



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

EL PRINCIPIO DE MÁXIMA ENTROPÍA
APLICADO AL ANÁLISIS DEL RIESGO
SISTÉMICO EN MÉXICO

T E S I S

QUE PARA OBTENER EL TÍTULO DE:
MATEMÁTICO

PRESENTA:
MARCO ANTONIO AVILA PONCE DE LEÓN

DIRECTOR DE TESIS:
DR. PABLO PADILLA LONGORIA



2012



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

<p>1. Datos del alumno Avila Ponce de León Marco Antonio 40 80 13 30 9 Universidad Nacional Autónoma de México Facultad de Ciencias Matemáticas</p>
<p>2. Datos del tutor Dr. Miguel Ángel Segoviano Basurto</p>
<p>3. Datos del sinodal 1 Dra. María del Rocío Raquel Elizondo Camejo</p>
<p>4. Datos del sinodal 2 Dr. Pablo Padilla Longoria</p>
<p>5. Datos del sinodal 3 M. en C. Jorge Humberto Del Castillo Spíndola</p>
<p>6. Datos del sinodal 4 Dr. Fernando Baltazar Larios</p>
<p>7. Datos del trabajo escrito El Principio de Máxima Entropía Aplicado al Análisis del Riesgo Sistémico en México. 104p 2012</p>

A mis padres, a mis hermanos.

Índice general

Introducción	IX
1. Una Definición del Riesgo Sistemico	1
1.1. Riesgo Sistemico	2
1.2. ¿Qué Aspecto del Riesgo Sistemico Mediremos?	4
2. Fundamentos Teóricos	9
2.1. Value at Risk y Expected Shortfall	9
2.1.1. Preliminares para el Value at Risk y Expected Shortfall	10
2.1.2. Definición: VaR y Expected Shortfall	11
2.2. Shapley Value	20
2.2.1. El Shapley Value de una Institución Financiera	20
2.2.2. Definición Matemática	21
2.2.3. Algunas Propiedades	22
2.2.4. Un Ejemplo Concreto	23
2.3. Cópulas y Correlación Spearman	25
2.3.1. Cópulas	25
2.3.2. Correlación Spearman	28
3. Entropía y CIMDO	31
3.1. La Teoría de la Probabilidad: Una Extensión de la Lógica	32
3.2. Entropía	35
3.2.1. Bayesianismo Objetivo	35
3.2.2. Derivaciones de la Entropía	36
3.3. El Principio de Máxima Entropía	43
3.3.1. Planteamiento de la Metodología	43
3.3.2. Una Propiedad de las Distribuciones que Maximizan la Entropía	47

3.4.	El Principio de Mínima Entropía Cruzada	49
3.4.1.	Planteamiento de la Metodología	49
3.4.2.	Caracterización Axiomática de la Entropía Cruzada: Caso Discreto	51
3.4.3.	Intuición: Principio de Mínima Entropía Cruzada . . .	60
3.5.	CIMDO	60
3.5.1.	Función Objetivo	61
3.5.2.	La Distribución A-priori	61
3.5.3.	Ecuaciones de Consistencia con la Información	62
3.5.4.	Umbral de Distress	62
3.5.5.	Planteamiento de la Metodología	63
3.5.6.	Intuición CIMDO	64
3.6.	CIMDO-Cópula	65
4.	El Sistema Financiero Mexicano	67
4.1.	Un Análisis del Sistema Mexicano	67
4.1.1.	Tamaño	67
4.1.2.	Correlación	68
4.1.3.	Probabilidad de Distress	69
4.2.	Estimación de la Importancia Sistémica	70
4.2.1.	Proceso de Estimación	70
4.2.2.	Resultados	71
5.	Conclusiones	85
5.1.	Importancia Sistémica Alta	85
5.2.	Importancia Sistémica Mediana	86
5.3.	Importancia Sistémica Baja	86
A.	Derivación: Solución Única CIMDO	87
B.	Propiedades del Shapley Value	89
	Bibliografía	93

Introducción

A raíz de la crisis del año 2008 surgió la necesidad de implementar un marco de supervisión macro-prudencial que permita la regulación del sistema financiero como un objeto riesgoso en sí mismo.

En este sentido uno de los principales objetivos de los reguladores contemporáneos es la apropiada medición del nivel de riesgo sistémico de un sistema financiero y la identificación de los principales contribuidores al mismo.

El riesgo sistémico de un sistema financiero depende principalmente del estado de tres factores de un sistema:

1. La granuralidad del sistema financiero.
2. El nivel de interdependencia entre las instituciones que la conforman.
3. El nivel (y plausibilidad) de las pérdidas en las que podría incurrir un sistema financiero, una vez haya sufrido una crisis.

Surge entonces la necesidad de encontrar maneras de modelar matemáticamente el comportamiento de dichas propiedades. Posteriormente necesitaremos utilizar dichos modelos para encontrar maneras de medir el nivel de riesgo sistémico y luego distribuir dicho nivel a las instituciones de acuerdo a sus respectivas contribuciones.

El objetivo del presente trabajo es por un lado desarrollar un marco metodológico de modelación de un sistema financiero y de medición de su riesgo sistémico así como justificar el uso de dicha metodología mediante la exposición de la teoría que se encuentra detrás de la misma y por otro lado aplicar la metodología a el caso del sistema financiero mexicano y analizar los resultados.

La organización es la siguiente:

1. En el capítulo 1 se presenta una definición general del concepto de riesgo sistémico. Esto con el fin de tener una definición con la cual podamos

empezar a trabajar al identificar los diversos aspectos que conforman al concepto.

2. En el capítulo 2 se presentan con detalle las herramientas que se utilizarán para medir el nivel de riesgo sistémico de un sistema financiero y la contribución de cada institución a la misma una vez se haya inferido la distribución de pérdidas del sistema.
3. En el capítulo 3 se aborda el problema de la inferencia de la distribución multivariada de pérdidas del sistema financiero, se presenta a Consistent Information Multivariate Density Optimization Methodology ¹ como una metodología sistemática de inferencia de dicha distribución y se justifica su uso exponiendo la teoría en la cual está basada.
4. Finalmente, en el capítulo 4 se presentan y analizan los resultados procedentes de una simulación de pérdidas del sistema basada en la distribución inferida por CIMDO, las cuales se utilizan como insumos para estimar el nivel de riesgo sistémico del sistema financiero y la contribución de cada institución a la misma.

¹En corto CIMDO.

Capítulo 1

Una Definición del Riesgo Sistémico

El concepto de riesgo sistémico, en un sentido general, no es para nada un fenómeno limitado a la economía o a un sistema financiero; tal vez la ilustración mas natural está en el área de salud y las enfermedades que provocan epidemias. En casos severos mediante un sistema de contagio, una enfermedad epidémica es capaz de eliminar una gran parte de cierta población. En el área de la economía se ha defendido el punto de que el riesgo sistémico es una característica de los sistemas financieros. Mientras que los efectos de contaminación ocurren en otros sectores de la economía, la plausibilidad y severidad de dichos efectos en el sistema financiero es considerablemente más alta, una crisis sistémica en el sistema financiero podría tener serias consecuencias adversas para la salud de la economía en general.

Como veremos más adelante el concepto de riesgo sistémico consta de varios elementos y por lo general los estudios en la literatura existente sobre el tema, se han encargado de medir ciertos aspectos del mismo. Nosotros nos daremos la tarea de medir la importancia sistémica de cada una de las instituciones que forman a un sistema financiero. El objetivo de este capítulo es el de proporcionar un marco teórico para el análisis económico del riesgo sistémico. Empezaremos proponiendo una terminología específica, aclarando los elementos importantes del concepto de riesgo sistémico y concluyendo con una definición general del mismo. Luego especificaremos claramente cuáles son las facetas del riesgo sistémico que pretendemos medir con las métricas que definiremos posteriormente.

1.1. Riesgo Sistémico

Con el fin de llegar a una definición del riesgo sistémico en un sistema financiero, aclararemos primero los conceptos necesarios para su entendimiento, los cuales se muestran en la tabla al final del capítulo. Definiremos a un evento sistémico en un **sentido estrecho** de la siguiente manera:

Definición 1.1.1. Un evento sistémico en el sentido estrecho sucede cuando el anuncio de ciertas «malas noticias» sobre una institución financiera, su quiebra, o el descalabro de un mercado financiero, conduzca a considerables efectos adversos sobre una o varias instituciones financieras o mercados.

La región sombreada en la tabla engloba los eventos sistémicos en el sentido estrecho. Lo esencial de la definición anterior es el «efecto dominó» de una institución a otra ó de un mercado financiero a otro. Provocado por un limitado (idiosincrático)¹ shock inicial. Un evento sistémico (los cuales están tachados en la región sombreada y la no-sombreada de la tabla al final del capítulo en el **sentido amplio** se define como sigue:

Definición 1.1.2. Un evento sistémico en el sentido amplio incluye los eventos sistémicos en el sentido estrecho y los eventos en los cuales ocurren efectos adversos simultáneos sobre una considerable cantidad de instituciones financieras o mercados. Provocados por shocks severos y generalizados (sistemáticos)².

Un evento sistémico en el sentido estrecho es fuerte si la institución o las instituciones perjudicadas en los efectos secundarios entran en quiebra como consecuencia del shock inicial a pesar de ser solventes anteriormente, o si el mercado o los mercados financieros se descalabran posterior al shock inicial y no lo hubieran hecho de no haber sido por el mismo. Llamaremos estas instancias fuertes de los eventos sistémicos en el sentido estrecho como contagio. De lo contrario, es decir si, los efectos del shock inicial no llevan a la quiebra o descalabro, llamaremos al evento sistémico en el sentido estrecho como débil. Análogamente, los eventos sistémicos relacionados a los shocks sistemáticos son fuertes si una considerable parte del sistema financiero o mercados afectados por ellos terminan entrando en quiebra o en descalabros

¹Con idiosincrático nos referimos a un shock que afecta a un banco o a un grupo muy pequeño de bancos.

²Con sistemático nos referimos a un shock que afecta a un grupo grande de bancos.

y débiles si no lo hacen. Basados en esta terminología, una crisis sistémica (en el sentido estrecho y amplio) se puede definir como sigue:

Definición 1.1.3. Una crisis sistémica es un evento sistémico que afecta una cantidad considerable de instituciones financieras o mercados en el sentido fuerte, causando estragos en el buen funcionamiento general (de una parte importante) del sistema financiero.

La distinción entre los sentidos estrecho y amplio del concepto de riesgo sistémico es importante, ya que las medidas de supervisión macro-prudencial, concentradas en atacar la raíz del problema podrían ser diferentes en el caso de un shock idiosincrático que podría contagiarse, comparado con el caso de un shock sistemático que podría tener un efecto de desequilibrio más amplio. Sin embargo, en situaciones prácticas de crisis tanto los shocks agregados como las quiebras causadas por un contagio a veces pueden sobreponerse, ya que las debacles macroeconómicas podrían debilitar a las instituciones financieras, haciendo el posible contagio de una institución financiera más plausible. De igual manera podría ser necesario que el shock inicial lastime a más de una institución para que los efectos de contagio se materialicen. El riesgo sistémico se puede definir entonces de la siguiente manera:

Definición 1.1.4. El riesgo sistémico de un sistema financiero es el riesgo de sufrir eventos sistémicos en el sentido fuerte.

En principio, el espectro del riesgo sistémico va desde el contagio a una institución o mercado hasta el riesgo de tener una crisis sistemática afectando a la mayor parte (incluso todo) del sistema financiero. El elemento clave en esta definición del riesgo sistémico, es decir, el evento sistémico, se compone de dos elementos importantes, shocks y mecanismos de propagación. Siguiendo la terminología de la teoría financiera, los shocks iniciales pueden ser idiosincráticos o sistemáticos. En el extremo, los shocks idiosincráticos son aquellos que inicialmente, afectan la salud de solamente una institución financiera o solamente el precio de un activo. Mientras que los shocks sistemáticos en el caso extremo afectan a toda la economía. Es decir, a todas las instituciones del sistema financiero al mismo tiempo. Un ejemplo de un shock idiosincrático a un sistema financiero nacional es la quiebra de una institución financiera por fraude interno. La repentina devaluación de una moneda internacional causada por un déficit del presupuesto doméstico puede ser visto como shock idiosincrático al sistema financiero mundial. Los shocks sistemáticos al sistema financiero nacional pueden ser, por ejemplo, la repentina alza en la tasa

de inflación. El descalabro bursátil de un mercado financiero en sí mismo actúa como un shock sistemático sobre varias instituciones financieras. Aunque, dadas las distintas exposiciones de las instituciones, las afecciones probablemente no serán uniformes. Por supuesto que existe un continuo de tipos de shock intermedios (por sector o por región) entre los extremos teóricos de los shocks idiosincráticos y los shocks sistemáticos. Los shocks idiosincráticos que no se propagan ampliamente son «asegurables», en el sentido que un inversionista se podría proteger contra ellos vía la diversificación de su portafolio, mientras que los shocks sistemáticos frecuentemente son «inasegurables», en el sentido que la diversificación del portafolio de un inversionista no hace nada para contrarrestar los efectos del shock. El segundo elemento clave en el sentido estrecho del riesgo sistémico es el mecanismo por el cual los shocks se propagan de una institución financiera o mercado a otro. La propagación de los shocks son en cierto sentido la parte más importante del concepto de riesgo sistémico y por lo tanto es importante definir que tipos de mecanismos podrían considerarse apropiados a la definición anterior de riesgo sistémico. Esto se desvía un poco del tema de la presente tesis. Referiremos al lector a consultar [16] y [17], quedándonos satisfechos con la definición general del concepto de riesgo sistémico.

1.2. ¿Qué Aspecto del Riesgo Sistémico Mediremos?

Visualizaremos al sistema financiero como un portafolio de instituciones financieras al cual le asociaremos o inferiremos una distribución multivariada de pérdidas. Donde cada marginal será la distribución de pérdidas de alguna de las instituciones que componen a dicho portafolio. Antes de poder hablar del nivel de riesgo sistémico, es necesario abordar el problema de proponer una metodología de inferencia sistemática y consistente de una distribución multivariada de pérdidas. Supongamos por el momento que tenemos en nuestras manos dicha distribución. Cuando el sistema financiero acaba de sufrir una crisis, independientemente de las razones por las cuales surgió, lo primero que se preguntan las autoridades es: ¿qué tan severa es la crisis?. O planteado de otra manera: ¿cuánto dinero se perdió por causa de la crisis?. En este sentido para medir el riesgo sistémico de un sistema financiero resulta natural concentrarse en el nivel de pérdidas a las cuales está sujeto dado que

ocurrió un evento sistémico. Resulta entonces natural que nuestra medida de nivel de riesgo sistémico sea el «*Value at Risk*» o «*Expected Shortfall*»³ a un cierto nivel de confianza del total de pérdidas de un sistema financiero.

Ahora, supongamos que ya calculamos el nivel de riesgo sistémico del sistema financiero. En la definición anterior de nivel de riesgo sistémico nos concentramos solamente en el nivel de pérdidas del sistema financiero como un portafolio de instituciones y no en la contribución de cada una de las instituciones a dicho nivel de pérdidas. Es evidente que el debacle de algunas instituciones, provocado ya sea por el shock inicial o por el sistema de contagio contribuirá en mayor medida que el debacle de otras. Es decir, lo que conoceremos posteriormente como importancia sistémica de algunas instituciones resulta ser mayor que la de otras. En este sentido, dado un cierto nivel de riesgo sistémico, para poder medir la importancia sistémica de las instituciones es necesario encontrar una manera de repartir justamente dicho nivel en pedazos correspondientes a cada institución de acuerdo a su contribución. Afortunadamente, ya existe una solución a este problema, Lloyd Shapley [18], en el contexto de los juegos cooperativos desarrolló una metodología la cual permite repartir justamente a cada institución su pedazo del pastel o puesto en el contexto en el que estamos, su pedazo del nivel de riesgo sistémico. La importancia sistémica de cada institución será entonces su correspondiente «*Shapley Value*»⁴. En efecto, en la presente tesis nos concentraremos en medir el nivel de riesgo sistémico y en calcular la importancia sistémica de cada institución financiera. Es importante notar en este momento que la importancia sistémica de cada institución engloba el estado de varios factores del banco en el sistema financiero. Captura principalmente lo siguiente:

1. La salud general de la institución financiera.
2. El tamaño de la institución como parte del sistema financiero.
3. La dependencia en el nivel de pérdidas de la institución financiera con respecto a las operaciones de las demás instituciones.

La importancia sistémica de una institución financiera mide tanto el nivel de toxicidad de la institución como el nivel de vulnerabilidad a ser contagiado por alguna institución y entonces formar parte de un evento sistémico. Es

³El «*Value at Risk*» y el «*Expected Shortfall*» son medidas de riesgo. Ver Capítulo 2.

⁴Término utilizado comúnmente en la literatura para referirse al valor que la metodología de Shapley le da a un cierto jugador.

decir, la contribución de la institución financiera en todos los aspectos al riesgo sistémico del sistema.

Tipo de shock inicial	Eventos Sistemicos Singulares (afectan solamente a una institución o a un mercado en el efecto de segundo orden)	Eventos Sistemicos Extensos (afectan a varias instituciones o mercados en el efecto de segundo orden)
-Shock leve que se propaga	Débil (no hay quiebra o caída)	Débil (no hay quiebra o caída)
-Shock Idiosincrático	Fuerte (quiebra de una institución)	Fuerte (quiebra de varias instituciones)
-Shock sistemático limitado	x contagio	x
-Shock sistemático extenso	x	x

contagio que lleva a una crisis sistemática

contagio que lleva a una crisis sistemática

crisis sistemática

Capítulo 2

Fundamentos Teóricos

La metodología que utilizaremos para medir la importancia sistémica y el nivel de riesgo sistémico consta de dos partes: la inferencia de la distribución de pérdidas del sistema financiero y la utilización de algunas herramientas matemáticas, las cuales tomando como insumo las pérdidas simuladas del sistema financiero arrojan las medidas de importancia sistémica y el nivel de riesgo sistémico en el sistema. En este capítulo, se presentarán con detalle las herramientas matemáticas que utilizaremos. En la primera sección desarrollaremos el concepto de «*Value at Risk*» y «*Expected Shortfall*» conceptos que se utilizarán para medir el nivel de riesgo sistémico en el sistema financiero. En la segunda sección presentaremos (en el contexto en el que lo utilizaremos) el *Shapley Value* de una institución financiera, la cual mide su importancia sistémica. Finalmente, introduciremos brevemente el concepto de cópulas y su relación con la correlación Spearman. Esto con el fin de poder presentar posteriormente la noción de CIMDO-Cópula, la cual es nuestro modelo de la dependencia que existe entre las instituciones del sistema financiero.

2.1. Value at Risk y Expected Shortfall

Las medidas de riesgo juegan un papel importante en problemas de optimización bajo incertidumbre, especialmente tratándose de pérdidas en las que podrían incurrir las instituciones financieras. Las pérdidas se pueden ver como una función $z = f(x, y)$ de un vector de decisión $x \in \mathbf{X} \subset \mathbb{R}^n$, el cual representa lo que generalmente llamamos un portafolio, con \mathbf{X} representando el espacio de las posibles decisiones que podría tomar una institución finan-

ciera y con y representando los futuros valores de ciertas variables económicas que afectan el comportamiento del portafolio de la institución como tasas de interés o el precio de un subyacente, etc. Si uno supone que y obedece una cierta distribución de probabilidad entonces z termina siendo una variable aleatoria con su distribución dependiente del vector de decisión x . Cualquier problema de optimización que involucre a z debería tomar en cuenta, no solo el valor esperado de las pérdidas al tomar la decisión x , sino también el riesgo que implica haber tomado esta decisión. En esta sección definiremos dos medidas de riesgo comúnmente usadas en el sector financiero: *Value at Risk*¹ y *Expected Shortfall*². En el camino enunciaremos algunas propiedades deseables que solamente cumple el *Expected Shortfall* con el fin de justificar el uso de éste sobre el *Value at Risk*.

2.1.1. Preliminares para el Value at Risk y Expected Shortfall

Espacio de Probabilidad

Sea Ω un **Espacio muestral** formado por todos los posibles resultados de un experimento aleatorio (por ejemplo, todas las posibles trayectorias del precio de un activo subyacente a lo largo del tiempo). Un subconjunto de Ω es un evento donde $\omega \in \Omega$ es un punto muestral.

Definición 2.1.1. Sea Ω un conjunto no vacío y \mathcal{F} una colección de subconjuntos de Ω . \mathcal{F} es una σ -álgebra si:

1. $\Omega \in \mathcal{F}$.
2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$.
3. Para $n \in \mathbb{N}$, si $A_n \in \mathcal{F} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

En términos financieros, \mathcal{F} representa el espacio de los eventos que son observables en el mercado.

Definición 2.1.2. Sea Ω un conjunto no vacío y \mathcal{F} una σ -álgebra de Ω . Una **Medida de Probabilidad** \mathbb{P} es una función que asigna un número entre $[0, 1]$ a todos los conjuntos $A \in \mathcal{F}$. Dicho número se llama “Probabilidad de A” y se escribe $\mathbb{P}(A)$. Dicha función debe cumplir:

¹Abreviado como VaR.

²Abreviado como ES.

1. $\mathbb{P}(A) \geq 0$.
2. $\mathbb{P}(\Omega) = 1$.
3. $A_1, A_2, \dots \in \mathcal{F}$ disjuntos $\Rightarrow \mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

A la terna $(\Omega, \mathcal{F}, \mathbb{P})$ se le conoce como **Espacio de Probabilidad**.

Variable Aleatoria y su Función de Distribución

Definición 2.1.3. Sea \mathcal{F} una σ -álgebra de Ω . Una función real X en Ω es **variable aleatoria** si los conjuntos $\{X \leq x\} := \{\omega \in \Omega : X(\omega) \leq x\} = X^{-1}([-\infty, x])$ pertenecen a \mathcal{F} , es decir, si $\{X \leq x\} \in \mathcal{F}$ para todo $x \in \mathbb{R}$.

Definición 2.1.4. Sea X una variable aleatoria. La **función de distribución** de X es la función real $\Psi(x) = \mathbb{P}[X \leq x]$.

La función de distribución de X satisface las siguientes propiedades:

1. Es continua por la derecha y no-decreciente.
2. $\lim_{x \rightarrow \infty} \Psi(x) = 1$.
3. $\lim_{x \rightarrow -\infty} \Psi(x) = 0$.

Estamos listos para poder definir el *VaR* y *Expected Shortfall*.

2.1.2. Definición: VaR y Expected Shortfall

En esta sección supondremos que el mercado es la variable aleatoria y , la cual está definida en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ independiente del vector de decisión x . Para cada x , denotamos por $\Psi(x, \cdot)$ sobre \mathbb{R} a la función de distribución resultante de la pérdida $z = f(x, y)$, es decir:

$$\Psi(x, \zeta) = \mathbb{P}\{y | f(x, y) \leq \zeta\},$$

haciendo las suposiciones técnicas que $f(x, y)$ es continua en x y medible en y , y que $\mathbb{E}[|f(x, y)|] < \infty$ para cada $x \in \mathbf{X}$. Denotemos por $\Psi(x, \zeta^-)$ al límite izquierdo de $\Psi(x, \cdot)$ en ζ , entonces:

$$\Psi(x, \zeta^-) = \mathbb{P}\{y | f(x, y) < \zeta\}.$$

Cuando la diferencia:

$$\Psi(x, \zeta) - \Psi(x, \zeta^-) = \mathbb{P}\{y | f(x, y) = \zeta\},$$

es positiva, de tal manera que hay un salto en ζ , se dice que hay un átomo de probabilidad en ζ . Consideremos un nivel de confianza $\alpha \in (0, 1)$ (comúnmente 0.95 ó 0.99), a dicho nivel de confianza hay un correspondiente VaR, el cual se define de la siguiente manera:

Definición 2.1.5. El α -VaR de la pérdida asociada al vector de decisión x es el valor:

$$\zeta_\alpha(x) = \text{mín}\{\zeta | \Psi(x, \zeta) \geq \alpha\}. \quad (2.1)$$

Proposición 2.1.1. El mínimo que define al α -VaR en (2.1.5) existe.

Demostración: Supongamos que el conjunto al cual le estamos sacando mínimo no está acotado por abajo. Entonces existe una sucesión $\{a_n\} \rightarrow -\infty$ cuando $n \rightarrow \infty$ tal que:

$$\Psi(x, a_n) \geq \alpha > 0.$$

De aquí es claro que

$$\lim_{n \rightarrow \infty} \Psi(x, a_n) \geq \alpha > 0,$$

lo cual contradice la propiedad 3 de una función de distribución. Esto prueba que el conjunto sobre el cual estamos tomando el mínimo está acotado por abajo. Para observar que en efecto existe el mínimo consideremos el ínfimo del conjunto $A = \{\zeta | \Psi(x, \zeta) \geq \alpha\}$, entonces existe una sucesión decreciente de elementos del conjunto A que converge a $\text{ínf}\{\zeta | \Psi(x, \zeta) \geq \alpha\}$. Debido a la continuidad por la derecha de una función de distribución llegamos a que

$$\Psi(x, \text{ínf}\{\zeta | \Psi(x, \zeta) \geq \alpha\}) \geq \alpha,$$

lo cual prueba que $\text{ínf}\{\zeta | \Psi(x, \zeta) \geq \alpha\}$ pertenece al conjunto y por lo tanto es un mínimo. *Q.E.D.*

Cuando $\Psi(x, \cdot)$ es continua y estrictamente creciente $\zeta_\alpha(x)$ es simplemente la única ζ que satisface $\Psi(x, \zeta) = \alpha$. De lo contrario, la ecuación puede no tener solución o tener un rango de soluciones. El caso de no tener solución corresponde a un hueco vertical en la gráfica de $\Psi(x, \cdot)$, como se ve en la

Figura 2.1, con α viviendo en un intervalo de niveles de confianza los cuales tienen el mismo *VaR*. Los puntos de inicio y final de este intervalo son:

$$\alpha^-(x) = \Psi(x, \zeta_\alpha^-(x)), \quad \alpha^+(x) = \Psi(x, \zeta_\alpha(x)). \quad (2.2)$$

El caso de tener todo un rango de soluciones corresponde a tener un segmento constante en la gráfica como se ve en la Figura 2.2. Las soluciones forman un intervalo con punto inicial $\zeta_\alpha(x)$ y punto final $\zeta_\alpha^+(x)$, el cual se definirá a continuación.

Definición 2.1.6. El α^+ -*VaR* de la pérdida asociada al vector de decisión x es el valor:

$$\zeta_\alpha^+(x) = \inf\{\zeta | \Psi(x, \zeta) > \alpha\}.$$

La relación $\zeta_\alpha(x) \leq \zeta_\alpha(x)^+$ siempre se mantiene dándose la desigualdad cuando $\Psi(x, \zeta)$ es constante con nivel de confianza α sobre un ζ -intervalo. Dicho intervalo es $[\zeta_\alpha(x), \zeta_\alpha^+(x)]$ ó $[\zeta_\alpha(x), \zeta_\alpha^+(x))$, dependiendo si en $\Psi(x, \cdot)$ existe un salto o no.

