



CIMAT

CENTRO DE INVESTIGACIÓN
EN MATEMÁTICAS A.C.

UNA APLICACIÓN DEL ANÁLISIS CONVEXO
A LA ESTADÍSTICA

T E S I S

QUE PARA OBTENER EL GRADO DE:

MAESTRO EN CIENCIAS CON ESPECIALIDAD EN
PROBABILIDAD Y ESTADÍSTICA

PRESENTA:

FABIAN PEREZ LOPEZ

TUTOR

DR. VÍCTOR MANUEL PÉREZ ABREU CARRIÓN



CONACYT



CIMAT
CENTRO DE INVESTIGACION
EN MATEMÁTICAS A. C.

Centro de Investigación en Matemáticas, A.C.

Acta de Examen de Grado

Acta No.: 132

Libro No.: 002

Foja No.: 132

En la Ciudad de Guanajuato, Gto., siendo las 18:00 horas del día 8 de febrero del año 2018, se reunieron los miembros del jurado integrado por los señores:

DR. MIGUEL NAKAMURA SAVOY
DR. ROGELIO RAMOS QUIROGA
DR. VÍCTOR MANUEL PÉREZ ABREU CARRIÓN

(CIMAT)

(CIMAT)

(CIMAT)

bajo la presidencia del primero y con carácter de secretario el segundo, para proceder a efectuar el examen que para obtener el grado de

MAESTRO EN CIENCIAS
CON ESPECIALIDAD EN PROBABILIDAD Y ESTADÍSTICA

Sustenta

FABIÁN PÉREZ LÓPEZ

en cumplimiento con lo establecido en los reglamentos y lineamientos de estudios de posgrado del Centro de Investigación en Matemáticas, A.C., mediante la presentación de la tesis

"UNA APLICACIÓN DEL ANÁLISIS CONVEXO A LA ESTADÍSTICA "

Los miembros del jurado examinaron alternadamente al (la) sustentante y después de deliberar entre sí resolvieron declararlo (a):

Aprobado



Miguel Nakamura Savoy

DR. MIGUEL NAKAMURA SAVOY
Presidente

Rogelio Ramos Quiroga

DR. ROGELIO RAMOS QUIROGA
Secretario

Victor Manuel Pérez Abreu Carrión

DR. VÍCTOR MANUEL PÉREZ ABREU CARRIÓN
Vocal



CIMAT
DIRECCIÓN
GENERAL

DR. VÍCTOR MANUEL RIVERO MERCADO
Director General

Agradecimientos

Quiero agradecer profundamente al *Consejo Nacional de Ciencia y Tecnología* (CONACYT) por la beca que me brindó durante el tiempo que requería el programa de *Maestría en Ciencias con especialidad en Probabilidad y Estadística*. Así mismo, agradezco a *CONACYT*, su visión de impulsar y fortalecer el desarrollo científico y tecnológico de México.

También agradezco al *Centro de Investigación en Matemáticas* (CIMAT), institución de enorme calidad, por todo el apoyo académico brindado durante mi estancia en él. Agradezco la valiosa labor de este centro por fortalecer la vinculación con los sectores público, privado y social mediante el desarrollo de proyectos de investigación aplicada, así como también por generar conocimiento científico invaluable a través de la investigación.

Mis agradecimientos para el Doctor Víctor Manuel Pérez Abreu Carrión por su guía y paciencia en este trabajo de tesis. Agradezco las magníficas clases que siempre nos impartió a mi y a mis compañeros durante la maestría, de las cuales todos aprendimos mucho.

Agradezco al Doctor Miguel Nakamura Savoy por el tiempo que dedicó a la lectura de este trabajo y por sus valiosas observaciones, las cuales fueron muy importantes para claridad y corrección de esta tesis.

Quiero agradecer también al Doctor Rogelio Ramos Quiroga por sus aportaciones durante mi examen profesional y a sus observaciones para este trabajo, que en gran medida ayudaron a pulirla. Deseo resaltar que gracias a los temas vistos en sus clases, se despertó un gran deseo de conocer sobre el *Análisis convexo*.

Agradezco a mis padres Tila López Pérez y José del Carmen Pérez Aguilar por la educación, acompañamiento y apoyo en todos mis estudios profesionales; por ser parte de mi vida y dirección en ella, Dios los bendiga siempre.

Gracias a todos mis amigos que siempre me recibieron en sus casas cuando los visitaba. Gracias a todos los que fueron y son parte de mi vida.

Bendito sea Dios ahora y por siempre. *Bendeciré al Señor con toda mi alma; no olvidaré ninguno de sus beneficios.*

Índice general

| | |
|--|-----------|
| Introducción | 1 |
| 1. Antecedentes de análisis convexo | 3 |
| 1.1. Relaciones monótonas y monótonas maximales | 3 |
| 1.2. Conexión entre relaciones monótonas maximales y funciones no decrecientes | 4 |
| 1.3. Subdiferenciación de funciones convexas | 5 |
| 1.4. Transformada de Legendre de funciones convexas | 6 |
| 1.5. Relación entre funciones convexas y segundas derivadas | 7 |
| 1.6. Convergencia de conjuntos en \mathbb{R}^2 | 7 |
| 2. Variables aleatorias y relaciones monótonas | 9 |
| 2.1. Representación de la esperanza en términos de la función cuantil. | 9 |
| 2.2. Relaciones monótonas inducidas por funciones de distribución y funciones cuantiles. | 10 |
| 2.3. Función supercuantil | 11 |
| 2.4. Función superesperanza | 12 |
| 2.5. Conexión entre la función supercuantil y superesperanza | 13 |
| 2.6. Convergencia en distribución y su relación con la convergencia gráfica | 15 |
| 2.7. Dominación estocástica | 16 |
| 2.8. Medidas de riesgo | 17 |
| 2.9. Supercuantiles como medida de riesgo | 18 |
| 3. Epi-splines | 19 |
| 3.1. Motivación | 19 |

| | | |
|-----------|--|-----------|
| 3.2. | Definición de epi-splines | 20 |
| 3.3. | Epi-splines de primer y segundo orden | 21 |
| 3.4. | Incorporación de información suave | 22 |
| 4. | Estimación de la función supercuantil | 23 |
| 4.1. | Función objetivo | 23 |
| 4.2. | Restricciones | 24 |
| 4.2.1. | Convexidad, continuidad y diferenciabilidad | 24 |
| 4.2.2. | Continuidad de la densidad | 25 |
| 4.2.3. | Puntos extremos | 25 |
| 4.2.4. | Cotas inferiores | 26 |
| 4.2.5. | Restricciones de cuantiles | 26 |
| 4.2.6. | Valores máximo y mínimo | 27 |
| 5. | Estimación de la función de densidad | 28 |
| 5.1. | Función objetivo | 28 |
| 5.2. | Restricciones | 29 |
| 5.2.1. | Unidad y no negatividad | 29 |
| 5.2.2. | Continuidad | 29 |
| 5.2.3. | Restricciones sobre cuantiles | 29 |
| 5.2.4. | Restricciones sobre supercuantiles | 30 |
| 5.3. | Implementación de las restricciones cuantil/supercuantil | 30 |
| 5.4. | Cuantificación del error de las estimaciones | 31 |
| | Conclusiones | 33 |
| | Bibliografía | 35 |

Introducción

Los supercuantiles son herramientas útiles para modelar y optimizar riesgos y son tan fundamentales para la descripción de una variable aleatoria como la función de distribución [8]. Tan es así, que se puede recuperar la función cuantil correspondiente a través de la diferenciación, y son duales a la función superesperanza, que son funciones convexas de las cuales se exponen también sus propiedades y su relación con los cuantiles.

La gran relevancia de los supercuantiles en el análisis de riesgos deriva de su ‘coherencia’ y ‘regularidad’ (Capítulo 2) cuando se considera como una medida del riesgo de una variable aleatoria. Estas propiedades hacen que los supercuantiles sean adecuados como representaciones escalares de una variable aleatoria en la toma de decisiones con aversión al riesgo.

En [5] se conjuntan estos conceptos, dando origen la función supercuantil y la función superesperanza, que son el tema central de este trabajo.

El propósito de esta tesis es exponer elementos del *Análisis Convexo* y mostrar algunas aplicaciones de esta teoría a la estadística, más concretamente en el ajuste de densidad a un conjunto de datos. La organización de este trabajo se expone a continuación.

El capítulo 1, tomado de [5] pone de manifiesto las relaciones monótonas y las relaciones monótonas maximales, además de la relación existente con funciones no decrecientes; en concreto, se estudia cómo obtener una función no decreciente dada una relación monótona maximal y viceversa.

En el Capítulo 2 se definen las funciones supercuantil y superesperanzas, las cuales juegan un papel muy importante dentro de este trabajo. Aunado a esto, se describe la relación existente entre ellas vía la transformada de Legendre. Se da un apartado de dominación estocástica y medidas de riesgo, y se presenta a la función supercuantil como una medida de riesgo.

En nuestro objetivo de aplicar estos conceptos al ajuste de densidad, en el Capítulo 3 se definen los *epi-splines*, que son una herramienta de aproximación de curvas dado un conjunto de datos. Este método se aplicará para aproximar la función supercuantil, y estas aproximaciones se utilizarán

para aproximar la función de densidad.

En el Capítulo 3 se definen los *epi-splines*, que son una herramienta de aproximación de curvas dado un conjunto de datos. Este método se aplicará para aproximar la función supercuantil, que a su vez, se utilizará para aproximar a la función de densidad subyacente de un conjunto de datos.

