



UNIVERSIDAD AUTÓNOMA METROPOLITANA

UNIDAD IZTAPALAPA

CIENCIAS BÁSICAS E INGENIERÍA

**CURVA ROC INCIDENTE/DINÁMICA VISTA DESDE EL ENFOQUE DE
LA CÓPULA NORMAL BIVARIADA**

PRESENTA:

DAVID VASQUEZ HERRERA

PARA OBTENER EL TÍTULO DE:

MAESTRÍA EN CIENCIAS (MATEMÁTICAS)

ASESOR:

Dr. GABRIEL ESCARELA PÉREZ

SINODALES

Dr. GABRIEL ESCARELA PÉREZ

Dra. HORTENSIA JOSEFINA REYES CERVANTES

Dr. GABRIEL NÚÑEZ ANTONIO

Ciudad de México, 6 de septiembre de 2019

El gran libro de la naturaleza se encuentra abierto ante nuestros ojos pero no lo podemos leer a menos que hayamos aprendido el lenguaje y los caracteres con los cuales está escrito, está escrito en lenguaje matemático y los caracteres son triángulos, círculos y otras figuras geométricas.

Galileo Galilei

Agradecimientos

Esta tesis va dedicada principalmente y de todo corazón a mis padres Carmen y Jaime, y a mis hermanos Javier y Noé, que de manera incondicional siempre han estado apoyándome en todos y cada uno de los proyectos de mi vida, a ellos les debo todo lo que tengo.

Agradezco de todo corazón a mi demás familia y amigos, quienes que me han abierto las puertas de sus casas y recibido con los brazos abiertos. Agradezco principalmente a mi abuelo quien a su avanzada edad me da momentos muy gratos y divertidos, y a Priscila Naranjo Baka quien ha sido un soporte fundamental en los últimos años de mi vida.

Agradezco a la UAM por haberme dado la oportunidad de estudiar la Maestría en Ciencias (Matemáticas), a mi asesor de tesis, el Dr. Gabriel Escarela Pérez, a todos mis sinodales quienes revisaron el presente trabajo, a todos mis profesores de la maestría y por último y no menos importante, agradezco a Conacyt quien me apoyó económicamente para realizar mis estudios de posgrado.

Índice general

1. Introducción	7
2. Cópulas Bivariadas	9
2.1. Introducción	9
2.2. Cópulas	10
2.3. Cópula Normal Bivariada	15
2.3.1. Cópula Normal Bivariada	15
2.3.2. Gráficas de la Cópula Normal Bivariada con Marginales Normales	23
2.4. Medida de Asociación Tau de Kendall ($\tau_{X,Y}$)	26
2.5. Cópula Normal Bivariada con Marginales Normal Asimétrica y Weibull	27
2.5.1. Representación Gráfica de la Cópula Normal Bivariada con Marginales Normal Asimétrica y Weibull	29
3. Inferencia Estadística sobre Cópulas Bivariadas	33
3.1. Introducción	33
3.2. Máxima Verosimilitud para Cópulas Bivariadas	35
3.3. Método Exacto de Máxima Verosimilitud para Cópulas Bivariadas	37
3.4. Método IFM	37
3.5. Ejemplo de Máxima verosimilitud en Cópulas Bivariadas	39
3.5.1. Cópula Normal Bivariada	39
4. Cuya ROC Acumulativa/Dinámica	43
4.1. Introducción	43
4.2. Introducción Histórica a la Curva ROC	43
4.3. Análisis de supervivencia	45
4.3.1. Función de Supervivencia	45
4.3.2. Función de Riesgo	46
4.3.3. Modelos	48

4.4. Aplicación de la Curva ROC al Análisis de Supervivencia	51
4.5. Curva ROC Acumulativa/Dinámica	55
4.6. Interpretación del Área Debajo de la Curva (AUC)	74
5. Sensibilidad Incidente y Especificidad Dinámica	77
5.1. Introducción	77
5.2. Sensibilidad Incidente y Especificidad Dinámica.	77
5.2.1. Distribución del Marcador $F(c)$ y del Tiempo $G(t)$	78
5.3. Función de Densidad Conjunta Cuando una de las Variables es Discreta.	80
5.3.1. Asignación de una Distribución al Indicador Binario	80
6. Curvas ROC Dependientes del Tiempo	85
6.1. Introducción	85
6.2. Curvas ROC Incidente-Dinámica	85
6.2.1. Curva $ROC_i^{I/D}(p)$ con Marcador Normal y Tiempo Log-Normal	88
6.2.2. Curvas $ROC_i^{I/D}(p)$ con Marcador Normal Asimétrica y Tiempo Weibull	91
7. Conclusiones	99
8. Apéndice	101
8.1. Cálculo de Probabilidades	101
Bibliografía	105

Capítulo 1

Introducción

La Curva Característica Operativa del Receptor o Curva ROC, por sus siglas en inglés (Receiver Operating Characteristic) es una representación gráfica de la Sensibilidad frente a la Especificidad para un sistema de clasificación binaria según se varía el umbral de discriminación.

La curva ROC fue desarrollada a principios de la década de 1940, para ser más exacto, se creó durante la segunda guerra mundial y su principal motivador para su creación fue el poder encontrar un método en el cual se podría decidir, de manera más exacta, si un borrón en la pantalla de un radar eran bombarderos dirigidos a través del Canal de la Mancha o bien, podría ser ruido aleatorio, lo cual quiere decir que los responsables encargados de las tomas de decisión se enfrentaron a un problema de clasificación binaria. Para comprender un poco más esta situación debemos hacer la aclaración de que el radar seguía siendo una tecnología secreta que se estaba desarrollando.

Para ver la potencia de un modelo de clasificación binaria como lo es la curva ROC se utiliza el AUC, por sus siglas en inglés (Area Under the Curve). Maximizar dicha área sigue siendo actualmente una medida muy buena de la potencia de un modelo. Los primeros documentos donde aparecen la Curva ROC y el AUC permanecen clasificados. No obstante, el desarrollo de la curva ROC se ha dado a través del tiempo y se han encontrado múltiples aplicaciones de la misma, una de ellas es en los modelos de supervivencia.