Tanto en la Figura 2.1 como en la Figura 2.2 se ilustran las situaciones que muestran las posibles dificultades en el tratamiento del *VaR* para las funciones de distribución en general. Estas se reflejan particularmente en las distribuciones discretas, las cuales son las utilizadas en la práctica estándar, pues las observaciones de las pérdidas de una institución financiera son finitas. En este caso $\Psi(x, \cdot)$ es una función escalonada (constante entre saltos)³ y no hay manera de evitar esta situación.⁴ Observar por ejemplo que la situación en la Figura 2.2 presenta una discontinuidad en el comportamiento del *VaR*. Un salto seguramente ocurrirá con el más pequeño aumento en el nivel de confianza. Este nivel de inestabilidad no es una propiedad deseable para una medida de riesgo en el cual enormes cantidades de dinero están en juego. Este y otros problemas hacen que el manejo del *VaR* como medida de riesgo resulte complicado, lo cual motiva la búsqueda y el uso de otras medidas de riesgo en la práctica distintas a él, como el *Expected Shortfall*.

Definición 2.1.7. El α -*Expected Shortfall* de la pérdida asociada al vector de decisión x es el valor:

$$\phi_\alpha(x) = \text{Esperanza de la distribución } \alpha\text{-cola de } z = f(x, y),$$

³Por lo tanto discontinua.

⁴Figuras tomadas de [4].

donde la distribución α -cola está definida de la siguiente manera:

$$\Psi_\alpha(x, \zeta) = \begin{cases} 0 & \text{si } \zeta < \zeta_\alpha(x) \\ \frac{\Psi(x, \zeta) - \alpha}{1 - \alpha} & \text{si } \zeta \geq \zeta_\alpha(x). \end{cases} \quad (2.3)$$

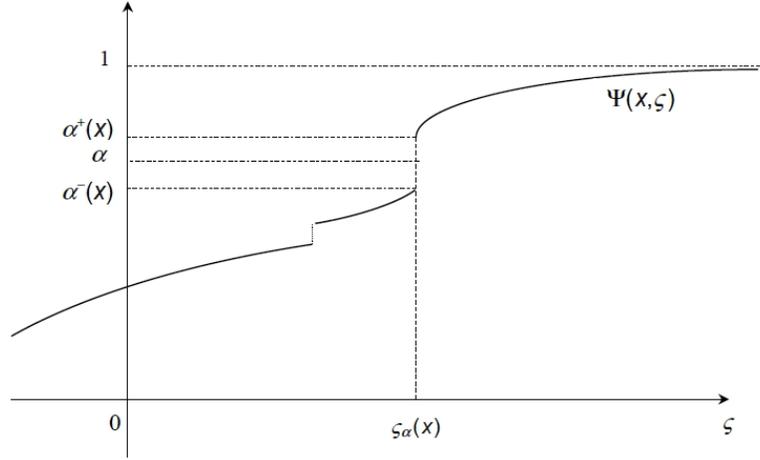


Figura 2.1: Ecuación $\Psi(x, \zeta) = \alpha$ no tiene solución para ζ .

Notar que $\Psi_\alpha(x, \cdot)$ es realmente otra función de distribución, pues $\Psi(x, \cdot)$, es no decreciente, continua por la derecha, y con $\Psi(x, \zeta) \rightarrow 1$ mientras $\zeta \rightarrow \infty$. De aquí es claro que la distribución α -cola en la definición 2.1.7 está bien definida a través de la ecuación (2.3). La sutileza en la definición 2.1.7 reside en el caso donde la pérdida con distribución $\Psi(x, \cdot)$ tiene un átomo de probabilidad en $\zeta_\alpha(x)$. En este caso el intervalo $[\zeta_\alpha(x), \infty)$ tiene probabilidad mayor a $1 - \alpha$, la probabilidad es mayor tanto como:

$$\Psi(x, \zeta_\alpha(x)^-) < \alpha \leq \Psi(x, \zeta_\alpha(x)) \text{ cuando } \Psi(x, \zeta_\alpha(x)^-) < \Psi(x, \zeta_\alpha(x)).$$

Entonces surge naturalmente la pregunta ¿a qué nos referimos con distribución α -cola?, ya que éste término debería referirse a la parte superior de la distribución ⁵. Este problema se resuelve al especificar la distribución a través de la ecuación (2.3), el cual se obtiene reescalando la parte de la gráfica

⁵Que tiene masa de $1 - \alpha$.

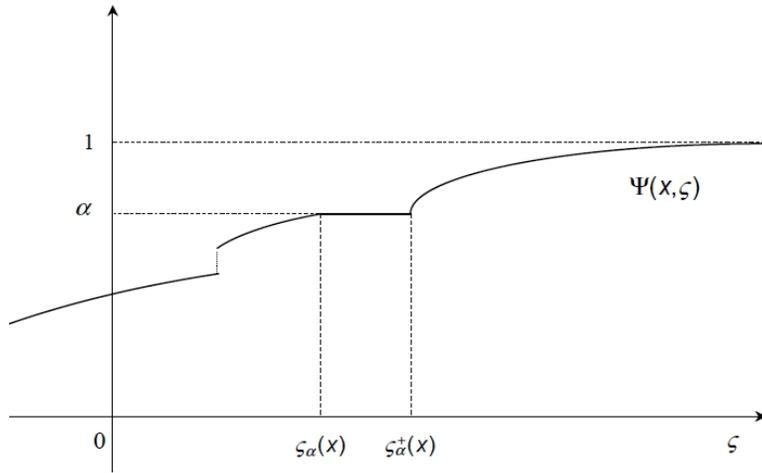


Figura 2.2: Ecuación $\Psi(x, \zeta) = \alpha$ tiene una infinidad de soluciones para ζ .

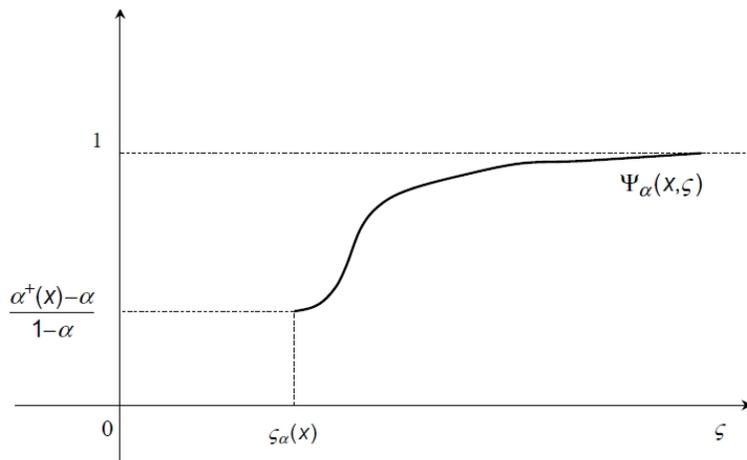


Figura 2.3: La distribución $\Psi_\alpha(x, \zeta)$ se obtiene reescalando la función $\Psi(x, \zeta)$ en el intervalo $[\alpha, 1]$

de la distribución original que se encuentra entre las líneas horizontales con niveles $1 - \alpha$ y 1 de tal manera que ahora recorra los valores entre 0 y 1 . Para el caso mostrado en la Figura 2.1 se muestra el resultado en la Figura 2.3. Las consecuencias de esta manipulación se examinarán mas adelante en

comparación con unas variantes, en las cuales nos enfocaremos en el intervalo $[\zeta_\alpha(x), \infty)$ o en su interior $(\zeta_\alpha(x), \infty)$.

Definición 2.1.8. El α^+ -*Expected Shortfall* de la pérdida asociada al vector de decisión x es el valor:

$$\phi_\alpha^+(x) = \mathbb{E} [f(x, y) | f(x, y) > \zeta_\alpha(x)], \quad (2.4)$$

mientras que el α^- -*Expected Shortfall* es el valor:

$$\phi_\alpha^-(x) = \mathbb{E} [f(x, y) | f(x, y) \geq \zeta_\alpha(x)]. \quad (2.5)$$

La esperanza condicional en la ecuación (2.5) está bien definida ya que

$$\mathbb{P}\{f(x, y) | f(x, y) \geq \zeta_\alpha(x)\} \geq 1 - \alpha > 0.$$

Ahora la esperanza en (2.4) solo tiene sentido cuando $\mathbb{P}\{f(x, y) > \zeta_\alpha(x)\} > 0$, es decir $\Psi(x, \zeta_\alpha(x)) < 1$. Lo cual no siempre es cierto, si solo asumimos que $\alpha \in (0, 1)$, ya que podría haber un átomo de probabilidad en $\zeta_\alpha(x)$.

Proposición 2.1.2. Si no existe un átomo de probabilidad en $\zeta_\alpha(x)$ entonces se tiene que:

$$\phi_\alpha^-(x) = \phi_\alpha(x) = \phi_\alpha^+(x). \quad (2.6)$$

Si existe un átomo de probabilidad en $\zeta_\alpha(x)$ entonces:

$$\phi_\alpha^-(x) < \phi_\alpha(x) = \phi_\alpha^+(x) \text{ cuando } \alpha = \Psi(x, \zeta_\alpha(x)).$$

En otro caso:

$$\phi_\alpha^-(x) = \phi_\alpha(x) \text{ cuando } \Psi(x, \zeta_\alpha(x)) = 1,$$

(con $\phi_\alpha^+(x)$ ahora mal definida). En el resto de los casos, caracterizados por:

$$\Psi(x, \zeta_\alpha(x)^-) < \alpha < \Psi(x, \zeta_\alpha(x)) < 1, \quad (2.7)$$

se tiene la desigualdad estricta:

$$\phi_\alpha^-(x) < \phi_\alpha(x) < \phi_\alpha^+(x).$$

Demostración: En comparación con la definición de $\phi_\alpha(x)$ en 2.1.7, el valor de $\phi_\alpha^+(x)$ en (2.4) es la media de la siguiente distribución de pérdidas:

$$\Psi_\alpha^+(x, \zeta) = \begin{cases} 0 & \text{si } \zeta < \zeta_\alpha(x) \\ \frac{\Psi(x, \zeta) - \alpha^+(x)}{1 - \alpha^+(x)} & \text{si } \zeta \geq \zeta_\alpha(x). \end{cases}$$

Mientras que el valor de $\phi_\alpha^-(x)$ en la ecuación (2.5) es la media de la siguiente distribución de pérdidas:

$$\Psi_\alpha^-(x, \zeta) = \begin{cases} 0 & \text{si } \zeta < \zeta_\alpha(x) \\ \frac{\Psi(x, \zeta) - \alpha^-(x)}{1 - \alpha^-(x)} & \text{si } \zeta \geq \zeta_\alpha(x). \end{cases}$$

Recordar que $\alpha^+(x)$ y $\alpha^-(x)$ definidas en la ecuación (2.2) son el principio y el fin del hueco vertical en $\zeta_\alpha(x)$ en la función de distribución original $\Psi(x, \cdot)$ (si hubiera un salto ahí), el caso en el que no hay átomo de probabilidad en $\zeta_\alpha(x)$ corresponde a tener que $\alpha^-(x) = \alpha^+(x) = \alpha \in (0, 1)$. De aquí que (2.6) claramente es cierta. Cuando sí existe un átomo de probabilidad pero $\alpha = \alpha^+(x)$ tenemos que $\alpha^-(x) < \alpha^+(x) < 1$ entonces :

$$\begin{aligned} \phi_\alpha^-(x) &= \int_0^1 1 - \Psi_\alpha^-(x, \zeta) d\zeta = \int_0^1 \frac{1 - \Psi(x, \zeta)}{1 - \alpha^-(x)} d\zeta \\ \phi_\alpha(x) &= \int_0^1 1 - \Psi_\alpha(x, \zeta) d\zeta = \int_0^1 \frac{1 - \Psi(x, \zeta)}{1 - \alpha(x)} d\zeta. \end{aligned}$$

Además sabemos que $\alpha^-(x) < \alpha$ entonces:

$$\frac{1 - \Psi(x, \zeta)}{1 - \alpha^-(x)} < \frac{1 - \Psi(x, \zeta)}{1 - \alpha},$$

lo cual implica que:

$$\phi_\alpha^-(x) < \phi_\alpha(x).$$

Mientras que si $\alpha^+(x) = 1$ tenemos que:

$$\begin{aligned} \phi_\alpha(x) &= \int_0^{\zeta_\alpha(x)} 1 d\zeta + \int_{\zeta_\alpha(x)}^\infty 0 d\zeta = \zeta_\alpha(x) \\ \phi_\alpha^-(x) &= \int_0^{\zeta_\alpha(x)} 1 d\zeta + \int_{\zeta_\alpha(x)}^\infty 0 d\zeta = \zeta_\alpha(x). \end{aligned}$$

Análogamente, en el otro caso (2.7) $\alpha^-(x) < \alpha^+(x)$ lo cual implica que $\phi_\alpha^-(x) < \phi_\alpha(x) < \phi_\alpha^+(x)$. *Q.E.D.*

El α -Expected Shortfall y el α^+ -Expected Shortfall de las pérdidas coinciden frecuentemente pero no siempre de acuerdo a la visto anteriormente. Otra perspectiva acerca de la conexión que existe entre estos dos valores se desarrollará en lo que sigue.

Proposición 2.1.3. Sea $\lambda_\alpha(x)$ la probabilidad asignada al nivel de pérdida $z = \zeta_\alpha(x)$ por la distribución α -cola definida en la ecuación (2.3):

$$\lambda_\alpha(x) = \frac{[\Psi(x, \zeta_\alpha(x) - \alpha)]}{1 - \alpha} \in [0, 1].$$

Si $\Psi(x, \zeta_\alpha(x)) < 1$, de tal manera que exista la posibilidad de pérdidas mayores que $\zeta_\alpha(x)$ entonces:

$$\phi_\alpha(x) = \lambda_\alpha(x)\zeta_\alpha(x) + [1 - \lambda_\alpha(x)]\phi_\alpha^+(x), \quad (2.8)$$

con $\lambda_\alpha(x) < 1$. Mientras que si $\Psi(x, \zeta_\alpha(x)) = 1$ de tal manera que $\zeta_\alpha(x)$ es la mayor pérdida posible ⁶ entonces se tiene que:

$$\phi_\alpha(x) = \zeta_\alpha(x). \quad (2.9)$$

Demostración: Si $\Psi(x, \zeta_\alpha(x)) < 1$, entonces $\mathbb{P}\{f(x, y) = \zeta_\alpha(x)\} > 0$ bajo la distribución α -cola. De aquí que:

$$\begin{aligned} \phi_\alpha(x) &= \lambda_\alpha\zeta_\alpha(x) + \mathbb{E}[f(x, y)|f(x, y) > \zeta_\alpha(x)](1 - \lambda_\alpha(x)) \\ &= \lambda_\alpha\zeta_\alpha(x) + (1 - \lambda_\alpha(x))\phi_\alpha^+(x). \end{aligned}$$

Si $\lambda_\alpha = 0$ entonces:

$$\phi_\alpha(x) = \int_0^{\zeta_\alpha(x)} 1dx = \zeta_\alpha(x).$$

Q.E.D.

Corolario 2.1.1. El α -Expected Shortfall domina al α -VaR: $\phi_\alpha(x) \geq \zeta_\alpha(x)$.

Demostración: Resulta evidente de la proposición anterior. *Q.E.D.*

Al representar al *Expected Shortfall* como un promedio ponderado del α -VaR y el α^+ -*Expected Shortfall*, la fórmula (2.8) luce sorprendente. Ninguno de α -VaR y α^+ -*Expected Shortfall* se comportan bien como medida de riesgo, sin embargo, *ES* tiene varias propiedades deseables. Lo inusual en la definición de *ES*, lo cual lo catapulta como una potencia entre las diversas medidas de riesgo existentes, está en la manera en la que los átomos de probabilidad se

⁶Entonces $\lambda_\alpha(x) = 1$ con $\phi_\alpha^+(x)$ mal definida.

parten en dos. Dicha partición se destaca en la fórmulas (2.8) y (2.9) de la proposición (2.1.3). Siguiendo la notación de $\alpha^+(x)$ y $\alpha^-(x)$ en la ecuación (2.2) y 2.1.7, y las circunstancias en la ecuación 2.7 donde $\alpha^-(x) < \alpha < \alpha^+(x)$, un átomo en $\zeta_\alpha(x)$ con probabilidad total de $\alpha^+(x) - \alpha^-(x)$ se parte efectivamente en dos partes con probabilidades $\alpha^+(x) - \alpha$ y $\alpha - \alpha^-(x)$, respectivamente. Solo la primera de las dos partes se encapsula en el intervalo $(\zeta_\alpha(x), \infty)$ el cual tiene probabilidad $1 - \alpha^+(x)$, para conseguir una probabilidad de $[1 - \alpha^+(x)] + [\alpha^+(x) - \alpha] = 1 - \alpha$. Mientras que, si el átomo no pudiera ser partido tendríamos que escoger entre los intervalos $[\zeta_\alpha(x), \infty)$ y $(\zeta_\alpha(x), \infty)$ ninguno de los cuales tiene probabilidad $1 - \alpha$. La partición de los átomos de probabilidad de ésta manera estabiliza la respuesta de α -Expected Shortfall a cambios en α , es decir α -Expected Shortfall es continua con respecto al nivel de confianza.

El siguiente resultado se enfoca al caso donde las distribuciones de pérdida en cuestión son discretas.

Proposición 2.1.4. *Supongamos que la medida de probabilidad \mathbb{P} se concentra en una cantidad finita de puntos y_k en \mathbf{Y} . De tal manera que para cada $x \in \mathbf{X}$ la distribución de pérdidas $z = f(x, y)$ es una función escalera con saltos en los puntos $z_1 < z_2 < \dots < z_N$ siendo $p_k > 0$ la probabilidad de z_k . Sea k_α el único índice tal que:*

$$\sum_{k=1}^{k_\alpha} p_k \geq \alpha > \sum_{k=1}^{k_\alpha-1} p_k. \quad (2.10)$$

El α -VaR de las pérdidas está dado por:

$$\phi_\alpha(x) = \frac{1}{1 - \alpha} \left[\left(\sum_{k=1}^{k_\alpha} p_k - \alpha \right) z_{k_\alpha} + \sum_{k=k_\alpha+1}^N p_k z_k \right].$$

Más aún, en esta situación:

$$\lambda_\alpha(x) = \frac{1}{1 - \alpha} \left(\sum_{k=1}^{k_\alpha} p_k - \alpha \right) \leq \frac{p_{k_\alpha}}{p_{k_\alpha} + \dots + p_N}. \quad (2.11)$$

Demostración: De acuerdo a la ecuación (2.10) tenemos que:

$$\Psi(x, \zeta_\alpha(x)) = \sum_{k=1}^{k_\alpha} p_k, \quad (2.12)$$

$$\Psi(x, \zeta_\alpha(x)^-) = \sum_{k=1}^{k_\alpha-1} p_k. \quad (2.13)$$

Restando (2.13) de (2.12) se tiene que:

$$\Psi(x, \zeta_\alpha(x)) - \Psi(x, \zeta_\alpha(x)^-) = p_{k_\alpha}.$$

La afirmación se sigue entonces de (2.3) y la proposición (2.1.3), excepto por la afirmación acerca de la cota por arriba para $\zeta_\alpha(x)$. Para entender esta última observación notar que la expresión para $\zeta_\alpha(x)$ en (2.11) decrece con respecto a α , el cual pertenece al intervalo en (2.10). *Q.E.D.*

2.2. Shapley Value

2.2.1. El Shapley Value de una Institución Financiera

La metodología *Shapley Value* fue desarrollada en el contexto de los juegos cooperativos [18], en el cual el esfuerzo colectivo de un grupo de jugadores genera un valor «compartido». Dado este valor, la metodología lo descompone con el propósito de dividirlo entre todos los jugadores de acuerdo a su contribución. La parte correspondiente del valor a cada jugador es el *Shapley Value* de éste. La metodología *Shapley Value* se puede aplicar directamente al caso de un sistema financiero [7]. En este contexto, los jugadores son las instituciones, que incurren en actividades riesgosas de manera interconectada, quienes generan un cierto nivel de riesgo sistémico. En este caso, el valor por dividirse es el *VaR* o *Expected Shortfall* de todo el sistema a un cierto nivel de confianza. Para poder aplicar la metodología *Shapley Value* es necesario definir una llamada «función característica». Esta función es la misma para cada sub-grupo de instituciones financieras y mapea cada sub-sistema a un cierto nivel de riesgo sistémico. En nuestro caso, la función característica mapeará cada sub-sistema a su correspondiente *VaR* o *Expected Shortfall* a un cierto nivel de confianza. Desarrollemos la metodología *Shapley Value* en el contexto del sistema financiero.

Por el momento supongamos que tenemos una medida de riesgo sistémico aplicable a cualquier sub-grupo de un sistema financiero \mathbf{S} . Supongamos que las \mathbf{N} instituciones financieras están ordenadas aleatoriamente en una fila y consideremos al sub-grupo \mathbf{S}_j que consta de todos los bancos que están

antes que la institución \mathbf{B}_j , incluyendo a la misma. La contribución de la institución \mathbf{B}_j al riesgo sistémico del sub-grupo \mathbf{S}_j es igual a el riesgo sistémico del sub-grupo \mathbf{S}_j menos el riesgo sistémico del sub-grupo formado por todas las instituciones que estaban en \mathbf{S}_j excepto \mathbf{B}_j . El *Shapley Value* de la institución \mathbf{B}_j es igual al promedio de dichas contribuciones sobre las $N!$ maneras de ordenar a las N instituciones financieras. La intuición presentada anteriormente la formalizaremos matemáticamente.

2.2.2. Definición Matemática

Consideremos el siguiente espacio de medida:

1. \mathbf{A} es el conjunto de todas las permutaciones de las N instituciones financieras que conforman todo el sistema.
2. \mathbf{A} es la σ -álgebra generada por \mathbf{A} .
3. \mathbf{P} es la medida uniforme en \mathbf{A} .

Llamemos a nuestra medida de riesgo sistémico \mathbf{V} , la cual es una función que va de \mathbf{A} a los reales ($\mathbf{V} : \mathbf{A} \rightarrow \mathbb{R}$). Definamos la variable aleatoria $\mathbf{X}_j : \mathbf{A} \rightarrow \mathbb{R}$ como:

Definición 2.2.1.

$$\mathbf{X}_j(a) = \mathbf{V}(\mathbf{S}_j) - \mathbf{V}(\mathbf{S}_j - \{\mathbf{B}_j\}),$$

donde a es un elemento de \mathbf{A} . Justamente vemos que la variable aleatoria \mathbf{X}_j representa la contribución de la institución \mathbf{B}_j , dada una formación aleatoria de las instituciones. Definimos entonces el *Shapley Value* de la institución \mathbf{B}_j como:

Definición 2.2.2. El *Shapley Value* de la institución financiera \mathbf{B}_j es:

$$\mathbf{Sh}B_j = \mathbb{E}[\mathbf{X}_j].$$

Es importante observar que la metodología reparte todo el nivel de riesgo sistémico en el sistema financiero. Es decir que:

$$\sum_{i=1}^N \mathbf{Sh}B_i = \mathbf{V}(\mathbf{S}).$$

La metodología *Shapley Value*, como se había mencionado anteriormente, divide el riesgo sistémico de todo el sistema financiero en \mathbf{N} partes correspondientes a cada institución de acuerdo a la contribución de cada una de ellas al nivel de riesgo sistémico de \mathbf{S} . Esta distribución captura la importancia sistémica de cada institución del sistema financiero.

2.2.3. Algunas Propiedades

A continuación enunciaremos algunas propiedades de la metodología *shapley value*

Propiedades de la Metodología Shapley Value

1. **Aditividad** : La suma de los *Shapley Value* de cada institución es igual al riesgo sistémico del sistema financiero.
2. **No Risk No Value**: Si un banco no corre riesgos entonces su *Shapley Value* será cero.
3. **Linealidad de las medidas de riesgo sistémico**: Supongamos que tenemos \mathbf{n} medidas de riesgo sistémico V_1, V_2, \dots, V_n cada una nos arroja sus respectivos *Shapley Value*. Definamos a una nueva medida de riesgo sistémico de la siguiente manera:

$$V = \sum_{i=1}^n \alpha_i V_i,$$

entonces, para toda institución financiera \mathbf{B}_j tenemos que:

$$Sh(\mathbf{B}_j, V) = \sum_{i=1}^n \alpha_i Sh(\mathbf{B}_j, V_i),$$

donde la nueva notación indica el *Shapley Value* correspondiente a la institución \mathbf{B}_j y la medida de riesgo sistémico correspondiente.

La propiedad 3 es de extrema importancia para el creador de una política de supervisión macroprudencial. Es muy complicado identificar con certeza, exactamente cuáles son los factores a los cuales el sistema financiero es más vulnerable. En general, las distintas medidas de riesgo sistémico que uno podría usar reflejan solo algunos factores que podrían afectar al sistema financiero. En este sentido, debido a la abundancia de factores que podrían

perturbar al sistema financiero y a la poca información que se tiene sobre la severidad de estas perturbaciones, resulta natural que una autoridad supervisora quisiera utilizar varias medidas de riesgo sistémico. La linealidad del *Shapley Value* permitiría a la autoridad supervisora incorporar todas estas medidas en una sola. Los coeficientes de la combinación lineal de las distintas medidas de riesgo sistémico estarían reflejando la relevancia que (tal vez intuitivamente) la autoridad supervisora considere apropiada para cada una. Otra propiedad muy importante del *Shapley Value Method* es que el incremento del *Shapley Value* de la institución \mathbf{B}_i causada por la presencia de la institución \mathbf{B}_k es igual al incremento del *Shapley Value* de la institución \mathbf{B}_k causada por la presencia de la institución \mathbf{B}_i . Más aún la propiedad sigue siendo válida si los *Shapley Values* son calculados para cualquier sub-grupo del sistema financiero. Puesto en términos matemáticos la propiedad dice lo siguiente :

$$Sh(\mathbf{B}_i, \mathbf{G}) - Sh(\mathbf{B}_i, \mathbf{G} - \{\mathbf{B}_k\}) = Sh(\mathbf{B}_k, \mathbf{G}) - Sh(\mathbf{B}_k, \mathbf{G} - \{\mathbf{B}_i\}),$$

para todo $\mathbf{B}_i, \mathbf{B}_k$ y todo \mathbf{G} subgrupo del sistema financiero tales que $\mathbf{B}_i, \mathbf{B}_k$ están en \mathbf{G} . Las pruebas de todas estas propiedades se encuentran en el Apéndice B.

2.2.4. Un Ejemplo Concreto

Consideremos el caso sencillo cuando el sistema financiero consta de tres instituciones $\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3$ y la siguiente función característica (V):

Sub-Grupo	Pérdida
ϕ	0
\mathbf{B}_1	1
\mathbf{B}_2	3
\mathbf{B}_3	5
$\mathbf{B}_1\mathbf{B}_2$	3.5
$\mathbf{B}_1\mathbf{B}_3$	5.5
$\mathbf{B}_2\mathbf{B}_3$	7
$\mathbf{B}_1\mathbf{B}_2\mathbf{B}_3$	8.5

Calcularemos el *Shapley Value* de la institución \mathbf{B}_1 . El cálculo del *Shapley Value* de las demás instituciones es completamente análogo. Lo primero que hay que hacer es calcular los valores que toma la variable aleatoria \mathbf{X}_1 sobre las 6 permutaciones de las tres instituciones financieras.