El Capítulo 4 describe explícitamente la forma de implementar la teoría expuesta en los Capítulos 1 y 2, además de aplicar los *epi-splines* en la aproximación de la función supercuantil asociada a un conjunto de datos.

El Capítulo 5 describe explícitamente cómo estimar la función de densidad subyacente de un conjunto de datos, a partir de la información obtenida de la función supercuantil y de ciertas restricciones que se irán imponiendo a dicha función de densidad.

Finalmente, se concluye este trabajo exponiendo las bondades de este método de aproximación de las funciones supercuantil y función de densidad de probabilidad, respectivamente. Asimismo, se exponen algunos aspectos que no se realizaron en esta tesis por la limitación de tiempo.

Capítulo 1

Antecedentes de análisis convexo

Un aspecto importante de las variables aleatorias son sus funciones de distribución. A su vez, estas funciones pueden ser identificadas con relaciones monótonas maximales. Tales relaciones son subdiferenciales de funciones convexas. El análisis de estas conexiones conlleva a nuevos conocimientos. La operación de inversión entre las funciones de distribución y las funciones cuantiles corresponde a la inversión gráfica de las relaciones monótonas. En términos de subdiferenciales, corresponde al paso para conjugar funciones convexas bajo la transformación de Legendre-Fenchel. Por tal motivo, en este capítulo se revisan algunos aspectos sobre *relaciones monótonas* y funciones convexas asociadas con ellas.

1.1. Relaciones monótonas y monótonas maximales

Definición 1 Un conjunto $\Gamma = \{(x, p) \mid (x, p) \in \mathbb{R} \times \mathbb{R}\}$ se dice que da una relación monótona si

$$(x_1 - x_2)(p_1 - p_2) \geq 0 \quad \forall (x_1, p_1), (x_2, p_2) \in \Gamma.$$

La simetría en los papeles de las dos variables de la relación monótona, tiene como consecuencia que si Γ es una relación monótona, entonces, la relación *inversa* $\Gamma^{-1} = \{(p, x) \mid (x, p) \in \Gamma\}$ también es una relación monótona.

A continuación se presenta el concepto de *relación monótona maximal*. Dicho concepto fue introducido en 1960 por Minty [1] en el estudio de las relaciones entre variables como corriente y

voltaje en redes eléctricas y sus análogos en otros tipos de redes.

Definición 2 Una relación monótona Γ es maximal si no existe relación monótona Γ' tal que $\Gamma \subset \Gamma'$ con $\Gamma \neq \Gamma'$.

En palabras, una relación monótona γ es maximal si no se puede ampliar sin destruir el orden total. Para una exposición más detallada de la teoría de relaciones monótonas maximales véase [2].

1.2. Conexión entre relaciones monótonas maximales y funciones no decrecientes

Suponga que $\gamma : (-\infty, +\infty) \rightarrow [-\infty, +\infty]$ es no decreciente y no idénticamente $-\infty$ ni idénticamente $+\infty$. Entonces, del cálculo diferencial se sabe que en cualquier punto de su dominio, los límites laterales existen

$$\gamma^-(x_0) = \lim_{x \rightarrow x_0^-} \gamma(x), \quad \gamma^+(x_0) = \lim_{x \rightarrow x_0^+} \gamma(x), \quad (1.1)$$

con $\gamma^-(x_0) \leq \gamma(x_0) \leq \gamma^+(x_0)$. γ^+ y γ^- definen funciones que son continuas por la izquierda y por la derecha, respectivamente.

En relación con lo anterior, se obtiene una relación monótona maximal Γ al tomar

$$\Gamma = \{(x, p) \in \mathbb{R} \times \mathbb{R} \mid \gamma^-(x) \leq p \leq \gamma^+(x)\}.$$

Recíprocamente, dada una relación monótona maximal Γ , se puede definir

$$\begin{aligned} \gamma^-(x) &= \min \{p \mid (x, p) \in \Gamma\} \text{ para } x \in \text{dom}\Gamma, \\ \gamma^+(x) &= \max \{p \mid (x, p) \in \Gamma\} \text{ para } x \in \text{dom}\Gamma, \\ \gamma^-(x) &= \gamma^+(x) = -\infty \text{ en puntos } x \text{ a la izquierda de } \text{dom}\Gamma \text{ (si los hay),} \\ \gamma^-(x) &= \gamma^+(x) = +\infty \text{ en puntos } x \text{ a la derecha de } \text{dom}\Gamma \text{ (si los hay),} \end{aligned}$$

para obtener un par de funciones no decrecientes γ^- y γ^+ , continuas por la izquierda y por la derecha, respectivamente.

1.3. Subdiferenciación de funciones convexas

En esta sección se explica la conexión entre las relaciones monótonas maximales y la subdiferenciación de funciones convexas.

Definición 3 Una función propiamente convexa sobre $(-\infty, +\infty)$ es una función $f : (-\infty, +\infty) \rightarrow (-\infty, +\infty]$, tal que $f \not\equiv +\infty$ y satisface

$$f([1 - \tau]x + \tau y) \leq [1 - \tau]f(x) + \tau f(y) \text{ para todo } \tau \in (0, 1) \text{ y todo } x, y.$$

Definición 4 Una función propiamente convexa es cerrada si sus conjuntos de nivel, a saber, $\{x \mid f(x) \leq c\}$, son conjuntos cerrados, para todo $c \in \mathbb{R}$.

Definición 5 El epigrafo de una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es el conjunto

$$\text{epi}(f) = \{(x, \mu) \mid x \in \mathbb{R}^n, \mu \in \mathbb{R}, f(x) \leq \mu < +\infty\}.$$

La relación entre funciones propiamente convexas y su epigrafo es el siguiente teorema.

Teorema 1 Una función propiamente convexa es cerrada si y solo si $\text{epi}(f)$ es un conjunto cerrado en \mathbb{R}^{n+1} .

Dado que una función $f : \mathbb{R} \rightarrow \mathbb{R}$ finita convexa sobre un intervalo abierto es necesariamente continua, una función propiamente convexa ha de ser continua excepto quizás en los puntos finales de $\text{dom} f$. Por lo tanto, que f sea cerrada sólo se refiere al comportamiento en esos extremos.

Para una función propiamente convexa f y cualquier $x_0 \in \text{dom} f$, las derivadas por la izquierda y por la derecha de f , a saber,

$$f'^-(x_0) = \lim_{x \rightarrow x_0^-} \frac{f(x) - f(x_0)}{x - x_0}, \quad f'^+(x_0) = \lim_{x \rightarrow x_0^+} \frac{f(x) - f(x_0)}{x - x_0},$$

existen, en virtud de que la función definida mediante la regla

$$g(x) = \frac{f(x) - f(x_0)}{x - x_0} \quad \forall x \neq x_0,$$

es no decreciente, y además $f'^-(x_0) \leq f'^+(x_0)$. Por otro lado, la función $\partial f(x)$ definida como

$$\partial f(x) = \begin{cases} \{p \in \mathbb{R} \mid f^-(x_0) \leq p \leq f^+(x_0)\} & \text{si } x \in \text{dom} f, \\ \emptyset & \text{si } x \notin \text{dom} f, \end{cases}$$

es llamada la *subdiferencial* de f . El hecho clave que relaciona el subdiferencial de una función y las relaciones monótonas, es el siguiente:

Teorema 2 *Para una función cerrada propiamente convexa, la gráfica de la función ∂f , a saber*

$$\text{gph}(\partial f) = \{(x, p) \mid p \in \partial f(x)\},$$

es una relación monótona maximal Γ . Aún más, cualquier relación monótona maximal Γ es la gráfica de ∂f , para alguna función cerrada propiamente convexa f , y tal función f está determinada únicamente salvo por una constante aditiva.

1.4. Transformada de Legendre de funciones convexas

El siguiente concepto es de gran relevancia el análisis convexo y se utilizará para mostrar la relación existente entre las funciones supercuantil y superesperanza.

Definición 6 *La transformada de Legendre de una función $f : I \rightarrow \mathbb{R}$ definida sobre un intervalo $I \subset \mathbb{R}$ es la función*

$$f^*(t) = \sup_{x \in I} \{tx - f(x)\}. \quad (1.2)$$

Como lo muestra el siguiente teorema, la transformada de Legendre de una función convexa es involutiva (su cuadrado es la transformación identidad).

Teorema 3 *Si en la ecuación (1.2) la función f es propiamente convexa sobre $(-\infty, +\infty)$, la transformada Legendre define una función cerrada propiamente convexa f^* sobre $(-\infty, +\infty)$ tal que*

$$f(x) = \sup_p \{xp - f^*(p)\} = (f^*)^*(x).$$

El paso de f a f^ es la transformación de Legendre.*

1.5. Relación entre funciones convexas y segundas derivadas

Que una función cerrada propiamente convexa f sea dos veces diferenciable en $x \in \text{dom} f$, significa que las funciones derivadas laterales, f'^- y f'^+ , coinciden en x . Gráficamente, en términos de la relación monótona $\Gamma = \partial f$, una formulación equivalente es: existe un único punto p tal que $(x, p) \in \Gamma$, en donde la línea tangente a Γ en ese punto es no vertical. Aquí, $f'(x) = p$ y la pendiente de la tangente es $f''(x)$. Por supuesto, esta pendiente debe ser no negativa.