Este trabajo de tesis se centra en llegar con buenos fundamentos matemáticos al estudio de la curva ROC Incidente/Dinámica, la cual se basa a su vez en los conceptos de Sensibilidad Incidente y Especificidad Dinámica, los cuales son dependientes del tiempo. Para lograr esto es necesario adentrarnos a la teoría de las cópulas bivariadas y al estudio de la curva ROC Acumulativa/Dinámica la cual se basa en la Sensibilidad Acumulativa y la Especificidad Dinámica. En síntesis, los principales objetivos de este trabajo de Tesis son los siguientes:

1. Desarrollar la teoría de cópulas bivariadas, así como el estudio detallado de la cópula normal bivariada.
2. Definir y estudiar a detalle la curva ROC Acumulativa/Dinámica la cual se basa en los conceptos de Sensibilidad Acumulativa y la Especificidad Dinámica
3. Extender el concepto de la curva ROC Acumulativa/Dinámica a la curva ROC Incidente/Dinámica la cual se basa en los conceptos de Sensibilidad Incidente y la Especificidad Dinámica. Se estudiará a detalle esta nueva definición y se mostrarán algunos ejemplos.

A continuación mostramos la estructura completa de este trabajo de Tesis.

En el Capítulo 2 damos una introducción al concepto de cópulas bivariadas, se enuncia el Teorema de Sklar, el cual fue demostrado en 1959, y es de gran importancia ya que hace referencia a que existe una relación funcional entre la función de distribución conjunta y las funciones de distribución marginales. Posteriormente encontramos las condiciones necesarias y suficientes bajo las cuales una cópula alcanza las cotas de Fréchet-Hoeffding. Obtuvimos la derivada parcial de la cópula para obtener de esta manera las funciones de densidad y de densidad conjunta. Se estudia a detalle la cópula normal bivariada con marginales normales estándar, así como la distribución condicionada de una cópula, esto último es de gran importancia en este trabajo de tesis porque será utilizado en capítulos posteriores. Finalmente se incluyeron algunas gráficas de la función de densidad y de sus respectivos contornos de la cópula normal bivariada con marginales normales estándar utilizando distintos parámetros de dependencia y de la cópula normal bivariada pero ahora utilizamos dos funciones marginales distintas, estas fueron la distribución normal asimétrica y la distribución Weibull, utilizando distintas medidas de asociación Tau de Kendall.

En el Capítulo 3 se aborda el tema de la inferencia estadística sobre cópulas bivariadas. Aquí mostramos que para la estimación de parámetros de una cópula por el método de máxima verosimilitud existen dos métodos, el primero es llamado el método exacto de máxima verosimilitud y al segundo se le conoce como el método IFM (Inference For the Margins).

En el Capítulo 4 se da una breve introducción al análisis de supervivencia, se define formalmente la curva ROC en términos de la Sensibilidad Acumulativa y la Especificidad Dinámica, se estudian a detalle los 3 casos extremos de la curva ROC y finalmente se define matemáticamente el AUC y se da su interpretación.

Finalmente, en el Capítulo 5 se definen formalmente la Sensibilidad Incidente y la Especificidad Dinámica. En este capítulo se le asigna una función de distribución al marcador y al tiempo, y ocupando la cópula normal bivariada obtenemos las expresiones explícitas de la Sensibilidad Incidente y la Especificidad Dinámica. Con base en esto, en el Capítulo 6, se define a la curva ROC Incidente/Dinámica y se muestran una serie de ejemplos teóricos y gráficos, en el primero de ellos ocupamos la cópula normal bivariada con marginales normal y log-normal, graficamos distintas curvas ROC y calculamos sus respectivas AUC. En los ejemplos posteriores procedemos de manera análoga pero ahora ocupando como marginales las distribuciones normal asimétrica y Weibull.

Añadimos que en este trabajo de Tesis se incluye un Apéndice, el cual contiene definiciones y teoremas relacionados al cálculo de probabilidades en un espacio bidimensional, también hacemos referencia a que todas las gráficas que vienen en este trabajo se desarrollaron por medio del programa R Project for Statistical Computing.

Capítulo 2

Cóputas Bivariadas

2.1. Introducción

La historia de las cópulas comienza con Fréchet, quien fue un matemático francés que hizo varias contribuciones importantes en el campo de la estadística y la probabilidad. Un problema que se planteó fue el siguiente: dadas las funciones de distribución F_1 y F_2 de dos variables aleatorias X_1 y X_2 definidas en el mismo espacio de probabilidad $(\Omega, \mathfrak{F}, \Pr)$, ¿qué se puede decir del conjunto de las funciones de distribución $\Gamma(F_1, F_2)$, cuyas marginales son F_1 y F_2 ? El conjunto $\Gamma(F_1, F_2)$ es distinto del vacío ya que, si X_1 y X_2 son independientes entonces la función de distribución $F(x_1, x_2) = F_1(x_1) \cdot F_2(x_2)$ pertenece a $\Gamma(F_1, F_2)$, sin embargo, no se sabía cuales otros elementos pertenecían a dicho conjunto. Para saber más acerca del espacio de probabilidad recomendamos ver el Apéndice.

En 1959 Sklar introduce la notación y el nombre de cópula, también enuncia y demuestra el teorema que ahora lleva su nombre, el cual hace referencia a que existe una relación funcional entre la función de distribución conjunta y las funciones de distribución marginales, lo que implica que conociendo las funciones marginales y la información de la función ($C > 0$) es posible reconstruir la función de distribución conjunta.

En este Capítulo damos una introducción al concepto de cópulas, se enuncia el Teorema de Sklar y derivado de ello obtenemos algunas propiedades importantes como lo son las cotas de Fréchet-Hoeffding y la derivada parcial de la cópula para obtener de esta manera las funciones de densidad y de densidad conjunta de una cópula. Se estudia a detalle la cópula normal bivariada con marginales normales estándar, así como la distribución condicionada de una cópula, esto último es de gran importancia en este trabajo de tesis porque será utilizado en capítulos posteriores.

Se incluyeron algunas gráficas de la función de densidad de la cópula normal bivariada con marginales normales estándar utilizando distintos parámetros de dependencia (correlación). Para hacer más visible como los parámetros de dependencia cambian la gráfica de la función de densidad sacamos los contornos de dicha función, aunado a esto, ocupamos la medida de asociación Tau de Kendall, la cual se aplicó a cada parámetro de dependencia y de nuevo obtuvimos los contornos de las funciones de densidad. Posteriormente, para hacer más interesante nuestro análisis, ocupamos nuevamente la cópula normal bivariada pero ahora utilizamos dos funciones marginales distintas, estas fueron la distribución normal asimétrica y la distribución Weibull, con medida de asociación Tau de Kendall y nuevamente obtuvimos las gráficas de las funciones de densidad y los contornos de cada una de ellas.