A continuación se presenta una tabla con el detalle de las cuentas necesarias para poder obtener \mathbf{X}_1 :

Permutación	Sub-Grupo: Instituciones antes de, e incluyendo a \mathbf{B}_1	Cálculo Detallado \mathbf{X}_1	Valor de \mathbf{X}_1
$\mathbf{B}_1\mathbf{B}_2\mathbf{B}_3$	\mathbf{B}_1	$V(\mathbf{B}_1)-V(\phi)$	1
$\mathbf{B}_1\mathbf{B}_3\mathbf{B}_2$	\mathbf{B}_1	$V(\mathbf{B}_1)-V(\phi)$	1
$\mathbf{B}_2\mathbf{B}_1\mathbf{B}_3$	$\mathbf{B}_2\mathbf{B}_1$	$V(\mathbf{B}_2, \mathbf{B}_1)-V(\mathbf{B}_2)=3.5-3$	0.5
$\mathbf{B}_3\mathbf{B}_1\mathbf{B}_2$	$\mathbf{B}_3\mathbf{B}_1$	$V(\mathbf{B}_3, \mathbf{B}_1)-V(\mathbf{B}_3)=5.5-5$	0.5
$\mathbf{B}_3\mathbf{B}_2\mathbf{B}_1$	$\mathbf{B}_3\mathbf{B}_2\mathbf{B}_1$	$V(\mathbf{B}_3, \mathbf{B}_2, \mathbf{B}_1)-$ $V(\mathbf{B}_3, \mathbf{B}_2)=8.5-7$	1.5
$\mathbf{B}_2\mathbf{B}_3\mathbf{B}_1$	$\mathbf{B}_2\mathbf{B}_3\mathbf{B}_1$	$V(\mathbf{B}_3, \mathbf{B}_2, \mathbf{B}_1)-$ $V(\mathbf{B}_3, \mathbf{B}_2)=8.5-7$	1.5
		$Sh\mathbf{B}_1$	= 1

Notar que el cálculo final del *Shapley Value* de \mathbf{B}_1 es el promedio de los valores de \mathbf{X}_1 sobre todas las permutaciones del sistema financiero. Para el caso sencillo, expuesto anteriormente, existe otra manera de expresar la definición del Shapley Value de la institución \mathbf{B}_1 que es la siguiente:

$$Sh\mathbf{B}_1 = \frac{1}{6}[2V(\mathbf{B}_1) - V(\Phi)] + (V(\mathbf{B}_2, \mathbf{B}_1) - V(\mathbf{B}_2)) + (V(\mathbf{B}_3, \mathbf{B}_1) - V(\mathbf{B}_3)) \\ + 2(V(\mathbf{B}_3, \mathbf{B}_2, \mathbf{B}_1) - V(\mathbf{B}_3, \mathbf{B}_2)),$$

aquí se está expresando al Shapley Value como un promedio ponderado sobre todos los sub-grupos del sistema financiero que contienen a la institución \mathbf{B}_1 . La expresión anterior del Shapley Value de una institución no es particular al ejemplo expuesto anteriormente. En el caso general la identidad es la siguiente:

$$Sh\mathbf{B}_1 = \sum_{m=1}^N \sum_{\substack{S \supset \mathbf{B}_1 \\ |S|=m}} \frac{(m-1)!(N-m)!}{N!} [V(S) - V(S - \{\mathbf{B}_1\})],$$

aquí el Shapley Value de \mathbf{B}_1 se expresa como un promedio ponderado sobre todos los sub-grupos del sistema que contienen a la institución \mathbf{B}_1 .

2.3. Cópulas y Correlación Spearman

Uno de los elementos del sistema financiero que quisiéramos modelar con respecto al tiempo es la estructura de dependencia que existe en el mismo. La correlación es una clásica medida de dependencia, pero su uso solamente es apropiado cuando la distribución de pérdidas es elíptica. Desafortunadamente, en la vida real, las distribuciones de pérdida de un sistema financiero no son elípticas. Surge entonces la necesidad de encontrar una nueva manera de caracterizar la dependencia que existe entre las distintas marginales y de encontrar medidas numéricas de asociación que traten de resumir en un valor el nivel de asociación que existe entre dos variables aleatorias. La cópula y la correlación Spearman [3] serán los nuevos conceptos que utilizaremos.

2.3.1. Cópulas

Empecemos con el caso particular de \mathbb{R}^2 , supongamos que tenemos un vector aleatorio (X, Y) con distribuciones marginales F y G y función de distribución conjunta H . La función de distribución conjunta H contiene tres tipos de información, información de la marginal F , información de la marginal G e información de la estructura de dependencia entre X y Y . Con el fin de desprender la información de la estructura de dependencia de la distribución conjunta H , hay que neutralizar la información que existe sobre las marginales. Dicha neutralización se hace componiendo las variables aleatorias X e Y con sus respectivas funciones de distribución. El efecto de esta composición es que ahora la distribución de las marginales $F(X)$ y $G(Y)$ es uniforme. Bajo esta distribución, las marginales toman valores en el intervalo $[0, 1]$ con la misma probabilidad y fuera del intervalo con probabilidad cero. Es por esto que la distribución uniforme se considera como la distribución no informativa ⁷. En este sentido la distribución conjunta de $F(X)$ y $G(Y)$ ahora solo contiene información sobre la dependencia que existe entre X y Y . A dicha distribución se le llama la cópula del vector aleatorio (X, Y) .

Formalmente tenemos lo siguiente:

Definición 2.3.1. Una cópula es una función tal que:

1. $C : [0, 1] \times [0, 1] \rightarrow [0, 1]$.

⁷Como veremos más adelante la distribución uniforme es la que maximiza la función de entropía; una medida de desinformación.

2. $C(u_1, 0) = C(0, u_2) = 0$.
3. $C(u_1, 1) = u_1$ $C(1, u_2) = u_2$.
4. $C(v_1, v_2) - C(u_1, v_2) \geq C(v_1, u_2) - C(u_1, u_2)$ para todo $u_1, u_2, v_1, v_2 \in [0, 1]$ con $u_1 \leq v_1$ y $u_2 \leq v_2$.

La propiedad 1 implica que la cópula actúa sobre los valores de las distribuciones marginales, las demás implican que la cópula es la función de distribución de dos variables aleatorias uniformes. Ahora es claro que hay muchas posibles distribuciones conjuntas para dos marginales uniformes, es decir, hay muchas posibles cópulas. Un importante resultado de Sklar muestra que las cópulas son únicas en un sentido muy preciso. La forma bivariada del Teorema de Sklar es la siguiente:

Teorema 2.3.1. *Dada cualquier función de distribución conjunta $F(X_1, X_2)$, existe una única cópula $C : [0, 1] \times [0, 1]$, tal que:*

$$F(X_1, X_2) = C(F_1(X_1), F_2(X_2)). \quad (2.14)$$

Del mismo modo, si C es una cópula y $F_1(X_1)$ y $F_2(X_2)$ son funciones de distribución, entonces $C(F_1(X_1), F_2(X_2))$ define una nueva función de distribución con distribuciones marginales $F_1(X_1)$ y $F_2(X_2)$. Diferenciando la ecuación (2.14) con respecto a X_1 y X_2 obtenemos una simple expresión para la densidad conjunta $f(X_1, X_2)$ en términos de las densidades marginales $f_1(X_1)$ y $f_2(X_2)$, si existen. Tenemos que

$$f(X_1, X_2) = f_1(X_1) f_2(X_2) c(F_1(X_1), F_2(X_2)), \quad (2.15)$$

donde

$$c(F_1(X_1), F_2(X_2)) = \frac{\partial^2 C(F_1(X_1), F_2(X_2))}{\partial F_1(X_1) \partial F_2(X_2)}.$$

Posteriormente utilizaremos la ecuación (2.15) para derivar la CIMDO-Cópula, es decir, el modelo de la estructura de dependencia en el sistema financiero inferido por la metodología CIMDO. La idea anterior se puede generalizar al caso de \mathbb{R}^n . Considera n variables aleatorias $X_1, X_2 \dots X_n$ con distribuciones marginales $F_1, F_2 \dots, F_n$.

Definición 2.3.2. Una cópula es una función monótona creciente que satisface las siguientes propiedades:

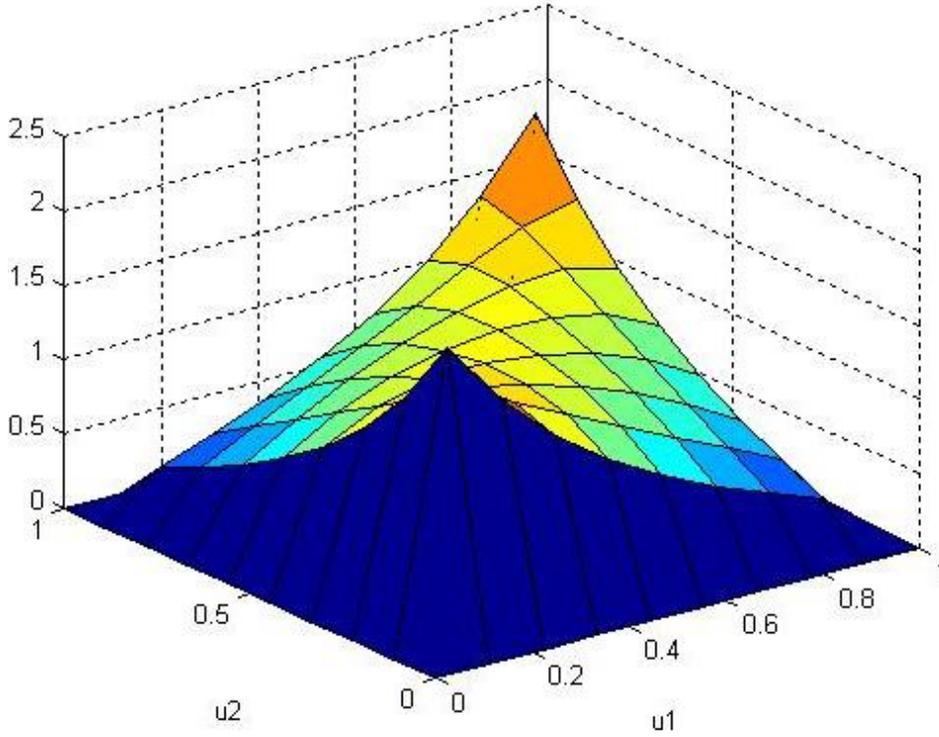


Figura 2.4: Densidad de la cópula de una t-student bivariada con $\rho = 0.5$ y $df = 10$

1. $C(u_1, \dots, u_n)$ es creciente en cada componente u_i .
2. $C(1, \dots, 1, u_i, \dots, 1) = u_i$ para todo $i \in \{1, \dots, n\}, u_i \in [0, 1]$.
3. $C(0, \dots, 0, u_i, \dots, 0) = 0$ para todo $i \in \{1, \dots, n\}, u_i \in [0, 1]$.
4. Para todo $(u_1, \dots, u_n), (v_1, \dots, v_n) \in [0, 1]^n$ tenemos que

$$(-1)^{i_1 + \dots + i_n} \sum_{i_1=1}^2 \dots \sum_{i_n=1}^2 C(x_{1i_1}, \dots, x_{ni_n}) \geq 0$$

donde $x_{j1} = u_j$ y $x_{j2} = v_j$ para toda $j \in \{1, \dots, n\}$.

El teorema de Sklar afirma que, dado un conjunto fijo de distribuciones marginales continuas, distintas cópulas definen distintas distribuciones conjuntas. Dada una densidad conjunta $F(X_1, X_2, \dots, X_n)$ con marginales continuas, podemos definir una cópula C tal que:

$$F(X_1, X_2, \dots, X_n) = C(F_1(X_1), F_2(X_2), \dots, F_n(X_n)).$$

En el caso que exista, la densidad de la cópula es la siguiente:

$$c(F_1(X_1), F_2(X_2), \dots, F_n(X_n)) = \frac{\partial^n C(F_1(X_1), F_2(X_2), \dots, F_n(X_n))}{\partial F_1(X_1) \dots \partial F_n(X_n)},$$

vista como función de $u_i = F_i(X_i)$.

Dada la densidad de la cópula y en caso que existan las densidades marginales $f_i(x) = F_i'(x)$, podemos obtener la densidad conjunta de las variables originales utilizando:

$$f(X_1, \dots, X_n) = f_1(X_1) \dots f_n(X_n) c(F_1(X_1), \dots, F_n(X_n)).$$

Fijando $u_i = F_i(X_i)$, cada distribución conjunta F define implícitamente una cópula de la forma:

$$C(u_1, u_2, \dots, u_n) = F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)),$$

donde los u_i pertenecen a $[0, 1]$. De aquí que existe una cópula implícita correspondiente a cada distribución conjunta.

Con la notación $u_i = F_i(X_i)$ la densidad se puede escribir como:

$$c(u_1, \dots, u_n) = \frac{\partial C(u_1, \dots, u_n)}{\partial u_1 \dots \partial u_n}$$

2.3.2. Correlación Spearman

Cuando no se asume que la distribución de pérdidas obedecen a una distribución elíptica, la correlación lineal Pearson deja de ser una medida apropiada de asociación entre dos variables aleatorias como se ha expuesto anteriormente. ¿Qué métrica deberíamos usar ahora?. Para contestar esta pregunta necesitamos encontrar una definición de asociación un poco más restrictiva.

Definición 2.3.3. Consideremos dos observaciones de las variables aleatorias continuas X_1 y X_2 denotadas por (x_1, y_1) y (x_2, y_2) , se dice que las observaciones son concordantes si $(x_1 - x_2)(y_1 - y_2) > 0$ y discordantes si $(x_1 - x_2)(y_1 - y_2) < 0$.

Definición 2.3.4. Una métrica de concordancia $m(X_1, X_2)$ es una medida numérica de asociación entre dos variables aleatorias (X_1, X_2) tal que:

1. $m(X_1, X_2) \in [-1, 1]$ y su valor dentro de este rango depende de $F(X_1, X_2)$, la función de distribución conjunta de X_1 y X_2 .
2. $m(X_1, X_1) = 1$ y $m(X_1, -X_1) = -1$.
3. $m(X_1, X_2) = m(X_2, X_1)$ y $m(X_1, -X_2) = -m(X_2, X_1)$.
4. Si X_1 y X_2 son independientes, entonces $m(X_1, X_2) = 0$.
5. Dadas dos posibles funciones de distribución conjunta $F(X_1, X_2)$ y $G(X_1, X_2)$, denotemos por $m_F(X_1, X_2)$ y $m_G(X_1, X_2)$ las medidas de concordancia bajo las dos distribuciones. Si $F(X_1, X_2) \geq G(X_1, X_2)$ entonces $m_F(X_1, X_2) \geq m_G(X_1, X_2)$.

El problema de la correlación Pearson es que no es una métrica de concordancia excepto, cuando las pérdidas obedecen una distribución elíptica, lo cual generalmente no es el caso en la práctica. Procedamos a definir la correlación Spearman :

Supongamos que tenemos n observaciones de un vector aleatorio (X_1, X_2) denotados por $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Para estimar la versión discreta de la correlación Spearman tenemos que ordenar la información para las dos series de observaciones ⁸ y luego sumar las diferencias al cuadrado entre las posiciones.

Definición 2.3.5. Sea d_i la diferencia entre el orden de x_i y el orden de y_i , sea $D = \sum_{i=1}^n d_i^2$ entonces la correlación Spearman discreta entre las variables aleatorias (X_1, X_2) es:

$$\varphi = 1 - \frac{6D}{n(n^2 - 1)}.$$

⁸Si dos o mas observaciones son iguales entonces cada una recibe el orden promedio.

La correlación Spearman de dos variables aleatorias X_1 y X_2 es una métrica de concordancia. Debido a que la cópula de las variables aleatorias X_1 y X_2 captura toda la estructura de dependencia entre X_1 y X_2 , resulta natural que exista una relación entre la cópula y la correlación Spearman de dos variables aleatorias X_1 y X_2 . En efecto uno podría definir analíticamente a la correlación Spearman de la siguiente manera:

Definición 2.3.6. Sea C la cópula de X_1 y X_2 , entonces la correlación Spearman es igual a:

$$\varphi = 12 \int_0^1 \int_0^1 C(u_1, u_2) du_1 du_2 - 3 = 12 \int_0^1 \int_0^1 [C(u_1, u_2) - u_1 u_2] du_1 du_2.$$

Intuitivamente la definición anterior permite visualizar a la correlación Spearman como el área promedio entre la cópula de X_1 y X_2 y la cópula de dos variables aleatorias independientes. Existe también una relación entre la correlación Spearman y la correlación Pearson. En efecto una definición alternativa de la correlación Pearson es la siguiente:

Definición 2.3.7. Si X_1 y X_2 son dos variables aleatorias con distribuciones marginales $F(X_1)$ y $G(X_2)$ la correlación Spearman de las variables aleatorias X_1 y X_2 se define como:

$$\varphi = \rho(F(X_1), G(X_2)).$$

donde ρ denota la correlación Pearson de dos variables aleatorias.

Capítulo 3

Entropía y CIMDO

Uno de los elementos mas importantes en nuestra metodología para medir la importancia sistémica es la distribución multivariada de pérdidas del sistema financiero. Desafortunadamente para la mayoría de los casos, (por ejemplo el caso de un regulador financiero) no existe información suficiente que nos permita una completa caracterización de la misma. Se han desarrollado métodos que nos permiten modelar la distribución de pérdidas en los cuales se hacen ciertas suposiciones paramétricas para tratar de resolver el problema de la falta de información. Aún así la propia calibración de estos modelos requieren de información que por lo general no existe o no está disponible para todo público (como es el caso de México). La falta de información hace imposible la correcta calibración de los modelos paramétricos; por lo tanto las distribuciones resultantes podrían no ser consistentes con el verdadero proceso generador de datos. Como consecuencia se podrían realizar inferencias estadísticas e interpretaciones económicas incorrectas. El enfoque de la teoría de la probabilidad propuesta por E.T. Jaynes [1] nos proporciona una metodología sistemática de inferencia ad-hoc para el problema que tenemos en nuestras manos. El objetivo del presente capítulo es el de desarrollar desde el principio y en un contexto general el método de inferencia conocido como «El Principio de Mínima Entropía Cruzada» con la intención de que al final de dicho desarrollo se pueda ver que CIMDO, que es la metodología sistemática de inferencia que utilizaremos no es más que una simple aplicación del principio de mínima entropía cruzada al caso del problema de inferencia de la distribución de pérdidas del sistema financiero que pretendemos resolver. Mas aún basados de nuevo en el desarrollo del marco metodológico de inferencia en el cual está basado CIMDO, pretendemos mostrar al lector que

CIMDO recupera las distribuciones de manera robusta a pesar de la falta de información que existe en la mayoría de los casos. En las secciones 3.1, 3.2, 3.3 y 3.4 se presentan con detalle tanto el enfoque de inferencia en el cual nos estamos basando como la implementación matemática de dicho enfoque. Finalmente en las secciones 3.5 y 3.6, se expone con detalle CIMDO como una aplicación de la teoría expuesta en las secciones anteriores.

3.1. La Teoría de la Probabilidad: Una Extensión de la Lógica

La teoría de la probabilidad propuesta principalmente por E. T. Jaynes surgió con su interés en el trabajo de Harold Jeffreys y el descubrimiento de los trabajos de R.T. Cox, Shannon y Pólya. George Pólya en su trabajo «Mathematics and Plausible Reasoning» desmenuzó nuestro «sentido común» intuitivo en un conjunto de axiomas elementales y demostró que los matemáticos los habían estado utilizando desde las primeras fases del descubrimiento, las cuales necesariamente preceden el encuentro de una prueba rigurosa. Los resultados fueron muy parecidos a los de James Bernoulli en su trabajo «Art of Conjecture», desarrollados analíticamente por Laplace en el siglo XVIII; pero Pólya pensaba que la semejanza era solamente cualitativa. Afortunadamente, los teoremas de consistencia de R.T. Cox fueron suficientes para atar cabos; cuando uno les añadía las condiciones cualitativas de Pólya el resultado fue una prueba de que, si los niveles de plausibilidad se representaran con números reales entonces existe un único conjunto de reglas cuantitativas para realizar inferencias. Los resultados finales no fueron más que las reglas estándares de la teoría de la probabilidad ya dadas por Daniel Bernoulli y Laplace; entonces ¿Por qué tanta faramalla? La nueva e importante diferencia es que ahora estas reglas fueron vistas como los únicos principios válidos de la lógica en general, sin hacer referencia a términos como «aleatoriedad» o «variable aleatoria» y haciendo así su rango de aplicación sumamente más amplio que lo que había sido supuesto en la teoría de la probabilidad convencional. Como resultado, la imaginaria distinción entre la «teoría de la probabilidad» y la «inferencia estadística» desaparece, y el nuevo campo consigue no solo unidad lógica y simplicidad, consigue también un mayor poder técnico y flexibilidad en las aplicaciones.

Después de muchos años de experiencias en las aplicaciones a los cientos

de problemas de la vida real, la nueva visión de la teoría de la probabilidad evolucionó hasta convertirse en algo un tanto complejo, el cual ya no se podía describir con términos tan simplistas como «pro-esto» o «anti-lo otro». Por ejemplo, el nuevo sistema de probabilidad no podría ser más distinto que el de Kolmogorov en estilo, filosofía y propósito. Lo que ahora se conoce como enteramente la otra mitad de la teoría de la probabilidad como se requiere en las aplicaciones más recientes ¹ -los principios de asignación de probabilidades mediante el análisis lógico de información incompleta- no está presente en el sistema de Kolmogorov. A pesar del hecho anterior, la teoría nueva se encuentra para la sorpresa incluso de Jaynes, de acuerdo con el sistema de Kolmogorov y en desacuerdo con sus críticos en prácticamente todas las cuestiones técnicas. Como podría verse en [1], cada uno de sus axiomas resultan ser, para todo propósito práctico, derivables de los axiomas Pólya-Cox de consistencia y raciocinio. En corto el nuevo sistema de probabilidad, no contradice al de Kolmogorov sino que busca unos cimientos lógicos más profundos los cuales permiten su extensión a las direcciones requeridas por las aplicaciones en los tiempos modernos. En los problemas de la vida real ninguno de los dos métodos, ni el frecuentista ni el bayesiano son universalmente aplicables. En este sentido, el nuevo sistema pretende visualizar las cosas de manera más amplia. El nuevo tema es simplemente «la teoría de la probabilidad como una extensión de la lógica». La nueva percepción consiste en entender que las reglas matemáticas de la probabilidad no son meramente reglas para calcular frecuencias de «variables aleatorias»; sino que también son las únicas reglas consistentes para realizar inferencias de cualquier tipo, y las aplicaremos en su generalidad total con ese fin.

Es cierto que todas las cuentas «Bayesianas» están incluidas automáticamente como casos particulares de las reglas; pero en este sentido, también lo están las cuentas frecuentistas. No obstante, las nuevas reglas básicas son más amplias que cualquiera de las otras dos y en muchas ocasiones las cuentas que surgen con la teoría nueva no pertenecen a ninguna categoría. Para explicar la situación como está vista en el presente diremos lo siguiente: los métodos frecuentistas tradicionales, solamente utilizan distribuciones muestrales, que son útiles en varios problemas particularmente simples e idealizados, sin embargo, dichos problemas solamente representan los casos más naturales (notorios) en la teoría de la probabilidad, porque presuponen condiciones (repeticiones independientes de un «fenómeno aleatorio» pero sin ningún ti-

¹Como CIMDO.

po de información «a-priori».) los cuales ocurren pocas veces en la vida real. En este sentido, el enfoque anterior resulta inadecuado para las recientes necesidades de la ciencia. Mas aún los métodos frecuentistas no proporcionan maneras para incorporar información «a-priori».

Todos estos defectos se corrigen con el uso de los métodos bayesianos, pero antes que alguno de estos dos metodos pueda ser usado es necesario que el problema en mano sea desarrollado más allá de la «fase exploratoria» al punto en el que tiene la suficiente estructura para determinar todo el aparato requerido (un modelo, espacio muestra, espacios de hipótesis, probabilidades a-priori, distribuciones muestrales). Casi todos los problemas científicos pasan por la fase exploratoria inicial en el cual no hay necesidad de hacer inferencia; pero las suposiciones frecuentistas no son válidas y el aparato bayesiano aún no está disponible. En efecto, algunos problemas nunca evolucionan al grado de pasar la fase exploratoria. Los problemas a este nivel piden métodos más primitivos de asignación de probabilidades directamente de la información parcial. Para este fin, el Principio de Máxima Entropía posee hasta el momento la justificación teórica más clara y es la más desarrollada computacionalmente con un aparato analítico tan poderoso y versátil que el bayesiano. Para aplicarlo necesitamos definir un espacio muestra pero no necesitamos definir una distribución muestral ni un modelo. En efecto, la maximización de la entropía nos crea un modelo directamente de la información, el cual ha probado ser el óptimo bajo tantos criterios que resulta difícil imaginarnos circunstancias mediante las cuales uno no quisiera usarlo en un problema donde tenemos un espacio muestra pero no un modelo.

Los métodos bayesianos y el de máxima entropía difieren en otro aspecto. Ambos procedimientos devuelven inferencias óptimas de la información, pero nosotros podemos escoger un modelo para el análisis bayesiano, esto consiste en expresar una cierta cantidad de conocimiento sobre el fenómeno en observación en alguna hipótesis. Usualmente estas hipótesis se extienden más allá de lo directamente observable en la información disponible y en este sentido podríamos decir que los métodos bayesianos son o al menos podrían ser especulativos. Si las hipótesis extras resultan ser ciertas entonces esperaríamos que los resultados bayesianos mejorarán los de la máxima entropía; si resultan ser falsos los resultados bayesianos muy probablemente serán peores.

Por otro lado, el principio de máxima entropía es un procedimiento no especulativo, en el sentido que no invoca ninguna hipótesis más allá del espacio muestra y la evidencia existente en la información. Consecuentemente predice solamente hechos observables en vez de valores de parámetros, los cuales

podrían existir solamente en nuestra imaginación. Es precisamente por esta razón que el principio de máxima entropía es la herramienta más apropiada (segura) cuando existe muy poca información más allá de la información pura; nos protege ante la posibilidad de llegar a conclusiones no soportadas por la información existente. CIMDO se basa en el principio de máxima entropía para realizar la inferencia de la distribución de pérdidas del sistema financiero. No es difícil ver que el problema de hacer dicha inferencia resulta ad-hoc para aplicar máxima entropía. La única información en la que nos podemos apoyar es en las probabilidades de distress ² de cada institución financiera, y es justamente dicha información la que utilizaremos para formular CIMDO.

