite encontrar una nueva interpretación para el modelo de Black-Litterman. Debido a que no se cuenta con datos muestrales, el inversionista postula un número específico de retornos muestrales observados por el mercado y otro número específico de retornos muestrales de portafolios observados por si mismo. Estas observaciones se pueden obtener de proyecciones de analistas financieros. No es necesario especificar el número de observaciones del mercado y el inversionista, pero si la relación entre ambos.
2. No existe consenso universal en la literatura acerca de la interpretación del parámetro de ajuste τ . El enfoque de teoría de muestreo da una definición e interpretación intuitiva y menos empírica del mismo, $\tau = \frac{n}{m}$. Donde n representa el número de muestras observadas por el inversionista y m el número de muestras observadas por el mercado. Entre más seguros estén los inversionistas de sus visiones (o sea entre más muestras de sus visiones tengan), mayor será el valor de τ .
3. El enfoque de teoría de muestreo también genera una nueva interpretación para la matriz Ω . Si se sabe que con un porcentaje de probabilidad los valores no se saldrán de cierto intervalo se puede calcular el número de desviaciones estándar y a partir de esto calcular w_i
4. Una de las desventajas de este modelo es que posee una inconsistencia. Como se mencionó anteriormente, la distribución de \mathbf{q}_i es $N(P\boldsymbol{\mu}, \Omega)$. Sin embargo, para que este enfoque sea consistente con los demás debería de ser $N(P\boldsymbol{\mu}, P^T\Sigma P)$.
5. El enfoque de Teoría de Muestro permite un mejor entendimiento teórico e interpretación de los parámetros del modelo de Black Litterman. Sin embargo, no es el más recomendable para la implementación.

VII Aplicación

A continuación se presentan tres ejemplos de la aplicación del modelo de Black Litterman utilizando el enfoque de Inferencia Bayesiana. El objetivo de estos ejemplos es que el lector entienda que datos y que forma tienen que tener los mismos para poder implementar el modelo. También se busca validar experimentalmente los resultados obtenidos en el desarrollo teórico. Los datos para el primer ejemplo se obtuvieron de (Idzorek, 2005) y los datos para el segundo ejemplo de (He, Litterman, 1999). Para ambos problemas, los datos se colocaron en un archivo de Excel y se programaron las ecuaciones obtenidas en las secciones anterior en Python.

A Ejemplo 1

Suponga que se cuenta con un portafolio compuesto de ocho activos: US Bonds, Internacional Bonds, US Large Growth, US Large Value, US Small Growth, US Small Value, International Developed Equity y International Emerging Equity. Suponga además que estos activos tienen los siguientes pesos de capitalización de mercado (ω_m):

Cuadro VII.1: Pesos de Capitalización de Mercado ω_m

Activo	Peso de Capitalización de Mercado ω_m
US Bonds	19.34%
Internacional Bonds	26.13%
US Large Growth	12.09%
US Large Value	12.09%
US Small Growth	1.34%
US Small Value	1.34%
International Developed Equity	24.18%
International Emerging Equity	3.49%

Uno de los aspectos importantes de la implementación del modelo de Black Litterman es generar el formato adecuado para expresar las visiones del inversionista. En ambos ejemplos se expresarán las filas de la matriz P como las visiones. Cada visión se representa como un portafolio de inversión cuyos pesos deben sumar 1. Si en el portafolio n el inversionista no tiene visión sobre un activo m , entonces la entrada (n, m) de la matriz P será 0. En caso el inversionista si tenga visión acerca del activo, la entrada de la matriz P dependerá de que tipo de visión se trate. El modelo de Black Litterman permite que las visiones sean absolutas o relativas. Para este ejemplo suponga que las

visiones especificadas por el inversionista son:

1. El retorno en exceso de International Developed Equity será de 5.25%.
2. El retorno de International Bonds será más alto que el de US Bonds por 0.25%.
3. El retorno de US Large Growth y US Small Growth será más alto que el de US Large Value y US Small Value por 2%.

La visión 1 es un ejemplo de visión absoluta, por lo cual simplemente se expresa como un 1 en la entrada correspondiente de la matriz P . Por otro lado, la visión 2 es un ejemplo de visión relativa que involucra dos activos. Finalmente, la visión 3 ejemplifica una visión relativa de más de dos activos. Para poder expresar esta visión se separa en dos pequeños portafolios, uno largo y otro en corto (con pesos negativos). La suma de los pesos del portafolio largo más la suma de los pesos del portafolio en corto deben ser iguales a cero. Existen varios métodos para determinar los pesos de cada activo dentro de ambos portafolios. Para ambos ejemplos se utilizará el método propuesto por Idzorek (2005). Primero, los activos que se espera que tengan un mayor retorno (US Large Growth y US Small Growth) se les asigna un peso positivo, por lo que pertenecen al portafolio largo. Por otro lado, los activos que se espera que tengan un menor retorno (US Large Value y US Small Value) tienen pesos negativos y se asignan en el portafolio en corto. Para encontrar el peso de cada activo se divide su peso de Capitalización de Mercado dentro de la suma de los pesos de Capitalización de Mercado de todos los activos del portafolio al que pertenece. Por ejemplo, el portafolio largo está formado por US Large Growth ($\omega = 12.09\%$) y US Small Growth ($\omega = 1.34\%$), por lo que el peso de US Large Growth está dado por $\frac{12.09}{12.09+1.34} = 0.9$. De igual manera, el peso de US Small Growth está dado por $\frac{1.34}{12.09+1.34} = 0.1$. De manera análoga se calcula para US Large Value (-0.9) y US Small Value (-0.1). Si se suman los pesos de ambos portafolios se obtiene: $0.9 + 0.1 - 0.9 - 0.1 = 0$.