2.2. Cópulas

Es importante notar que para hacer el cálculo de probabilidades con un vector aleatorio bivariado es necesario conocer su función de distribución conjunta, así mismo, la función de densidad conjunta se puede deducir de las funciones de distribución marginales de las variables aleatorias involucradas. Esto se muestra en el Apéndice de este trabajo de Tesis.

Conociendo la función de distribución conjunta, se puede *marginalizar* y así obtener las funciones de distribución univariadas individuales de las variables involucradas, sin embargo, el regreso no está garantizado, es decir, conociendo las marginales no siempre es posible reconstruir la función de distribución conjunta, esto quiere decir que al *marginalizar* se pierde algún tipo de información que ya no permite regresar. Es en esta parte donde entra el Teorema de Sklar 2.1, el cual se demostró en 1959 y menciona que existe una función C , en este caso bivariada, asociada a un vector aleatorio también bivariado, es decir: $C : [0, 1] \times [0, 1] \rightarrow [0, 1]$, tal que $F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$, es decir, que existe una relación funcional entre la función de distribución conjunta y las funciones de distribución marginales, lo que implica que conociendo las funciones marginales y la información de la función C es posible reconstruir la función de distribución conjunta.

Dada la ecuación $F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$ es posible obtener la función C , tal y como se muestra a continuación:

Primero se aplica un cambio de variable, a saber, sea $U = F_1(x_1)$ y $V = F_2(x_2)$, sustituyendo U y V en la ecuación $F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$ se obtiene la siguiente función: $F(F_1^{-1}(U), F_2^{-1}(V)) = C(U, V)$. Ahora se aplican algunas de las propiedades de las funciones marginales mencionadas en el Apéndice del presente trabajo.

$$\begin{aligned} F_1(x_1) &= \lim_{x_2 \rightarrow \infty} F(x_1, x_2) \\ &= \lim_{x_2 \rightarrow \infty} C(F_1(x_1), F_2(x_2)) = C(F_1(x_1), 1) \end{aligned}$$

procedemos de manera análoga para el otro caso

$$\begin{aligned} F_2(x_2) &= \lim_{x_1 \rightarrow \infty} F(x_1, x_2) \\ &= \lim_{x_1 \rightarrow \infty} C(F_1(x_1), F_2(x_2)) = C(1, F_2(x_2)) \end{aligned}$$

por lo tanto, se han obtenido dos propiedades

$$U = C(F_1(x_1), 1) \quad \text{y} \quad V = C(1, F_2(x_2)).$$

De manera similar se procede cuando las variables tienden ahora a menos infinito, es decir:

$$\begin{aligned} \lim_{x_1 \rightarrow -\infty} F(x_1, x_2) &= \lim_{x_1 \rightarrow -\infty} \Pr(X_1 \leq x_1, X_2 \leq x_2) \\ &= \lim_{x_1 \rightarrow -\infty} C(F_1(x_1), F_2(x_2)) = C(0, F_2(x_2)) = 0 \end{aligned}$$

y

$$\begin{aligned} \lim_{x_2 \rightarrow -\infty} F(x_1, x_2) &= \lim_{x_2 \rightarrow -\infty} \Pr(X_1 \leq x_1, X_2 \leq x_2) \\ &= \lim_{x_2 \rightarrow -\infty} C(F_1(x_1), F_2(x_2)) = C(F_1(x_1), 0) = 0. \end{aligned}$$

Nuevamente hemos obtenido dos propiedades

$$C(0, F_2(x_2)) = 0 \quad \text{y} \quad C(F_1(x_1), 0) = 0.$$

Finalmente, realizando el correspondiente cambio de variable en la propiedad 4 del Teorema 8.1, el cual se muestra en el Apéndice, se obtiene

$$0 \leq C(U_2, V_2) - C(U_2, V_1) - C(U_1, V_2) + C(U_1, V_1).$$

El análisis realizado anteriormente da la motivación para la definición de Cópula.

Definición 2.1 (Definición de Cópula) Una función $C : [0, 1]^2 \rightarrow [0, 1]$ es llamada una función cópula biva-riada si y solo si cumple con las siguientes condiciones

1. **Condiciones de frontera**

- a.) $\forall U, V \in [0, 1], C(U, 0) = C(0, V) = 0.$
 b.) $\forall U, V \in [0, 1], C(U, 1) = U; C(1, V) = V.$

2. **Propiedad 2-Creciente**

- a.) $\forall U_1, U_2, V_1, V_2 \in [0, 1]$ tal que $U_1 \leq U_2$ y $V_1 \leq V_2$, $C(U_2, V_2) - C(U_2, V_1) - C(U_1, V_2) + C(U_1, V_1) \geq 0.$

Ahora se enunciará el Teorema de Sklar, el cual dio la motivación para la Definición de Cópula 2.1.

Teorema 2.1 (Teorema de Sklar) Sea H una función de distribución conjunta con marginales F y G , entonces existe una cópula C tal que para todo x y y en \mathbb{R} ,

$$H(x, y) = C(F(x), G(y)). \quad (2.1)$$

Si F y G son continuas, entonces C es única, en otro caso, C está únicamente determinada sobre $\text{Ran}F \times \text{Ran}G$. Inversamente, si C es una cópula con F y G funciones de distribución, entonces la función H definida en (2.1) es una función de distribución conjunta con marginales F y G .

La demostración de este teorema se muestra en [17].

Se mostrará a continuación un par de ejemplos de cópulas, las cuales serán de gran utilidad en un posterior teorema.

Ejemplo 2.1 La función $M(u, v) \equiv \min\{u, v\}$ con $u, v \in [0, 1]^2$, cumple con las condiciones dadas en la Definición de Cópula (2.1), ya que:

a) *Condiciones de Frontera*

$$M(u, 0) = \min\{u, 0\} = 0 = \min\{0, v\} = M(u, 0)$$

$$M(u, 1) = \min\{u, 1\} = u; M(1, v) = \min\{1, v\} = v.$$

b) *Propiedad 2-Creciente*

$\forall u_1, u_2, v_1, v_2 \in [0, 1]$ tal que $u_1 \leq u_2$ y $v_1 \leq v_2$, sin pérdida de generalidad supongamos además que $u_1 \leq v_1 \leq u_2 \leq v_2$. Entonces

$$M(u_2, v_2) - M(u_2, v_1) - M(u_1, v_2) + M(u_1, v_1) = u_2 - v_1 - u_1 + u_1 = u_2 - v_1 \geq 0.$$

Por lo tanto, la función $M(u, v) \equiv \min\{u, v\}$ resulta ser una cópula.