3.2. Entropía

3.2.1. Bayesionismo Objetivo

Como se mencionó en la sección anterior, el proceso de asignación de una distribución de probabilidad a un fenómeno aleatorio representa meramente la otra mitad de la teoría de la probabilidad de la concepción de Jaynes. Ninguno de los métodos frecuentistas o bayesianos proporcionan una solución completa a este problema. Cada uno de los métodos funciona perfectamente para ciertos tipos de problemas de inferencia muy particulares. Los métodos frecuentistas solo aplican cuando podemos repetir un evento una infinidad de veces. Los métodos bayesianos funcionan cuando existe información a-priori suficiente como para poder hacer hipótesis iniciales correctas sin utilizar nada más que los datos disponibles. El Principio de Máxima Entropía fue propuesto por Jaynes para generar distribuciones de probabilidad de forma sistemática y objetiva para el caso (o más bien especialmente para el caso) en que tenemos muy poca información más allá de los datos mismos. Se puede describir de la siguiente manera: de entre todas aquellas distribuciones de probabilidad compatibles con cierta clase de información, escoger la que conlleve a una mayor incertidumbre. Desde luego la ambigüedad que falta resolver en la descripción anterior es nuestra definición de incertidumbre de una distribución de probabilidad, es decir, la entropía de una distribución. Con el fin de fijar ideas hablaremos un poco sobre las ideas expuestas en la sección pero puestas en términos de probabilidades. La postura expuesta anteriormente considera a la probabilidad como una medida de «plausibilidad»

²Por «distress» nos referimos a un estado de mala salud.

que le deberíamos otorgar a cierta proposición A , dada cierta información previa I :

$$P(A/I).$$

En otras palabras, $P(A/I)$ corresponde a un estado del conocimiento, al estado de la información que poseemos sobre A dado I . Esta definición es completamente objetiva, en el sentido de que dos personas que posean exactamente la misma información previa I , deberían otorgarle exactamente la misma plausibilidad a A . Hay que notar dos cosas: a) La probabilidad se aplica a proposiciones, y no a conjuntos (aunque por supuesto estas proposiciones pueden referirse a conjuntos) y b) la probabilidad condicional no se define en términos de un objeto más simple sino que ellas mismas son los elementos fundamentales de la teoría. Ahora bien, si la probabilidad condicional no se define en términos de algo más, entonces, ¿cómo se supone que le vamos a asignar valores numéricos en situaciones concretas? Hay dos posibilidades: o bien la expresión se puede descomponer (usando las reglas básicas de la suma y el producto) en expresiones presumiblemente más simples, o bien tenemos una proposición que podríamos llamar «atómica». El segundo caso es justamente el que proporciona el nexo con la realidad, pues aquí debemos entrar en consideraciones lógicas que tengan que ver con el problema a tratar. Pues bien, el principio de máxima entropía nos proporciona justamente un principio para asignar valores numéricos a las probabilidades de estas proposiciones atómicas. Hay que señalar que no se hace la afirmación de que éste sea el único principio válido en todas las situaciones; el propio Jaynes comenta que este problema, el de encontrar asignaciones únicas de probabilidad dada cierta información previa, representa una buena parte faltante («la otra mitad») de la teoría de la probabilidad, y que conforme se vayan descubriendo (o inventando) nuevos principios se irá extendiendo el dominio de aplicación de la probabilidad a nuevas esferas de la realidad.

3.2.2. Derivaciones de la Entropía

Nos damos a la tarea de definir la entropía de una distribución de probabilidad. El concepto de entropía se puede derivar de distintas formas y basándose en diversos principios. Fue nuestro caso que después de haber digerido las tres derivaciones, alguna le pareció más «intuitiva» que las otras. Desde luego que esto es cuestión de gustos y es en este sentido que decidimos presentar tres posibles derivaciones del concepto de entropía con el fin de

dejar al juicio del lector su «derivación favorita».

La Derivación de Shannon

Claude E. Shannon se pregunta si es posible definir una cantidad que mida, en algún sentido, la cantidad de información que es producida por un proceso de Markov. Específicamente, supongamos que tenemos ciertos eventos con probabilidades de ocurrencia p_1, \dots, p_n conocidas. ¿Podemos encontrar una medida de qué tanta «elección» está involucrada en la selección de un evento, o de qué tan incierto estemos del resultado? Si existiera una medida consistente de la «cantidad de incertidumbre» para distribuciones de probabilidad discretas, entonces, tendría que satisfacer las siguientes condiciones:

1. Asumiremos que existe una medida numérica $H_n(p_1, p_2, \dots, p_n)$ que represente la cantidad de incertidumbre de la distribución, es decir, existe una asociación entre la cantidad de incertidumbre y los reales.
2. Asumiremos continuidad: H_n es una función continua de las p_i .
3. Intuitivamente queda claro que mientras más posibilidades existan en una distribución de probabilidad hay mayor cantidad de incertidumbre, ésta intuición la traduciremos matemáticamente a que:

$$h(n) = H_n \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n} \right)$$

es una función monótona creciente de n .

4. Necesitamos que H_n sea consistente, en el sentido que si existen más de una manera de calcular H_n , entonces cada camino nos lleva al mismo resultado.

Utilicemos las condiciones anteriores para determinar la función H_n . Supongamos que tenemos una distribución de probabilidad que le asigna probabilidades p_1, p_2, p_3 a los eventos A_1, A_2, A_3 . Un posible camino para calcular $H_3(p_1, p_2, p_3)$ es el siguiente:

Primero decidamos que si el evento A_1 resultó ser cierto o no, esta decisión nos cuesta una cantidad de incertidumbre de $H_2(p_1, p_2 + p_3)$. Luego con probabilidad $p_2 + p_3$ el evento A_1 no es cierto llevándonos a tener que decidir

entre los eventos A_2 y A_3 , lo cual nos cuesta una cantidad de incertidumbre de $H_2\left(\frac{p_2}{p_2+p_3}, \frac{p_3}{p_2+p_3}\right)$, el análisis anterior nos lleva a la siguiente ecuación:

$$H_3(p_1, p_2, p_3) = H_2(p_1, q) + qH_2\left(\frac{p_2}{q}, \frac{p_3}{q}\right), \quad (3.1)$$

donde $q = p_2 + p_3$. Es importante notar que en realidad estamos tratando de determinar las funciones H_n para toda n en los naturales y que además en la ecuación anterior, supusimos que las H_n son aditivas, condición que deberíamos añadir a la lista anterior de condiciones para las H_n .

Supongamos que ahora tenemos A_1, A_2, \dots, A_n eventos mutuamente excluyentes a las cuales le asignamos la distribución de probabilidad (p_1, p_2, \dots, p_n) . Agrupemos los eventos en k grupos de longitud a_i para $1 \leq i \leq k$,³ de tal forma que el primer grupo tenga los primeros a_1 eventos con probabilidad total w_1 , el segundo grupo tenga los siguientes a_2 eventos con probabilidad total w_2 y así sucesivamente. La cantidad de incertidumbre de las proposiciones compuestas es $H(w_1, w_2, \dots, w_k)$. Luego damos las probabilidades condicionales $(p_1/w_1, \dots, p_{a_1}/w_1)$ para las proposiciones (A_1, \dots, A_{a_1}) , dado que la proposición compuesta es cierta. La incertidumbre adicional, obtenida con probabilidad w_1 , es $H(p_1/w_1, \dots, p_{a_1}/w_1)$. Llevando a cabo este proceso para las proposiciones $(A_{a_r+1}, \dots, A_{a_r+a_2})$ para todas las $1 \leq r \leq k$ llegamos finalmente al mismo estado de conocimiento que si nos dieran las (p_1, \dots, p_n) directamente; la consistencia requiere que estos cálculos nos arrojen la misma incertidumbre, sin importar como se hayan agrupado las posibilidades. Entonces tenemos la ecuación de Shannon que dice lo siguiente:

$$\begin{aligned} H(p_1, p_2, \dots, p_n) &= H(w_1, w_2, \dots, w_n) + w_1 H\left(\frac{p_1}{w_1}, \dots, \frac{p_{a_1}}{w_1}\right) \\ &+ w_2 H\left(\frac{p_{a_1+1}}{w_2}, \dots, \frac{p_{a_1+a_2}}{w_2}\right) \\ &+ \dots, \end{aligned} \quad (3.2)$$

que es el caso general de la ecuación funcional en la ecuación (3.1). Por ejemplo:

$$H\left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right) = H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2}H\left(\frac{2}{3}, \frac{1}{3}\right).$$

³Obviamente $a_1 + a_2 + \dots + a_k = n$.

Por la continuidad de H basta determinar los valores que toma H en los racionales de la forma:

$$p_i = \frac{n_i}{\sum n_i}.$$

Pero entonces la ecuación (3.2) expresa estos valores de H en términos de los valores de $h(n) \equiv H(1/n, 1/n, \dots, 1/n)$ los cuales miden la cantidad de incertidumbre para el caso de tener n alternativas equiprobables. Ya que podemos ver la elección de los eventos (A_1, A_2, \dots, A_n) como el primer paso en la elección de uno de las:

$$\sum_{i=1}^n n_i$$

alternativas de igual probabilidad. Luego el segundo paso es la elección entre n_i posibilidades de igual probabilidad. Por ejemplo, con $n = 3$, podríamos escoger $n_1 = 3, n_2 = 4, n_3 = 2$. Para este caso la ecuación (3.2) nos da:

$$h(9) = H\left(\frac{3}{9}, \frac{4}{9}, \frac{2}{9}\right) + \frac{3}{9}h(3) + \frac{4}{9}h(4) + \frac{2}{9}h(2).$$

Para una elección en general de las n_i , la ecuación (3.2) se convierte en:

$$h\left(\sum_i n_i\right) = h(p_1, p_2, \dots, p_n) + \sum_i p_i h(n_i). \quad (3.3)$$

Si suponemos que todos los $n_i = m$ entonces la ecuación anterior se reduce a:

$$h(mn) = h(m) + h(n), \quad (3.4)$$

la cual tiene como solución:

$$h(n) = K \log(n),$$

donde K es una constante. ¿Pero será ésta la única solución? Las técnicas estándar de cálculo no son aplicables en este caso. En realidad ver que existen una infinidad de soluciones para la ecuación (3.4) se vuelve un divertido ejercicio de teoría de números. Basta con ver que como todo entero N se puede expresar de manera única como producto de primos, entonces por la ecuación (3.4) tenemos que $h(N)$ se puede expresar como $\sum_i m_i h(q_i)$ donde q_i son los primos y m_i son los exponentes no negativos correspondientes a los primos en

la descomposición de N . Al definir los valores de $h(q_i)$ arbitrariamente para todos los primos q_i tenemos una solución a la ecuación (3.4). Para obtener una solución única para $h(n)$ necesitamos la condición 3, el hecho que $h(n)$ es una función monótona creciente de n . Para ver esto, notar primero que (3.4) se puede extender por inducción a:

$$h(nmr \dots) = h(n) + h(m) + h(r) + \dots,$$

e igualando los factores de la extensión al orden k nos da:

$$h(n^k) = kh(n). \quad (3.5)$$

Ahora sean t y s dos enteros cualesquiera no menores que dos. Entonces para toda n arbitrariamente grande, existe un entero m tal que:

$$\frac{m}{n} \leq \frac{\log(t)}{\log(s)} < \frac{m+1}{n} \quad s^m \leq t^n < s^{m+1}, \quad (3.6)$$

Como h es monótona creciente tenemos que $h(s^m) \leq h(t^n) \leq h(s^{m+1})$. O más bien de (3.5) tenemos que:

$$mh(s) \leq nh(t) \leq (m+1)h(s), \quad (3.7)$$

lo cual se puede escribir como :

$$\frac{m}{n} \leq \frac{h(t)}{h(s)} \leq \frac{m+1}{n},$$

comparando (3.6) y (3.7), vemos que:

$$\left| \frac{h(t)}{h(s)} - \frac{\log(t)}{\log(s)} \right| \leq \frac{1}{n} \quad \left| \frac{h(t)}{\log(t)} - \frac{h(s)}{\log(s)} \right| \leq \epsilon.$$

donde

$$\epsilon \equiv \frac{h(s)}{n \log(t)},$$

es arbitrariamente pequeño. Luego entonces $\frac{h(t)}{\log(t)}$ es una constante y queda probada la unicidad de $h(n)$. Ahora los distintos valores de K solamente cambian la base sobre la cual se está tomando el logaritmo, por lo que podríamos sin pérdida de generalidad tomar $h(n) = \log(n)$. Sustituyendo esto

en la ecuación (3.3) tenemos que la única función que satisface las condiciones impuestas por nosotros sobre una medida razonable de la cantidad de incertidumbre es:

$$H(p_1, p_2, \dots, p_n) = - \sum_{i=1}^n p_i \log(p_i). \quad (3.8)$$

Aceptando esta interpretación, se sigue que la distribución (p_1, \dots, p_n) que maximiza a la expresión (3.8), sujeta a las restricciones impuestas por la información disponible, representará nuestra descripción más honesta acerca de lo que nosotros sabemos sobre las proposiciones (A_1, \dots, A_n) . La única ambigüedad es que tenemos la opción de tomar el logaritmo sobre la base que queramos, correspondiente a una constante que multiplica a H . Esto, por supuesto, no tiene ningún efecto sobre los valores (p_1, \dots, p_n) que maximizan H .

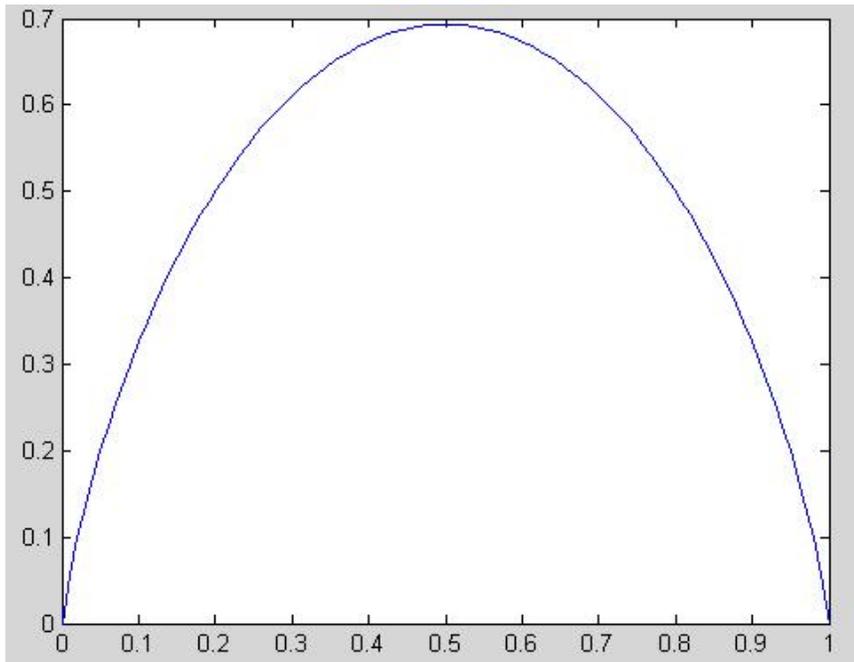


Figura 3.1: Entropía de una distribución Bernoulli(p).

La Derivación de Wallis

Supongamos que nos es dada cierta información I , la cual vamos a utilizar para asignar probabilidades $\{p_1, \dots, p_m\}$ a m distintas posibilidades. Por otra parte, al hacer esta asignación queremos utilizar únicamente la información I y las reglas de la probabilidad. Procederemos de la siguiente forma: Escojamos un entero n mucho mayor que m , y supongamos que tenemos n pequeñas «unidades» de probabilidad, de magnitud $1/n$, que podemos distribuir como mejor nos parezca entre las m posibilidades. Esta asignación se puede representar por medio de un pequeño juego: tenemos m cajas y n canicas; la idea es arrojar al azar las canicas. Para prevenir cualquier sesgo de parte del que tira las canicas, podemos pensar que utilizamos un grupo de monos, o bien podríamos tirarlas nosotros mismos con los ojos vendados. Después de tirar toda las canicas, contamos cuantas han caído en cada caja. Diremos que el experimento ha generado la distribución de probabilidades:

$$p_i = \frac{n_i}{n} \quad i = 1, \dots, m.$$

donde n_i representa el número de canicas en la i -ésima caja ⁴. La probabilidad de que caigan exactamente n_1 canicas en la primera caja, n_2 en la segunda, etc. se puede calcular con la distribución multinomial:

$$m^{-n} \frac{n!}{n_1! \cdots n_m!}. \quad (3.9)$$

Cada vez que realizamos el experimento, examinamos la asignación de probabilidades. Si es compatible con la información I , la aceptamos, en caso contrario la rechazamos y repetimos el experimento. Continuamos de esta forma hasta obtener una distribución compatible con I .

Ahora bien, ¿Qué distribución podemos esperar obtener de este juego?

En ausencia de restricciones, y dado que cada canica tiene la misma probabilidad de terminar en cualquiera de las m cajas, es natural suponer que la distribución más probable es la uniforme.

En general (i.e. cuando tenemos restricciones I), la distribución resultante tendría que ser aquella que maximice (3.9), y que por lo tanto maximice:

$$W = \frac{n!}{n_1! \cdots n_m!},$$

⁴Claramente $\sum_{i=1}^m n_i = n$.

sujeta a las restricciones que nos imponga I .

Podemos refinar este procedimiento, escogiendo una cantidad mayor de canicas n de un tamaño cada vez más chico. Utilicemos la aproximación de Stirling:

$$\log(n!) \approx n \log(n) - n + \sqrt{2\pi n} + \frac{1}{12n} + o\left(\frac{1}{n^2}\right).$$

Tomando el logaritmo y dividiendo entre n tenemos

$$\begin{aligned} \frac{1}{n} \log(W) &= \frac{1}{n} \left[n \log(n) - n + \sqrt{2\pi n} + \frac{1}{12n} + o\left(\frac{1}{n^2}\right) \right. \\ &\quad \left. - \sum_i \left(n_i \log(n_i) - n_i + \sqrt{2\pi n_i} + \frac{1}{12n_i} + o\left(\frac{1}{n_i^2}\right) \right) \right] \\ &= \frac{1}{n} \left[n \log(n) - \sum_i n_i \log(n_i) \right] + \frac{1}{n} \left[\sqrt{2\pi} \left(\sqrt{n} - \sum_i \sqrt{n_i} \right) \right] \\ &\quad + \frac{1}{12n} \left[\frac{1}{n} - \sum_i \frac{1}{n_i} \right] + \frac{1}{n} \left[o\left(\frac{1}{n^2}\right) - \sum_i \frac{1}{n_i} \right]. \end{aligned}$$

Tomando límite cuando $n \rightarrow \infty$, $n_i \rightarrow \infty$ de tal manera que $\frac{n_i}{n} \rightarrow p_i$ donde p_i es constante para toda i tenemos que $\frac{1}{n} \log(W) \rightarrow \sum_i p_i \log p_i$.

Desde un punto de vista conceptual, la derivación de Wallis es bastante atractiva. Es enteramente independiente de las ecuaciones funcionales de Shannon (3.2); no requiere de ningún postulado acerca de las conexiones entre la probabilidad y la frecuencia; ni presupone que las distintas posibilidades $\{1, 2, \dots, m\}$ son ellas mismas resultado de algún experimento aleatorio repetible, más aún, nos lleva automáticamente a la prescripción de que hay que maximizar H (y no tratarlo de otra manera).

3.3. El Principio de Máxima Entropía

3.3.1. Planteamiento de la Metodología

Supongamos que una variable x puede tomar n diferentes valores discretos (x_1, \dots, x_n) , los cuales corresponden a las proposiciones (A_1, A_2, \dots, A_n) ; y que hay m funciones distintas de x ,

$$f_k(x), \quad 1 \leq k \leq m,$$

que deseamos tengan esperanzas:

$$E[f_k(x)] = F_k \quad 1 \leq k \leq m, \quad (3.10)$$

donde los $\{F_k\}$ son números dados en el planteamiento del problema ⁵.
¿Qué probabilidades (p_1, \dots, p_n) asignaremos a las posibilidades (x_1, \dots, x_n) ?
Tendremos que

$$F_k = E[f_k(x)] = \sum_{i=1}^n p_i f_k(x_i),$$

y para encontrar el conjunto de las p_i 's que maximizan la entropía sujeto a todas las restricciones simultáneamente, introduciremos los multiplicadores de Lagrange necesarios:

$$\delta \left[H - (\lambda_0 - 1) \sum_i p_i - \sum_{j=1}^m \lambda_j \sum_i p_i f_j(x_i) \right] = \sum_i \left[\frac{\partial H}{\partial p_i} - (\lambda_0 - 1) - \sum_{j=1}^m \lambda_j f_j(x_i) \right] \delta p_i.$$

Como

$$\frac{\partial H}{\partial p_i} = -\log p_i - 1$$

nuestra solución es la siguiente:

$$p_i = \exp\left\{-\lambda_0 - \sum_{j=1}^m \lambda_j f_j(x_i)\right\}. \quad (3.11)$$

La suma de las probabilidades tiene que ser uno, así que:

$$1 = \sum_i p_i = \exp\{-\lambda_0\} \sum_i \exp\left\{-\sum_{j=1}^m \lambda_j f_j(x_i)\right\}. \quad (3.12)$$

Si definimos ahora la siguiente función:

$$Z(\lambda_1, \dots, \lambda_m) \equiv \sum_{i=1}^n \exp\left\{\sum_{j=1}^m \lambda_j f_j(x_i)\right\}, \quad (3.13)$$

entonces la ecuación (3.12) se reduce a:

⁵Las ecuaciones en (3.10) encapsulan la información previa que tenemos.

$$\lambda_0 = \log Z(\lambda_1, \dots, \lambda_m).$$

El valor esperado F_k tiene que ser igual a la esperanza de $f_k(x)$ sobre la distribución de probabilidad, es decir:

$$F_k = \exp\{-\lambda_0\} \sum_i f_k(x_i) \exp\left\{-\sum_{j=1}^m \lambda_j f_j(x_i)\right\},$$

o utilizando la ecuación (3.13)

$$F_k = -\frac{\partial \log Z(\lambda_1, \dots, \lambda_m)}{\partial \lambda_k}. \quad (3.14)$$

El valor máximo de la entropía es:

$$H_{max} = \left[-\sum_{i=1}^n p_i \log p_i \right]_{max},$$

y de la ecuación (3.11) tenemos que:

$$H_{max} = \lambda_0 + \sum_{j=1}^m \lambda_j F_j.$$

Ahora estos resultados abren camino a tantas nuevas aplicaciones que es importante tener una prueba lo más rigurosa posible. Pero resolver un problema de maximización por métodos variacionales, como acabamos de hacer, no es 100 % riguroso. Nuestro argumento de multiplicadores de Lagrange tiene la atractiva particularidad que nos da la respuesta instantáneamente. Tiene la mala particularidad que una vez que la hicimos, no estamos del todo seguros que el resultado es en realidad la solución. Ahora daremos una derivación totalmente diferente que es fuerte justo donde el argumento variacional es débil. Para esto necesitamos el siguiente resultado.

Lema 3.3.1. Sean $\{p_i\}_{i=1}^n$ cualquier conjunto de números que podrían representar una posible distribución de probabilidad, es decir,

$$\sum_{i=1}^n p_i = 1 \quad p_i \geq 0;$$

y sean $\{u_i\}$ otra posible distribución de probabilidad de nuevo tal que

$$\sum_{i=1}^n u_i = 1 \quad u_i \geq 0.$$

Entonces tenemos que

$$H(p_1, p_2, \dots, p_n) \leq \sum_{i=1}^n p_i \log \left(\frac{1}{u_i} \right)$$

con igualdad si y solo si $p_i = u_i$, $i=1, \dots, n$.

Demostración: Tenemos que,

$$\log(x) \leq (x - 1), \quad 0 \leq x < \infty$$

con igualdad si y solo si $x=1$. Entonces

$$\sum_{i=1}^n p_i \log \left(\frac{u_i}{p_i} \right) \leq \sum_{i=1}^n p_i \left(\frac{u_i}{p_i} - 1 \right) = 0.$$

Entonces tenemos que

$$-\sum_{i=1}^n p_i \log p_i \leq \sum_{i=1}^n p_i \log \left(\frac{1}{u_i} \right).$$

Q.E.D.

Ahora nos inventamos una distribución u_i «de la nada»:

$$u_i \equiv \frac{1}{Z(\lambda_1, \dots, \lambda_m)} \exp \left\{ - \sum_{j=1}^m \lambda_j f_j(x_i) \right\}, \quad (3.15)$$

donde $Z(\lambda_1, \dots, \lambda_m)$ está definida en la ecuación (3.13). Sin importar en este momento el por qué de la distribución anterior, aplicando el lema (3.3.1) tenemos que:

$$H \leq \sum_{i=1}^n p_i \left[\log Z(\lambda_1, \dots, \lambda_m) + \sum_{j=1}^m \lambda_j f_j(x_i) \right],$$

o equivalentemente

$$H \leq \log Z(\lambda_1, \dots, \lambda_m) + \sum_{j=1}^m \lambda_j E[f_j(x_i)]. \quad (3.16)$$

Ahora dejamos que las p_i varíen sobre la clase de distribuciones de probabilidad que satisfacen las restricciones en (3.10). El lado derecho de la ecuación (3.16) se mantiene constante. Nuestro lema nos dice que H obtiene su máximo en H_{max} , haciendo a la desigualdad en (3.16) una igualdad si y solo si las p_i se escogen como la distribución de pérdidas definida en (3.15).

Esta es nuestra prueba rigurosa, la cual es independiente de cualquier cosa que nos podría pasar si tratamos de probarlo mediante métodos variacionales. Este argumento es como podemos ver, fuerte en donde el método variacional es débil. Por otro lado este argumento es débil justo en donde el método variacional es fuerte, porque tuvimos que sacar la respuesta prácticamente de la nada en (3.15). Teníamos que saber la respuesta antes de probarla. Si tuviéramos ambos argumentos en nuestras manos, entonces tendríamos toda la historia.

3.3.2. Una Propiedad de las Distribuciones que Maximizan la Entropía

El máximo valor posible para H al fijar los valores de las esperanzas en (3.10) depende, por supuesto de los valores esperados que fijamos,

$$H_{max} = S(F_1, \dots, F_m) = \log Z(\lambda_1, \dots, \lambda_m) + \sum_{k=1}^m \lambda_k F_k. \quad (3.17)$$

Podemos ver a H como una medida de «la cantidad de incertidumbre» para cualquier distribución de probabilidad. Una vez que la hayamos maximizado, se convierte en una función de los datos del problema $\{F_i\}$, y llamaremos a este máximo $S(F_1, \dots, F_m)$. Sigue siendo una medida de «incertidumbre» pero es incertidumbre cuando toda la información que tenemos consiste en solamente estos números. Es completamente «objetiva» en el sentido que solamente depende de los datos del problema, y para nada en la personalidad o los deseos de alguien. Si S ha de ser una función solamente de (F_1, \dots, F_m) , entonces en (3.17) los $Z(\lambda_1, \dots, \lambda_m)$ necesariamente tienen que ser vistos

como funciones de (F_1, \dots, F_m) . Al principio, las λ_i 's no eran más que multiplicadores de Lagrange no especificados, pero eventualmente quisiéramos saber qué son.

Si escogiéramos distintas λ_k 's, estaríamos escribiendo diferentes distribuciones de probabilidad en (3.15), y vimos en la ecuación (3.14) que las esperanzas sobre estas distribuciones concuerdan con los valores esperados F_k si:

$$F_k = -\frac{\partial \log Z(\lambda_1, \dots, \lambda_m)}{\partial \lambda_k}, \quad k = 1, 2, \dots, m. \quad (3.18)$$

Las ecuaciones en (3.18) son un conjunto de m ecuaciones no-lineales simultáneas, las cuales necesitan ser resueltas para las λ_k 's en términos de las F_k . Por lo general, en un problema no trivial, no es práctico resolver para las λ_k 's explícitamente. Dejamos las λ_k 's como están, expresando las cosas de forma paramétrica. Sin embargo, si podemos evaluar la función $S(F_1, \dots, F_m)$ explícitamente, entonces podemos expresar las λ_k 's como funciones explícitas de las $\{F_k\}$ como sigue: Supongamos que realizamos un pequeño cambio en una de las F_k . ¿De qué manera cambia esto el máximo obtenible para H ? Tenemos de la ecuación (3.17) que:

$$\frac{\partial S(F_1, \dots, F_m)}{\partial F_k} = \sum_{j=1}^m \left[\frac{\partial \log Z(\lambda_1, \dots, \lambda_m)}{\lambda_j} \right] \left[\frac{\partial \lambda_j}{\partial F_k} \right] + \sum_{j=1}^m \frac{\partial \lambda_j}{\partial F_k} + \lambda_k,$$

el cual gracias a (3.18) se colapsa en:

$$\lambda_k = \frac{\partial S(F_1, \dots, F_m)}{\partial F_k}, \quad (3.19)$$

en el cual la λ_k se da explícitamente.