Definición 7 *Un punto (x, p) de Γ es llamado no singular si existe una línea tangente en dicho punto que es no vertical y no horizontal.*

Esto significa que f tiene una segunda derivada diferente de cero en x .

Teorema 4 *Las siguientes afirmaciones son equivalentes:*

(i) f es derivable en un punto x tal que $f'(x) = p$ y $f''(x) > 0$,

(ii) (x, p) es un punto no singular de Γ ,

(iii) (p, x) es un punto no singular de $\Delta = \Gamma^{-1}$,

(iv) f^* cumple $(f^*)'(p) = x$, $(f^*)''(p) > 0$ y además $(f^*)''(p) = 1/f''(x)$.

1.6. Convergencia de conjuntos en \mathbb{R}^2

Finalmente, se exponen las nociones de convergencia que conciernen al análisis convexo. Esta teoría de convergencia se expone para el caso \mathbb{R}^2 porque es todo lo que se requiere para este trabajo, sin embargo, una teoría completa en dimensiones finitas se expone en [2].

Definición 8 *Para un subconjunto cerrado no vacío $S \subseteq \mathbb{R}^2$ y $u \in \mathbb{R}^2$, se define la distancia del conjunto S a u como*

$$d_S(u) = \min_{w \in S} \|u - w\| \text{ para la norma euclidiana } \|\cdot\|. \quad (1.3)$$

Se está interesado en decidir cuándo sucesiones de subconjuntos cerrados no vacíos $S_k \subseteq \mathbb{R}^2$ convergen a un subconjunto $S \subseteq \mathbb{R}^2$ cuando $k \rightarrow \infty$; esta convergencia se dice ser en el sentido de Kuratowski-Painvelé.

Aunque hay numerosas caracterizaciones de este tipo de convergencia, llamada convergencia de conjuntos o convergencia gráfica [2], es suficiente concentrarse en la siguiente definición particular:

Definición 9 $\lim_{k \rightarrow \infty} S_k = S \iff \lim_{k \rightarrow \infty} d_{S_k}(u) = d_S(u)$ para cualquier $u \in \mathbb{R}^2$.

Capítulo 2

VARIABLES ALEATORIAS Y RELACIONES MONÓTONAS

Este capítulo muestra que la convergencia en distribución de una sucesión de variables aleatorias es equivalente a la convergencia gráfica de las relaciones monótonas. Asimismo, se presentan dos nuevas funciones que tienen el protagonismo de este trabajo, la función supercuantil y la función superesperanza, cuya relación también se exponen. Asimismo se aplican los resultados expuestos en el Capítulo 1 a variables aleatorias. En lo que resta de este trabajo, a menos que se diga lo contrario, se asume que la esperanza de cualquier variable aleatoria X existe y es finita.

2.1. Representación de la esperanza en términos de la función cuantil.

El valor esperado de una variable aleatoria X sobre el espacio de probabilidad (Ω, \mathcal{F}, P) es la integral de X respecto a la medida P ,

$$EX = \int X dP = \int_{\Omega} X(\omega) P(d\omega).$$

Suponga que X tiene distribución F . Si g es una función real de variable real, por la fórmula de cambio de variable expuesta en [3]

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x)F(dx).$$

Una expresión de la misma esperanza en términos de la función cuantil Q_X en lugar de la función de distribución F es:

$$E[g(X)] = \int_0^1 g(Q_X(p))dp,$$

en virtud del teorema [3]¹. En particular,

$$\begin{aligned} EX &= \int_{-\infty}^{+\infty} xF(dx) = \int_0^1 Q_X(p)dp, \\ E|X|^r &= \int_{-\infty}^{+\infty} |x|^r F(dx) = \int_0^1 |Q_X(p)|^r dp, \text{ para } r \geq 1. \end{aligned}$$

2.2. Relaciones monótonas inducidas por funciones de distribución y funciones cuantiles.

La función de distribución F_X de una variable aleatoria es continua por la derecha y no decreciente; su contraparte Q_X , es no decreciente y continua por la izquierda. De la Sección 1.2, la relación monótona

$$\Gamma_X = \{(x, p) \in \mathbb{R} \times \mathbb{R} \mid F_X^-(x) \leq p \leq F_X(x)\}$$

es maximal.

Si se extiende Q_X fuera del intervalo $(0, 1)$ mediante

$$Q_X(1) = \lim_{p \rightarrow 1^-} Q_X(p), \quad Q_X(p) = +\infty \text{ para } p > 1, \quad Q_X(p) = -\infty \text{ para } p \leq 0,$$

¹Capítulo 3, Teorema 16.13.

se obtiene una función no decreciente continua por la izquierda sobre \mathbb{R} que además cumple que

$$\begin{aligned} Q_X^+(0) &= \lim_{p \rightarrow 0^+} Q_X(p) \\ Q_X^+(1) &= \infty. \end{aligned}$$

Así, la relación

$$\Delta_X = \{(p, x) \in \mathbb{R} \times \mathbb{R} \mid Q_X(p) \leq x \leq Q_X^+(p)\}$$

es una relación monótona maximal.

Estas relaciones, Γ_X y Δ_X son inversas, es decir:

$$(x, p) \in \Gamma_X \iff (p, x) \in \Delta_X, \text{ es decir, } \Delta_X = \Gamma_X^{-1} \text{ y } \Gamma_X = \Delta_X^{-1}.$$

2.3. Función supercuantil

Comúnmente una variable aleatoria real X se caracteriza por su *función de distribución* $F_X : \mathbb{R} \rightarrow [0, 1]$, la cual es continua por la derecha y está dada por

$$F_X(x) = P(X \leq x) \text{ para } x \in \mathbb{R}.$$

Una caracterización equivalente se da en términos de la función *cuantil* $Q_X : (0, 1) \rightarrow \mathbb{R}$ continua por la izquierda expresada como

$$Q_X(p) = \min \{x \mid F_X(x) \geq p\} \text{ para } p \in (0, 1).$$

Si F_X es estrictamente creciente, es claro que $Q_X(p) = F_X^{-1}(p)$ ya que su inversa existe y además, F_X se puede construir a partir del conocimiento de Q_X por medio de la fórmula

$$F_X(x) = \begin{cases} \max \{p \mid Q_X(p) \leq x\} & \text{para } x \in (\inf Q_X, \sup Q_X], \\ 1 & \text{para } x > \sup Q_X, \\ 0 & \text{para } x \leq \inf Q_X. \end{cases}$$

Otra descripción de una variable aleatoria X se da en términos de la *función supercuantil*

$\bar{Q}_X : (0, 1) \rightarrow (-\infty, +\infty]$ donde

$\bar{Q}_X(p)$ = esperanza en la distribución de la p-cola superior de X , para $p \in (0, 1)$.

Aquí la esperanza se refiere a la distribución de probabilidad sobre $[Q_X(p), +\infty)$, la cual en el caso de que $F_X(Q_X(p)) = p$, es la distribución condicional de X sujeto a $X \geq Q_X(p)$.

De forma equivalente, la función supercuantil está dada por

$$\bar{Q}_X(p) = \frac{1}{1-p} \int_p^1 Q_X(y) dy \text{ para } p \in (0, 1). \quad (2.1)$$

Esta expresión pone de relieve una conexión entre los cuantiles y los supercuantiles, siendo esta última una “media” de la primera. Para una explicación de la equivalencia entre las dos fórmulas consúltese a Rockafellar y Royset [5]. Como Q_X es no decreciente, continua por la izquierda con a lo más una cantidad numerable de saltos, Q_X es Lebesgue medible. Además, el integrando en (2.1) está acotado inferiormente por $Q_X(p)$, y por tanto la integral está bien definida, aunque el valor puede ser infinito. Para $p = 0$ y $p = 1$ es natural extender la definición al hacer

$$\bar{Q}_X(0) = E[X] \text{ y } \bar{Q}_X(1) = \sup X,$$

lo que nos lleva a una función continua en $[0, 1]$, donde la primera afirmación requiere que la esperanza esté bien definida y la segunda implique el supremo esencial².

2.4. Función superesperanza

Se define la función *superesperanza* $E_X : \mathbb{R} \rightarrow \mathbb{R}$ asociada con una variable aleatoria X por

$$E_X(x) = E[\text{máx}\{x, X\}] = \int_{-\infty}^{+\infty} \text{máx}\{x, y\} F_X(dy) = \int_0^1 \text{máx}\{x, Q_X(p)\} dp,$$

²Sea $(\Omega, \mathcal{A}, \mu)$ un espacio de medida y $f : \Omega \rightarrow \bar{\mathbb{R}}$ una función Borel medible. Sea $a \in \mathbb{R}$ y defina

$$\begin{aligned} M_a &= \{x \mid f(x) > a\} \\ A_0 &= \{a \in \mathbb{R} \mid \mu(M_a) = 0\}. \end{aligned}$$

El supremo esencial de f es

$$\text{ess sup } f := \inf A_0.$$

donde el valor $E_X(x)$ se denomina la *superesperanza* de X en el nivel x y la última igualdad se sigue de la fórmula de cambio de variable expuesta en [3].

Del siguiente teorema, es clara la conexión entre la función superesperanza E_X y Γ_X ; aquí se utiliza la notación ∂E_X para denotar el subdiferencial de E_X y $\text{gph}(\partial E_X)$ la gráfica correspondiente.