De manera análoga se demuestra que la función $W(u, v)$ dada por $W(u, v) \equiv \max\{u + v - 1, 0\}$ es una cópula.

Las funciones M y W juegan un papel muy importante en el estudio de las cópulas, tal y como lo muestra el siguiente teorema.

Teorema 2.2 Sea C una cópula bivariada, entonces para todo $(u, v) \in [0, 1]^2$ se cumple

$$W(u, v) \equiv \max\{u + v - 1, 0\} \leq C(u, v) \leq \min\{u, v\} \equiv M(u, v). \quad (2.2)$$

La demostración de este teorema se muestra en [17].

El teorema anterior menciona que toda cópula $C(u, v)$ está acotada superior e inferiormente por una cópula $M(u, v)$ y $W(u, v)$ respectivamente. Esto motiva a dar la siguiente definición.

Definición 2.2 Dado que toda cópula $C(u, v)$, con $(u, v) \in [0, 1]^2$, está acotada por las cópulas $M(u, v)$ y $W(u, v)$, éstas serán nombradas como cotas de Fréchet-Hoeffding. Específicamente a $M(u, v)$ se le llamará cota superior de Fréchet-Hoeffding y a $W(u, v)$ se le llamará la cota inferior de Fréchet-Hoeffding.

Una de las consecuencias del Teorema de Sklar 2.1 es que si se tienen X y Y variables aleatorias con una función de distribución conjunta H con marginales F y G respectivamente, entonces para toda x y y en \mathbb{R} se tiene

$$\max\{F(x) + G(y) - 1, 0\} \leq H(x, y) \leq \min\{F(x), G(y)\}. \quad (2.3)$$

Esto se debe a que M y W son cópulas. Una observación importante es notar que las cotas, mostradas en la Ecuación (2.3), son funciones de distribución conjuntas y son llamadas las cotas de Fréchet-Hoeffding para la función de distribución conjunta H con marginales F y G .

Resulta de gran interés saber las condiciones bajo las cuales la función de distribución conjunta H alcanza las cotas de Fréchet-Hoeffding, ya que esto va a ser de gran utilidad en la Sección 4 Curva ROC Acumulativa/Dinámica. Estas condiciones las da el siguiente teorema.

Teorema 2.3 Sean X y Y variables aleatorias con función de distribución conjunta H , entonces H es igual a su cota superior de Fréchet-Hoeffding sí y solo sí para todo (x, y) en \mathbb{R}^2 se cumple que $\Pr[X > x, Y \leq y] = 0$ o $\Pr[X \leq x, Y > y] = 0$.

Demostración 2.3.1 De la Ley de Probabilidad Total se tienen las siguientes ecuaciones

$$\Pr[X \leq x] = \Pr[X \leq x, Y \leq y] + \Pr[X \leq x, Y > y]$$

$$\Pr[Y \leq y] = \Pr[X \leq x, Y \leq y] + \Pr[X > x, Y \leq y].$$

Por lo tanto, se tiene lo siguiente

$$\begin{aligned} F(x) &= \Pr[X \leq x] = \Pr[X \leq x, Y \leq y] + \Pr[X \leq x, Y > y] \\ &= H(x, y) + \Pr[X \leq x, Y > y] \end{aligned}$$

y

$$\begin{aligned} G(y) &= \Pr[Y \leq y] = \Pr[X \leq x, Y \leq y] + \Pr[X > x, Y \leq y] \\ &= H(x, y) + \Pr[X > x, Y \leq y]. \end{aligned}$$

Por lo tanto, $H(x, y) = M(F(x), G(y))$ si y solo si $\min\{\Pr[X \leq x, Y > y], \Pr[X > x, Y \leq y]\} = 0$, lo cual completa la demostración.

Del Teorema de Sklar se obtuvo la siguiente expresión:

$$H(x, y) = C(F(x), G(y)). \quad (2.4)$$

Una pregunta de gran interés es, a saber, ¿cuándo existen las derivadas parciales de la cópula?. La respuesta a esta pregunta la da el siguiente teorema [17].

Teorema 2.4 Sea C una cópula. Entonces, para cualquier $v \in [0, 1]$, la derivada parcial $\frac{\partial C(u, v)}{\partial u}$ existe para casi toda u , y para tal u y v se tiene:

$$0 \leq \frac{\partial C(u, v)}{\partial u} \leq 1. \quad (2.5)$$

De manera similar se tiene que para cualquier $u \in [0, 1]$, la derivada parcial $\frac{\partial C(u, v)}{\partial v}$ existe para casi toda v , y para tal u y v se tiene:

$$0 \leq \frac{\partial C(u, v)}{\partial v} \leq 1. \quad (2.6)$$

Aún más, las funciones $u \mapsto \frac{\partial C(u, v)}{\partial v}$ y $v \mapsto \frac{\partial C(u, v)}{\partial u}$ están definidas y son no decrecientes en casi todas partes en $[0, 1]$

Aplicando el Teorema 2.4, obtenemos que al derivar la expresión $H(x, y) = C(F(x), G(y))$ con respecto a las variables x y y se obtiene una expresión para la función de densidad conjunta en términos de las funciones de densidad marginales, es decir

$$\begin{aligned} \frac{\partial^2}{\partial x \partial y} (H(x, y)) &= \frac{\partial^2}{\partial x \partial y} (C(F(x), G(y))) \\ &= \frac{\partial^2 C(F(x), G(y))}{\partial x \partial y} \cdot \frac{\partial F(x)}{\partial x} \cdot \frac{\partial G(y)}{\partial y} \end{aligned} \quad (2.7)$$

sea

$$c(F(x), G(y)) = \frac{\partial^2 C(F(x), G(y))}{\partial x \partial y}. \quad (2.8)$$

Entonces se puede escribir la ecuación anterior de la siguiente manera:

$$\frac{\partial^2}{\partial x \partial y} (H(x, y)) = c(F(x), G(y)) f(x) g(y). \quad (2.9)$$

Si hacemos la sustitución $F(x) = u$ y $G(y) = v$, se obtiene la siguiente expresión:

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}. \quad (2.10)$$

La cual es la función de densidad de la cópula.