Comparemos las ecuaciones (3.19) con (3.18), una nos da F_k explícitamente de las λ_k , la otra nos da las λ_k explícitamente en términos de las F_k . Especificando $\log Z(\lambda_1, \dots, \lambda_m)$ o $S(\lambda_1, \dots, \lambda_m)$ es equivalente en el sentido que cada una nos da toda la información acerca de la distribución de probabilidad.

3.4. El Principio de Mínima Entropía Cruzada

3.4.1. Planteamiento de la Metodología

Una extensión al razonamiento de Jaynes es el principio de mínima entropía cruzada desarrollado por Kullback [19]. En esta generalización del principio de máxima entropía, se asume que en adición a la información proporcionada por los datos en forma de esperanzas, existe una noción intuitiva acerca de la forma de la distribución a-posteriori, el cual se puede expresar como una distribución a-priori q . En contraste al principio de máxima entropía y su metodología, el objetivo ahora será minimizar la entropía cruzada entre la distribución a-posteriori p y la a-priori q . La función objetivo del problema de optimización, es decir, la entropía cruzada de dos distribuciones $p = \{p_1, p_2, \dots, p_n\}$ y $q = \{q_1, q_2, \dots, q_n\}$ se define como sigue:

$$C[p, q] = \sum_{i=1}^n p_i \ln \left[\frac{p_i}{q_i} \right], \quad (3.20)$$

el principio de mínima entropía cruzada propone seleccionar la distribución a-posteriori que minimiza la función objetivo sujeto a las restricciones:

$$\sum_{i=1}^n p_i f_k(x_i) = F_k \quad y \quad \sum_{i=1}^n p_i = 1 \quad para \quad 1 \leq k \leq m < n.$$

Consecuentemente, el vector de probabilidad p se puede recuperar minimizando el siguiente lagrangiano:

$$L = \sum_{i=1}^n p_i \ln \left[\frac{p_i}{q_i} \right] + \sum_{k=1}^m \lambda_k \left[F_k - \sum_{i=1}^n p_i f_k(x_i) \right] + \mu \left[1 - \sum_{i=1}^n p_i \right],$$

donde las λ_k y μ representan los multiplicadores de Lagrange. Utilizando el método de multiplicadores de Lagrange, la solución óptima del problema de entropía cruzada es:

$$\hat{p}_i = \frac{q_i}{\sum_{i=1}^n q_i \exp \left[\sum_{k=1}^m \hat{\lambda}_k f_k(x_i) \right]} \exp \left[\sum_{k=1}^m \hat{\lambda}_k f_k(x_i) \right]. \quad (3.21)$$

La entropía H , presentada en la ecuación (3.8), como vimos anteriormente, se puede interpretar como una medida del nivel de incertidumbre en las

distribuciones de probabilidad, la entropía H toma su máximo cuando la distribución se dispersa al máximo, el cual es el caso de la distribución uniforme en el rango de la distribución. Vemos entonces que el principio de máxima entropía es un caso particular del principio de mínima entropía cruzada en el sentido que maximizar la entropía H sujeto a la información que proporcionan los datos disponibles es equivalente a minimizar la entropía cruzada entre la distribución a-posteriori y la distribución uniforme en el rango de las distribuciones en cuestión. Es decir, el principio de máxima entropía no es más que el principio de mínima entropía cruzada en el caso en el que la distribución a-priori es la distribución uniforme. Lo anterior se puede ver también en como la ecuación (3.21) se reduce a la ecuación (3.11) cuando $q(x)$ es constante, es decir cuando q es una distribución uniforme.

El principio de mínima entropía cruzada se puede plantear también en el caso continuo de la siguiente manera:

La entropía cruzada para distribuciones continuas ⁶ se define de la siguiente manera:

$$C[p, q] = \int \int p(x, y) \ln \left[\frac{p(x, y)}{q(x, y)} \right] dx dy, \quad (3.22)$$

se desea recuperar la distribución a-posteriori p que minimice la función en (3.22) sujeta a las m restricciones:

$$\int f_k(x, y) p(x, y) dx dy = F_k \quad y \quad \int p(x, y) dx dy = 1.$$

Consecuentemente podemos recuperar la distribución continua p al minimizar el siguiente lagrangiano:

$$\begin{aligned} L[p, q] &= \int \int p(x, y) \ln p(x, y) dx dy - \int \int p(x, y) \ln q(x, y) dx dy \\ &+ \sum_{k=1}^m \lambda_k \left[F_k - \int \int f_k(x, y) p(x, y) dx dy \right] \\ &+ \mu \left[1 - \int \int p(x, y) dx dy \right], \end{aligned}$$

donde las λ_k y μ representan los multiplicadores de Lagrange. Utilizando el

⁶Por simplicidad nos concentramos en el caso de \mathbb{R}^2 .

cálculo de variaciones llegamos a que la solución óptima es ⁷:

$$\hat{p}(x, y) = q(x, y) \exp \left\{ - \left[1 + \mu + \sum_{k=1}^m \lambda_k f_k(x, y) \right] \right\}.$$

3.4.2. Caracterización Axiomática de la Entropía Cruzada: Caso Discreto

Para el principio de máxima entropía derivamos la entropía como una medida del nivel de incertidumbre de una distribución de probabilidad. Desde luego que la definición de entropía cruzada en la ecuación (3.20) no fue «sacada de la manga». En esta sección mostraremos una derivación axiomática de la función de entropía cruzada para el caso discreto muy al estilo de Shannon, desarrollada por Arthur Hobson en [2].

Interpretación Intuitiva de la Entropía Cruzada

Para realizar la derivación de la expresión en (3.20), primero tenemos que dar una interpretación intuitiva acerca de lo que quisiéramos que la entropía cruzada mida:

Consideremos un observador que está interesado en algún experimento con posibles resultados: z_1, z_2, \dots, z_n , pero que no sabe el preciso resultado z_i . Supongamos que el «estado de conocimiento» inicial del observador acerca del experimento se describe con las probabilidades a-priori p_1^0, \dots, p_n^0 . Ahora, supongamos que a nuestro observador se le entrega un mensaje que contiene nuevos datos acerca del resultado, y que el «estado de conocimiento» después de haber recibido el mensaje se describe con las probabilidades posteriores p_1, \dots, p_n . Buscamos una expresión para medir la cantidad de información I en el mensaje. La cantidad de información debería ser una función $I(p_1, \dots, p_n; p_1^0, \dots, p_n^0)$ de la distribución a-priori y la a-posteriori.

Caracterización de la Entropía Cruzada

Hobson propuso los siguientes axiomas que debería de cumplir I :

1. I es una función continua de sus $2n$ variables.

⁷Para mayor detalle ver el Apéndice A.

2. Para cualesquier par de distribuciones a-priori p_1^0, \dots, p_n^0 y a-posteriori p_1, \dots, p_n se cumple que para toda $1 \leq j \leq k \leq n$ se tiene que:

$$I(p_1, \dots, p_j, \dots, p_k, \dots, p_n; p_1^0, \dots, p_j^0, \dots, p_k^0, \dots, p_n^0) = I(p_1, \dots, p_k, \dots, p_j, \dots, p_n; p_1^0, \dots, p_k^0, \dots, p_j^0, \dots, p_n^0).$$

3. $I = 0$ si $p = p_i^0$ para toda $1 \leq i \leq n$.
4. Para cualquier pareja de enteros n, n_0 tales que $n_0 \geq n$, la expresión $I(1/n, \dots, 1/n, 0, \dots, 0; 1/n_0, \dots, 1/n_0)$ es una función creciente de n_0 y una decreciente de n .
5. Sea $S = \{z_1, \dots, z_n\}$ el espacio de resultados del experimento en cuestión. Supongamos que S se divide en dos eventos $E_1 = \{z_1, \dots, z_r\}$ y $E_2 = \{z_{r+1}, \dots, z_n\}$. Las probabilidades a-posteriori de estos eventos son:

$$P(E_1) = q_1 = p_1 + \dots + p_r, \quad P(E_2) = q_2 = p_{r+1} + \dots + p_n, \quad (3.23)$$

mientras que las probabilidades a-priori de E_1 y E_2 son:

$$P^0(E_1) = q_1^0 = p_1^0 + \dots + p_r^0, \quad P^0(E_2) = q_2^0 = p_{r+1}^0 + \dots + p_n^0.$$

Las probabilidades a-posteriori de las z_i , dado que ocurrió el evento E_1 , son:

$$P^0(z_i|E_1) = \frac{p_i^0}{q_1^0} \quad \text{para toda } 1 \leq i \leq r. \quad (3.24)$$

De igual manera:

$$P(z_i|E_2) = \frac{p_i}{q_2} \quad \text{para toda } 1 \leq i \leq r \quad (3.25)$$

$$P^0(z_i|E_2) = \frac{p_i^0}{q_2^0} \quad \text{para toda } 1 \leq i \leq r. \quad (3.26)$$

Ahora, la información acerca del resultado del experimento se puede dar especificando las probabilidades p_1, \dots, p_n directamente o especificando las probabilidades en (3.23) de los dos eventos E_1 y E_2 , y luego dar las probabilidades condicionales en (3.24), (3.25) y (3.26). Estas dos

alternativas se muestran gráficamente en las Figuras 3.2 y 3.3. Ahora damos el siguiente axioma:

$$\begin{aligned} & I(p_1, \dots, p_r, p_{r+1}, \dots, p_n; p_1^0, \dots, p_r^0, p_{r+1}^0, \dots, p_n^0) \\ &= I(q_1, q_2; q_1^0, q_2^0) + q_1 I(p_1/q_1, \dots, p_r/q_1; p_1^0/q_1^0, \dots, p_r^0/q_1^0) \\ & \quad + q_2 I(p_{r+1}/q_2, \dots, p_n/q_2; p_{r+1}^0/q_2^0, \dots, p_n^0/q_2^0). \end{aligned}$$

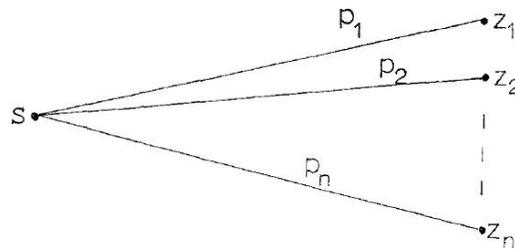


Figura 3.2: Alternativa 1

Intuitivamente los axiomas anteriores dicen lo siguiente:

1. El axioma 1 dice que la información cambia solamente un poco cuando las probabilidades se cambian también un poco.
2. El axioma 2 dice que la información no depende del orden en el que se acomoden los posibles resultados z_i .
3. El axioma 3 dice que no hay información obtenida si el mensaje no cambia las probabilidades.
4. El axioma 4 dice que la cantidad de información es mayor cuando la cantidad n_0 de posibilidades a-priori es mayor y/o la cantidad de probabilidades a-posteriori es menor.
5. El axioma 5 dice que la cantidad de información en la alternativa, representada por la Figura 3.2, es igual a la cantidad de información en la alternativa equivalente de la Figura 3.3, más aún, que el lado derecho de la igualdad en el axioma 5 es una expresión razonable para la cantidad de información en la alternativa equivalente de la Figura 3.3. La

idea de este axioma es la suposición que la información obtenida en la alternativa de la Figura 3.3, es la suma de la información $I(q_1, q_2; q_1^0, q_2^0)$ correspondiente al primer paso (es decir, correspondiente a cuál de los eventos E_1 y E_2 ocurrió) más la información ponderada (por los pesos E_1 y E_2) del segundo paso. Esto suena bastante razonable si uno lo analiza un rato; o mínimo suena más razonable que cualquier otra expresión que represente la información obtenida en la alternativa de la Figura 3.3.

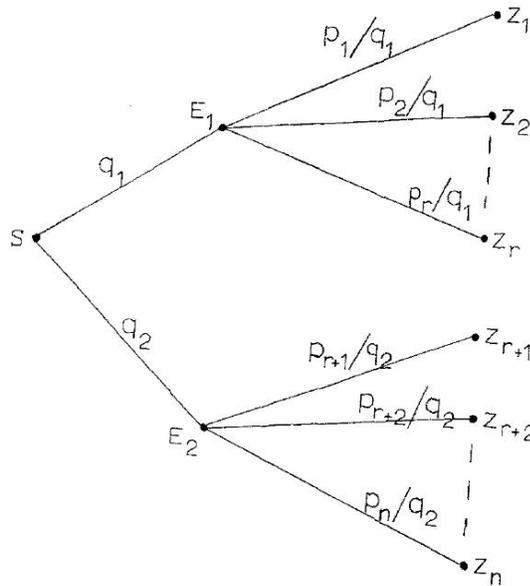


Figura 3.3: Alternativa 2

Tenemos ya la interpretación que deseamos acerca de lo que quisiéramos que nuestra función I satisficiera. Utilizando esta intuición postulamos algunos axiomas los cuales nos quedan claro a-priori que I tiene que cumplir. Utilizando estas propiedades caracterizamos nuestra medida de información I , mediante el siguiente teorema:

Teorema 3.4.1. *Sea $I(p_1, \dots, p_n; p_1^0, \dots, p_n^0)$ una función definida para cualquier par de distribuciones de probabilidad (p_1, \dots, p_n) y (p_1^0, \dots, p_n^0) sobre un espacio de probabilidad finito $S = \{z_1, \dots, z_n\}$, y supongamos que I satisface*

las propiedades 1-5. Entonces tenemos que:

$$I = (p_1, \dots, p_n; p_1^0, \dots, p_n^0) = k \sum_{i=1}^n p_i \ln \left(\frac{p_i}{q_i} \right), \quad (3.27)$$

donde k es una constante positiva.

El teorema anterior dice que (3.27) es la única expresión que tiene las propiedades deseables 1-5. El teorema de unicidad de Shannon para la expresión:

$$H(p_1, p_2, \dots, p_n) = - \sum_{i=1}^n p_i \ln p_i, \quad (3.28)$$

es similar al teorema de unicidad de arriba. Sin embargo, el teorema de Shannon (y su prueba) es un tanto menos complicado que el presente teorema, ya que (3.28) involucra solamente una distribución de probabilidad, mientras que (3.27) involucra dos distribuciones.

Primero probaremos que: ⁸

$$\begin{aligned} I(q_1, \dots, q_{n-1}, q_n; \dots) &= I(q_1, \dots, q_{n-2}, q_{n-1} + q_n; \dots) \\ &+ (q_{n-1} + q_n) I \left(\frac{q_{n-1}}{q_{n-1} + q_n}, \frac{q_n}{q_{n-1} + q_n}; \dots \right). \end{aligned} \quad (3.29)$$

Demostración: Usando el axioma 5 con $Q \equiv q_1 + \dots + q_{n-2}$, tenemos que:

$$\begin{aligned} I(q_1, \dots, q_n; \dots) &= I(Q, q_{n-1} + q_n; \dots) + Q I \left(\frac{q_1}{Q}, \dots, \frac{q_{n-2}}{Q}; \dots \right) \\ &+ (q_{n-1} + q_n) I \left(\frac{q_{n-1}}{q_{n-1} + q_n}, \frac{q_n}{q_{n-1} + q_n}; \dots \right). \end{aligned} \quad (3.30)$$

De nuevo utilizando el axioma 5 tenemos que:

$$\begin{aligned} I(q_1, \dots, q_{n-2}, q_{n-1} + q_n; \dots) &= I(Q, q_{n-1} + q_n; \dots) \\ &+ Q I \left(\frac{q_1}{Q}, \dots, \frac{q_{n-2}}{Q}; \dots \right) + (q_{n-1} + q_n) I(1; 1), \end{aligned} \quad (3.31)$$

⁸Guiones que aparezcan a la derecha del punto y coma siempre significarán que las variables a la derecha del punto y coma son idénticas a las que están a la izquierda, nadamás que con el superíndice «0» anexado. Por ejemplo, $I(p_1, p_2; \dots)$ significa $I(p_1, p_2; p_1^0, p_2^0)$.

y utilizando el axioma 3 tenemos que $I(1; 1) = 0$. Combinando (3.30) y (3.31) llegamos a (3.29). *Q.E.D.*

Ahora probaremos que:

$$I(p_{11}, \dots, p_{1m}, p_{n1}, \dots, p_{nm}; \dots) = I(q_1, \dots, q_n; \dots) \quad (3.32)$$

$$+ \sum_{i=1}^n q_i I\left(\frac{p_{i1}}{q_i}, \dots, \frac{p_{im}}{q_i}; \dots\right),$$

donde $q_i \equiv p_{i1} + \dots + p_{im}$.

Demostración: Del axioma 5 tenemos que (3.32) se vale para $n = 2$. Procederemos por inducción. Supongamos que (3.32) se vale para n , tenemos entonces que:

$$I(p_{11}, \dots, p_{1m}, \dots, p_{n+1,1}, \dots, p_{n+1,m}; \dots) = \quad (3.33)$$

$$I(q_1, \dots, q_{n-1}, q_n + q_{n+1}; \dots) + \sum_{i=1}^n q_i I\left(\frac{p_{i1}}{q_i}, \dots, \frac{p_{im}}{q_i}; \dots\right)$$

$$+ \left(\frac{p_{n1}}{q_n + q_{n+1}}, \dots, \frac{p_{nm}}{q_n + q_{n+1}}, \frac{p_{n+1,1}}{q_n + q_{n+1}}, \dots, \frac{p_{n+1,m}}{q_n + q_{n+1}}; \dots\right).$$

De (3.29), el primer término en el lado derecho de (3.33) se puede escribir como:

$$I(q_1, \dots, q_n, q_{n+1}; \dots) - (q_n + q_{n+1}) I\left(\frac{q_n}{q_n + q_{n+1}}, \frac{q_{n+1}}{q_n + q_{n+1}}; \dots\right). \quad (3.34)$$

Del axioma 5, el último término del lado derecho de (3.33) se puede escribir como:

$$(q_n + q_{n+1}) I\left(\frac{q_n}{q_n + q_{n+1}}, \frac{q_{n+1}}{q_n + q_{n+1}}; \dots\right)$$

$$+ q_n I\left(\frac{p_{n1}}{q_n}, \dots, \frac{p_{nm}}{q_n}; \dots\right)$$

$$+ q_{n+1} I\left(\frac{p_{n+1,1}}{q_{n+1}}, \dots, \frac{p_{n+1,m}}{q_{n+1}}; \dots\right).$$

Insertando estas dos expresiones en (3.33), obtenemos que la ecuación (3.32) es válida para $n + 1$; de aquí que la ecuación (3.32) es válida para toda n . *Q.E.D.*

Denotaremos: ⁹

$$f(r, r_0) \equiv I(1/r, \dots, 1/r, 0, \dots, 0; 1/r_0, \dots, 1/r_0). \quad (3.35)$$

De (3.32), con $m = r_0$ y $n = s_0$, tenemos que:

$$\begin{aligned} I(p_{11} &= 1/rs, \dots, p_{1r} = 1/rs, p_{1,r+1} = 0, \dots, p_{1r_0} = 0, \dots, \\ & p_{s1} = 1/rs, \dots, p_{sr} = 1/rs, p_{s,r+1} = 0, \dots, \\ & p_{sr_0} = 0, p_{s+1,1} = 0, \dots, p_{s_0r_0} = 0; 1/r_0s_0, \dots, 1/r_0s_0) \\ &= I(1/s, \dots, 1/s, 0, \dots, 0; 1/s_0, \dots, 1/s_0) \\ &+ s \left(\frac{1}{s} \right) I(1/r, \dots, 1/r, 0, \dots, 0; 1/r_0, \dots, 1/r_0). \end{aligned}$$

De (3.35) y el axioma 2, la ecuación anterior se convierte en:

$$f(rs, r_0s_0) = f(s, s_0) + f(r, r_0). \quad (3.36)$$

De (3.36) y el axioma 2:

$$f(rs, rs_0) = f(s, s_0) + f(r, r) = f(s, s_0). \quad (3.37)$$

Dejando que $\frac{r_0}{r} = \frac{r'_0}{r'} \equiv u$, (3.37) implica que:

$$f(r, r_0) = f(r, ur) = f(r_0, r'_0r/r') = f(r'_0r, r'_0r) = f(r', r'_0). \quad (3.38)$$

La ecuación (3.38) nos dice que $f(r, r_0)$ depende solamente de la razón r_0/r . Entonces

$$g(r_0/r) \equiv f(r, r_0), \quad (3.39)$$

define una nueva función $g(x)$, para todo racional $x \geq 1$. De (3.39) y (3.37), definiendo $x = \frac{r_0}{r}$ y $x' = \frac{r'_0}{r'}$,

$$g(xx') = g(r_0r'_0/rr') = f(rr'/r_0r'_0) = f(r, r_0) + f(r', r'_0) = g(r_0/r) + g(r'_0/r').$$

Entonces:

$$g(xx') = g(x) + g(x'), \quad (3.40)$$

para todo racional $x \geq 1$ y $x' \geq 1$. Del axioma 4, $g(x)$ es función monótona creciente de x .

⁹Los símbolos r , r_0 , s y s_0 siempre serán enteros positivos, con $r_0 \geq r$ y $s_0 \geq s$.

Ahora probaremos que:

$$f(r, r_0) = g(r_0/r) = k \ln(r_0/r), \quad (3.41)$$

donde k es una constante positiva.

Demostración: De (3.40) tenemos que:

$$g(r_0^n) = ng(r_0). \quad (3.42)$$

Sean n , r_0 y s_0 enteros definidos arbitrariamente, y sea m el único entero positivo tal que:

$$m \leq n \frac{\ln s_0}{\ln r_0} < m + 1, \quad (3.43)$$

de donde:

$$r_0^m \leq s_0^n \leq r_0^{m+1} \quad (3.44)$$

$$\frac{m}{n} \leq \frac{\ln s_0}{\ln r_0} \leq \frac{m+1}{n}. \quad (3.45)$$

Como g es una función monótona creciente, (3.42) y (3.44) implican que:

$$m g(r_0) \leq ng(s_0) \leq (m+1)g(r_0), \quad (3.46)$$

de tal manera que:

$$\frac{m}{n} \leq \frac{g(s_0)}{g(r_0)} \leq \frac{m+1}{n}. \quad (3.47)$$

De (3.45) y (3.47) tenemos que:

$$\left| \frac{g(s_0)}{g(r_0)} - \frac{\ln s_0}{\ln r_0} \right| \leq \frac{1}{n}. \quad (3.48)$$

Como el lado izquierdo de (3.48) es independiente de n , y como la n puede ser escogida arbitrariamente grande. La ecuación (3.48) implica que:

$$\frac{g(s_0)}{g(r_0)} = \frac{\ln s_0}{\ln r_0}, \quad (3.49)$$

o más bien

$$g(r_0) = k \ln r_0. \quad (3.50)$$

Como g es una función monótona creciente, la constante k tiene que ser positiva. Esto prueba (3.41) para $r = 1$ y cualquier r_0 . Sustituyendo r_0 por $\frac{r+1}{r}$

y s_0 por $\frac{s+1}{s}$ y utilizando exactamente el mismo razonamiento que llevamos en las ecuaciones (3.42)-(3.50) se puede ver que:

$$g\left(\frac{r+1}{r}\right) = k \ln\left(\ln\frac{r+1}{r}\right). \quad (3.51)$$

Ahora probaremos (3.41) por inducción sobre r . Supongamos que (3.41) se vale para (r_0, r) , probémoslo para $(r_0, r+1)$. De las ecuaciones (3.40) y (3.51) tenemos que:

$$\begin{aligned} g\left(\frac{r_0}{r}\right) &= g\left(\frac{r_0}{r+1} \frac{r+1}{r}\right) = g\left(\frac{r_0}{r+1}\right) + g\left(\frac{r+1}{r}\right), \\ g\left(\frac{r_0}{r+1}\right) &= g\left(\frac{r_0}{r}\right) - g\left(\frac{r+1}{r}\right) = k \ln\frac{r_0}{r} - k \ln\frac{r+1}{r} = k \ln\left(\frac{r_0}{r+1}\right). \end{aligned}$$

Entonces, si (3.41) se vale para (r_0, r) , también se vale para $(r_0, r+1)$. Pero por (3.50), la ecuación (3.41) se vale para $(r_0, 1)$. De aquí que (3.41) se vale para todo (r_0, r) .

Ahora sean p y p_0 dos números racionales menores o iguales a 1: $p = \frac{r}{s}$, $p_0 = \frac{r_0}{s_0} \leq 1$. Si las tres condiciones $r_0 \geq r, s_0 \geq s$ y $s_0 - r_0 \geq s - r$, no se valen para (r, s, r_0, s_0) , entonces sustituiremos (r_0, s_0) por (nr_0, ns_0) , con n suficientemente grande de tal manera que se valgan las tres condiciones anteriores para (r, s, nr_0, ns_0) . Este reemplazo no afecta al valor de p ni al de p_0 . Entonces se puede asumir sin pérdida de generalidad que $r_0 \geq r, s_0 \geq s$ y $s_0 - r_0 \geq s - r$. De el axioma 1 y (3.32), con un agrupamiento apropiado de las variables:

$$\begin{aligned} I(1/s, \dots, 1/s, 0, \dots, 0; 1/s_0, \dots, 1/s_0) &= I\left[\frac{r}{s}, \frac{(s-r)}{s}, \frac{r_0}{s_0}, \frac{(s_0-r_0)}{s_0}\right] \\ &+ \left(\frac{r}{s}\right) I(1/r, \dots, 1/r, 0, \dots, 0; 1/r_0, \dots, 1/r_0) \\ &+ [(s-r)/s] I\left[\frac{1}{(s-r)}, \dots, \frac{1}{(s-r)}, 0, \dots, 0; \frac{1}{(s_0-r_0)}, \dots, \frac{1}{(s_0-r_0)}\right]. \end{aligned}$$

De (3.35) y (3.41), la ecuación anterior se convierte en:

$$\begin{aligned} I(p, 1-p; p^0, 1-p^0) &= f(s, s_0) - pf(r, r_0) - (1-p)f(s-r, s_0-r_0) \\ &= kp \ln \frac{p}{p_0} + k(1-p) \ln \left[\frac{1-p}{1-p_0} \right]. \end{aligned}$$

Del axioma 1 se extiende el resultado para todo irracional p . Esto prueba el teorema para $n = 2$. Utilizando (3.29) el teorema se vale por inducción para toda n . *Q.E.D.*

3.4.3. Intuición: Principio de Mínima Entropía Cruzada

La interpretación de la función $I(p, q)$ que dimos en la sección pasada conceptualiza a I como la cantidad de información obtenida al pasar de una distribución de probabilidad a-priori q a una distribución a-posteriori p . El principio de mínima entropía cruzada, de manera análoga al raciocinio de Jaynes, propone seleccionar la distribución a-posteriori p como la distribución que, sujeta a las condiciones impuestas por la información que proporcionan los datos disponibles al observador, minimice la cantidad de información obtenida al pasar de la distribución de probabilidad q a la distribución a-posteriori p . Notar que el principio de mínima entropía cruzada generaliza el principio de máxima entropía. El principio de mínima entropía cruzada se convierte en el principio de máxima entropía en el caso en el que la distribución a-priori es una uniforme. En este sentido, los argumentos que soportan el uso del principio de máxima entropía son igual de aplicables para este nuevo.