Teorema 5 (*Caracterización de superesperanza*) *La función superesperanza E_X para una variable aleatoria X tal que $E|X| < +\infty$ es una función convexa finita no decreciente en \mathbb{R} con las siguientes propiedades:*

(i) $\Gamma_X = \text{gph}(\partial E_X)$,

(ii) $F_X(x) = E'_X(x+)^3$,

(iii) $E_X(x) - x \geq 0$, $\lim_{x \rightarrow +\infty} [E_X(x) - x] = 0$, y $\lim_{x \rightarrow -\infty} E_X(x) = E[X]$.

(iv) Para cualesquiera variables aleatorias X_0, X_1 con $E|X_0| < +\infty$ y $E|X_1| < +\infty$,

$$E_X(x) \leq (1 - \lambda)E_{X_0}(x) + \lambda E_{X_1}(x) \text{ cuando } X = (1 - \lambda)X_0 + \lambda X_1, \quad 0 < \lambda < 1.$$

Además, cualquier función convexa f en \mathbb{R} con las propiedades

$$f(x) - x \geq 0, \quad \lim_{x \rightarrow +\infty} [f(x) - x] = 0, \quad \lim_{x \rightarrow -\infty} f(x) = \text{un valor finito},$$

es una función superesperanza para una variable aleatoria X tal que $E|X| < +\infty$.

2.5. Conexión entre la función supercuantil y superesperanza

La conexión entre la función superesperanza y la función supercuantil surge de la transformada de Legendre. El siguiente teorema proporciona una expresión para la función conjugada de E_X y sus propiedades. En esencia, el resultado se estableció en [6] y [7].

³ $E'_X(x+)$ es la derivada por la derecha de E_X .

Teorema 6 (Dual de las superesperanzas) La función cerrada propiamente convexa E_X^* sobre \mathbb{R} que es conjugada de la función superesperanza E_X para una variable aleatoria X tal que $E|X| < +\infty$ está dada por

$$E_X^*(p) = \begin{cases} -(1-p)\bar{Q}_X(p) & \text{si } p \in (0, 1), \\ -E[X] & \text{si } p = 0, \\ 0 & \text{si } p = 1, \\ \infty & \text{si } p \notin [0, 1]. \end{cases}$$

E_X^* tiene las siguientes propiedades:

(i) $\Delta_X = \text{gph}(\partial E_X^*)$.

(ii) $Q_X(p) =$ derivada por la izquierda de E_X^* en p .

(iii) E_X^* es continua en $[0, 1]$ con

$$\lim_{p \rightarrow 1^-} (1-p)\bar{Q}_X(p) = 0, \quad \lim_{p \rightarrow 0^+} \bar{Q}_X(p) = E[X].$$

Además, cualquier función finita g en \mathbb{R} que sea convexa y continua en $[0, 1]$ que satisfaga

$$g(p) = \begin{cases} 0 & \text{si } p = 1 \\ \infty & \text{si } p \notin [0, 1], \end{cases}$$

es la conjugada de una función superesperanza para alguna variable aleatoria X .

La conjugada E_X^* está únicamente determinada por la función supercuantil \bar{Q}_X . No sólo eso, sino que también E_X , F_X , Q_X , Γ_X y Δ_X se pueden reconstruir a partir del conocimiento de \bar{Q}_X (véase [7]). Además, las siguientes propiedades de una función \bar{g} sobre $(0, 1)$ son necesarias y suficientes para que $\bar{g} = \bar{Q}_X$ para una variable aleatoria X tal que $E|X| < \infty$:

$$(1-p)\bar{g}(p) \text{ es cóncava en } p \text{ con } \lim_{p \rightarrow 1^-} (1-p)\bar{g}(p) = 0, \quad \lim_{p \rightarrow 0^+} \bar{g}(p) = \text{un valor finito.}$$

El siguiente ejemplo es de ayuda para ilustrar los conceptos [5].

Sea X distribuída exponencialmente con parámetro $\lambda > 0$. Entonces la función de distribución es $F_X(x) = 1 - e^{-\lambda x}$, la función superesperanza es

$$E_X(x) = \begin{cases} x + \frac{1}{\lambda}e^{-\lambda x} & \text{si } x \geq 0 \\ \frac{1}{\lambda} & \text{si } x < 0, \end{cases}$$

y la función superesperanza conjugada es

$$E_X^*(p) = (1/\lambda)(p-1)(1-\log(1-p)) \text{ para } p \in [0, 1).$$

Las funciones cuantil y supercuantil en $(0, 1)$ están dadas por

$$Q_X(p) = -(1/\lambda)\log(1-p), \quad \bar{Q}_X(p) = (1/\lambda)[1-\log(1-p)].$$

Estos resultados proporcionan nuevas estimaciones para trabajar con supercuantiles, como lo muestra el siguiente teorema expuesto en [5].

Teorema 7 (*Estimaciones de supercuantiles*) Para $p \in [0, 1)$ se tiene

$$(i) \quad |\bar{Q}_X(p) - \bar{Q}_Y(p)| \leq \frac{1}{1-p}E|X - Y| \text{ siempre que } E|X| < +\infty, E|Y| < +\infty.$$

$$(ii) \quad E[X] \leq \bar{Q}_X(p) \leq E[X] + \frac{1}{\sqrt{1-p}}\sigma(X) \text{ siempre que } E[X^2] < +\infty, \sigma(X) = \text{desviación estándar.}$$

El desarrollo anterior permite caracterizaciones alternativas de la noción clásica de convergencia en distribución para una sucesión de variables aleatorias. De nuevo, la función supercuantil juega un nuevo papel.

2.6. Convergencia en distribución y su relación con la convergencia gráfica

A modo de repaso, una sucesión de variables aleatorias X_k converge en distribución a una variable aleatoria X si $F_{X_k}(x) \rightarrow F_X(x)$ en todos los puntos de continuidad x de F_X .

El siguiente teorema expone la equivalencia de distintos tipos de convergencia, como lo es la convergencia gráfica y la convergencia en distribución.

Teorema 8 (*Caracterización de convergencia en distribución*) Sea X, X_k variables aleatorias con $E|X| < +\infty, E|X_k| < +\infty, k = 1, 2, \dots$. Entonces, las siguientes condiciones son equivalentes:

- (i) X_k converge en distribución a X .
- (ii) Γ_{X_k} converge gráficamente a Γ_X .
- (iii) Δ_{X_k} converge gráficamente a Δ_X .
- (iv) $Q_{X_k}(p) \rightarrow Q_X(p)$ en todos los puntos de continuidad p de Q_X en $(0, 1)$.
- (v) $E_{X_k} \rightarrow E_X(x)$ para todo $x \in \mathbb{R}$.
- (vi) $\bar{Q}_{X_k}(p) \rightarrow \bar{Q}_X(p)$ para todo $p \in (0, 1)$.

2.7. Dominación estocástica

A continuación se expone una breve discusión de dominación estocástica, donde sobresale su relación con las funciones supercuantil y superesperanzas (véase [8, 9]).

Definición 10 *La dominación estocástica de primer orden de X sobre Y , denotada como $X \leq_1 Y$, se define como*

$$X \leq_1 Y \iff E[g(X)] \leq E[g(Y)] \text{ para una función } g \text{ continua, creciente y acotada.}$$

La dominación estocástica de segundo orden denotada como $X \leq_2 Y$, significa que

$$X \leq_2 Y \iff E[g(X)] \leq E[g(Y)] \text{ para una función } g \text{ creciente, convexa y finita.}$$

La conexión entre estas nociones y las funciones de distribución, cuantil, superesperanza y supercuantil se enuncia en el siguiente teorema, de donde se tiene que la función superesperanza al igual que la función supercuantil caracterizan la dominación estocástica de segundo orden.

Teorema 9 (Dominación estocástica)

$$X \leq_1 Y \iff F_X \geq F_Y \iff Q_X \leq Q_Y, \tag{2.2}$$

$$X \leq_2 Y \iff E_X \leq E_Y \iff \bar{Q}_X \leq \bar{Q}_Y. \tag{2.3}$$

2.8. Medidas de riesgo

Una medida de riesgo \mathcal{R} es un funcional que asigna a una variable aleatoria X un valor $\mathcal{R}(X)$ en $(-\infty, \infty]$, por lo que la comparación entre dos variables aleatorias X y Y se puede reducir a comparar los números reales $\mathcal{R}(X)$ y $\mathcal{R}(Y)$.

Existe una demanda de orientación sobre lo que constituiría una buena y útil medida de riesgo. Hay dos conceptos que destacan en este aspecto: coherencia y regularidad.

Una medida de riesgo \mathcal{R} es *coherente* en el sentido de Artzner y otros [10] (ver también Delbaen [11]) si satisface los siguientes axiomas:

- (i) $\mathcal{R}(C) = C$ para variables aleatorias constantes $X \equiv C$,
- (ii) $\mathcal{R}(X) \leq \mathcal{R}(Y)$ cuando $X \leq Y$ a.s. (monotonía),
- (iii) $\mathcal{R}(X + Y) \leq \mathcal{R}(X) + \mathcal{R}(Y)$ (subaditividad),
- (iv) $\mathcal{R}(\lambda X) = \lambda \mathcal{R}(X)$ para $\lambda > 0$ (homogeneidad positiva).

Ejemplos de medidas coherentes son:

- (i) $\mathcal{R}(X) = \mathbb{E}[X]$,
- (ii) $\mathcal{R}(X) = \sup X$ (supremo esencial),
- (iii) $\mathcal{R}(X) = \bar{Q}_X(p)$ (función supercuantil).