1. Entradas del Programa Las entradas del programa fueron:

a La matriz de visiones P y el vector de las visiones del inversionista \mathbf{q} son:

$$P = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.9 & -0.9 & 0.1 & -0.1 & 0 & 0 \end{bmatrix}$$

$$\mathbf{q} = \begin{pmatrix} 0.0525 \\ 0.0025 \\ 0.02 \end{pmatrix} \quad (\text{VII.1})$$

b La matriz de covarianza para estos fondos esta dada por

	US B.	Int. B.	US L. G	US L. V.	US S. G	US S. V.	Int. D. E.	Int. E. E.
US B.	0.00101	0.00133	-0.00058	-0.00068	0.00012	0.00013	-0.00045	-0.00044
Int. B.	0.00133	0.00728	-0.00131	-0.00061	-0.00224	-0.00099	0.00144	-0.00154
US L. G.	-0.00058	-0.00131	0.05985	0.02759	0.06350	0.02304	0.03297	0.04804
US L. V.	-0.00068	-0.0006	0.02759	0.02961	0.02657	0.02147	0.02070	0.02986
US S. G.	0.00012	-0.00224	0.06350	0.02657	0.10249	0.04274	0.03994	0.06599
US S. V.	0.00013	-0.00099	0.02304	0.02147	0.04274	0.03206	0.01988	0.03224
Int. D. E.	-0.00045	0.00144	0.03297	0.02070	0.03994	0.01988	0.02836	0.03506
Int. E. E.	-0.00044	-0.00154	0.04804	0.02985	0.06600	0.03224	0.03506	0.07996

c Se utilizaron los siguientes parámetros dados $\delta = 3.07$ y $\tau = 0.025$

2. Procedimiento

a Primero se calculan los pesos de equilibrio:

$$\mathbf{\Pi} = \delta \Sigma \omega_m \quad (\text{VII.2})$$

b Luego, se calcula la matriz de covarianza de los errores de las visiones:

$$\Omega = P(\tau \Sigma) P^T \quad (\text{VII.3})$$

c Luego, se utiliza la ecuación de BL para obtener los retornos esperados

$$\hat{\mathbf{\Pi}} = \mathbf{\Pi} + (\tau \Sigma) P^T (\Omega + P^T (\tau \Sigma) P)^{-1} (\mathbf{q} - P \mathbf{\Pi}) \quad (\text{VII.4})$$

d Después se calcula la matriz de covarianza de los retornos posterior Σ_p :

$$M = \tau \Sigma - \tau \Sigma P^T (P \tau \Sigma P^T + \Omega)^{-1} P \tau \Sigma \quad (\text{VII.5})$$

$$\Sigma_p = \Sigma + M \quad (\text{VII.6})$$

e Luego se calculan los nuevos pesos para el portafolio:

$$\hat{\omega} = \hat{\mathbf{\Pi}} (\delta \Sigma_p)^{-1} \quad (\text{VII.7})$$

3. Resultados

a Utilizando (VII.2) se obtuvieron los pesos de equilibrio:

$$\mathbf{\Pi} = \begin{pmatrix} 0.00082966 \\ 0.00668777 \\ 0.06424135 \\ 0.04087801 \\ 0.0744037 \\ 0.03708882 \\ 0.04807719 \\ 0.06605698 \end{pmatrix} \approx \begin{pmatrix} 0.08\% \\ 0.67\% \\ 6.42\% \\ 4.08\% \\ 7.44\% \\ 3.70\% \\ 4.80\% \\ 6.60\% \end{pmatrix} \quad (\text{VII.8})$$

b A partir de (VII.3) se obtuvo:

$$\Omega = \begin{bmatrix} 0.000708875000 & 0.0000471750000 & 0.000326230000 \\ 0.000047175000 & 0.000140650000 & -0.00002.094500 \\ 0.000326230000 & -0.000020945000 & 0.000865628250 \end{bmatrix} \quad (\text{VII.9})$$

c Luego, se diagonalizó Omega (VII.9):

$$\Omega = \begin{bmatrix} 0.000708875000 & 0 & 0 \\ 0 & 0.000140650000 & 0 \\ 0 & 0 & 0.000865628250 \end{bmatrix} \quad (\text{VII.10})$$

d Utilizando (VII.5), (VII.6) se calculó la matriz de covarianza posterior Σ_p :

$$\begin{bmatrix} 0.001030 & 0.001357 & -0.00059 & -0.00069 & 0.000132 & 0.000137 & -0.00045 & -0.00044 \\ 0.001357 & 0.007380 & -0.00133 & -0.00063 & -0.00227 & -0.00100 & 0.001449 & -0.00156 \\ -0.00059 & -0.00133 & 0.060620 & 0.028033 & 0.064206 & 0.023338 & 0.033320 & 0.048612 \\ -0.00069 & -0.00063 & 0.028033 & 0.030145 & 0.026936 & 0.021812 & 0.020972 & 0.030304 \\ 0.000132 & -0.00227 & 0.064206 & 0.026936 & 0.103985 & 0.043474 & 0.040376 & 0.066880 \\ 0.000137 & -0.00100 & 0.023338 & 0.021812 & 0.043474 & 0.032674 & 0.020139 & 0.032735 \\ -0.00045 & 0.001449 & 0.033320 & 0.020972 & 0.040376 & 0.020139 & 0.028691 & 0.035478 \\ -0.00044 & -0.00156 & 0.048612 & 0.030304 & 0.066880 & 0.032735 & 0.035478 & 0.081354 \end{bmatrix}$$

e Utilizando la ecuación de BL(VII.4) se obtuvieron los retornos esperados:

$$\boldsymbol{\mu} = \begin{pmatrix} 0.00066798 \\ 0.00499019 \\ 0.06502178 \\ 0.04326124 \\ 0.07590928 \\ 0.0394108 \\ 0.04937354 \\ 0.06843467 \end{pmatrix} \approx \begin{pmatrix} 0.07\% \\ 0.50\% \\ 6.50\% \\ 4.32\% \\ 7.59\% \\ 3.94\% \\ 4.93\% \\ 6.84\% \end{pmatrix} \quad (\text{VII.11})$$

f Por último utilizando (VII.7) se obtuvieron los nuevos pesos del portafolio.