Es muy importante notar que en el análisis que se hizo anteriormente es necesario que las funciones marginales $F(x)$ y $G(y)$ sean funciones continuas y diferenciables. En caso contrario, es decir, en el caso en que $F(x)$ y $G(y)$ no fueran diferenciables, no es posible llevar a cabo dicho procedimiento.

Todo lo anterior motiva a dar la siguiente definición.

Definición 2.3 La función de densidad h de una función de distribución bivariada H , si es que existe, se define como

$$h(x, y) = \frac{\partial^2}{\partial x \partial y} (H(x, y))$$

para la cópula, la función de densidad está dada por:

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}.$$

Donde $u, v \in [0, 1]$

Aún más, si en la anterior fórmula hacemos la sustitución $F(x) = u$ y $G(y) = v$, donde las funciones marginales $F(x)$ y $G(y)$ son continuas y diferenciables, entonces la función de densidad conjunta h de una función de distribución bivariada H se puede expresar como

$$h(x, y) = c(F(x), G(y)) f(x) g(y).$$

Lo último quiere decir que, dada la función de densidad de la cópula y las funciones de densidad marginales, si es que existen, tenemos la función de densidad conjunta de las variables originales.

$$h(x, y) = c(F(x), G(y)) f(x) g(y).$$

Por último, notamos que al hacer el cambio de variable $u = F(x)$ y $v = G(y)$, donde F y G son funciones de distribución, entonces, toda función de distribución bivariada, H , define una cópula implícita

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)). \quad (2.11)$$

2.3. Cópula Normal Bivariada

Como se vio anteriormente, se puede deducir una cópula a partir de cualquier función de distribución bivalente. La cópula implícita que se obtiene a partir de dicha distribución nos da la dependencia entre las variables aleatorias que se consideran, las cuales pueden tener cualquier tipo de distribución, como veremos más adelante. Por lo tanto, se puede combinar cualquier tipo de cópula con cualquier tipo de distribuciones marginales, que no necesariamente necesitan ser iguales para todas las variables.

A manera de ejemplo se detallarán algunos aspectos importantes de la Cópula Normal Bivariada.

2.3.1. Cópula Normal Bivariada

Comenzamos definiendo las funciones de densidad y de distribución para las distribuciones normal estándar y normal estándar bivariada.

Para la función de distribución normal estándar, se definen

$$\phi(x) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad \Phi(j) := \int_{-\infty}^j \phi(x) dx \quad (2.12)$$

la función de densidad y la función de distribución respectivamente.

Para la función de distribución normal estándar bivariada, se definen

$$\phi_2(x, y; \rho) := \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right), \quad (2.13)$$

$$\Phi_2(j, k; \rho) := \int_{-\infty}^j \int_{-\infty}^k \phi_2(x, y; \rho) dx dy \quad (2.14)$$

la función de densidad y la función de distribución respectivamente, con parámetro de correlación o dependencia $\rho \in (-1, 1)$.

La Cópula Normal Bivariada o Cópula Gaussiana Bivariada se deriva a partir de una distribución Normal Bivariada Φ_2 para la dependencia, es decir, la Cópula Normal Bivariada es una función de dependencia asociada a la distribución Normal Bivariada, junto con distribuciones marginales Φ , donde $\Phi(x) = \Pr(X < x)$, para $X \sim N(0, 1)$. Entonces se puede ver a la Cópula Normal como

$$C(u, v; \Gamma) := \Phi_2\left(\Phi^{-1}(u), \Phi^{-1}(v)\right). \quad (2.15)$$

Donde Γ es la matriz de varianzas-covarianzas. Para el caso bivariado Γ tiene la siguiente forma:

$$\Gamma := \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Donde ρ es el parámetro de correlación, con $\rho \in (-1, 1)$.

También notemos que:

$$\Phi_2 \left(\Phi^{-1}(u), \Phi^{-1}(v) \right) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \exp \left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)} \right) dx dy. \quad (2.16)$$

Al realizar la derivada parcial de la expresión anterior, respecto a v y u , se obtiene la densidad de la Cópula Normal, es decir:

$$c(u, v, \Gamma) = \frac{\partial^2 C(u, v; \Gamma)}{\partial u \partial v} = \frac{1}{\sqrt{1-\rho^2}} \exp \left(-\frac{\rho^2 q_1^2 - 2\rho q_1 q_2 + \rho^2 q_2^2}{2(1-\rho^2)} \right). \quad (2.17)$$

Donde $q_1 = \Phi^{-1}(u)$ y $q_2 = \Phi^{-1}(v)$, los cuales son los cuantiles de variables normales estándar.

Recalamos que la pasada derivación será demostrada más adelante, cuando se vea la derivada parcial de la cópula respecto a una sola de sus variables.

Formalmente, la densidad de cópula normal bivariada tiene la siguiente forma:

$$c(u, v, \Gamma) = \frac{\partial^2 C(u, v; \Gamma)}{\partial u \partial v} = |\Gamma|^{-1/2} \exp \left(-\frac{1}{2} \xi^T (\Gamma^{-1} - I_2) \xi \right) \quad (2.18)$$

donde I_2 es la matriz identidad de 2×2 y $\xi^T = (q_1, q_2)^T$, donde q_1 es el cuantil u y q_2 es el cuantil v , ambos cuantiles de la distribución normal estándar, es decir:

$$u = \Pr(X < q_1) \quad y \quad v = \Pr(X < q_2), \quad \text{con} \quad X \sim N(0, 1) \quad (2.19)$$

y $|\Gamma|$ es el determinante de la matriz Γ .

Lo anterior nos lleva a dar la siguiente definición.

Definición 2.4 Sea Γ una matriz simétrica, definida positiva con $\text{diag}\Gamma = 1$ y Φ_2 la distribución normal estándar bivariada. Entonces se define a la Cópula normal como:

$$C(u, v; \Gamma) := \Phi_2 \left(\Phi^{-1}(u), \Phi^{-1}(v) \right)$$

y su función de densidad conjunta es:

$$c(u, v, \Gamma) = |\Gamma|^{-1/2} \exp \left(-\frac{1}{2} \xi^T (\Gamma^{-1} - I_2) \xi \right).$$

Donde I_2 es la matriz identidad de 2×2 , $\xi^T := (q_1 \quad q_2)^T$, con $q_1 = \Phi^{-1}(u)$ y $q_2 = \Phi^{-1}(v)$ y $|\Gamma|$ es el determinante de la matriz Γ .