Por último y sin entrar en detalles técnicos, una medida de riesgo \mathcal{R} es *regular* en el sentido de Rockafellar y Uryasev [2] si satisface los siguientes cuatro axiomas:

- (i) $\mathcal{R}(C) = C$ para variables aleatorias constantes $X = C$,
- (ii) $\mathcal{R}((1 - \tau)X + \tau X') \leq (1 - \tau)\mathcal{R}(X) + \tau\mathcal{R}(X')$ para todo X, X' y $\tau \in (0, 1)$ (convexidad),
- (iii) $\{X \mid \mathcal{R}(X) \leq C\}$ es cerrado para todo $C \in \mathbb{R}$,
- (iv) $\mathcal{R}(X) > E[X]$ para X diferente de una constante (aversión).

2.9. Supercuantiles como medida de riesgo

Los métodos mediante los cuales se interpreta, mide y comunica la incertidumbre constituyen la base para la cuantificación del riesgo. La probabilidad de superar un valor umbral es una de esas medidas de riesgo, que en este caso utiliza un valor de cuantil particular como medida de riesgo. Los cuantiles, sin embargo, son sólo una de muchas de estas medidas de riesgo que pueden ser usadas. De particular importancia son los supercuantiles, que poseen varias características que los hacen particularmente bien adaptados como medidas de riesgo. Conocidos más comúnmente como ‘condicional value-at-risk’, ‘value-at-risk’, ‘tail value-at-risk’ y ‘expected shortfall’ de sus aplicaciones en el análisis financiero, muchos analistas han llegado a preferir supercuantiles por sus propiedades deseables de coherencia y regularidad en la evaluación del riesgo bajo modelos probabilísticos incompletos o inexactos (véase [7, 9]).

Dada una variable aleatoria X , podemos calcular el cuantil de X en un cierto percentil p como valor umbral por debajo del cual existe una proporción p de la función de densidad subyacente. Si X es continua, entonces la función cuantil $Q(p)$ es simplemente la inversa de la función de distribución. Un supercuantil es el valor esperado siempre que se sabe que se está operando en la p -cola superior de la distribución de probabilidad subyacente.

Siempre que se conozca el supercuantil de una distribución en un percentil dado, buscaremos aprovechar dicha información para hacer mejores estimaciones de la cola de la densidad de dicha distribución. Al llegar a mejores estimaciones de valores supercuantiles para p en rangos más cercanos a uno, esperamos evaluar con mayor precisión la probabilidad de superar un umbral de riesgo particular.

Capítulo 3

Epi-splines

Aunque los epi-splines están relacionados estructuralmente con los splines ‘clásicos’, en lugar de ser una herramienta de interpolación, son primordialmente una herramienta de aproximación. Esto les permite acomodar más fácilmente una clase más rica de información acerca de la función, o sistema, que se esté tratando de aproximar, produciendo resultados muy buenos en algunos casos. La teoría que apoya sus propiedades está en gran medida basada en el análisis variacional y las implementaciones se basan casi siempre en la disponibilidad de rutinas de optimización que se han desarrollado con tanto éxito en el último medio siglo [19].

3.1. Motivación

Tradicionalmente, el ajuste de una curva h cuyos valores son conocidos en un número finito de puntos se basó en la aproximación polinomial, usualmente en el mejor ajuste cuadrático medio, una técnica ahora incorporada en prácticamente todos los paquetes matemáticos estándar. La Figura 3.1 da un ejemplo donde h se aproxima por un polinomio $q(x) = a_n x^n + \dots + a_1 x + a_0$ de grado 5 dado que conocemos los valores de $h : [0, 1] \rightarrow \mathbb{R}$ en un número finito de puntos $k/20$ para $k = 1, \dots, 20$; polinomios de grado $n \neq 5$ generan ajustes aún menos aceptables, especialmente cuando el rango va ligeramente más allá del intervalo $[0, 1]$. La aproximación parece demasiado áspera lo que condujo al desarrollo de una técnica, si no universalmente exacta, que proporcionaría interpolaciones más precisas.

Estructuralmente los epi-splines están relacionados con los splines, pero el objetivo perseguido es fundamentalmente diferente. De hecho, los splines clásicos de interpolación pueden ser vistos como

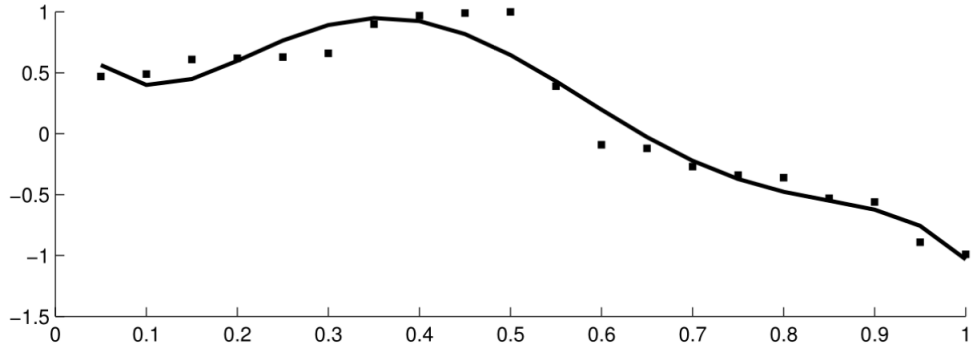


Figura 3.1: Ajuste de polinomio de grado 5.

una subfamilia estratégica de epi-splines que son específicamente adecuados para tratar problemas en los que la noción de aproximación se formula en términos de ‘igualar’ los valores de una función y posiblemente algunas derivadas en un número finito de puntos. Pero en esencia, en lugar de buscar una interpolación ‘óptima’, los epi-splines se ocupan de encontrar una aproximación ‘óptima’ con respecto a algún criterio (restricciones, por ejemplo), que puede variar de una aplicación a otra. Pero esta función de aproximación o más generalmente el sistema, también será capaz de tener en cuenta la información externa que podría estar disponible sobre la función o el sistema que se intenta aproximar.

Bajo el nombre de epi-curvas, los epi-splines fueron utilizados por primera vez para obtener *curvas cero*¹ asociadas con una familia de instrumentos financieros [12]. Su uso en el contexto de la estimación de funciones de densidad, donde la experimentación computacional desempeñó un papel importante [13, 14, 15], llevó a algunas investigaciones relacionadas con la mejora de la ley de los grandes números de funciones aleatorias semi continuas inferiores (random lower semi-continuous functions) [16].

3.2. Definición de epi-splines

Casi inevitablemente, encontrar una función que resuelva un sistema de ecuaciones funcionales de dimensión infinita requiere reducir el sistema o el problema de optimización a una versión de dimensión finita que esperamos genere una solución aproximada. Generalmente, esto significa que la solución de aproximación va a ser determinada mediante la fijación de un número finito de

¹Una curva cero es una serie de factores de descuento que representan el valor actual de un dólar recibido en el futuro.

parámetros o va a ser una combinación lineal de una colección finita de funciones seleccionadas cuidadosamente o será directamente determinada por esta colección finita de parámetros.

A continuación se presenta la definición de epi-splines.

Definición 11 *Un epi-spline $s : [m_0, m_N] \subset \mathbb{R} \rightarrow \mathbb{R}$ con malla $m = \{m_k\}_{k=0}^N$ que particiona su dominio, es de orden $p \in \mathbb{N}_0$ ² si*

(i) *En cada intervalo (no vacío, abierto) (m_{k-1}, m_k) para $k = 1, \dots, N$, s es un polinomio de grado p .*

(ii) *El polinomio s es de valores finitos sobre m .*

La familia de todos estos epi-splines se denota como $e - \text{spl}^p(m)$.

Dada una malla m , $|m| = \max_{1 \leq k \leq N} (m_k - m_{k-1})$ denotará el grado de la malla. La malla m está anidada en la malla \bar{m} si $m \subseteq \bar{m}$.

3.3. Epi-splines de primer y segundo orden

Mediante la modificación de un conjunto de restricciones que definen el problema ya sea a través de información empírica a partir de los datos en sí o de información externa suave, se puede desarrollar un marco que identifica los valores de los coeficientes que definen la función epi-spline por partes. Los epi-splines son adecuados en el manejo de múltiples restricciones de forma y en la aproximación de funciones de densidad mediante la maximización de una función de verosimilitud o función de entropía, incluso para pequeños conjuntos de datos como se muestra en [17].