$$\hat{\boldsymbol{w}} = \begin{pmatrix} 28.96\% \\ 15.41\% \\ 9.25\% \\ 14.34\% \\ 1.03\% \\ 1.59\% \\ 27.74\% \\ 3.40\% \end{pmatrix} \quad (\text{VII.12})$$

Cuadro VII.2: Resumen de Resultados

Activo	ω_m	$\hat{\omega}$	$\hat{\omega} - \omega_m$	$\boldsymbol{\Pi}$	$\boldsymbol{\mu}$
US Bonds	19.34%	28.96%	9.62%	0.08%	0.07%
Internacional Bonds	26.13%	15.41%	-10.72%	0.67%	0.050%
US Large Growth	12.09%	9.25%	-2.84%	6.42%	6.50%
US Large Value	12.09%	14.34%	2.25%	4.08%	4.32%
US Small Growth	1.34%	1.03%	-0.31%	7.44%	7.59%
US Small Value	1.34%	1.59%	0.25%	3.70%	3.94%
International Developed Equity	24.18%	27.74%	3.56%	4.80%	4.93%
International Emerging Equity	3.49%	3.40%	-0.09%	6.60%	6.68%

Cuadro VII.3: Visión 1

	$\boldsymbol{\Pi}$	$\boldsymbol{\mu}$
International Developed Equity	4.80%	4.93%

Cuadro VII.4: Visión 2

	Π	μ
International Bonds	0.67%	0.50%
US Bonds	0.08%	0.07%
Diferencia	0.59%	0.43%

Cuadro VII.5: Visión 3

	Π	μ
US Large Growth	6.41%	6.50%
US Large Value	4.08%	4.32%
Diferencia	2.33%	2.18%

Cuadro VII.6: Visión 3

	Π	μ
US Small Growth	7.43%	7.59%
US Small Value	3.70%	3.94%
Diferencia	3.73%	3.65%

4. Discusión de resultados Lo primero que se debe notar de los resultados obtenidos (VII.2) es el único activo sobre el cual no se tenía ninguna visión (International Emerging Equity) fue el que tuvo menor cambio de retorno esperado (-0.09%) y peso asignado (0.08%) al aplicarse el modelo de Black Litterman. Esto es consistente con lo predicho por la teoría (C.1.1), ya que si el inversionista no tiene opinión respecto al activo, el retorno del activo debería ser el retorno de equilibrio.

Luego es importante notar que los resultados obtenidos con el modelo de Black Litterman son consistentes con las visiones del inversionista. Por ejemplo, la primera visión decía que el retorno de International Developed Equity sería de 5.25% . El retorno predicho por el mercado es de 4.80% , mientras que el retorno predicho por Black Litterman es 4.93% , lo cual es más cercano a lo esperado por el inversionista. La visión dos decía que el retorno de International Bonds sería más alto que el de US Bond por 0.25% . Como se puede observar en (VII.4) la diferencia de retornos esperados más cercana a lo esperado por el inversionista es la predicha por el modelo de Black Litterman. Por último la visión tres decía que US Large Growth tendría un rendimiento mayor al de US Large Value por 2% , al igual que US Small Growth con US Small Value. Como se puede observar en las tablas (VII.5) y (VII.6) los resultados obtenidos con el modelo de Black Litterman también son más apegados a lo esperado por el inversionista.

B Ejemplo 2

Debido a que los datos utilizados en el Ejemplo 1 eran del 2005, se repitió el ejemplo investigando datos actualizados (del 2007 al 2016). Los pesos de capitalización de mercado obtenidos fueron:

Cuadro VII.7: Pesos de Capitalización de Mercado ω_m

Activo	Peso de Capitalización de Mercado ω_m
US Bonds	19.58%
Internacional Bonds	40.00%
US Large Growth	8.55%
US Large Value	6.64%
US Small Growth	0.22%
US Small Value	0.22%
International Developed Equity	15.23%
International Emerging Equity	9.55%

Las visiones del inversionista fueron:

1. El retorno en exceso de International Developed Equity será de 4.00%.
2. El retorno de Internacional Bonds será más alto que el de US Bonds por 1.25%.
3. El retorno de US Large Growth y US Small Growth será más alto que el de US Large Value y US Small Value por 2.50%

Las entradas de la matriz P se calcularon de la misma manera que en el Ejemplo 1. Como el portafolio largo está conformado por US Large Growth y US Small Growth, los pesos obtenidos fueron: para US Large Growth $\frac{8.55}{8.55+0.22} = 0.9751$ y para US Small Growth $\frac{0.22}{8.55+0.22} = 0.0249$. Por otro lado, el portafolio corto está conformado por US Large Value y US Small Value. Los pesos obtenidos fueron: para US Large Value $\frac{6.64}{6.64+0.22} = 0.9673$ y para US Small Value $\frac{0.22}{6.64+0.22} = 0.0327$.