En este trabajo de tesis se define a Γ de la siguiente forma:

$$\Gamma := \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Donde ρ es el parámetro de correlación, con $\rho \in (-1, 1)$.

En nuestro caso, a la función de densidad conjunta, definida anteriormente, se le puede ver como a continuación se indica.

Definición 2.5 Sean $F(x)$ y $G(y)$ funciones de distribución acumulada, continuas y diferenciales, y Γ la matriz de varianzas-covarianzas. Entonces, la función de densidad conjunta de una Cópula Normal Bivariada está dada de la siguiente manera:

$$\begin{aligned} h(x, y; \Gamma) &= |\Gamma|^{-1/2} \exp\left(-\frac{1}{2} \xi^T (\Gamma^{-1} - I_2) \xi\right) f(x) g(y) \\ &= |\Gamma|^{-1/2} \exp\left(-\frac{1}{2} \left(\Phi^{-1}[F(x)], \Phi^{-1}[G(y)]\right)^T (\Gamma^{-1} - I_2) \left(\Phi^{-1}[F(x)], \Phi^{-1}[G(y)]\right)\right) \\ &\quad \cdot f(x) g(y). \end{aligned} \tag{2.20}$$

Una cuestión muy importante en el estudio de las cópulas es la derivada de la misma pero respecto a una sola de sus variables. Para hacer esto en el caso bivariado, necesitamos de la función de distribución condicional para U dado $V = v$, donde (U, V) son variables aleatorias cuya función de distribución conjunta es C y $u, v \in [0, 1]$, entonces se emplea la fórmula del siguiente modo:

$$\begin{aligned} H_{U/V=v}(u) &= \Pr[U \leq u \mid V = v] \\ &= \lim_{\Delta v \rightarrow 0} \frac{C(u, v + \Delta v) - C(u, v)}{\Delta v} \\ &= \frac{\partial C(u, v)}{\partial v}. \end{aligned} \tag{2.21}$$

A esto se le denomina la distribución condicionada de una cópula.

Definición 2.6 (Distribución condicionada de una Cópula): Si la distribución conjunta de X y Y es $C(F(x), G(y))$, entonces la distribución condicionada $X = x/Y$ es:

$$H_{X=x/Y}(y) = \frac{\partial C(F(x), G(y))}{\partial F(x)} = \frac{\partial C(u, v)}{\partial u}.$$

Donde $F(x) = u$ y $G(y) = v$.

Análogamente, definimos la distribución condicionada $X/Y = y$ como

$$H_{X/Y=y}(x) = \frac{\partial C(F(x), G(y))}{\partial G(y)} = \frac{\partial C(u, v)}{\partial v}.$$

A continuación se realizará este análisis en la cópula normal.

Si (X, Y) tiene una distribución normal estándar bivariada con parámetro de correlación ρ , entonces Y condicionada a $X = x$ está normalmente distribuida, aún más, está dada por la siguiente fórmula:

$$\frac{\partial C(u, v; \Gamma)}{\partial u} = \Phi \left(\frac{\Phi^{-1}(v) - \rho \Phi^{-1}(u)}{\sqrt{1 - \rho^2}} \right). \quad (2.22)$$

Donde $F(x) = u$ y $G(y) = v$.

Para demostrar esto, primero se definen a q_1 y q_2 como los cuantiles de las variables normales estándar u y v respectivamente, es decir, $q_1 := \Phi^{-1}(u)$ y $q_2 := \Phi^{-1}(v)$, entonces se obtiene la siguiente expresión para la derivada parcial de la cópula:

$$\begin{aligned} \frac{\partial C(u, v; \Gamma)}{\partial u} &= \frac{\partial}{\partial u} \frac{1}{2\pi\sqrt{|\Gamma|}} \int_{-\infty}^{q_1} \int_{-\infty}^{q_2} \exp\left(-\frac{1}{2}x^T \Gamma^{-1}x\right) dx \\ &= \frac{\partial}{\partial u} \frac{1}{2\pi\sqrt{|\Gamma|}} \int_{-\infty}^{q_1} \int_{-\infty}^{q_2} \exp\left(-\frac{1}{2} \begin{pmatrix} x_1 & x_2 \end{pmatrix} \Gamma^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) dx. \end{aligned}$$

Ahora aplicamos el Teorema Fundamental del Cálculo, el cual dice:

Teorema 2.5 Dada una función f integrable sobre el intervalo $[a, b]$, definimos F sobre $[a, b]$ por $F(x) = \int_a^x f(t) dt$. Si f es continua en $c \in (a, b)$, entonces F es derivable en c y $F'(c) = f(c)$.

Si se aplica la regla de la cadena se obtiene el primer teorema fundamental del cálculo, obteniendo así la siguiente expresión:

$$\frac{d}{dx} \int_{a(x)}^{b(x)} f(t) dt = f(b(x)) b'(x) - f(a(x)) a'(x). \quad (2.23)$$

Donde $f(t)$ es una función integrable sobre el intervalo $[a(x), b(x)]$, con $a(x)$ y $b(x)$ funciones derivables.

Entonces

$$\frac{\partial C(u, v; \Gamma)}{\partial u} = \frac{1}{2\pi\sqrt{1 - \rho^2}} \int_{-\infty}^{q_2} \exp\left(-\frac{1}{2} \begin{pmatrix} q_1 & x_2 \end{pmatrix} \Gamma^{-1} \begin{pmatrix} q_1 \\ x_2 \end{pmatrix}\right) dx_2 \cdot \frac{\partial}{\partial u} q_1.$$

Ahora aplicamos el hecho de que $\frac{\partial}{\partial u} q_1 = \sqrt{2\pi} \exp\left(-\frac{q_1^2}{2}\right)$, el cual fue demostrado por Steinbrecher y Shaw en [22], obteniendo de este modo que

$$\frac{\partial C(u, v; \Gamma)}{\partial u} = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{q_2} \exp\left(-\frac{1}{2} \begin{pmatrix} q_1 & x_2 \end{pmatrix} \Gamma^{-1} \begin{pmatrix} q_1 \\ x_2 \end{pmatrix}\right) dx_2 \cdot \sqrt{2\pi} \exp\left(\frac{q_1^2}{2}\right)$$