Dado un intervalo cerrado $[l, u]$, se divide el intervalo en segmentos de mallas uniformemente espaciados (Figura 3.2)³. Aquí se distingue entre una malla de epi-splines k de segundo orden utilizado para la estimación de la dual superesperanza (en adelante DSE), de una malla m de primer orden utilizado para la estimación de la densidad. Como tal, se cuenta con K segmentos de malla de segundo orden uniformemente espaciados, por lo que se define p como $p = \{p^k | k = 1, 2, \dots, K\}$; además, se tiene M segmentos de malla de primer orden espaciados uniformemente, con lo que $x = \{x^m | m = 1, 2, \dots, M\}$. Los segmentos de malla son continuos a la derecha, con puntos finales

² $\mathbb{N} = \{1, 2, \dots\}$ y $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$

³Sabol J. (2016). Dual approach to superquantile estimation and applications to density fitting [Tesis]. Recuperado de <https://www.semanticscholar.org/paper/Dual-Approach-to-Superquantile-Estimation-and-to-Sabol-Royset/ad777ca96feeee562a3c099eb9d731f9614bf354>

definidos como izquierdos (p_L, x_L) o derechos (p_R, x_R) respectivamente, como se ve en la Figura 3.2. Los segmentos epi-splines se definen como polinomios de primer o segundo orden según su asignación de malla.

$$\begin{aligned} \text{Segundo orden: } \hat{\mathbb{E}}_X^*(p) &= a_0^k + a_1^k p + a_2^k p^2 \quad \text{para } p \in [p_L^k, p_R^k), \\ \text{Primer orden: } \hat{f}(x) &= b_0^m + b_1^m x \quad \text{para } x \in [x_L^m, x_R^m). \end{aligned} \tag{3.1}$$

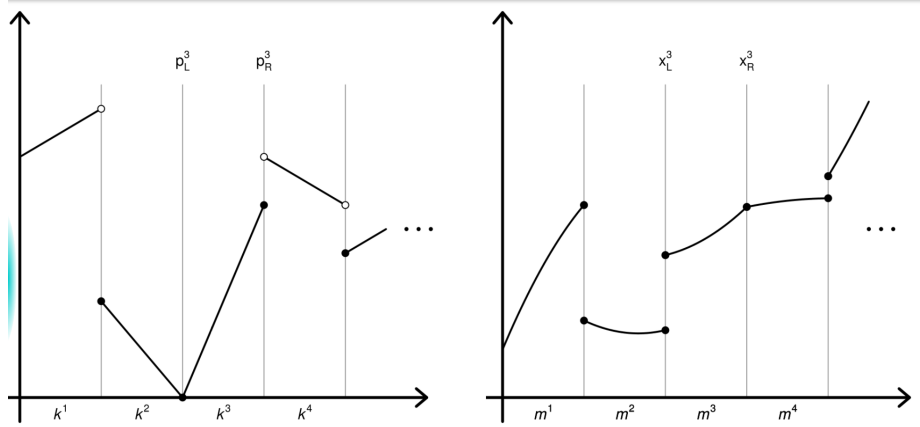


Figura 3.2: Epi-splines de primer (izquierda) y segundo (derecha) orden con mallas asociadas. Al aumentar la resolución de malla, se logran aproximaciones arbitrariamente cercanas a prácticamente cualquier curva utilizando epi-splines de bajo orden.

Para diferenciar la notación entre formulaciones, se utilizan coeficientes ‘ a ’ para la optimización de segundo orden definida a través del índice de malla k , y coeficientes ‘ b ’ para la optimización de primer orden definida a través del índice de malla m . Los detalles de las implementaciones de epi-spline se tratan más adelante.

3.4. Incorporación de información suave

A menudo, las deficiencias en el tamaño de la muestra se pueden mitigar parcialmente a través de la incorporación de información suave, que proporciona información cualitativa sobre la función de densidad subyacente, además de la información empírica proporcionada por los propios datos. Ejemplos de información suave incluyen características de densidad tales como monotonicidad, convexidad de cola, unimodalidad y muchas otras. Combinaciones de información suave a través de la optimización restringida conduce a mejores estimaciones de la densidad para limitados conjuntos de datos cuando se implementa en un marco inteligente, como se muestra en [13, 18].

Capítulo 4

Estimación de la función supercuantil

En adelante y a menos que se diga lo contrario, se asume que todos los datos son una muestra independiente e idénticamente distribuída (en adelante IID) de una función de densidad desconocida, con media y varianza finita. La distribución no tiene que pertenecer a un modelo paramétrico, ya que los epi-splines se adaptan bien a la estimación no paramétrica, sin embargo, sin un segundo momento finito en la distribución subyacente, estos métodos tendrán una aplicación inadecuada debido a los intentos de estimar tanto la media como la varianza para uso en selectas restricciones dentro de las optimizaciones. Dado que la muestra (datos) es IID, no poseen dependencia temporal u otra estructura de ordenación.

En este capítulo se expone cómo se estima un supercuantil deseado a partir de optimización restringida de epi-splines de segundo orden. Utilizando las características conocidas de la DSE y su relación con supercuantiles, se realiza la optimización restringida para llegar a estimaciones de $E_X^*(p)$. Luego se convierten en estimaciones de $\bar{Q}_X(p)$ a través de la relación descrita en la Sección 2.5.

Dado que se sabe que $E_X^*(p)$ solo existe para $p \in [0, 1]$ se define libremente la longitud de los segmentos epi-splines de segundo orden conforme una resolución de malla deseada.

4.1. Función objetivo

Se desea aproximaciones conservadoras para la cola derecha de la densidad de tal forma que nuestras estimaciones para cuantiles y supercuantiles cercanos a $p = 1$ sean al menos tan altas como los valores reales. Para lograr esto, se selecciona una función objetivo que maximiza la curvatura

de los segmentos individuales de los epi-splines, denotados como $\kappa(k)$. Con el fin de promover la curvatura sobre toda la malla (en vez de algunos segmentos individuales) se aplica un término de suavidad adicional que penaliza los cambios de a_2 entre epi-splines consecutivos. El grado de penalización se rige por un parámetro de suavidad ρ que se inicializa en cero y se aumenta según sea necesario para obtener estimaciones de DSE suaves basadas en la *observación visual*. Por lo tanto, dado el segundo orden de nuestros epi-splines y el objetivo de maximizar la curvatura a través de toda la malla, se llega a una formulación objetiva definida por

$$\begin{aligned} \max_{a_0, a_1, a_2} \left\{ (1 - \rho) \sum_k \kappa(k) - \rho \int \left(\frac{d^2}{dp^2} E_X^*(p) \right)^2 dp \right\}, & \text{ Función objetivo: DSE} \\ \kappa(k) = \frac{2a_2^k}{(1 + \sqrt{a_1^k + 2a_2^k})^{1.5}}, & \text{ Función de Curvatura} \\ \rho \int \left(\frac{d^2}{dp^2} E_X^*(p) \right)^2 dp = \rho (a_2^k - a_2^{k+1})^2. & \text{ Penalización de suavidad} \end{aligned} \quad (4.1)$$

Se hace notar en [4] que para estimar $E_X^*(p)$ la maximización de la suma de curvaturas muestra una respuesta de forma razonable, aunque se sustenta que otras funciones objetivo podrían ser utilizadas. Además, se menciona que valores del parámetro de penalización $\rho \approx 0.01 - 0.1$ funcionaban razonablemente bien cuando se requería suavidad.

4.2. Restricciones

Se imponen restricciones de acuerdo con criterios conocidos, así como información externa disponible. A continuación se examinan estas restricciones, aunque esto no constituye una lista exhaustiva.

4.2.1. Convexidad, continuidad y diferenciabilidad

Del Teorema 5 de la caracterización de superesperanzas, se sabe que $E_X^*(p)$ es convexa. Por tanto, se impone convexidad en cada segmento epi-spline requiriendo

$$a_2^k \geq 0 \quad \forall k \in K. \quad (4.2)$$

Este requisito además asegura que todos los valores de curvatura serán positivos, evitando así cualquier cancelación de potencias a través de la suma de segmentos. Si además se necesita continuidad a través de los segmentos uniendo los extremos de mallas juntas e imponiendo pendientes iguales

en estas intersecciones, reforzamos la convexidad a través de toda la función epi-spline. De nuevo, para p_R y p_L correspondientes a los extremos derecho e izquierdo del segmento de malla, se llega a

$$a_0^k + a_1^k p_R^k + a_2^k (p_R^k)^2 = a_0^{k+1} + a_1^{k+1} p_L^{k+1} + a_2^{k+1} (p_L^{k+1})^2, \quad (4.3)$$

$$a_1^k + 2a_2^k p_R^k = a_1^{k+1} + 2a_2^{k+1} p_L^{k+1}. \quad (4.4)$$

4.2.2. Continuidad de la densidad

Si se asume que la densidad subyacente es continua, se puede pedir que a_2 sea estrictamente positivo a través de la malla. De nuevo, esto se basa en la relación DSE-cuantil de *ii*) del Teorema 6, de manera que si el cuantil es constante a través de un intervalo p , hay un salto vertical a ese valor cuantil en la CDF. Por tal motivo se puede ajustar la restricción de convexidad a una desigualdad estricta para que

$$a_2^k > 0 \quad \forall k \in K. \quad (4.5)$$

4.2.3. Puntos extremos

Del Teorema 6 se sabe que $E_X^*(0) = -E[X]$ y $E_X^*(1) = 0$. Siempre que no se conoce $E[X]$, se estima a partir de los datos muestrales, ya sea por un estimador puntual o un intervalo de confianza de la media muestral \bar{X} . La cota de confianza inferior (LCB por sus siglas en inglés) y la cota de confianza superior (UCB) para \bar{X} se puede determinar vía un número de técnicas estadísticas tales como la distribución t de student o muestreo bootstrap dependiendo del nivel de confianza deseado. Así se puede imponer una restricción puntual inicial en $p = 0$ dependiendo del conocimiento de X y el grado de flexibilidad que se desee proporcionar a la formulación, por tanto se ha de hacer

$$a_0^1 = -\mathbb{E}[X], \quad (4.6)$$

$$\bar{X}_{LCB} \leq a_0^1 \leq \bar{X}_{UCB}. \quad (4.7)$$

Adicionalmente, el punto final siempre es fijo, proporcionando así una restricción para el último segmento epi-spline

$$a_0^k + a_1^k + a_2^k = 0 \text{ para } k = K. \quad (4.8)$$

4.2.4. Cotas inferiores

De acuerdo con el Teorema 7, se tiene una cota inferior en $E_X^*(p)$ ya sea que se tenga o no una media conocida, varianza conocida $\sigma^2(X)$, o alguna combinación de los mismos. Por lo tanto

$$-(1-p) \left(\mathbb{E}[X] + \frac{\sigma_X}{\sqrt{1-p}} \right) \leq a_0^k + a_1^k p + a_2^k p^2 \quad (4.9)$$

$$-(1-p) \left(\bar{X} + \frac{s}{\sqrt{1-p}} \right) \leq a_0^k + a_1^k p + a_2^k p^2 \quad (4.10)$$

donde s se refiere a la desviación estándar muestral. Si $\mathbb{E}[X]$ o $\sigma^2(X)$ son desconocidas, sus valores se estiman dentro de un rango de un intervalo de confianza.