3.5. CIMDO

Inferir la distribución multivariada de pérdidas del sistema financiero es uno de nuestros principales objetivos. Por lo general para hacer esta inferencia solamente contamos con información parcial acerca de la distribución; comunmente la única información con la que contamos son las probabilidades de distress de cada una de las instituciones financieras. Resulta en este momento natural utilizar el enfoque de la entropía cruzada de Kullback para hacer la inferencia. CIMDO basado en el principio de mínima entropía cruzada, nos proporciona una metodología para inferir la distribución multivariada de pérdidas, $p(l^1, l^2, \dots, l^M) \in \mathbb{R}^M$ de un sistema financiero que consta de M instituciones financieras, a la cual de ahora en adelante llamaremos «distribución a-posteriori». Para dicho objetivo, bajo una hipótesis inicial se asumirá que la distribución obedece una cierta forma paramétrica $q(l^1, l^2, \dots, l^M) \in \mathbb{R}^M$ que de ahora en adelante llamaremos «distribución a-

priori». Es importante hacer notar que la hipótesis inicial se hace utilizando nuestra intuición económica y no necesariamente mediante información empírica. Finalmente siguiendo el principio de mínima entropía cruzada, CIMDO recupera la distribución a-posteriori. Esta distribución es la más cercana a la distribución a-priori consistente con las probabilidades de distress observables en el mercado.

3.5.1. Función Objetivo

Para un portafolio de M instituciones financieras l^1, l^2, \dots, l^M definiremos la función objetivo como:

$$C [p, q] = \int_{l^M} \cdots \int_{l^1} p(l^1, \dots, l^M) \ln \left[\frac{p(l^1, \dots, l^M)}{q(l^1, \dots, l^M)} \right] dl^1 \dots dl^M, \quad (3.52)$$

donde $q(l^1, \dots, l^M)$ la distribución a-priori y $p(l^1, \dots, l^M)$ la distribución a-posteriori $\in \mathbb{R}$.

Por simplicidad, de ahora en adelante, desarrollaremos el caso multivariado como uno bivariado, no obstante los resultados son aplicables para el caso de \mathbb{R}^M cuando $M \geq 2$. Nos enfocaremos en un portafolio de dos instituciones financieras X y Y . Podemos entonces definir nuestra función objetivo como:

$$C [p, q] = \int \int p(x, y) \ln \left[\frac{p(x, y)}{q(x, y)} \right] dx dy,$$

donde $q(x, y)$ la distribución a-priori y $p(x, y)$ la distribución a-posteriori $\in \mathbb{R}$.

3.5.2. La Distribución A-priori

Tomando como parámetro inicial el modelo estructural de Merton [20] y asumiendo que sus principales suposiciones e intuición económica son correctas, fijaremos como hipótesis inicial que el portafolio $q(x, y) \in \mathbb{R}$ obedece una distribución t-student bivariada con cinco grados de libertad. Enfatizamos aquí que la hipótesis inicial suena razonable en un momento previo al análisis de información que tenemos disponible. Desde luego que después de haber analizado la información es posible que la distribución a-priori $q(x, y)$ no sea consistente con dicha información.

3.5.3. Ecuaciones de Consistencia con la Información

Como se mencionó arriba, la distribución a-priori se escoge debido a su consistencia con argumentos teóricos e intuición económica y no necesariamente mediante observaciones empíricas. Es por esto que la información proporcionada por las probabilidades de distress son de extrema importancia para la recuperación de la distribución a-posteriori. Para poder incorporar esta información a la distribución a-posteriori formulamos las ecuaciones de consistencia. Intuitivamente quisiéramos actualizar la forma de la distribución del sistema financiero en cada período en el tiempo, de tal manera que la distribución a-posteriori sea consistente con las probabilidades de distress de las instituciones. En otras palabras, la distribución satisface un conjunto de restricciones impuestas en sus distribuciones marginales. Las ecuaciones de consistencia son las siguientes:

$$\int_{-\infty}^{\infty} \int_{X_d^x}^{\infty} p(x, y) dx dy = PoD_t^x, \quad (3.53)$$

$$\int_{-\infty}^{\infty} \int_{X_d^y}^{\infty} p(x, y) dx dy = PoD_t^y, \quad (3.54)$$

donde X_d^x, X_d^y representa los umbrales de default. Estas restricciones implican que en cada período en el tiempo la región de default para cada institución financiera tiene un peso consistente con su respectiva probabilidad de distress. Notar que la ecuaciones de consistencia contienen toda la información que tenemos disponible para cada institución financiera. Esta información se incluye en las ecuaciones de consistencia vía el umbral de distress y las probabilidades distress empíricas. La imposición de estas restricciones al problema de optimización garantiza que la distribución a-posteriori se infiera utilizando la información que tenemos disponible.

3.5.4. Umbral de Distress

Dada la distribución a-priori asumida caracterizamos el umbral de default de cada institución como:

$$X_d^x = \Phi^{-1}(\bar{\alpha}^x) \text{ y } X_d^y = \Phi^{-1}(\bar{\alpha}^y),$$

donde $\bar{\alpha}^x = 1 - PoD^x$ y $\bar{\alpha}^y = 1 - PoD^y$ ya que en nuestro modelo la región de default para cada institución se encuentra en la parte superior de

la distribución y ϕ es la función de distribución de una t-student bivariada con cinco grados de libertad.

3.5.5. Planteamiento de la Metodología

El Consistent Information Multivariate Density Optimizing recupera la distribución a-posteriori $p(x, y) \in \mathbb{R}$ minimizando la función objetivo definida en (3.52) sujeta a las condiciones impuestas por las ecuaciones de consistencia definidas en (3.53) y (3.54). Con el fin de asegurar que $p(x, y)$ representa una distribución válida, haremos que las condiciones $p(x, y) \geq 0$ y $\int p(x, y) = 1$, también se satisfagan.

La densidad CIMDO se recupera minimizando el siguiente funcional:

$$\begin{aligned}
 L[p, q] &= \int \int p(x, y) \ln p(x, y) dx dy \\
 &- \int \int p(x, y) \ln q(x, y) dx dy \\
 &+ \lambda_1 \left[\int \int p(x, y) \chi_{[X_d^x, \infty)} dx dy - PoD_t^x \right] \\
 &+ \lambda_2 \left[\int \int p(x, y) \chi_{[X_d^y, \infty)} dx dy - PoD_t^y \right] \\
 &+ \mu \left[\int \int p(x, y) dx dy - 1 \right],
 \end{aligned}$$

donde $p(x, y)$ es la distribución multivariada a-posteriori, la variable por recuperarse, y $q(x, y)$ es la distribución multivariada a-priori. Las λ_1 y λ_2 representan los multiplicadores de Lagrange de las ecuaciones de consistencia definidas en (3.53) y (3.54), μ representa el multiplicador de Lagrange de las condiciones impuestas para garantizar que $p(x, y)$ sea una distribución válida.

El funcional anterior se puede reescribir como:

$$\begin{aligned}
 L[p, q] &= \int \int p(x, y) [\ln p(x, y) - \ln q(x, y)] dx dy \\
 &+ \int \int p(x, y) \left[\lambda_1 \chi_{[X_d^x, \infty)} + \lambda_2 \chi_{[X_d^y, \infty)} + \mu \right] dx dy \\
 &- \lambda_1 PoD_t^x - \lambda_2 PoD_t^y - \mu.
 \end{aligned}$$

Al usar el cálculo de variaciones, el procedimiento de optimización se puede realizar calculando la siguiente variación ¹⁰:

$$\delta L = \frac{\partial L [p(x, y) + \epsilon \gamma(x, y), q(x, y)]}{\partial \epsilon} \Big|_{\epsilon=0} = 0,$$

donde ϵ es una cantidad pequeña y $\gamma(x, y)$ es una función continua y arbitraria que vale cero en la frontera de integración con varianza finita. Podemos suponer sin pérdida de generalidad que $\gamma(x, y)$ está acotada. La solución óptima se representa con la siguiente densidad multivariada a-posteriori:

$$\hat{p}(x, y) = q(x, y) \exp \left\{ - \left[1 + \hat{\mu} + \left(\hat{\lambda}_1 \chi_{[X_d^x, \infty)} \right) + \left(\hat{\lambda}_2 \chi_{[X_d^y, \infty)} \right) \right] \right\}.$$

3.5.6. Intuición CIMDO

Analizando el funcional definido en (3.52), es claro que el «Consistent Information Multivariate Density Optimizing» recupera la distribución que minimiza la divergencia probabilística o mejor dicho «distancia entrópica» de la distribución a-priori y que es consistente con la información contenida en las ecuaciones de consistencia. Entonces, de todas las distribuciones que satisfagan las ecuaciones de consistencia, el procedimiento propuesto nos proporciona una metodología sistemática en el cual seleccionamos la distribución a-posteriori que está lo más cercana posible a la a-priori, resolviendo la ambigüedad que existía al querer determinar la distribución desconocida multivariada con la información parcial proporcionada por las probabilidades de distress. La intuición detrás de este procedimiento de optimización se puede entender analizando la función objetivo del enfoque de la entropía cruzada definida en (3.22). Esta función objetivo es una extensión de la función pura de entropía definida en la sección 3.4. Al maximizar la función de entropía pura sujeta a ciertas restricciones obtenemos la distribución a-posteriori más cercana a la distribución uniforme sujeta a un conjunto de restricciones. Sin embargo, de existir una hipótesis inicial acerca de la forma de la distribución a-priori, ahora, el problema se puede formular minimizando la divergencia probabilística entre la distribución a-priori y la a-posteriori. De todas las distribuciones de probabilidad que satisfacen las ecuaciones de consistencia la solución es la a-posteriori más cercana a la a-priori. A pesar de que la distribución a-priori se basa en la intuición económica usualmente, es inconsistente con las observaciones. Entonces, utilizando la solución del enfoque

¹⁰Para mayor detalle ver el Apéndice A.

de la entropía cruzada, resolvemos esta inconsistencia, reconciliándola de la mejor manera posible al recuperar la distribución lo más cercana posible a la a-priori, pero consistente con las observaciones empíricas. Al usar CIMDO para encontrar $p(x, y)$, el problema se convierte en uno de deducción matemática a uno de inferencia a través de un problema de optimización. Esto es porque el enfoque de la entropía cruzada inmerso en CIMDO invierte el proceso de modelación de información. En vez de asumir una cierta forma paramétrica para caracterizar información contenida en los datos, este enfoque utiliza la información para inferir valores de la desconocida distribución de probabilidad. Utilizando este procedimiento buscamos realizar las mejores predicciones posibles de la escasa información que tenemos.

3.6. CIMDO-Cópula

Una de las propiedades deseables de la metodología CIMDO es que captura dependencias tanto lineales como no lineales. En efecto, la cópula de la distribución de pérdidas inferida por CIMDO se ajusta conforme pasa el tiempo y empieza a surgir más información disponible. A continuación se calculará explícitamente la cópula de la distribución a-posteriori recuperada por CIMDO. En la sección anterior se vio que si la distribución a-priori es $q(x, y)$ entonces la a-posteriori se distribuye de la siguiente forma:

$$\hat{p}(x, y) = q(x, y) \exp \left\{ - \left[1 + \hat{\mu} + \left(\hat{\lambda}_1 \chi_{[X_d^x, \infty)} \right) + \left(\hat{\lambda}_2 \chi_{[X_d^y, \infty)} \right) \right] \right\},$$

con marginales:

$$f(x) = \int_{-\infty}^{\infty} p(x, y) dy, \quad g(y) = \int_{-\infty}^{\infty} p(x, y) dx.$$

Sean F y G las funciones de distribución de las marginales f y g , utilizando el corolario (2.15) la densidad de la CIMDO-Cópula es la siguiente:

$$\begin{aligned} C(u, v) &= \frac{q(F^{-1}(u), G^{-1}(v)) A}{\int_{-\infty}^{\infty} q(F^{-1}(u), y) A dy \int_{-\infty}^{\infty} q(x, G^{-1}(v)) A dx} \\ &= \frac{q(F^{-1}(u), G^{-1}(v)) \exp \{1 + \hat{\mu}\}}{\int_{-\infty}^{\infty} q(F^{-1}(u), y) B dy \int_{-\infty}^{\infty} q(x, G^{-1}(v)) C dx}, \end{aligned}$$

donde

$$\begin{aligned} A &= \exp \left\{ - \left[1 + \hat{\mu} + \left(\hat{\lambda}_1 \chi_{[X_d^x, \infty)} \right) + \left(\hat{\lambda}_2 \chi_{[X_d^y, \infty)} \right) \right] \right\}, \\ B &= \exp \left\{ - \hat{\lambda}_2 \chi_{[X_d^y, \infty)} \right\}, \\ C &= \exp \left\{ - \hat{\lambda}_1 \chi_{[X_d^x, \infty)} \right\}. \end{aligned}$$

Recordar que la metodología CIMDO cambia conforme pasa el tiempo y se va obteniendo más información. Cada vez que se obtiene más información las condiciones en el problema variacional cambian, lo cual provocará cambios en los multiplicadores de Lagrange resultantes del problema. La expresión de la CIMDO-Cópula expuesta arriba muestra como ésta se va ajustando conforme pasa el tiempo reflejadas con el cambio de los multiplicadores de Lagrange del presente problema de optimización.

Capítulo 4

El Sistema Financiero Mexicano

En este capítulo mostramos los resultados de la implementación de la metodología expuesta en los capítulos anteriores para el caso del sistema financiero mexicano. Es importante hacer un análisis de los diversos factores que afectan intuitivamente los valores del nivel del riesgo sistémico y la importancia sistémica de cada institución financiera, esto se hace con el fin de tener alguna noción intuitiva de los resultados finales de nuestras estimaciones. En este sentido, en la sección 4.1 se realiza dicho análisis tomando en cuenta los siguientes factores: Tamaño, Correlación y Probabilidad de Distress. Finalmente en la sección 4.2 se exhiben los resultados.

4.1. Un Análisis del Sistema Mexicano

4.1.1. Tamaño

El sistema financiero tiene un nivel de granularidad bajo, como se puede ver en la Figura 4.1 :

Guiándonos en estas gráficas podemos dividir al sistema financiero según el tamaño en tres bloques como sigue: ¹

- **Tamaño Grande:** Banco 1, Banco 2, Banco 3.
- **Tamaño Mediano:** Banco 4 y Banco 5.

¹La división la hicimos utilizando el tamaño promedio de cada banco y definiendo los cuartiles con dicha información.

- **Tamaño Chico:** Banco 6, Banco 7, Banco 8, Banco 9, Banco 10, Banco 11 y Banco 12.

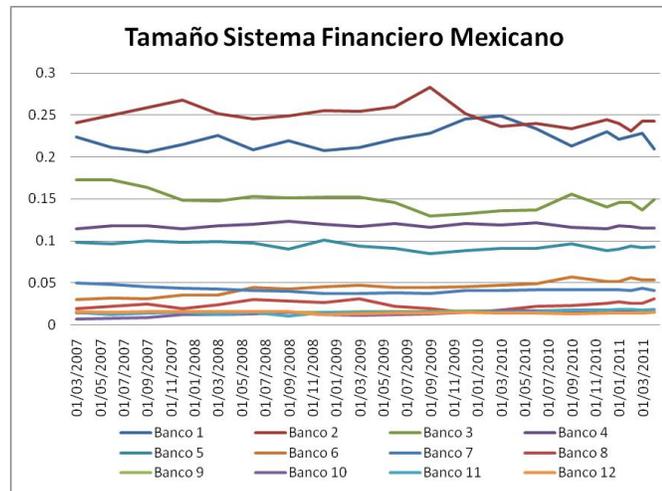


Figura 4.1: Tamaño de bancos como porcentaje del total de activos del sistema de cada institución.

El bloque 1 controla el 60 % del total de activos en el sistema, el bloque 2 controla el 23 %. El bloque 3 que consta de siete bancos controla solamente el 17 % del total de activos en el sistema.

Tomando en cuenta solamente el tamaño de las instituciones resulta intuitivos pensar que el bloque 1 contribuirá a la mayor parte del nivel de riesgo sistémico del sistema financiero.

4.1.2. Correlación

La importancia sistémica de una institución aumenta cuando incrementa el nivel de correlación de sus pérdidas con respecto a las pérdidas de las demás instituciones. Esto es porque aumenta la plausibilidad de dicha institución a participar en eventos sistémicos, es decir, aumenta la probabilidad de que surjan incumplimientos simultáneos y por lo tanto aumenta el nivel de pérdidas no esperadas. Las correlaciones se estimaron calculando la correlación Spearman de los vectores de pérdidas de cada banco que salen del

proceso de estimación de pérdidas basada en CIMDO ². Para el caso del sistema financiero mexicano, los bancos cinco y nueve tienen un nivel más alto de correlación que las demás. Los bancos diez y once tienen un nivel menor de correlación como se puede ver en la Figura 4.2:

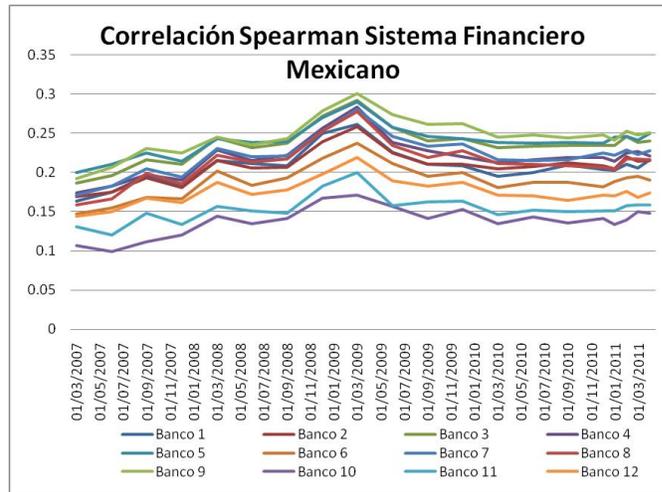


Figura 4.2: Promedio de las correlaciones entre las pérdidas que tiene cada institución con las demás.

4.1.3. Probabilidad de Distress

La importancia sistémica de una institución aumenta cuando su estado de salud en general no es buena. Las pérdidas no esperadas de una institución no saludable aumentan y por ende aumenta la contribución de dicha institución al nivel de riesgo sistémico del sistema financiero. Para el caso del sistema financiero mexicano los dos bancos más saludables según el nivel de sus probabilidades de distress son el banco 1 y banco 2, luego el banco 4 y el banco 7 tienen un nivel de probabilidad de distress medio. Los demás bancos tienen un nivel de probabilidad de distress más alta.

²Ver la subsección **Proceso de Estimación**.

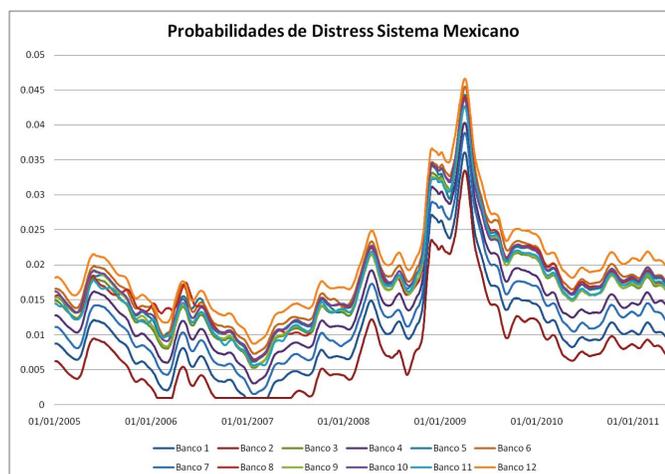


Figura 4.3: Probabilidades de distress de las instituciones mexicanas

4.2. Estimación de la Importancia Sistémica

4.2.1. Proceso de Estimación

Los pasos a seguir para la medición de la importancia sistémica de cada institución financiera y el nivel de riesgo sistémico del sistema financiero son los siguientes:

1. Visualización del sistema financiero como un portafolio de instituciones.
2. Obtención de la probabilidades de distress de cada una de las instituciones financieras.
3. Utilización de CIMDO para obtener la distribución de pérdidas del sistema financiero.
4. Simulación (basada en la distribución CIMDO) de las pérdidas sistémicas del sistema financiero
5. Estimación de la importancia sistémica de cada institución financiera y el nivel de riesgo sistémico del sistema financiero.

Las estimaciones se hicieron mediante procesos ejecutados en Matlab. Los únicos insumos necesarios para realizar las estimaciones son las probabilidades de distress de cada una de las instituciones del sistema financiero las

cuales se muestran en la Figura 4.3. Utilizando estas probabilidades se realiza una simulación basada en la distribución multivariada CIMDO de 10,000 observaciones de las pérdidas de cada una de las instituciones financieras y de todo el sistema. Dichas simulaciones se utilizan como insumo para estimar la importancia sistémica de cada institución y el nivel de riesgo sistémico del sistema financiero.

4.2.2. Resultados

A continuación presentamos los resultados de las estimaciones de las importancias sistémicas de las instituciones del sistema financiero. Los resultados se muestran en forma de dos gráficas para cada banco que contienen lo siguiente:

Gráficas Marginales

1. El promedio de las pérdidas sobre los activos totales.
2. El expected shortfall al 99.5 % sobre los activos totales.
3. El nivel de reservas sobre los activos totales.
4. El nivel de capital sobre los activos totales.

Estas gráficas pretenden mostrar la solvencia que tienen los bancos al enfrentar las pérdidas esperadas y no esperadas.

Gráficas Sistémicas

1. La probabilidad de distress del banco.
2. El promedio de la correlación spearman de las pérdidas de una institución con respecto a las de las demás instituciones (promedio de la columna correspondiente a la institución de la matriz de correlaciones spearman de las pérdidas de las instituciones).
3. El tamaño del banco como un porcentaje del total de activos en el sistema.
4. La importancia sistémica.

Estas gráficas pretenden mostrar la contribución al riesgo sistémico de las instituciones financieras y su interacción con los diversos factores que afectan su valor.

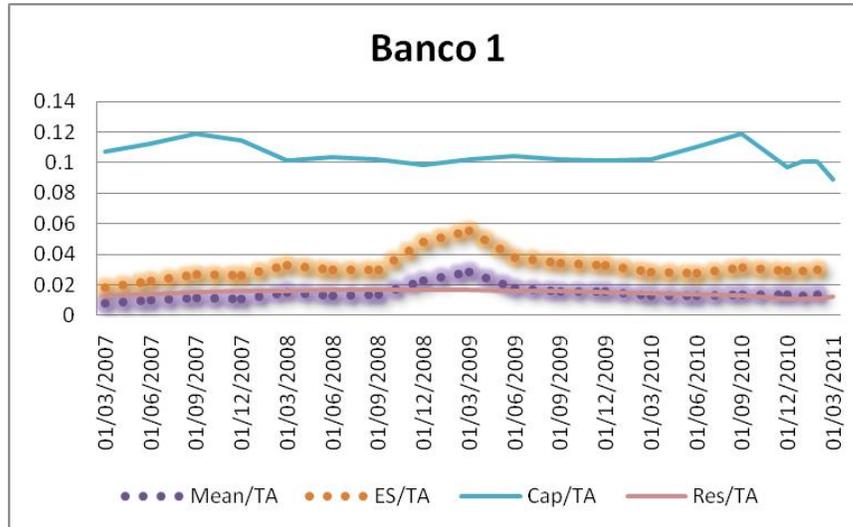


Figura 4.4: Pérdidas esperadas y no esperadas de la institución uno junto con sus respectivos buffers.

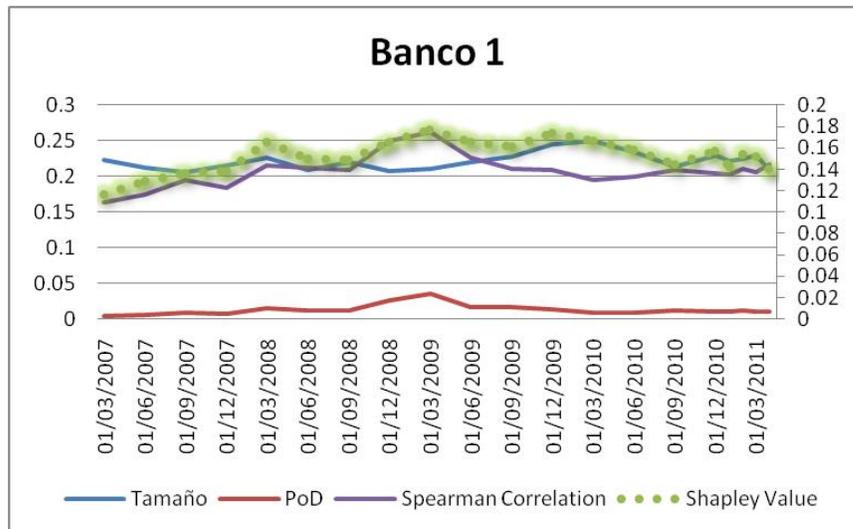


Figura 4.5: Importancia sistémica de la institución uno junto con sus factores influyentes.

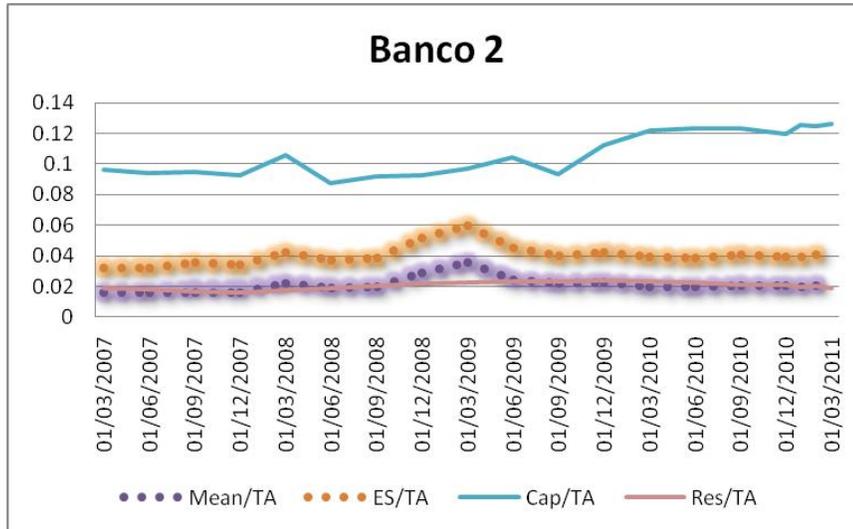


Figura 4.6: Pérdidas esperadas y no esperadas de la institución dos junto con sus respectivos buffers.