Recordemos que $\Gamma := \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, entonces $\Gamma^{-1} := \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$, por lo tanto:

$$\begin{aligned} -\frac{1}{2} \begin{pmatrix} q_1 & x_2 \end{pmatrix} \Gamma^{-1} \begin{pmatrix} q_1 \\ x_2 \end{pmatrix} &= -\frac{1}{2(1-\rho^2)} \begin{pmatrix} q_1 & x_2 \end{pmatrix} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \begin{pmatrix} q_1 \\ x_2 \end{pmatrix} \\ &= -\frac{1}{2(1-\rho^2)} (q_1^2 - 2x_2\rho q_1 + x_2^2). \end{aligned}$$

Entonces:

$$\begin{aligned} \frac{\partial C(u, v; \Gamma)}{\partial u} &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{q_2} \exp\left(-\frac{1}{2} \begin{pmatrix} q_1 & x_2 \end{pmatrix} \Gamma^{-1} \begin{pmatrix} q_1 \\ x_2 \end{pmatrix}\right) dx_2 \cdot \sqrt{2\pi} \exp\left(\frac{q_1^2}{2}\right) \\ &= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \int_{-\infty}^{q_2} \exp\left(-\frac{1}{2(1-\rho^2)} (q_1^2 - 2x_2\rho q_1 + x_2^2)\right) dx_2 \cdot \exp\left(\frac{q_1^2}{2}\right) \\ &= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \int_{-\infty}^{q_2} \exp\left(-\frac{(q_1^2 - 2x_2\rho q_1 + x_2^2)}{2(1-\rho^2)}\right) \cdot \exp\left(\frac{q_1^2}{2}\right) dx_2 \\ &= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \int_{-\infty}^{q_2} \exp\left(-\frac{(q_1^2 - 2x_2\rho q_1 + x_2^2)}{2(1-\rho^2)} + \frac{q_1^2}{2}\right) dx_2 \\ &= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \int_{-\infty}^{q_2} \exp\left(\frac{-2(q_1^2 - 2x_2\rho q_1 + x_2^2) + 2q_1^2(1-\rho^2)}{4(1-\rho^2)}\right) dx_2 \\ &= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \int_{-\infty}^{q_2} \exp\left(\frac{-2(\rho^2 q_1^2 - 2x_2\rho q_1 + x_2^2)}{4(1-\rho^2)}\right) dx_2 \\ &= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \int_{-\infty}^{q_2} \exp\left(-\frac{1}{2(1-\rho^2)} (\rho q_1 - x_2)^2\right) dx_2. \end{aligned} \tag{2.24}$$

Ahora procedemos a realizar el cambio de variable dado por $x_2 = z\sqrt{1-\rho^2} + \rho q_1$. Al derivar esta expresión obtenemos las siguientes fórmulas

$$\frac{dx_2}{dz} = \sqrt{1-\rho^2} \quad \text{por lo tanto} \quad dz = \frac{dx_2}{\sqrt{1-\rho^2}}. \quad (2.25)$$

Aún más, al despejar a z de $x_2 = z\sqrt{1-\rho^2} + \rho q_1$ se obtiene: $z = \frac{x_2 - \rho q_1}{\sqrt{1-\rho^2}}$. Al sustituir estas fórmulas en las pasadas ecuaciones obtenemos lo siguiente:

$$\begin{aligned} \frac{\partial C(u, v; \Gamma)}{\partial u} &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{q_2 - \rho q_1}{\sqrt{1-\rho^2}}} \exp\left(-\frac{1}{2(1-\rho^2)} \left(\rho q_1 - \left(z\sqrt{1-\rho^2} + \rho q_1\right)\right)^2\right) \frac{dx_2}{(1-\rho^2)} \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{q_2 - \rho q_1}{\sqrt{1-\rho^2}}} \exp\left(-\frac{1}{2} z^2\right) dz \\ &= \Phi\left(\frac{q_2 - \rho q_1}{\sqrt{1-\rho^2}}\right) \\ &= \Phi\left(\frac{\Phi^{-1}(v) - \rho \Phi^{-1}(u)}{\sqrt{1-\rho^2}}\right). \end{aligned} \quad (2.26)$$

Del procedimiento que acabamos de realizar vamos a hacer tres observaciones. La primera de ellas es que podemos derivar parcialmente a la cópula dos veces, es decir, primero derivamos parcialmente con respecto a v y después con respecto con u , y con ayuda de las Ecuaciones que aparecen en (2.24) obtenemos las siguientes expresiones:

$$\begin{aligned} \frac{\partial^2 C(u, v; \Gamma)}{\partial u \partial v} &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{(q_1^2 - 2\rho q_1 q_2 + q_2^2)}{2(1-\rho^2)}\right) \cdot \sqrt{2\pi} \exp\left(\frac{q_1^2}{2}\right) \sqrt{2\pi} \exp\left(\frac{q_2^2}{2}\right) \\ &= \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{(q_1^2 - 2\rho q_1 q_2 + q_2^2)}{2(1-\rho^2)} + \left(\frac{q_1^2}{2} + \frac{q_2^2}{2}\right)\right) \\ &= \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{(\rho^2 q_1^2 - 2\rho q_1 q_2 + \rho^2 q_2^2)}{2(1-\rho^2)}\right) \\ &= \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2} \xi^T (\Gamma^{-1} - I_2) \xi\right). \end{aligned} \quad (2.27)$$

Donde I_2 es la matriz identidad de 2×2 , $\Gamma := \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, entonces $\Gamma^{-1} := \begin{pmatrix} \frac{1}{1-\rho^2} & \frac{-\rho}{1-\rho^2} \\ \frac{-\rho}{1-\rho^2} & \frac{1}{1-\rho^2} \end{pmatrix}$ y $\xi^T := (q_1 \quad q_2)$, con $q_1 := \Phi^{-1}(u)$ y $q_2 := \Phi^{-1}(v)$.