Aunque también existe la correspondiente cota superior para $\mathbb{E}_X^*(p)$, debido al requerimiento de convexidad, se convierte en una restricción redundante, y por tanto se omite.

4.2.5. Restricciones de cuantiles

De acuerdo con el Teorema 6, si se conoce los valores cuantiles para X en algún p específico, se aplica que

$$\frac{d}{dp} \hat{\mathbb{E}}_X^*(p) = a_1^k + 2a_2^k p = Q_X(p) \text{ para } p \in [p_L^k, p_R^k], \quad (4.11)$$

dentro de cada uno de los segmentos epi-spline que contenga tal p . Además, aunque la media y la desviación estándar son notoriamente difíciles de estimar para distribuciones pequeñas y asimétricas, los cuantiles son robustos y pueden estimarse muy fácilmente vía *intervalos de confianza no paramétricos binomiales*. Aún en muestras relativamente pequeñas, intervalos razonables para la mediana, cuantil 25 % y 75 % se pueden lograr a niveles de confianza modestos. Utilizando esta técnica no paramétrica, se modifican las restricciones para incluir estimaciones de cotas superiores e inferiores de $\hat{Q}(p)$ para cualquier p deseado

$$\hat{Q}_{LCB}(p) \leq a_1^k + 2a_2^k p \leq \hat{Q}_{UCB}(p) \text{ para } p \in [p_L^k, p_R^k]. \quad (4.12)$$

En [4] se utilizaron los cuantiles 25 %, 50 % y 75 % para acotar pendientes spi-splines para la estimación de DSE.

4.2.6. Valores máximo y mínimo

Puesto que el cuantil 0 % de la distribución debe ser a lo más tan pequeño como el mínimo de las observaciones, y que su cuantil 100 % debe ser al menos tan grande como la observación máxima de la muestra, se pueden desarrollar restricciones sobre las pendientes inicial y final de acuerdo al Teorema 6, e implementarlas como

$$a_1^k + 2a_2^k \geq \text{máx}\{x_i\} \text{ para } k = K, \quad (4.13)$$

$$a_1^k \leq \text{mín}\{x_i\} \text{ para } k = 1. \quad (4.14)$$

Similarmente, si se conocen los valores mínimo y máximo de la distribución (o tal vez alguna cota razonable de ellos), se pueden aplicar estas ecuaciones como igualdades o desigualdades sobre el valor conocido en la forma de

$$a_1^k + 2a_2^k = \text{máx}\{f(x)\} \text{ para } k = K, \quad (4.15)$$

$$a_1^k = \text{mín}\{f(x)\} \text{ para } k = 1, \quad (4.16)$$

$$a_1^k + 2a_2^k \leq f(x)_{UCB} \text{ para } k = K, \quad (4.17)$$

$$a_1^k \geq f(x)_{LCB} \text{ para } k = 1. \quad (4.18)$$

$$(4.19)$$

Capítulo 5

Estimación de la función de densidad

Típicamente para la estimación de la densidad no se conoce de antemano los puntos finales de la distribución subyacente (si es que son finitos). Como tal, la malla x se define tanto por una resolución M al igual que por puntos extremos l y u . A continuación se expone la forma de obtener estimaciones para la densidad mediante los métodos expuestos en los primeros capítulos.

5.1. Función objetivo

De acuerdo con [4] las formulaciones generales para la densidad son

$$\begin{aligned} \text{MLP: } & \max_{\hat{f}} \sum_{i=1}^n \log(\hat{f}(x_i)) \quad \text{tal que } \hat{f} \in F, \\ \text{MEP: } & \max_{\hat{f}} \int -\hat{f}(x) \log(\hat{f}(x)) dx \quad \text{tal que } \hat{f} \in F, \end{aligned} \quad (5.1)$$

donde el conjunto de restricciones F están determinadas por las propiedades de las funciones de densidad (no negatividad, que integre 1), al igual que información suave, tal como continuidad y restricciones de forma. Con los epi-splines de primer orden, se optimizan los coeficientes en los segmentos de acuerdo a una de las siguientes ecuaciones

$$\text{MLP: } = \max_{b_0, b_1} \sum_{i=1}^n \log(b_0^m + b_1^m x_i) \quad \forall i, m | x_i \in m, \quad (5.2)$$

$$\text{MEP: } = \max_{b_0, b_1} \int -(b_0^m + b_1^m x) \log(b_0^m + b_1^m x) dx \quad (5.3)$$

$$= \max_{b_0, b_1} \sum_m \left[z_R^m \log(z_R^m) + z_L^m \log(z_L^m) + 2(z_R^m + z_L^m) \log\left(\frac{z_R^m + z_L^m}{2}\right) \right], \quad (5.4)$$

donde $z_R^m = b_0^m + b_1^m x_R^m$ y $z_L^m = b_0^m + b_1^m x_L^m$. Se utiliza la regla de Simpson para aproximar la integración contenida en MEP.

5.2. Restricciones

Se menciona un conjunto mínimo de restricciones que definen una función de densidad, así como aquellas que tratan sobre estimaciones de cuantiles y supercuantiles.

5.2.1. Unidad y no negatividad

Por definición, la integral sobre todo \mathbb{R} de cualquier función de densidad es igual a uno. Por tal motivo, esta restricción se formula de la siguiente manera: si $\Delta_x = (u - l)/M$ es la resolución de la malla, entonces

$$\sum_m \Delta_x \left[b_0^m + b_1^m \left(\frac{x_L^m + x_R^m}{2} \right) \right] = 1. \quad (5.5)$$

Finalmente, por definición de función de densidad, las estimaciones deben contener todos sus valores no negativos, así que

$$b_0^m + b_1^m x_L^m \geq 0 \quad \forall m \in M, \quad (5.6)$$

$$b_0^m + b_1^m x_R^m \geq 0 \quad \forall m \in M. \quad (5.7)$$

5.2.2. Continuidad

Se asume que las densidades exploradas provienen de una familia de funciones continuas tal que

$$b_0^m + b_1^m x_R^m = b_0^{m+1} + b_1^{m+1} x_L^{m+1} \quad \forall m | m < M. \quad (5.8)$$

5.2.3. Restricciones sobre cuantiles

Si se conoce (o se cree haber estimado precisamente) el valor de un cuantil particular, se puede formular restricciones a partir de la ecuación (5.5), requerir que algún porcentaje de la densidad ocupe la región izquierda o derecha del valor en sí mismo. Esto funciona no sólo para cuantiles

estimados provenientes de los mismos datos, sino también para valores de cuantiles cercanos a $p = 0$ o $p = 1$ al estimar la pendiente de la función DSE. En uno u otro caso, las restricciones del cuantil para un p específico puede ser implementado como

$$\sum_{m|x_L^m \geq Q(p)} \Delta_x \left[b_0^m + b_1^m \left(\frac{x_R^m + x_L^m}{2} \right) \right] = 1 - p. \quad (5.9)$$

5.2.4. Restricciones sobre supercuantiles

Se desea incorporar una restricción utilizando las estimaciones de supercuantiles derivadas a partir de la optimización previa. Usando la definición de supercuantiles esbozada y aprovechando la regla de Simpson para simplificar la expresión, se obtiene una restricción supercuantil para un cuantil particular p como

$$\frac{1}{1-p} \sum_{m|x_L^m \geq Q(p)} \frac{\Delta_x}{6} h(x) = \bar{Q}(p) \quad (5.10)$$

$$h(x) = 3b_0^m(x_L^m + x_R^m) + b_1^m((x_R^m)^2 + (x_L^m)^2) + b_1^m(x_R^m + x_L^m)^2. \quad (5.11)$$

5.3. Implementación de las restricciones cuantil/supercuantil

Las restricciones expuestas con anterioridad se han formulado en forma de igualdades. En la práctica, dada la naturaleza de las aproximaciones de los valores de cuantiles/supercuantiles de la estimación DSE, estas restricciones se implmentan como cotas superiores o inferiores, cuya elección depende de la función objetivo para la densidad (MLP vs. MEP).

Comúnmente, MLP (Maximum Likelihood Prediction) emplea cotas inferiores para ‘empujar’ la densidad a la cola, mientras que MEP (Maximum Entropy Probability) emplea cotas superiores para ‘controlar’ la densidad de las colas.