1. Entradas del Programa Las entradas del programa fueron:

a La matriz de visiones P y el vector de las visiones del inversionista \mathbf{q} son:

$$P = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.9751 & -0.9673 & 0.0249 & -0.0327 & 0 & 0 \end{pmatrix} \quad (\text{VII.13})$$

$$\mathbf{q} = \begin{pmatrix} 0.04 \\ 0.0125 \\ 0.025 \end{pmatrix} \quad (\text{VII.14})$$

b La matriz de covarianza para estos fondos esta dada por

	US B.	Int. B.	US L. G	US L. V.	US S. G	US S. V.	Int. D. E.	Int. E. E.
US B.	0.00108	0.00017	0.00025	0.00017	0.00016	0.00011	0.00040	-0.00031
Int. B.	0.00017	0.00891	0.00228	0.00268	0.00252	0.00319	0.00619	0.000596
US L. G.	0.00025	0.00228	0.02144	0.02008	0.02701	0.02414	0.02186	-0.00013
US L. V.	0.00017	0.00268	0.02008	0.02113	0.02580	0.02475	0.02225	0.000450
US S. G.	0.00016	0.00252	0.02701	0.02580	0.04056	0.03560	0.02769	0.000585
US S. V.	0.00011	0.00319	0.02414	0.02475	0.03560	0.03441	0.02606	0.001151
Int. D. E.	0.00040	0.00619	0.02186	0.02225	0.02769	0.02606	0.02991	-0.00108
Int. E. E.	-0.0003	0.00060	-0.0001	0.00045	0.00058	0.00115	-0.00108	0.039647

c Se utilizaron los siguientes parámetros $\delta = 3.00$ y $\tau = 0.025$.

2. Resultados Los resultados obtenidos se resumen en la siguiente tabla:

Cuadro VII.8: Resumen de Resultados

Activo	ω_m	$\hat{\omega}$	$\hat{\omega} - \omega_m$	Π	μ
US Bonds	19.58%	18.39%	-1.19%	0.10%	0.15%
Internacional Bonds	40.00%	39.73%	-0.27%	1.49%	1.39%
US Large Growth	8.55%	182.01%	173.46%	2.27%	3.39%
US Large Value	6.64%	-165.80%	-172.44%	2.33%	2.17%
US Small Growth	0.22%	4.64%	4.42%	2.85%	4.01%
US Small Value	0.22%	-5.60%	-5.82%	2.77%	2.88%
International Developed Equity	15.23%	21.04%	5.81%	3.14%	3.47%
International Emerging Equity	9.55%	9.32%	-0.23%	1.15%	0.81%

Cuadro VII.9: Visión 1

	Π	μ
International Developed Equity	3.14%	3.47%

Cuadro VII.10: Visión 2

	Π	μ
International Bonds	1.49%	1.39%
US Bonds	0.10%	0.15%
Diferencia	1.39%	1.24%

Cuadro VII.11: Visión 3

	Π	μ
US Large Growth	2.27%	3.39%
US Large Value	2.33%	2.17%
Diferencia	-0.06%	1.22%

Cuadro VII.12: Visión 3

	Π	μ
US Small Growth	2.85%	4.01%
US Small Value	2.77%	2.88%
Diferencia	0.08%	1.13%

3. Discusión de resultados Al igual que en el Ejemplo 1 el único activo sobre el cual no se tenía ninguna visión (International Emerging Equity) fue el que tuvo menor cambio de retorno esperado (-0.23%) y peso asignado (0.34%) al aplicarse el modelo de Black Litterman. Esto es consistente con lo predicho por la teoría (C.1.1), ya que si el inversionista no tiene opinión respecto al activo, el retorno del activo debería ser el retorno de equilibrio.

Luego es importante notar que los resultados obtenidos con el modelo de Black Litterman son consistentes con las visiones del inversionista. Por ejemplo, la primera visión decía que el retorno de International Developed Equity sería de 4.00%. El retorno predicho por el mercado es de 3.14%, mientras que el retorno predicho por Black Litterman es 3.47%, lo cual es más cercano a lo esperado por el inversionista. La visión dos decía que el retorno de International Bonds sería más alto que el de US Bond por 1.25%. Como se puede observar en (VII.10) la diferencia de retornos esperados más cercana a lo esperado por el inversionista es la predicha por el modelo de Black Litterman. Además, la visión tres decía que US Large Growth tendría un rendimiento mayor al de US Large Value por 2.50%, al igual que US Small Growth con US Small Value. Como se puede observar en las tablas (VII.11) y (VII.12) los resultados obtenidos con el modelo de Black Litterman también son más apegados a lo esperado por el inversionista.” Sin embargo, es importante notar que se obtuvo un valor atípico para el peso de capitalización de mercado estimado de US Large Growth

(182.01%) y para US Large Value (−165.80%).

C Ejemplo 3

Suponga que el mercado consiste de los índices de patrimonio neto de siete países industriales: Australia, Canadá, Francia, Alemania, Japón, Reino Unido y Estados Unidos. Este ejemplo se diferencia del anterior, ya que en lugar de conocerse la covarianza entre los activos, se conoce su índice de correlación y las desviaciones estándar de cada uno. Por lo que se debe agregar un paso previo al procedimiento detallado en la sección anterior. Suponga además que se tienen las siguientes visiones:

1. El retorno de Alemania será mayor al de Francia y al de Reino Unido por 5%
2. El retorno de Canadá será mayor al de EEUU por 3%.

Ambas visiones son relativas. En la visión uno, como solo hay un activo en el portafolio largo de la visión 1, su entrada en la fila correspondiente a la primera visión es 1. Por otra parte, para encontrar los pesos del portafolio en corto se utiliza el mismo método utilizado en el ejemplo 1, obteniendo que el peso de Francia es −0.295 y Reino Unido es −0.705.

1. Entradas Las entradas para este ejemplo son:

a La matriz de visiones P y el vector de retornos esperados \mathbf{q} son:

$$P = \begin{bmatrix} 0 & 0 & -0.295 & 1 & 0 & -0.705 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

$$\mathbf{q} = \begin{pmatrix} 0.05 \\ 0.03 \end{pmatrix}$$

b La matriz de índices de correlación está dada por

	Australia	Canada	Francia	Alemania	Japon	RU	EEUU
Australia	1.0	0.488	0.478	0.515	0.439	0.512	0.491
Canada	0.488	1.0	0.664	0.655	0.31	0.608	0.779
Francia	0.478	0.664	1.0	0.861	0.355	0.783	0.668
Alemania	0.515	0.655	0.861	1.0	0.354	0.777	0.653
Japon	0.439	0.31	0.355	0.354	1.0	0.405	0.306
RU	0.512	0.608	0.783	0.777	0.405	1.0	0.652
EEUU	0.491	0.779	0.668	0.653	0.306	0.652	1

c Como parámetros se utilizaron $\delta = 2.5$ y $\tau = 0.05$.