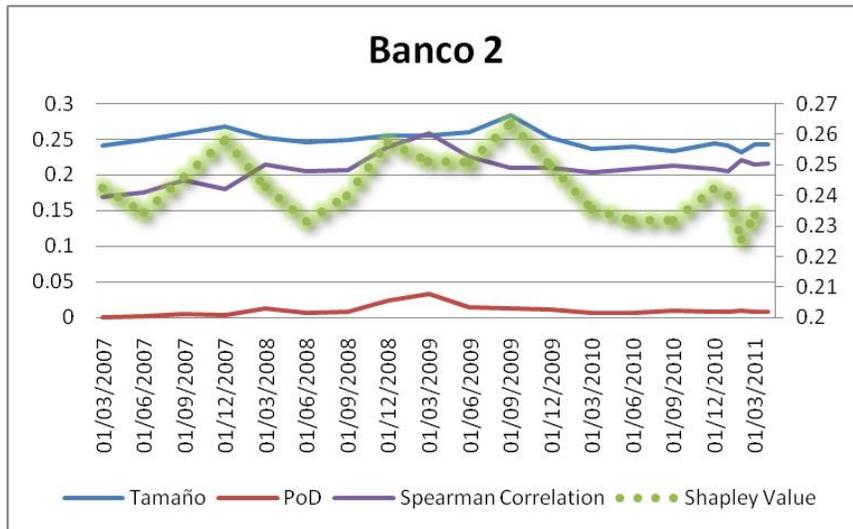


Figura 4.7: Importancia sistémica de la institución dos junto con sus factores influyentes.

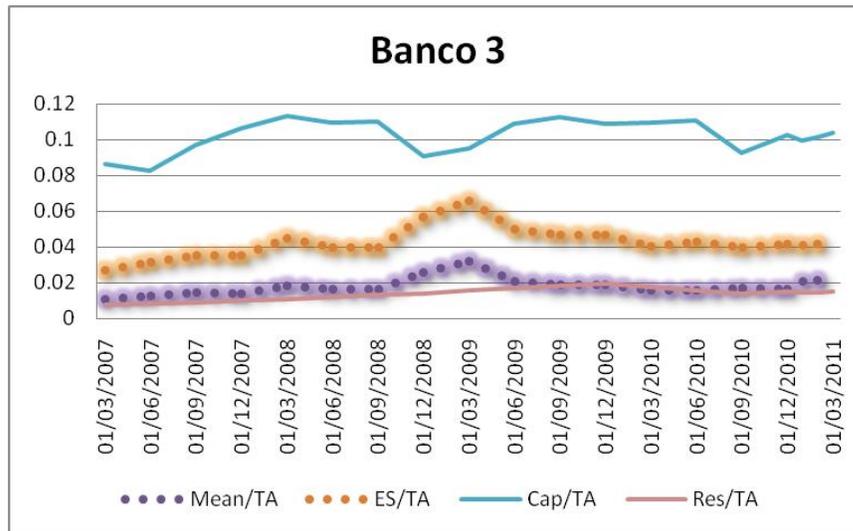


Figura 4.8: Pérdidas esperadas y no esperadas de la institución tres junto con sus respectivos buffers.

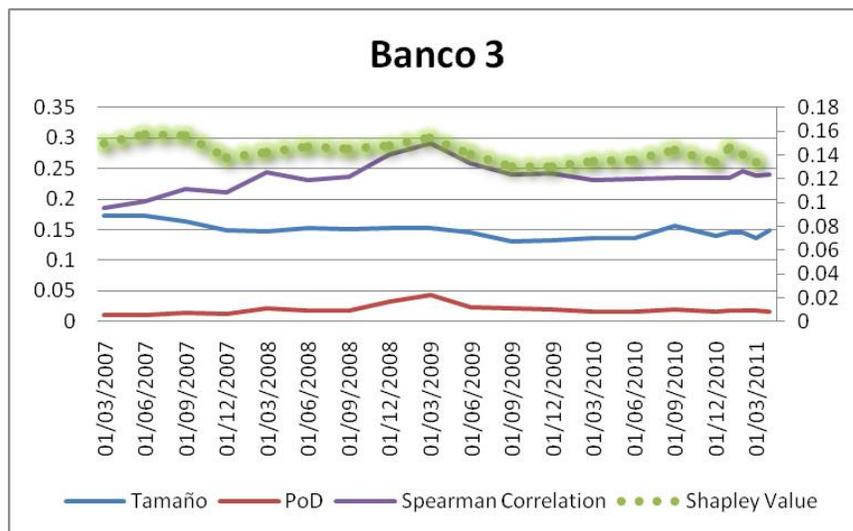


Figura 4.9: Importancia sistémica de la institución tres junto con sus factores influyentes.

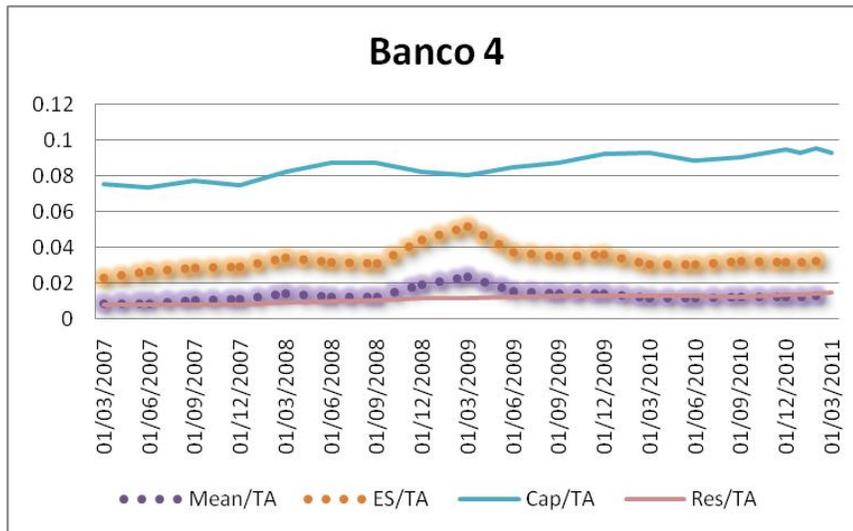


Figura 4.10: Pérdidas esperadas y no esperadas de la institución cuatro junto con sus respectivos buffers.

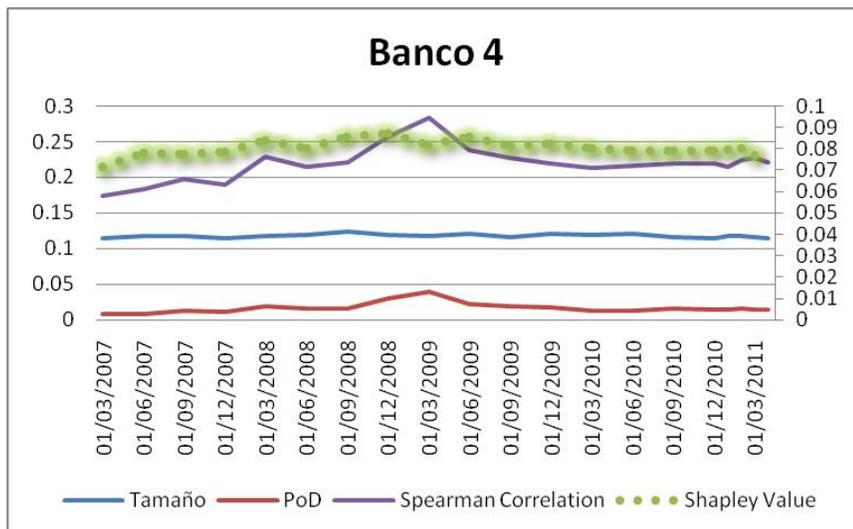


Figura 4.11: Importancia sistémica de la institución cuatro junto con sus factores influyentes.

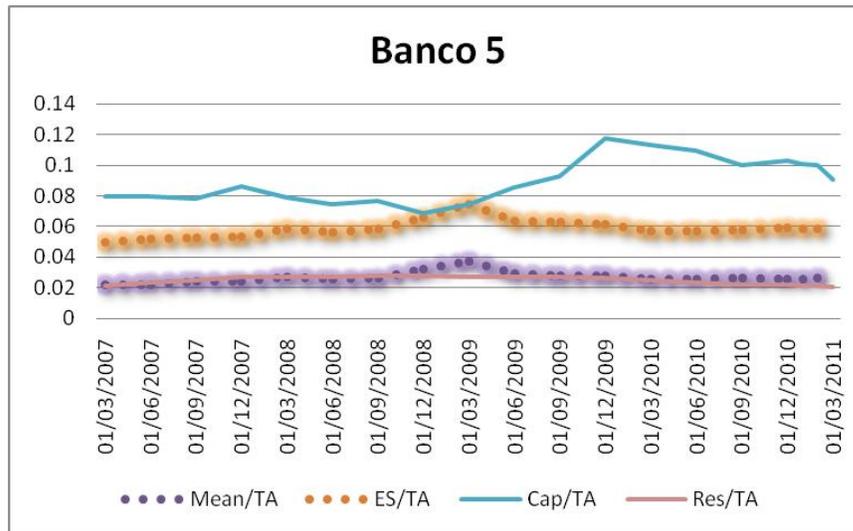


Figura 4.12: Pérdidas esperadas y no esperadas de la institución cinco junto con sus respectivos buffers.

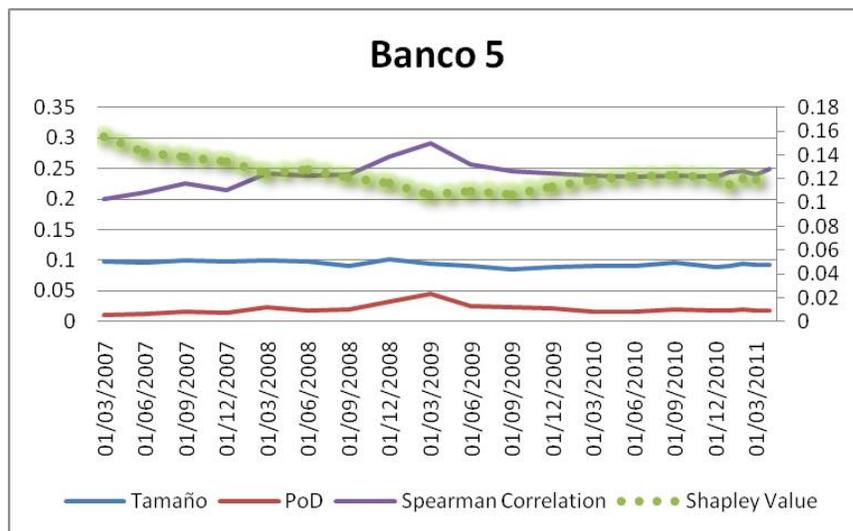


Figura 4.13: Importancia sistémica de la institución cinco junto con sus factores influyentes.

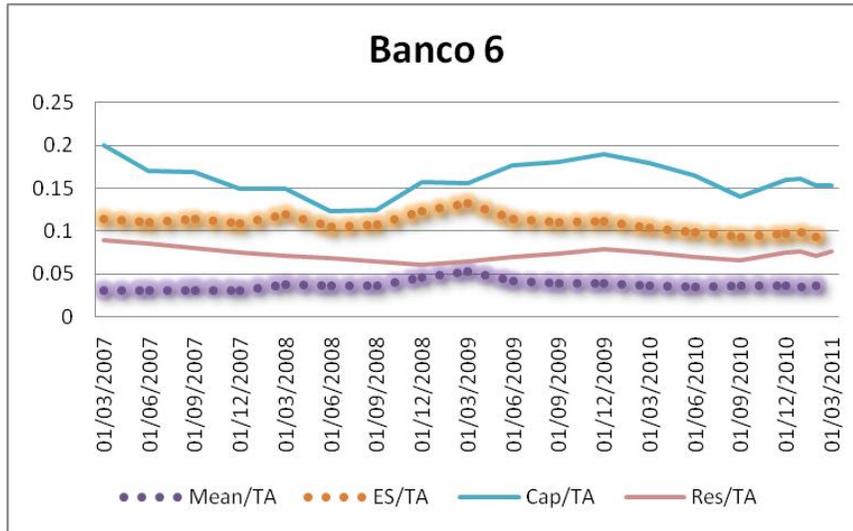


Figura 4.14: Pérdidas esperadas y no esperadas de la institución seis junto con sus respectivos buffers.

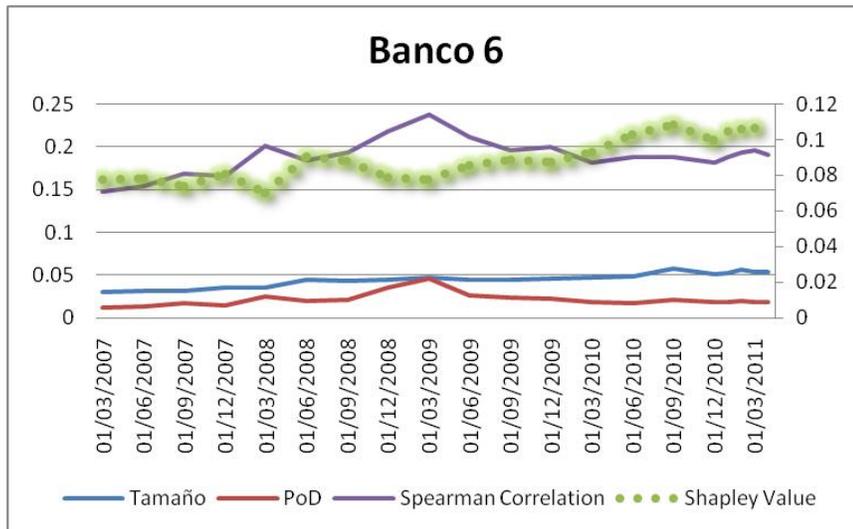


Figura 4.15: Importancia sistémica de la institución seis junto con sus factores influyentes.

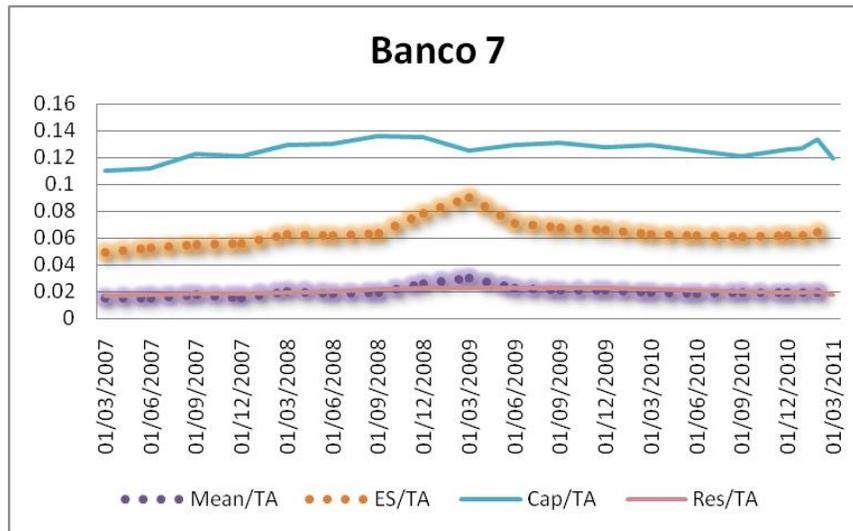


Figura 4.16: Pérdidas esperadas y no esperadas de la institución siete junto con sus respectivos buffers.

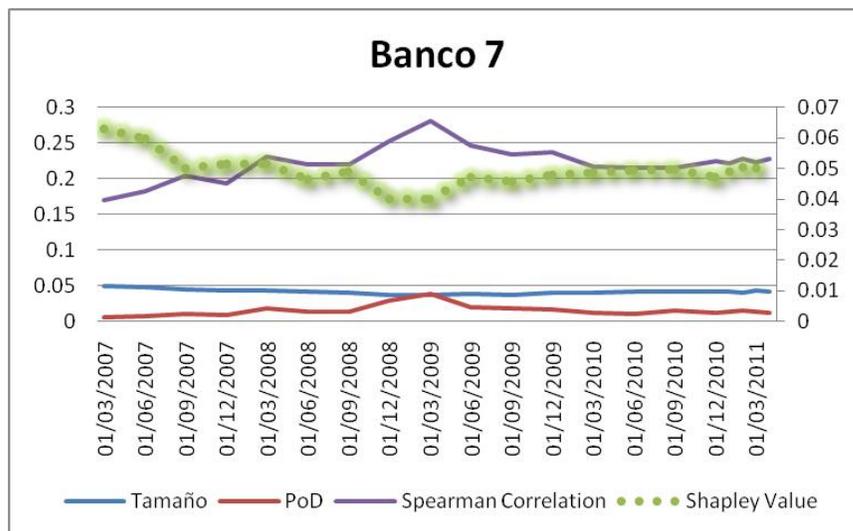


Figura 4.17: Importancia sistémica de la institución siete junto con sus factores influyentes.

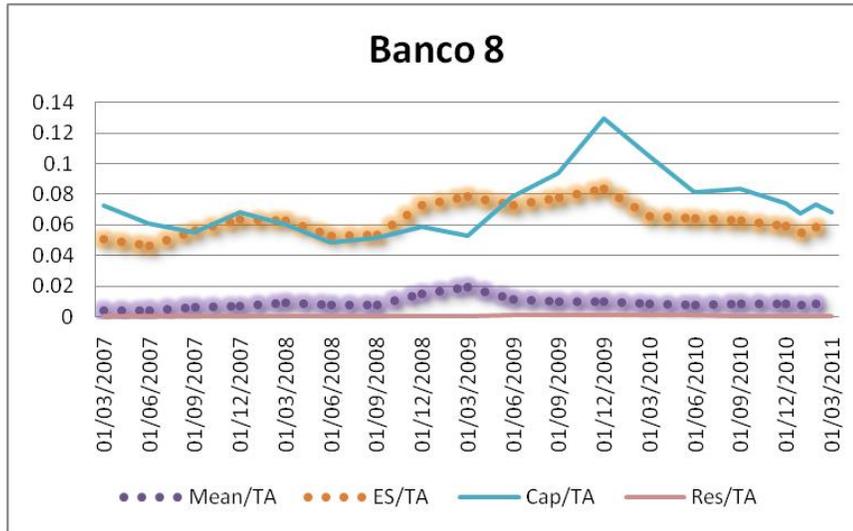


Figura 4.18: Pérdidas esperadas y no esperadas de la institución ocho junto con sus respectivos buffers.

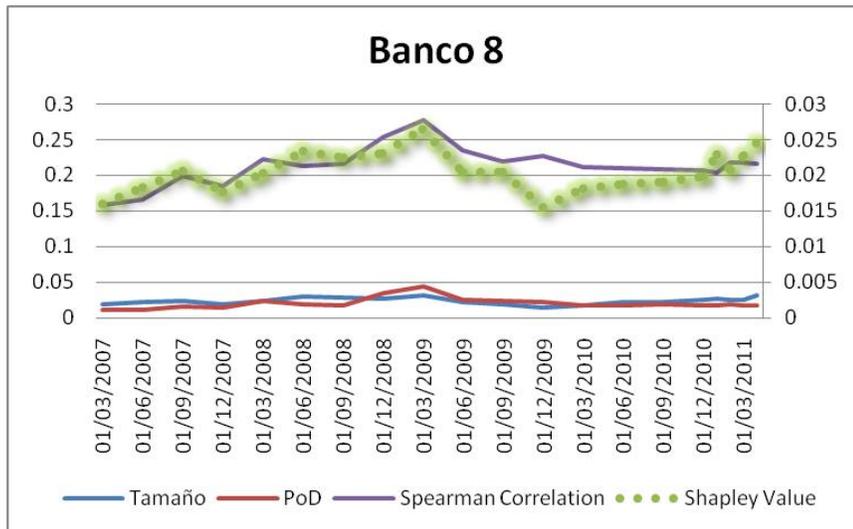


Figura 4.19: Importancia sistémica de la institución ocho junto con sus factores influyentes.

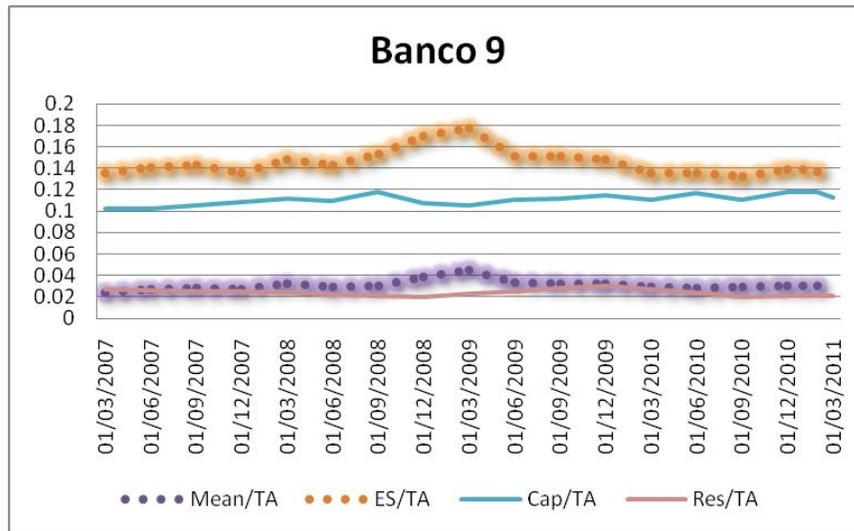


Figura 4.20: Pérdidas esperadas y no esperadas de la institución nueve junto con sus respectivos buffers.

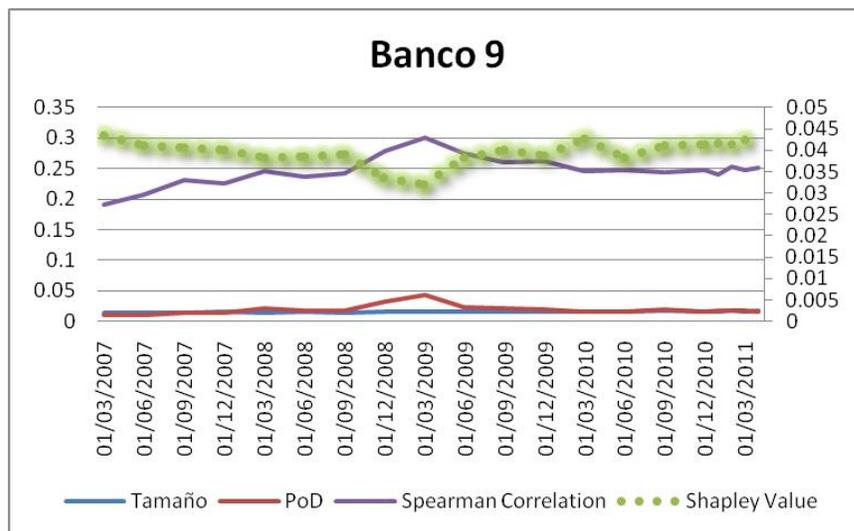


Figura 4.21: Importancia sistémica de la institución nueve junto con sus factores influyentes.

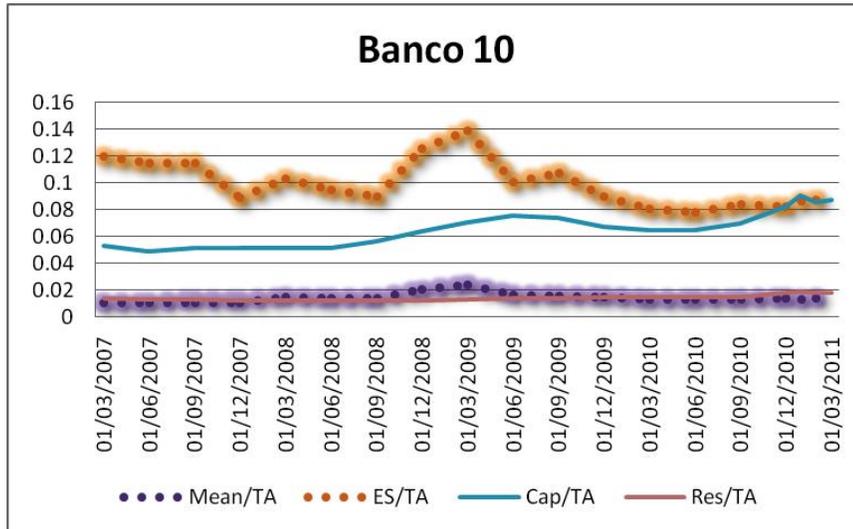


Figura 4.22: Pérdidas esperadas y no esperadas de la institución diez junto con sus respectivos buffers.

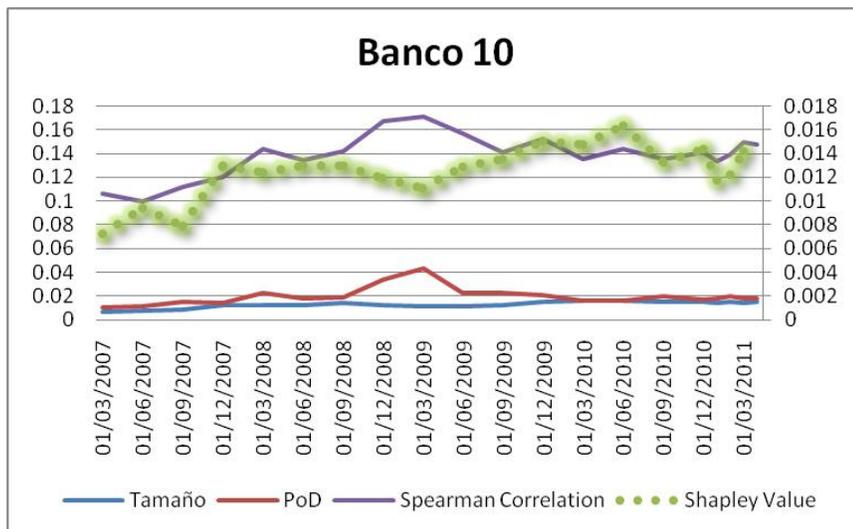


Figura 4.23: Importancia sistémica de la institución diez junto con sus factores influyentes.

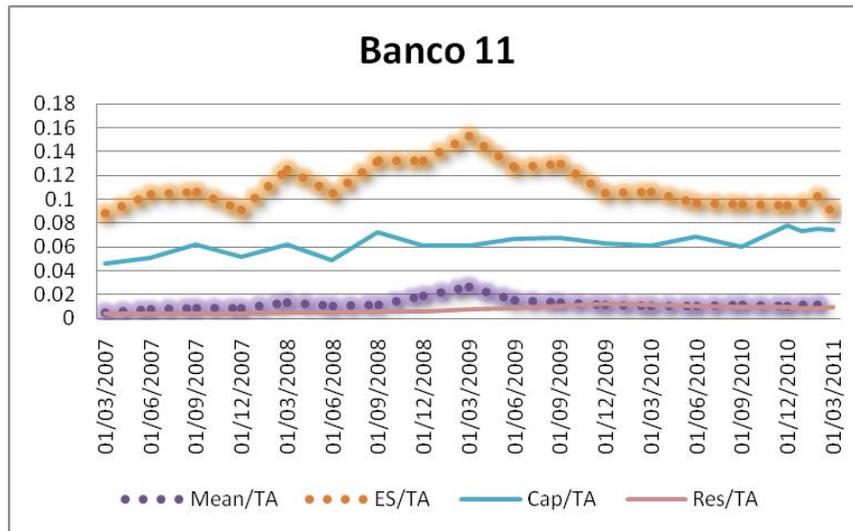


Figura 4.24: Pérdidas esperadas y no esperadas de la institución once junto con sus respectivos buffers.