Por el contrario, si se cree que los valores obtenidos mediante la aproximación DSE sobreestima los verdaderos valores supercuantiles entonces la implementación de la restricción supercuantil como cota superior puede ser más apropiada. Cuando no se indica lo contrario, se asume que la formulación MLP incorpora restricciones de valores esperados, cuantiles y supercuantiles via cotas inferiores de las densidades de la cola derecha. Con esto en mente, se representan las siguientes

restricciones de cuantiles y supercuantiles, donde $\hat{Q}(p)$ y $\hat{\bar{Q}}(p)$ son los resultados de la estimación de la DSE calculados con anterioridad

$$\begin{aligned}
\sum_{m|x_L^m \geq \hat{Q}(p)} \Delta_x \left[b_0^m + b_1^m \left(\frac{x_R^m + x_L^m}{2} \right) \right] &\leq 1 - p && \text{cota superior del cuantil} \\
\sum_{m|x_L^m \geq \hat{Q}(p)} \Delta_x \left[b_0^m + b_1^m \left(\frac{x_R^m + x_L^m}{2} \right) \right] &\geq 1 - p && \text{cota inferior del cuantil} \\
\frac{1}{1 - p} \sum_{m|x_L^m \geq \hat{Q}(p)} \frac{\Delta_x}{6} h(x) &\leq \hat{Q}(p) && \text{cota superior del supercuantil} \\
\frac{1}{1 - p} \sum_{m|x_L^m \geq \hat{Q}(p)} \frac{\Delta_x}{6} h(x) &\geq \hat{Q}(p) && \text{cota inferior del supercuantil.}
\end{aligned}$$

5.4. Cuantificación del error de las estimaciones

Para evaluar la validez de este método, se debe cuantificar el error de las estimaciones numéricamente. Para la evaluación de las estimaciones de cuantiles y supercuantiles, se utiliza la desviación absoluta media (AAD-Average Absolute Deviation) y la desviación absoluta mediana (MAD) entre los valores estimados y conocidos en varios p en todas las iteraciones de optimización ($j \in J$). Se formulan las métricas de error de la DSE como

$$\begin{aligned}
\text{AAD}(p) &= \frac{1}{J} \sum_j \left| \hat{Q}(p) - \bar{Q}(p) \right|, \\
\text{MAD}(p) &= \text{mediana}_j \left\{ \left| \hat{Q}(p) - \bar{Q}(p) \right| \right\},
\end{aligned}$$

con errores de cuantiles calculados de la misma manera.

Para la estimación de la densidad se proponen dos métricas. La primera es una medida de ajuste general donde se suman los errores al cuadrado (SSE) en los puntos finales del segmento epi-spline de toda la malla. La segunda métrica utiliza la misma medida, solo que a través de puntos finales dentro de la región de la cola

$$\begin{aligned}\text{SSE} &= \sum_m \left(\hat{f}(x) - f(x) \right)^2, \\ \text{SSTE} &= \sum_{m|x_L^m \geq Q(p)} \left(\hat{f}(x) - f(x) \right)^2.\end{aligned}$$

Conclusiones

En esta tesis se han presentado dos nuevos conceptos con propiedades interesantes, la función supercuantil y superesperanza, de las cuales se expuso su relación entre ellas vía la transformada de Legendre. De entre todas las propiedades que se expusieron se puede recalcar que la convergencia en distribución de una sucesión de variables aleatorias es equivalente a la convergencia gráfica de las correspondientes relaciones monótonas maximales de cada una de las variables aleatorias de la sucesión en juego a la relación monótona maximal de la variable aleatoria límite (Teorema 8).

Por otra parte, el Teorema 7 establece cotas para la diferencia entre las funciones supercuantiles de dos variables aleatorias. Dichas desigualdades son implementadas en el proceso de aproximación de la función de supercuantil vía optimización restringida. Tales desigualdades figuran como unas de varias restricciones.

Un hecho importante que se estableció en el Teorema 6 dice de qué manera se relacionan las funciones superesperanzas y la función supercuantil; y es que vía la transformada de Legendre de la función superesperanza, la función supercuantil hace su aparición. Este resultado no solo se queda en lo bonito, sino que es la base para aproximar la función supercuantil, que se hace en el capítulo 4.

Los métodos para la aproximación de la función supercuantil y de densidad fueron tomados de [4]. En tal trabajo, estos métodos han sido presentados como nuevos para mejorar las estimaciones de densidad para datos provenientes de distribuciones desconocidas o de las cuales se sospecha son de cola pesada. Se presentan un conjunto adicional de restricciones dentro de un marco de optimización de densidad, que hasta donde se sabe por el autor y los artículos colocados en la bibliografía, aún no se habían intentado. Al incorporar todas las fuentes potenciales de información disponible, se pueden proporcionar formulaciones personalizadas para aproximaciones de densidad que mejoran la fidelidad de las estimaciones incluso para muestras pequeñas.

Las estimaciones de la función de densidad se logran mediante la implementación de epi-splines

utilizando las estimaciones de supercuantiles previamente también estimados; además, se cuantifica el error de las estimaciones en el sentido de calidad de aproximación numérica.

En el capítulo 4 de [4] su método fué aplicado a un conjunto de datos no paramétricos donde, a saber por ellos, fue comparado con una técnica de estimación de densidad existente, a saber, kernel. En este caso no se expone cómo es que miden la calidad de su aproximación, y más aún, el autor dice enfocarse principalmente en una evaluación cualitativa de la densidad, evitando una comparación cuantitativa de las regiones de la cola, lo cual no parece tener sentido (evitar sus mismas métricas de calidad de aproximación), pues queda como vaga.

Por último, dada la falta de tiempo, a un servidor no le fue posible experimentar con los métodos de aproximación de densidad y supercuantiles, para tener una postura definitiva sobre de que este método tiene algunas inconsistencias como la descrita en el párrafo anterior y qué de nuevo y ventajas aporta a la tarea de aproximación de densidades.

Bibliografía

- [1] Minty, G. J., *Monotone networks*, Proceedings Royal Soc. London **A** 257, 194-212, 1960.
- [2] Rockafellar, R. T., Wets, R. J-B, *Variational Analysis*, Springer Verlag, New York, 1998.
- [3] Billingsley, P., *Probability and Measure*, Wiley, anniversary edition, 2012.
- [4] Sabol, John J., *Dual approach to superquantile estimation and applications to density fitting* (tesis de maestría). Naval postgraduate school, Monterey, California, 2016.
- [5] R. T. Rockafellar, J. O. Royset. *Random variables, monotone relations, and convex analysis*. Mathematical Programming, Series B. 2014.
- [6] W. Ogryczak, A. Ruszczyński. *Dual stochastic dominance and related mean-risk models*. SIAM Journal on Optimization 13(1), 60-78, 2002.
- [7] D. Dentcheva, G. Martinez. *Two-stage stochastic optimization problems with stochastic ordering constraints on the recourse*. European Journal of Operational Research, 219(1), 1-8, 2012.
- [8] Rockafellar, R. Tyrrell. Superquantiles and their applications to risk, random variables and regression. Tutorial in operations research, 159, 2013.
- [9] D. Dentcheva, A. Ruszczyński. *Stochastic optimization with dominance constraints*. SIAM Journal on Optimization 14, 548-566, 2003.
- [10] P. Artzner, F. Delbaen, J.-M. Eber, D. Heath. *Coherent measures of risk*. Mathematical Finance, 9(3), 203-227, 1999.

- [11] F. Delbaen. *Coherent measures of risk on general probability spaces*. P. J. Schonbucher, K. Sandmann, eds. *Advances in Finance and Stochastics: Essays in Honor of Dieter Sondermann*. Springer, Berlin, 1-37, 2002.
- [12] R. Wets, S. Bianchi. *Term and volatility structures*. In S. Zenios and W. Ziemba, editors, *Handbook of Asset and Liability Management*, 26–68. Elsevier, 2006.
- [13] R. Sood, R. Wets. Information fusion. <http://www.math.ucdavis.edu/prop01>, 2011.
- [14] R. Wets. Exploiting non-data information. In L. Dümbgen and J. Wellner, editors, *Qualitative Assumptions and Regularization in High-Dimensions*, 3, 2953–3006, <http://www.math.ucdavis.edu/rjbw/mypage/Statistics.html>. European Mathematical Society, 2006.
- [15] R. Wets. *Optimization technology in statistical estimation: fusion of soft and hard information*. In D. Hlubinka, editor, *Nonparametrics and Geometry*, 1–60, <http://www.math.ucdavis.edu/rjbw/mypage/Statistics.html>. Czech Mathematical Society, 2011.
- [16] M. Casey and R. Wets. *Density estimation: exploiting non-data information*. Technical report, Mathematics, University of California, Davis, 2010.
- [17] J. O. Royset, R. J.-B. Wets, *From Data to Assessments and Decisions: Epi-Spline Technology*. *Informs*, 3, 27–53. Available: <http://pubsonline.informs.org/doi/abs/10.1287/educ.2014.0126>
- [18] R. Wets. *Exploiting non-data information*. In L. Dübgen and J. Wellner, editors, *Qualitative Assumptions and Regularization in High-Dimensions*, volume 3, pages 2953–3006, <http://www.math.ucdavis.edu/rjbw/mypage/Statistics.html>, 2006. European Mathematical Society.
- [19] O. Royset Johannes, J-B Wets Roger, *Epi-splines and exponential epi-splines: pliable approximation tools*. Technical report, Naval Postgraduate School, 2013.