2. Procedimiento La única diferencia con el procedimiento del Ejemplo 1 es que en este caso la matriz de covarianza Σ no era dada, en lugar se tenían los coeficientes de correlación y las desviaciones estándar de cada activo. Por tanto, lo primero que se realizó fue construir Σ sabiendo que:

$$cov(X, Y) = cor(X, Y)des(X)des(Y) \quad (VII.15)$$

Donde

- $cor(X, y)$ se refiere a la correlación entre X, Y .
- $cov(X, y)$ se refiere a la covarianza entre X, Y .
- $des(X)$ se refiere a la desviación estándar de X .

Por tanto, se obtuvo que Σ está dada por:

$$\begin{array}{|ccccccc|} \hline 0.02560 & 0.015850 & 0.01897 & 0.0223304 & 0.0147504 & 0.016384 & 0.01469072 \\ \hline 0.01585 & 0.041209 & 0.03343 & 0.0360335 & 0.0132153 & 0.024685 & 0.02957162 \\ \hline 0.01897 & 0.033428 & 0.06150 & 0.0578661 & 0.0184884 & 0.038837 & 0.03097917 \\ \hline 0.02233 & 0.036034 & 0.05787 & 0.0734410 & 0.0201461 & 0.042113 & 0.03309208 \\ \hline 0.01475 & 0.013215 & 0.01849 & 0.0201461 & 0.0441000 & 0.017010 & 0.01201662 \\ \hline 0.01638 & 0.024685 & 0.03884 & 0.0421134 & 0.0170100 & 0.040000 & 0.02438480 \\ \hline 0.01469 & 0.029572 & 0.03098 & 0.0330921 & 0.0120166 & 0.024385 & 0.03496900 \\ \hline \end{array}$$

3. Resultados

a Los pesos de equilibrio obtenidos por (VII.2) fueron:

$$\mathbf{\Pi} = \begin{pmatrix} 0.03937555 \\ 0.06915190 \\ 0.08358087 \\ 0.09027240 \\ 0.04302810 \\ 0.06767693 \\ 0.07560047 \end{pmatrix} \approx \begin{pmatrix} 3.94\% \\ 6.92\% \\ 8.35\% \\ 9.03\% \\ 4.30\% \\ 6.77\% \\ 7.56\% \end{pmatrix} \quad (VII.16)$$

b Luego, se obtuvo Ω (V.40):

$$\mathbf{\Omega} = \begin{bmatrix} 0.00106538 & 0.00010037 \\ 0.00010037 & 0.00085174 \end{bmatrix} \quad (VII.17)$$

c Se diagonalizó (VII.17):

$$\Omega = \begin{bmatrix} 0.00106538 & 0 \\ 0 & 0.00085174 \end{bmatrix} \quad (\text{VII.18})$$

d Luego, utilizando (VII.4) se obtuvieron los retornos esperados:

$$\boldsymbol{\mu} = \begin{pmatrix} 0.04422145 \\ 0.08729864 \\ 0.09479745 \\ 0.11209947 \\ 0.04616347 \\ 0.06971660 \\ 0.07481560 \end{pmatrix} \approx \begin{pmatrix} 4.42\% \\ 8.73\% \\ 9.48\% \\ 11.21\% \\ 4.62\% \\ 6.97\% \\ 7.48\% \end{pmatrix} \quad (\text{VII.19})$$

e Se calculó la matriz de covarianza posterior Σ_p :

$$\begin{vmatrix} 0.02684723 & 0.01657429 & 0.01983780 & 0.02328236 & 0.01547005 & 0.01718801 & 0.01539182 \\ 0.01657429 & 0.04299381 & 0.03494199 & 0.03753285 & 0.01383059 & 0.02589072 & 0.03107504 \\ 0.01983780 & 0.03494199 & 0.06439533 & 0.06036839 & 0.01937078 & 0.04074256 & 0.03244592 \\ 0.02328236 & 0.03753285 & 0.06036839 & 0.07627367 & 0.02106627 & 0.04414171 & 0.03454918 \\ 0.01547005 & 0.01383059 & 0.01937078 & 0.02106627 & 0.04629476 & 0.01785243 & 0.01260504 \\ 0.01718801 & 0.02589072 & 0.04074256 & 0.04414171 & 0.01785243 & 0.04199287 & 0.02558610 \\ 0.01539182 & 0.03107504 & 0.03244592 & 0.03454918 & 0.01260504 & 0.02558610 & 0.03661572 \end{vmatrix}$$

f Por último, se calcularon los nuevos pesos para el portafolio (VII.7):

$$\hat{\boldsymbol{w}} = \begin{pmatrix} 1.52\% \\ 41.86\% \\ -3.41\% \\ 33.58\% \\ 11.05\% \\ -8.17\% \\ 18.80\% \end{pmatrix} \quad (\text{VII.20})$$

Cuadro VII.13: Pesos de Capitalización de Mercado comparando con los Pesos obtenidos con BL

Activo	ω_m	$\hat{\omega}$
Australia	1.60%	1.52%
Canada	2.20%	41.86%
Francia	5.20%	-3.41%
Alemania	5.50%	33.58%
Japón	11.60%	11.05%
Reino Unido	12.40%	-8.17%
Estados Unidos	61.50%	18.80%