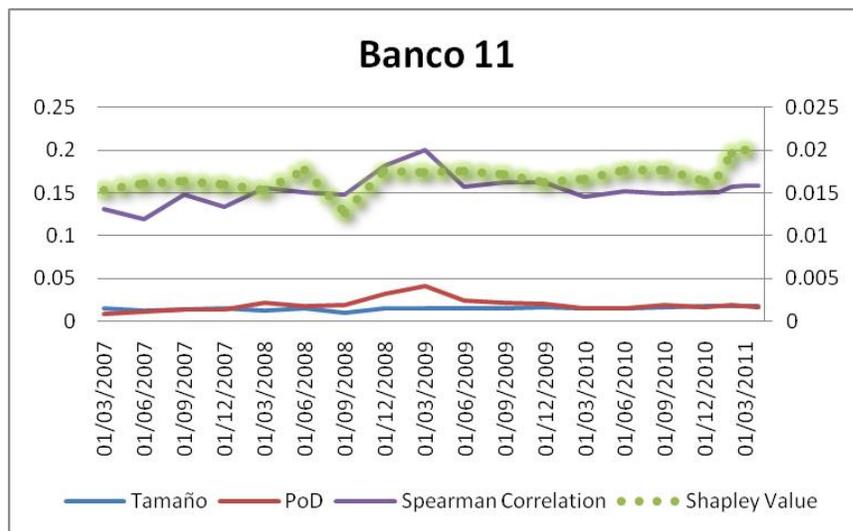


Figura 4.25: Importancia sistémica de la institución once junto con sus factores influyentes.

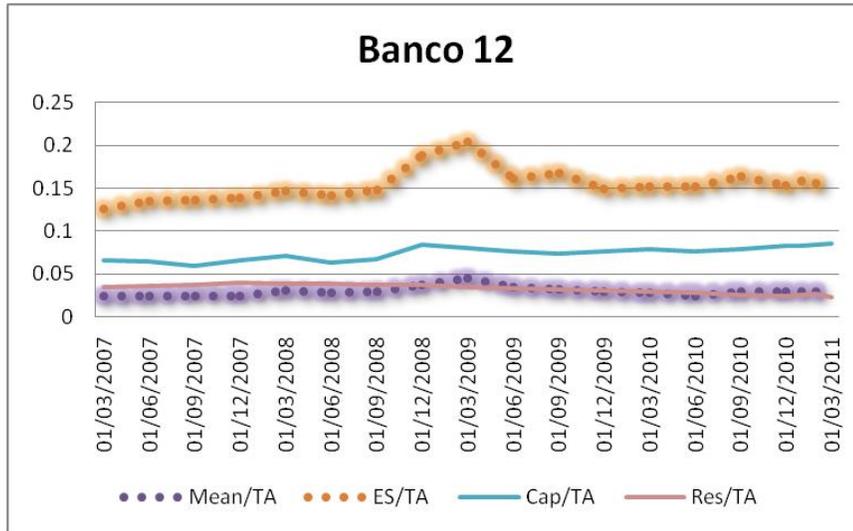


Figura 4.26: Pérdidas esperadas y no esperadas de la institución doce junto con sus respectivos buffers.

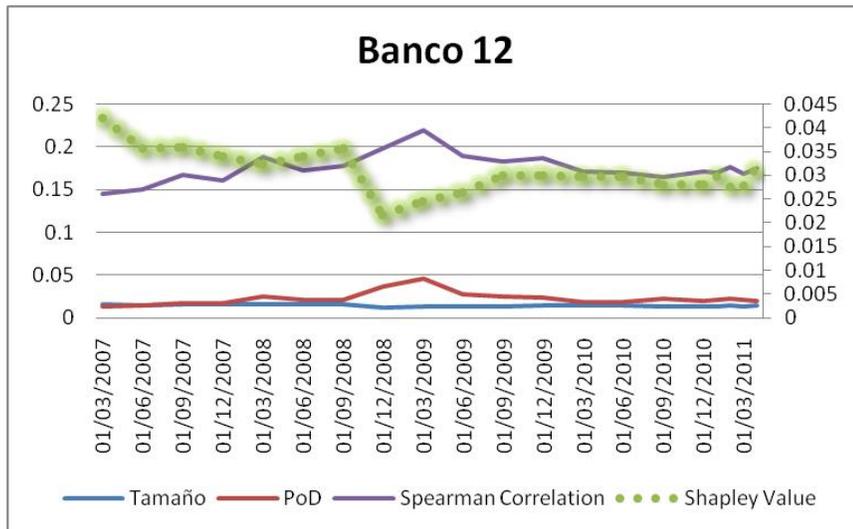


Figura 4.27: Importancia sistémica de la institución doce junto con sus factores influyentes.

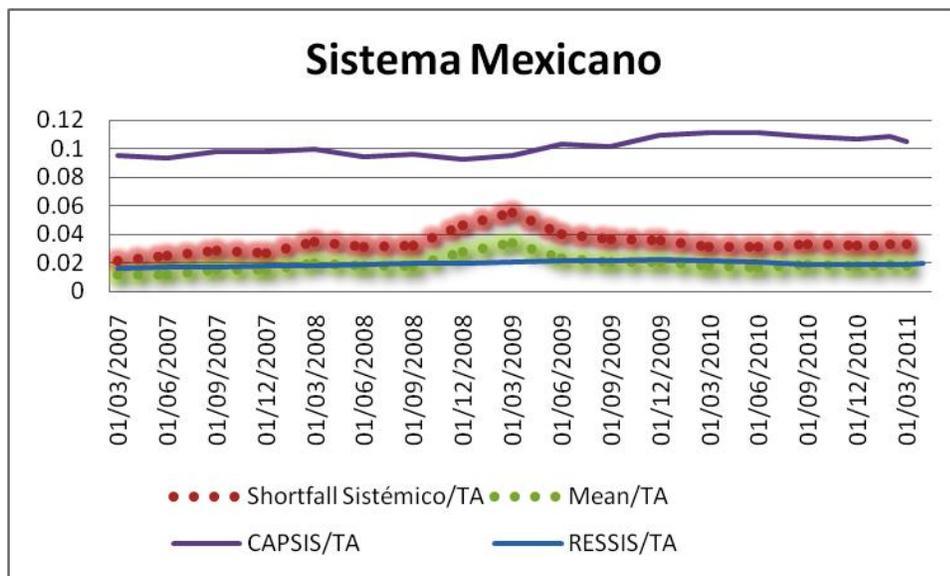


Figura 4.28: Pérdidas esperadas y no esperadas del sistema financiero junto con sus respectivos buffers.

Capítulo 5

Conclusiones

El análisis de los resultados sugiere una división de las instituciones del sistema financiero mexicano respecto a la importancia sistémica en tres bloques:

- **Importancia Sistémica Alta:** Banco 1 y Banco 2.
- **Importancia Sistémica Mediana:** Banco 3, Banco 5 y Banco 6.
- **Importancia Sistémica Baja:** Banco 4, Banco 7, Banco 8, Banco 9, Banco 10, Banco 11 y Banco 12.

Para cada grupo sugerimos las siguientes políticas de supervisión macroprudencial:

5.1. Importancia Sistémica Alta

Los bancos de este bloque cuentan con niveles sanos de capital y de reservas respecto a sus niveles de pérdidas esperadas y pérdidas no esperadas. Esto es consistente con sus bajos niveles de probabilidad de distress. Lo cual implica que sus altos niveles de importancia sistémica se deben principalmente a sus tamaños respecto a los activos totales en el sistema financiero. En efecto, los integrantes de este bloque operan en promedio el 47% de los activos totales del sistema. Se ha visto empíricamente [7] que en épocas de crisis la importancia sistémica se traslada a las instituciones con mayor tamaño. Por lo cual, desde un punto de vista macroprudencial, es importante

tener un sistema financiero altamente granular. En este sentido se sugiere para este bloque una política de supervisión que castigue el aumento en el tamaño de estas instituciones. Esto con el fin de incentivar el aumento en el nivel de granularidad del sistema financiero.

5.2. Importancia Sistémica Mediana

En las últimas fechas se han observado para los bancos de este bloque, niveles sanos de capital y de reservas respecto a sus niveles de pérdidas esperadas y no esperadas. Esto es consistente con el bajo nivel de probabilidad de distress observado a partir del 2010 (alrededor del 1 %). En promedio los bancos de este bloque tienen un tamaño mediano respecto al sistema financiero. Por lo tanto, con el fin de aumentar la granularidad del sistema financiero, se propone seguir una política de supervisión que incentive el crecimiento en tamaño de los bancos de este bloque. Sin embargo, es importante condicionar esta política al mantenimiento de los niveles sanos de capital y de reservas que se han observado últimamente.

5.3. Importancia Sistémica Baja

En general la mayor parte de los bancos de este bloque son de un tamaño pequeño respecto del sistema financiero y poseen un nivel de correlación alta. Cuentan con niveles no sanos de capital y de reservas respecto a sus niveles de pérdidas esperadas y pérdidas no esperadas. Esto es consistente con los altos niveles de probabilidad de distress que se han observado. Debido a sus bajos niveles de importancia sistémica ¹ estos bancos no representan un riesgo sistémico. Por lo tanto, para este bloque se propone llevar una política de supervisión microprudencial obligando a estos bancos a aumentar sus niveles de capital y de reservas.

¹En promedio tienen un nivel de importancia sistémica del 3.57%.

Apéndice A

Calculemos el siguiente límite:

$$\lim_{\epsilon \rightarrow 0} \frac{\mathcal{L}[p(x, y) + \epsilon\gamma(x, y), q(x, y)] - \mathcal{L}[p(x, y), q(x, y)]}{\epsilon} = \lim_{\epsilon \rightarrow 0} \frac{A}{\epsilon}$$

Desarrollando A y relajando la notación tenemos que:

$$\begin{aligned} A &= \int \int (p + \epsilon\gamma) [\ln(p + \epsilon\gamma) - \ln(q)] \, dx dy \\ &+ \int \int (p + \epsilon\gamma) \left[\lambda_1 \chi_{[x_d^x, \infty)} + \lambda_2 \chi_{[x_d^y, \infty)} + \mu \right] \, dx dy \\ &- \lambda_1 PoD_t^x - \lambda_2 PoD_t^y - \mu \\ &- \int \int p [\ln(p) - \ln(q)] \, dx dy \\ &- \int \int p \left[\lambda_1 \chi_{[x_d^x, \infty)} + \lambda_2 \chi_{[x_d^y, \infty)} + \mu \right] \, dx dy \\ &+ \lambda_1 PoD_t^x + \lambda_2 PoD_t^y + \mu \end{aligned}$$

Simplificando llegamos a que:

$$\begin{aligned} A &= \int \int (p + \epsilon\gamma) [\ln(p + \epsilon\gamma) - \ln(q)] \, dx dy \\ &- \int \int p [\ln(p) - \ln(q)] \, dx dy \\ &+ \epsilon\gamma \left[\lambda_1 \chi_{[x_d^x, \infty)} + \lambda_2 \chi_{[x_d^y, \infty)} + \mu \right] \end{aligned}$$

Utilizando expansiones en series de Taylor para $\ln(x)$ centrado en $p(x, y)$ tenemos que:

$$\begin{aligned} A &= \iint (p + \epsilon\gamma) \left[\ln(p) + \frac{\epsilon\gamma}{p} + o(\epsilon^2) - \ln(q) \right] dx dy \\ &\quad - \iint p [\ln(p) - \ln(q)] dx dy \\ &\quad + \iint \epsilon\gamma \left[\lambda_1 \chi_{[x_d^x, \infty)} + \lambda_2 \chi_{[x_d^y, \infty)} + \mu \right] \end{aligned}$$

Agrupando términos:

$$\begin{aligned} A &= \iint \left[\epsilon\gamma + p o(\epsilon^2) + \epsilon \ln(p)\gamma + \epsilon^2 \frac{\gamma^2}{p} + \gamma o(\epsilon^3) - \epsilon\gamma \ln(q) \right] dx dy \\ &\quad + \iint \epsilon\gamma \left[\lambda_1 \chi_{[x_d^x, \infty)} + \lambda_2 \chi_{[x_d^y, \infty)} + \mu \right] \end{aligned}$$

Multiplicando por $\frac{1}{\epsilon}$ y juntando las integrales:

$$A = \iint \left[\gamma + \ln(p)\gamma + \epsilon \frac{\gamma^2}{p} + \gamma o(\epsilon^2) - \gamma \ln(q) + \gamma \left[\lambda_1 \chi_{[x_d^x, \infty)} + \lambda_2 \chi_{[x_d^y, \infty)} + \mu \right] \right] dx dy$$

Tomando el límite cuando $\epsilon \rightarrow 0$:

$$\lim_{\epsilon \rightarrow 0} \frac{A}{\epsilon} = \iint \gamma \left[1 + \ln(p) - \ln(q) + \left[\lambda_1 \chi_{[x_d^x, \infty)} + \lambda_2 \chi_{[x_d^y, \infty)} + \mu \right] \right] dx dy$$

Choro:

$$\left[1 + \ln(p) - \ln(q) + \left[\lambda_1 \chi_{[x_d^x, \infty)} + \lambda_2 \chi_{[x_d^y, \infty)} + \mu \right] \right] dx dy = 0$$

Por lo que:

$$\ln \widehat{p}(x, y) = \ln(q(x, y)) - 1 - \lambda_1 \chi_{[x_d^x, \infty)} - \lambda_2 \chi_{[x_d^y, \infty)} - \mu$$

Tomando exponencial de ambos lados:

$$\widehat{p}(x, y) = q(x, y) \exp \left\{ - \left[1 + \widehat{\mu} + \left(\widehat{\lambda}_1 \chi_{[x_d^x, \infty)} \right) + \left(\widehat{\lambda}_2 \chi_{[x_d^y, \infty)} \right) \right] \right\}$$

Apéndice B

Totalidad

Demostración: Queremos probar que:

$$\sum_{i=1}^N \mathbf{ShB}_i = \mathbf{V}(\mathbf{S}).$$

Ahora:

$$\begin{aligned} \sum_{i=1}^N \mathbf{ShB}_i &= \sum_{i=1}^N \mathbb{E}[\mathbf{X}_i] \\ &= \mathbb{E} \left[\sum_{i=1}^N \mathbf{V}(\mathbf{S}_i) - \mathbf{V}(\mathbf{S}_i - \{\mathbf{B}_i\}) \right] \end{aligned}$$

Sea a cualquier elemento de A . Tenemos que:

$$\begin{aligned} \sum_{i=1}^N \mathbf{V}(\mathbf{S}_i) - \mathbf{V}(\mathbf{S}_i - \{\mathbf{B}_i\}) &= \sum_a \mathbf{V}(\mathbf{S}_{a_i}) - \mathbf{V}(\mathbf{S}_{a_i} - \{\mathbf{B}_{a_i}\}) \\ &= \mathbf{V}(\mathbf{S}) - \mathbf{V}(\emptyset) \\ &= \mathbf{V}(\mathbf{S}), \end{aligned}$$

donde la segunda suma se está tomando en el orden de la permutación a . La segunda igualdad se da ya que al tomar la suma en el orden de a se obtiene una suma telescópica. De aquí es claro que:

$$\sum_{i=1}^N \mathbf{ShB}_i = \mathbb{E}[\mathbf{V}(\mathbf{S})] = \mathbf{V}(\mathbf{S})$$

Q.E.D.

Aditividad

Demostración: Sean \mathbf{V}_1 y \mathbf{V}_2 dos medidas de riesgo sistémico y definamos $\mathbf{V}_3 = \alpha\mathbf{V}_1 + \beta\mathbf{V}_2$. Entonces tenemos que para toda $i \in 1, \dots, N$:

$$\begin{aligned} X_i^3(a) &= \mathbf{V}_3(\mathbf{S}_i) - \mathbf{V}_3(\mathbf{S}_i - \{i\}) \\ &= \alpha\mathbf{V}_1(\mathbf{S}_i) + \beta\mathbf{V}_2(\mathbf{S}_i) \\ &\quad - \alpha\mathbf{V}_1(\mathbf{S}_i - \{i\}) + \beta\mathbf{V}_2(\mathbf{S}_i - \{i\}) \\ &= \alpha X_i^1(a) + \beta X_i^2(a) \end{aligned}$$

Calculando esperanzas de ambos lados de la igualdad anterior llegamos a que:

$$\text{Shp}(\mathbf{V}_3, \mathbf{B}_i) = \alpha\text{Shp}(\mathbf{V}_1, \mathbf{B}_i) + \beta\text{Shp}(\mathbf{V}_2, \mathbf{B}_i)$$

Lo cual es lo que queríamos probar. Es claro de aquí la generalización para n medidas de riesgo sistémico. *Q.E.D.*

Igualdad de incrementos

Demostración: Queremos probar que:

$$\text{Shp}(\mathbf{B}_i, \mathbf{G}) - \text{Shp}(\mathbf{B}_i, \mathbf{G} - \{\mathbf{B}_k\}) = \text{Shp}(\mathbf{B}_i, \mathbf{G}) - \text{Shp}(\mathbf{B}_i, \mathbf{G} - \{\mathbf{B}_k\})$$

Para todo \mathbf{B}_i y \mathbf{B}_k y todo subgrupo del sistema financiero tales que $\mathbf{B}_i, \mathbf{B}_k$ están en \mathbf{G} .

La prueba se dará para el caso cuando la medida \mathbb{P} que actúa en A es la uniforme.

Sea a una permutación de \mathbf{G} y a_k la permutación de $\mathbf{G} - \{\mathbf{B}_k\}$ que se genera al quitar la institución \mathbf{B}_k manteniendo el orden formado por a . Tenemos entonces que:

$$\begin{aligned} \text{Shp}(\mathbf{B}_i, \mathbf{G}) &- \text{Shp}(\mathbf{B}_i, \mathbf{G} - \{\mathbf{B}_k\}) \\ &= \frac{1}{|\mathbf{G}|!} \sum_A (\mathbf{V}(\mathbf{S}_i^a) - \mathbf{V}(\mathbf{S}_i^a - \{\mathbf{B}_i\})) \\ &\quad - (\mathbf{V}(\mathbf{S}_i^{a_k}) - \mathbf{V}(\mathbf{S}_i^{a_k} - \{\mathbf{B}_i\})) \end{aligned}$$

De aquí es claro que para demostrar lo que queremos basta con probar que:

$$\begin{aligned}
& \sum_A (\mathbf{V}(\mathbf{S}_i^a) - \mathbf{V}(\mathbf{S}_i^a - \{\mathbf{B}_i\})) - (\mathbf{V}(\mathbf{S}_i^{a_k}) - \mathbf{V}(\mathbf{S}_i^{a_k} - \{\mathbf{B}_i\})) \\
&= \sum_A (\mathbf{V}(\mathbf{S}_k^a) - \mathbf{V}(\mathbf{S}_k^a - \{\mathbf{B}_k\})) - (\mathbf{V}(\mathbf{S}_k^{a_i}) - \mathbf{V}(\mathbf{S}_k^{a_i} - \{\mathbf{B}_k\}))
\end{aligned}$$

Notar que en la suma de la izquierda solo hay que tomar en cuenta las permutaciones en las que \mathbf{B}_k está antes que \mathbf{B}_i ya que los términos generados por las otras permutaciones son cero. Análogamente para la segunda suma solo basta tomar las permutaciones en las que \mathbf{B}_i está antes que \mathbf{B}_k .

En lo que sigue de la prueba checaremos que para cada término en la suma de la izquierda hay un término igual en la suma de la derecha y vice versa.

Sea a una permutación de \mathbf{G} en la que \mathbf{B}_k está antes que \mathbf{B}_i (acordarse que estas son las que tenemos que tomar en cuenta para la suma de la izquierda) y sea \mathbf{a}_i^k la permutación generada por a al intercambiar las instituciones \mathbf{B}_i por \mathbf{B}_k ; \mathbf{a}_i^k es de las permutaciones que tenemos que tomar en cuenta para la suma de la derecha. Fijémonos en el término generado por a en la suma de la izquierda:

$$\begin{aligned}
& \mathbf{V}(\mathbf{S}_i^a) - \mathbf{V}(\mathbf{S}_i^a - \{\mathbf{B}_i\}) - (\mathbf{V}(\mathbf{S}_i^{a_k}) - \mathbf{V}(\mathbf{S}_i^{a_k} - \{\mathbf{B}_i\})) \\
&= \mathbf{V}(\mathbf{S}_i^a) - \mathbf{V}(\mathbf{S}_i^a - \{\mathbf{B}_i\}) \\
&\quad - (\mathbf{V}(\mathbf{S}_i^a - \{\mathbf{B}_k\}) - \mathbf{V}(\mathbf{S}_i^a - \{\mathbf{B}_i\} - \{\mathbf{B}_k\}))
\end{aligned}$$

El término se reduce a la expresión en la igualdad anterior.

$$\mathbf{V}(\mathbf{S}_i^a) - \mathbf{V}(\mathbf{S}_i^a - \{\mathbf{B}_i\}) - (\mathbf{V}(\mathbf{S}_i^a - \{\mathbf{B}_k\}) - \mathbf{V}(\mathbf{S}_i^a - \{\mathbf{B}_i\} - \{\mathbf{B}_k\}))$$

Análogamente el término generado por \mathbf{a}_i^k en la suma de la derecha se reduce a:

$$\mathbf{V}(\mathbf{S}_i^{\mathbf{a}_i^k}) - \mathbf{V}(\mathbf{S}_i^{\mathbf{a}_i^k} - \{\mathbf{B}_k\}) - (\mathbf{V}(\mathbf{S}_i^{\mathbf{a}_i^k} - \{\mathbf{B}_i\}) - \mathbf{V}(\mathbf{S}_i^{\mathbf{a}_i^k} - \{\mathbf{B}_k\} - \{\mathbf{B}_i\}))$$

Notar ahora que:

$$\mathbf{V}(\mathbf{S}_i^a) = \mathbf{V}(\mathbf{S}_i^{\mathbf{a}_i^k}), \quad \mathbf{V}(\mathbf{S}_i^a - \mathbf{B}_i) = \mathbf{V}(\mathbf{S}_i^{\mathbf{a}_i^k} - \{\mathbf{B}_i\})$$

Y que:

$$\mathbf{V}(\mathbf{S}_i^a - \{\mathbf{B}_k\}) = \mathbf{V}(\mathbf{S}_i^{\mathbf{a}_i^k} - \{\mathbf{B}_k\}),$$

$$\mathbf{V}(\mathbf{S}_i^a - \{\mathbf{B}_i\} - \{\mathbf{B}_k\}) = \mathbf{V}(\mathbf{S}_i^{a_k} - \{\mathbf{B}_k\} - \{\mathbf{B}_i\})$$

Es decir los dos términos que analizamos (uno en la suma izquierda y otra en la suma derecha) son iguales. Con esto hemos verificado que para cada término en la suma izquierda, hay un término en la suma derecha y viceversa. Hemos probado lo que queremos.

Un comentario acerca de la siguiente igualdad que utilizamos en la prueba:

$$\begin{aligned} \text{Shp}(\mathbf{B}_i, \mathbf{G}) &= \text{Shp}(\mathbf{B}_i, \mathbf{G} - \{\mathbf{B}_k\}) \\ &= \frac{1}{|\mathbf{G}|!} \sum_A (\mathbf{V}(\mathbf{S}_i^a) - \mathbf{V}(\mathbf{S}_i^a - \{\mathbf{B}_i\})) \\ &= (\mathbf{V}(\mathbf{S}_i^{a_k}) - \mathbf{V}(\mathbf{S}_i^{a_k} - \{\mathbf{B}_i\})) \end{aligned}$$

Notar como para calcular $\text{Shp}(\mathbf{B}_i, \mathbf{G} - \{\mathbf{B}_k\})$ necesitamos calcular términos de la forma:

$$\mathbf{V}(\mathbf{S}_i^a - \{\mathbf{B}_i\}) - (\mathbf{V}(\mathbf{S}_i^{a_k}) - \mathbf{V}(\mathbf{S}_i^{a_k} - \{\mathbf{B}_i\})),$$

los cuales están siendo tomados en cuenta en la suma anterior. Cada permutación $\mathbf{G} - \{\mathbf{B}_k\}$ nos genera un término como el puesto anteriormente. En la sumatoria de la igualdad anterior van a existir $|\mathbf{G}|$ permutaciones de \mathbf{G} tales que al quitarle la institución \mathbf{B}_k nos da la misma permutación de $\mathbf{G} - \{\mathbf{B}_k\}$. Entonces en la suma de la igualdad expuesta anteriormente cada de $\mathbf{G} - \{\mathbf{B}_k\}$ la estamos contando $|\mathbf{G}|$ veces lo cual se compensa por el hecho que estamos dividiendo por $|\mathbf{G}|!$, lo cual en realidad nos estaría dando la suma de todos los términos generados por cada permutación $\mathbf{G} - \{\mathbf{B}_k\}$ dividido por $(|\mathbf{G}| - 1)!$ el cual es el factor de la medida uniforme en la σ álgebra generada por las permutaciones de $\mathbf{G} - \{\mathbf{B}_k\}$, la medida que estamos usando para calcular la esperanza que nos arroja los shapley values correspondientes. De aquí que la igualdad expuesta anteriormente en efecto se da. *Q.E.D.*

Bibliografía

- [1] Jaynes E.T., *Probability Theory: The logic of science*, Second Edition. Cambridge University Press, 2003.
- [2] Hobson A., *A new theorem of information theory*, Journal of Statistical Physics. 1 (1969).
- [3] Nelsen R., *An introduction to copulas*, Second Edition, Springer, 2006.
- [4] Rockafellar R. T., Uryasev S., *Conditional Value at Risk for general loss distributions*, Risk Management and Financial Engineering Lab Research REPORT No. 2001-5, 2001.
- [5] Segoviano M., *Consistent Information Multivariate Density Optimizing Methodology*, Financial Markets Group, LSE, Discussion paper 557.
- [6] Segoviano M., Goodhart C-, *Banking Stability Measures*, IMF WP/09/04, 2009.
- [7] Tarashev N., Borio C., Tsatsaronis K., *Attributing systemic risk to individual institutions*, BIS WP No.308, 2010.
- [8] De Bandt O., Hartman P., *Systemic risk: a survey*, Journal of Computing in Science & Engineering 11 (3) 6-15.
- [9] Embrechts P., Mcneil A., Strauman D., *Correlation and Dependency in risk management: properties and pitfalls*, Working Paper Risk Lab 1998.
- [10] Artzner P., Delbaen F., Eber J., Heath D., *Coherent measures of risk*, Mathematical Finance 9(1999), 203-228.
- [11] Segoviano M., Espinoza R., *Probabilities of default and the market price of risk in a distressed economy*, IMF WP-Forthcoming.

- [12] Espinosa J., *Black Scholes y más allá*, J. Tesis de licenciatura Universidad Nacional Autónoma de México 2011.
- [13] Alexander C., *Market Risk Analysis Volume 2: Practical Financial Econometrics*, First Edition, John Wiley & Sons, 2008.
- [14] Nigham D., Nigham N. J., *Matlab Guide*, SIAM, Second Edition, 2005.
- [15] Alexander C., *Market Risk Analysis Volume 4: Value at risk models*, John Wiley & Sons, 2008.
- [16] Goodhart C., Sunirand P., Tsomocos D. P., *A model to analyse financial fragility: applications*, Journal of Financial Stability 1 (2004) 1-30.
- [17] Calomiris C. W., Gorton G. *The origins of banking panics: models, facts and bank regulation*, Financial markets and financial crises, The University of Chicago Press, 1991.
- [18] Barron E.N. *Game Theory: An introduction*, John Wiley & Sons, 2008.
- [19] Kullback S. *Information Theory and Statistics*, Dover Publications, 1968.
- [20] Merton Robert C. *On the pricing of corporate debt: the risk structure of interest rates*, Journal of Finance, 1970.