Cuadro VII.14: Resumen de Resultados

Activo	ω_m	$\hat{\omega}$	$\hat{\omega} - \omega_m$	Π	μ
Australia	1.60%	1.52%	-0.08%	3.94%	4.42%
Canadá	2.20%	41.86%	39.66%	6.92%	8.73%
Francia	5.20%	-3.41%	-8.61%	8.35%	9.48%
Alemania	5.50%	33.58%	28.08%	9.03%	11.21%
Japón	11.60%	11.05%	-0.55%	4.30%	4.62%
Reino Unido	12.40%	-8.17%	-20.57%	6.77%	6.97%
Estados Unidos	61.50%	18.80%	-42.70%	7.56%	7.48%

Cuadro VII.15: Visión 1

	Π	μ
Canadá	6.92%	8.73%
Estados Unidos	7.56%	7.48%
Diferencia	-0.64%	1.25%

Cuadro VII.16: Visión 2

	Π	μ
Alemania	9.03%	11.21%
Francia	8.35%	9.48%
Diferencia	0.68%	1.73%
Alemania	9.03%	11.21%
Reino Unido	6.77%	6.97%
Diferencia	2.26%	4.24%

4. Discusión de resultados Como se había mencionado en el ejemplo anterior, en la Tabla (VII.14) se puede ver que los pesos asignados por el modelo de Black Litterman de los activos que no poseen una visión por parte del inversionista (Australia y Japón) permanecen casi constantes. Por otro lado, se analizó que tanto se acercaban los retornos obtenidos con el modelo de Black Litterman a las visiones del inversionista. La primera visión decía que Alemania tendrá un retorno mayor a Francia y Reino Unido por 5%. Como se puede observar en la tabla (VII.16) el resultado obtenido con el modelo de Black Litterman es bastante más cercano al esperado por el inversionista (4.24%) que el predicho por el equilibrio (2.26%). La segunda visión decía que Canadá tendrá un retorno mayor a EEUU por 3%. A pesar que el resultado obtenido no es tan cercano (1.25%), se acerca más que el resultado de equilibrio (-0.64%).

VIII Conclusiones

1. Se cumplió con el objetivo principal de derivar el modelo de Black Litterman desde los enfoques de Inferencia Bayesiana, Optimización Convexa y Teoría de Muestreo; encontrando que la perspectiva de Bayes es la más adecuada para la implementación, la de Optimización Convexa para la generalización y refinamiento del modelo y la de Teoría de Muestreo para el mejor entendimiento de los parámetros del modelo.
2. El modelo de Black Litterman a diferencia del modelo tradicional de Markowitz, permite incorporar las visiones personales de retorno y riesgo que pueden llegar a tener los inversionistas en el proceso de optimización de portafolios.
3. El enfoque de Bayes aplicado a la optimización de portafolios permite emplear información a priori útil sobre cantidades de interés (en el caso de Black Litterman las visiones del inversionista). Además, permite estimar riesgo e incertidumbre del modelo y se presta a el uso algoritmos numéricos rápidos e intuitivos para simular cantidades económicas complejas.
4. El enfoque de Optimización Convexa elimina la necesidad de un modelo estadístico, lo cual evita hacer supuestos estadísticos que no siempre se cumplen.
5. Los resultados obtenidos con el enfoque de Optimización Convexa pueden ser utilizados como base para refinar el modelo de Black Litterman, de manera que el inversionista pueda incorporar sus visiones acerca de la volatilidad y dinámicas del mercado.
6. La ventaja que presenta el enfoque de Teoría de Muestreo sobre los otros enfoques es la interpretación intuitiva de algunos de los parámetros, por ejemplo τ . Una de las principales desventajas es que no es tan práctico para la implementación, debido a que se necesita una cantidad considerable de muestras.
7. El modelo de Black Litterman bajo la perspectiva de Teoría de Muestreo presenta una inconsistencia, ya que la distribución de \mathbf{q}_i debería de ser $N(P\boldsymbol{\mu}, P^T\Sigma P)$, en lugar de $N(P\boldsymbol{\mu}, \Omega)$.
8. Si se quisiera refinar la implementación del modelo de Black Litterman se podría ajustar el parámetro τ respecto a las características anticipadas de riesgo-retorno y trabajar la seguridad de los inversionistas acerca de sus visiones. También se podría utilizar una medida de riesgo más general, como lo es VaR o CVaR.
9. Se obtuvieron los resultados esperados al implementar el modelo de Black Litterman con datos numéricos obtenidos en (Idzorek, 2005) y (He, Litterman, 1999).

IX Bibliografía

- [1] Avramov, D., Zhou, G. (2009) *Bayesian Portfolio Analysis*. American Economic Association G11, G12, C11.
- [2] Bertsekas, D. (1998). *Nonlinear Programming* Segunda edición. Athena Scientific, Belmont, Massachusetts. Massachusetts, Estados Unidos.
- [3] Bertsimas, D., Gupta, V., Paschalidis, I. (2012). *Inverse Optimization: A New Perspective on the Black-Litterman Model* Operations Research Vol. 60, No.6 pp.1389-1403
- [4] Christodoulakes, G. (2002). *Bayesian Optimal Portfolio Selection: the Black-Litterman Approach* Quantitative Asset Pricing notes, City University. Londres, Reino Unido.
- [5] Boyd, S., Vandenberghe, L. (2004). *Convex Optimization* Séptima Edición. Cambridge University Press. Cambridge, Reino Unido.
- [6] Delbaen, F. (2000) *Coherent Risk Measures* ETH. Zurich, Suiza.
- [7] Gelman, A. (2014). *Bayesian Data Analysis*. Tercera edición. CRC Press. Nueva York, Estados Unidos.
- [8] He, G., Litterman, R. (1999) *The Intuition behind Black-Litterman Model Portfolios* Goldman Sachs Asset Management. Nueva York, Estados Unidos
- [9] Idzorek, T. (2005). *A Step-by-Step Guide to the Black-Litterman Model: Incorporating user-specified confidence levels*. Ibbotson Associates. Chigago Illinois, Estados Unidos.
- [10] Kreyszig, E. (1970). *Introductory Mathematical Statistics: Principles and Methods* Editorial John Wiley and Sons, inc. Nueva York, Estados Unidos
- [11] Mankert, C., Seiler, M. (2011). *Mathematical Derivations and Practical Implications for the use of the Black-Litterman Model* Journal of Real Estate Portfolio Management.

X Apéndice

A Apéndice para Inferencia Bayesiana

Teorema A.1. *Propiedad de matrices inversas*

Sean $G \in R^{m \times m}$, $H \in R^{m \times n}$ y $F \in R^{n \times n}$ entonces:

$$(G + HFH^T)^{-1} = G^{-1} - G^{-1}H(H^T G^{-1}H + F^{-1})H^T G^{-1} \quad (\text{X.1})$$

Demostración.

Sea M una matriz $n \times n$, la cual se puede escribir como $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$. Donde A es $p \times p$, D es $q \times q$, B es $p \times q$, C es $q \times p$ y $p + q = n$. Si se tiene el sistema:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{c} \\ \mathbf{d} \end{pmatrix} \quad (\text{X.2})$$

Resolviendo por complementos de Schur (II.69 y II.71) obtenemos:

$$\mathbf{x} = (A - BD^{-1}C)^{-1}(\mathbf{c} - BD^{-1}\mathbf{d}) \quad (\text{X.3})$$

$$\mathbf{y} = D^{-1}(\mathbf{d} - C(A - BD^{-1}C)^{-1}(\mathbf{c} - BD^{-1}\mathbf{d})) \quad (\text{X.4})$$

Modificando X.3 y X.4 se obtiene:

$$\mathbf{x} = (A - BD^{-1}C)^{-1}\mathbf{c} - (A - BD^{-1}C)^{-1}BD^{-1}\mathbf{d} \quad (\text{X.5})$$

$$\mathbf{y} = -D^{-1}C(A - BD^{-1}C)^{-1}\mathbf{c} + (D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1})\mathbf{d} \quad (\text{X.6})$$

Entonces como:

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = M^{-1} \begin{pmatrix} \mathbf{c} \\ \mathbf{d} \end{pmatrix} \quad (\text{X.7})$$

Se tiene que:

$$M^{-1} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & (D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1}) \end{pmatrix} \quad (\text{X.8})$$

La ecuación X.8 se puede reescribir como:

$$\begin{aligned} \begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} &= \begin{pmatrix} (A - BD^{-1}C)^{-1} & 0 \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} \end{pmatrix} \begin{pmatrix} I & -BD^{-1} \\ 0 & I \end{pmatrix} \\ \begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} &= \begin{pmatrix} I & 0 \\ -D^{-1}C & I \end{pmatrix} \begin{pmatrix} (A - BD^{-1}C)^{-1} & 0 \\ 0 & D^{-1} \end{pmatrix} \begin{pmatrix} I & -BD^{-1} \\ 0 & I \end{pmatrix} \end{aligned} \quad (\text{X.9})$$

Por tanto,

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} I & BD^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} I & 0 \\ D^{-1}C & I \end{pmatrix} \quad (\text{X.10})$$

Además, podemos usar el complemento de Schur $D - CA^{-1}B$ de A para obtener la siguiente factorización de M :

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} I & 0 \\ CA^{-1} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{pmatrix} \begin{pmatrix} I & A^{-1}B \\ 0 & I \end{pmatrix} \quad (\text{X.11})$$

Si $D - CA^{-1}B$ es invertible, entonces al invertir las tres matrices de X.11 se obtiene:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{pmatrix} \quad (\text{X.12})$$

Comparando las ecuaciones X.8 y X.12 se obtiene la siguiente identidad:

$$(A - BD^{-1}C)^{-1} = A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} \quad (\text{X.13})$$

Ahora, al hacer las sustituciones $A = G$, $B = -H$, $D = F^{-1}$ y $C = H^T$ se obtiene la identidad deseada:

$$(G + HFH^T)^{-1} = G^{-1} - G^{-1}H(H^TG^{-1}H + F^{-1})H^TG^{-1} \quad (\text{X.14})$$

Teorema A.2. *Equivalencia de la fórmula de Black Litterman*

Se mostrará que las ecuaciones de Black Litterman X.15 y X.16 son equivalentes.

$$E(\mathbf{r}) = [(\tau\Sigma)^{-1} + P^T\Omega^{-1}P]^{-1} [(\tau\Sigma)^{-1}\mathbf{\Pi} + P^T\Omega^{-1}\mathbf{q}] \quad (\text{X.15})$$

$$E(\mathbf{r}) = \mathbf{\Pi} + (\tau\Sigma)P^T[P(\tau\Sigma)P^T + \Omega]^{-1}(\mathbf{q} - P\mathbf{\Pi}) \quad (\text{X.16})$$

Demostración.

Primero se modifica X.15 utilizando X.1

$$\begin{aligned}
E[\mathbf{r}] &= \left[(\tau\Sigma) - (\tau\Sigma)P^T [P(\tau\Sigma)P^T + \Omega]^{-1} P(\tau\Sigma) \right] [(\tau\Sigma)^{-1}\mathbf{\Pi} + P^T\Omega^{-1}\mathbf{q}] \\
&= \mathbf{\Pi} - (\tau\Sigma)P^T [P(\tau\Sigma)P^T + \Omega]^{-1} P\mathbf{\Pi} + (\tau\Sigma)P^T\Omega^{-1}\mathbf{q} - (\tau\Sigma)P^T [P\Sigma P^T + \Omega]^{-1} P(\tau\Sigma)P^T\Omega^{-1}\mathbf{q} \\
&= \mathbf{\Pi} - (\tau\Sigma)P^T [P(\tau\Sigma)P^T + \Omega]^{-1} P\mathbf{\Pi} + (\tau\Sigma)P^T [I - (P\Sigma P^T + \Omega)^{-1}P(\tau\Sigma)P^T] \Omega^{-1}\mathbf{q} \\
&= \mathbf{\Pi} - (\tau\Sigma)P^T [P(\tau\Sigma)P^T + \Omega]^{-1} P\mathbf{\Pi} + \\
&\quad (\tau\Sigma)P^T [(P(\tau\Sigma)P^T + \Omega)^{-1}(P(\tau\Sigma)P^T + \Omega) - (P\Sigma P^T + \Omega)^{-1}P(\tau\Sigma)P^T] \Omega^{-1}\mathbf{q} \\
&= \mathbf{\Pi} - (\tau\Sigma)P^T [P(\tau\Sigma)P^T + \Omega]^{-1} P\mathbf{\Pi} + (\tau\Sigma)P^T [(P(\tau\Sigma)P^T + \Omega)^{-1}(P(\tau\Sigma)P^T + \Omega - P(\tau\Sigma)P^T)] \Omega^{-1}\mathbf{q} \\
&= \mathbf{\Pi} - (\tau\Sigma)P^T [P(\tau\Sigma)P^T + \Omega]^{-1} P\mathbf{\Pi} + (\tau\Sigma)P^T [P(\tau\Sigma)P^T + \Omega]^{-1} \Omega\Omega^{-1}\mathbf{q} \\
&= \mathbf{\Pi} + (\tau\Sigma)P^T [P(\tau\Sigma)P^T + \Omega]^{-1}(\mathbf{q} - P\mathbf{\Pi})
\end{aligned} \tag{X.17}$$

Por tanto, se obtiene:

$$E[\mathbf{r}] = \mathbf{\Pi} + (\tau\Sigma)P^T [P(\tau\Sigma)P^T + \Omega]^{-1}(\mathbf{q} - P\mathbf{\Pi}) \tag{X.18}$$

B Apéndice para Optimización Convexa**Definición B.1.** *Medidas Coherentes de Riesgo*

Una medida coherente de riesgo es una medida M (A.2) que posee las siguientes propiedades:

1. Monotonicidad
2. Homogeneidad positiva
3. Subaditividad
4. Invarianza

Definición B.2. *Monotonicidad*

Sean X, Y activos y sean r_X, r_Y retornos aleatorios de estos activos, tal que $r_X \leq r_Y$, entonces:

$$M(r_Y) \leq M(r_X) \tag{X.19}$$

Definición B.3. *Homogeneidad positiva*

Sea X un portafolio y r_X su retorno. Si el retorno del portafolio se incrementa por un factor $d \geq 0$, entonces el riesgo del portafolio también se incrementa por ese factor d :

$$M(dr_X) = dM(r_X) \quad (\text{X.20})$$

Definición B.4. *Sub-aditividad*

Sean X, Y activos, λ el peso del activo X y $(1 - \lambda)$ el peso del activo Y , entonces el riesgo del portafolio formado por X, Y no es más grande que la suma de los riesgos de los retornos aleatorios.

$$M[\lambda r_X + (1 - \lambda)r_Y] \leq M[\lambda r_X] + M[(1 - \lambda)r_Y] \quad (\text{X.21})$$

Nota.

Utilizando la propiedad de homogeneidad y la propiedad de sub-aditividad se obtiene la propiedad de Convexidad:

$$M[\lambda r_X + (1 - \lambda)r_Y] \leq \lambda M[r_X] + (1 - \lambda)M[r_Y] \quad (\text{X.22})$$

Definición B.5. *Invarianza*

Si se agrega un activo libre de riesgo al portafolio, entonces el riesgo del portafolio deberá reducirse.

Definición B.6. *α -cuantil*

Sea X una variable aleatoria y $\alpha \in [0, 1]$ q es un α -cuantil si:

$$P[X < q] \leq \alpha \leq P[X \leq q] \quad (\text{X.23})$$

El mayor α -cuantil es:

$$q_\alpha(X) = \inf\{x | P[X \leq x] \geq \alpha\} \quad (\text{X.24})$$

El menor α -cuantil es:

$$q_\alpha^-(X) = \inf\{x | P[X \leq x] \geq \alpha\} \quad (\text{X.25})$$

q es un α -cuantil ssi $q_\alpha^- \leq q \leq q_\alpha$

Definición B.7. *VaR (Valor en Riesgo)*

VaR se define como la pérdida máxima que con algún porcentaje de certeza no será excedida en un período de tiempo delimitado. Matemáticamente VaR se define como:

Sea X una posición y $\alpha \in [0, 1]$, entonces se define:

$$VaR_\alpha(X) = -q_\alpha(X) \tag{X.26}$$

X se denomina VaR-aceptable si $VaR_\alpha(X) \leq 0$ o equivalentemente si $q_\alpha \geq 0$

Nota.

Si se fija una probabilidad $\alpha \in [0, 1]$, una posición X se define como VaR-aceptable ssi la probabilidad de irse a la banca rota es menor a α .

Nota.

VaR no es una medida de riesgo coherente, ya que no cumple la sub-aditividad.

Definición B.8. *CVaR (Valor en Riesgo Condicional)*

CVaR se define como el tamaño promedio de la pérdida que se puede esperar cuando se excede el nivel de pérdida del VaR. Matemáticamente CVaR se define de la siguiente manera:

Sea X una variable aleatoria continua que representa las pérdidas y $\alpha \in (0, 1)$, entonces $CVaR_\alpha$ de X es:

$$CVaR_\alpha(X) = E[X|X \geq VaR_\alpha(X)] \tag{X.27}$$

Nota.

CVaR si es una medida de riesgo coherente.