Por lo tanto, hemos podido conseguir la función $c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}$ y por la Definición 2.3 nos falta multiplicar a $c(u, v)$ por las funciones de densidad marginales las cuales tienen una distribución $N(0, 1)$ para formar la función de densidad conjunta h de una función de distribución bivariada H como a continuación se muestra

$$\begin{aligned}
 h(x, y) &= \frac{\partial^2}{\partial x \partial y} (H(x, y)) \\
 &= \frac{\partial^2}{\partial x \partial y} (C(F(x), G(y))) \\
 &= \frac{\partial^2 C(F(x), G(y))}{\partial x \partial y} \cdot \frac{\partial F(x)}{\partial x} \cdot \frac{\partial G(y)}{\partial y} \\
 &= c(F(x), G(y)) f(x) g(y).
 \end{aligned} \tag{2.28}$$

Por lo tanto

$$\begin{aligned}
 h(x, y) &= \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{(\rho^2 q_1^2 - 2\rho q_1 q_2 + \rho^2 q_2^2)}{2(1-\rho^2)}\right) \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{q_1^2}{2}\right) \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{q_2^2}{2}\right) \\
 &= \frac{1}{2\pi} \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{(\rho^2 q_1^2 - 2\rho q_1 q_2 + \rho^2 q_2^2)}{2(1-\rho^2)} - \left(\frac{q_1^2 + q_2^2}{2}\right)\right) \\
 &= \frac{1}{2\pi} \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{(q_1^2 - 2\rho q_1 q_2 + q_2^2)}{2(1-\rho^2)}\right).
 \end{aligned} \tag{2.29}$$

Si hacemos el cambio de variable tal que $x = q_1 := \Phi^{-1}(u)$ y $y = q_2 := \Phi^{-1}(v)$, entonces obtenemos

$$h(x, y) = \frac{1}{2\pi} \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{(x^2 - 2\rho xy + y^2)}{2(1-\rho^2)}\right). \tag{2.30}$$

Esta última ecuación es la misma que la Ecuación (2.13), la cual es la función de densidad de la distribución normal estándar bivariada $\phi_2(x, y; \rho)$.

Por lo tanto, hemos obtenido una propiedad de la Cópula Normal Bivariada, la cual nos dice que dada la función de densidad de la cópula y las funciones de densidad marginales, si existen, tenemos la función de densidad conjunta de las variables originales, tal y como se muestra a continuación

$$\begin{aligned} h(x, y) &= c(F(x), G(y)) f(x) g(y) \\ &= \phi_2(x, y; \rho). \end{aligned} \quad (2.31)$$

Las dos últimas Expresiones obtenidas en (2.27) son las mismas que las Ecuaciones (2.17) y en (2.18) respectivamente. Las cuales nos llevaron a dar la Definición 2.4.

La segunda observación que haremos, es que, mediante un procedimiento análogo al anterior, se obtiene la derivada parcial de la cópula respecto a su otra variable, es decir:

$$\begin{aligned} \frac{\partial C(u, v; \Gamma)}{\partial v} &= \Phi\left(\frac{q_1 - \rho q_2}{\sqrt{1 - \rho^2}}\right) \\ &= \Phi\left(\frac{\Phi^{-1}(u) - \rho \Phi^{-1}(v)}{\sqrt{1 - \rho^2}}\right). \end{aligned} \quad (2.32)$$

Por lo tanto, obtenemos las siguientes fórmulas:

$$\frac{\partial C(u, v; \Gamma)}{\partial u} = \Phi\left(\frac{\Phi^{-1}(v) - \rho \Phi^{-1}(u)}{\sqrt{1 - \rho^2}}\right) \quad (2.33)$$

$$\frac{\partial C(u, v; \Gamma)}{\partial v} = \Phi\left(\frac{\Phi^{-1}(u) - \rho \Phi^{-1}(v)}{\sqrt{1 - \rho^2}}\right). \quad (2.34)$$

Finalmente, la tercera observación que haremos tiene que ver con la aplicación de la Regla de la Cadena a las ecuaciones anteriores. La Regla de la Cadena nos dice que si se tiene una función $y = h(u)$, la cual es una función derivable de u y si además $u = f(x)$ es una función derivable de x , entonces $y = h(f(x))$ es una función derivable, con:

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}. \quad (2.35)$$

La ecuación anterior también puede ser vista de la siguiente manera:

$$\frac{d}{dx} h(f(x)) = h'(f(x)) \cdot f'(x). \quad (2.36)$$

Si aplicamos esto a las Ecuaciones (2.33) y (2.34), donde $u = F(x)$, con F una función continua y derivable con respecto a x y $v = G(y)$, con G una función continua y derivable con respecto a y , entonces obtenemos las siguientes expresiones:

$$\begin{aligned} \frac{\partial C(u, v; \Gamma)}{\partial x} &= \frac{\partial C(u, v; \Gamma)}{\partial u} \cdot \frac{\partial u}{\partial x} \\ &= \frac{\partial C(F(x), G(y); \Gamma)}{\partial F(x)} \cdot \frac{\partial}{\partial x} F(x) \\ &= \Phi \left(\frac{\Phi^{-1}(v) - \rho \Phi^{-1}(u)}{\sqrt{1 - \rho^2}} \right) \cdot f(x) \end{aligned} \quad (2.37)$$

y

$$\begin{aligned} \frac{\partial C(u, v; \Gamma)}{\partial y} &= \frac{\partial C(u, v; \Gamma)}{\partial v} \cdot \frac{\partial v}{\partial y} \\ &= \frac{\partial C(F(x), G(y); \Gamma)}{\partial G(y)} \cdot \frac{\partial}{\partial y} G(y) \\ &= \Phi \left(\frac{\Phi^{-1}(u) - \rho \Phi^{-1}(v)}{\sqrt{1 - \rho^2}} \right) \cdot g(y). \end{aligned} \quad (2.38)$$

Procedemos a mostrar de manera gráfica la forma que tiene la función de densidad de la cópula normal bivariada.

2.3.2. Gráficas de la Cópula Normal Bivariada con Marginales Normales

Para ilustrar la forma que tiene una cópula normal bivariada mostramos a continuación imágenes de ésta, con el énfasis en que hacemos variar el parámetro de dependencia ρ , el cual va a tomar 3 valores en este trabajo de tesis, estos valores son -0.8 , -0.6 y -0.4 .

En las Figuras 2.1, 2.2 y 2.3 se muestran las funciones de densidad y de distribución acumulativa de una cópula normal bivariada la cual tiene ambas marginales normales estándar y con parámetros de dependencia $\rho = -0.8$, $\rho = -0.6$ y $\rho = -0.4$ respectivamente. Finalmente, en la Figura 2.4 se muestran los contornos de la cópula normal bivariada con los tres valores de ρ que hemos usado y notamos como al variar este parámetro cambia la forma de la función de densidad de la cópula normal, lo cual motiva a introducir la medida de asociación Tau de Kendall.

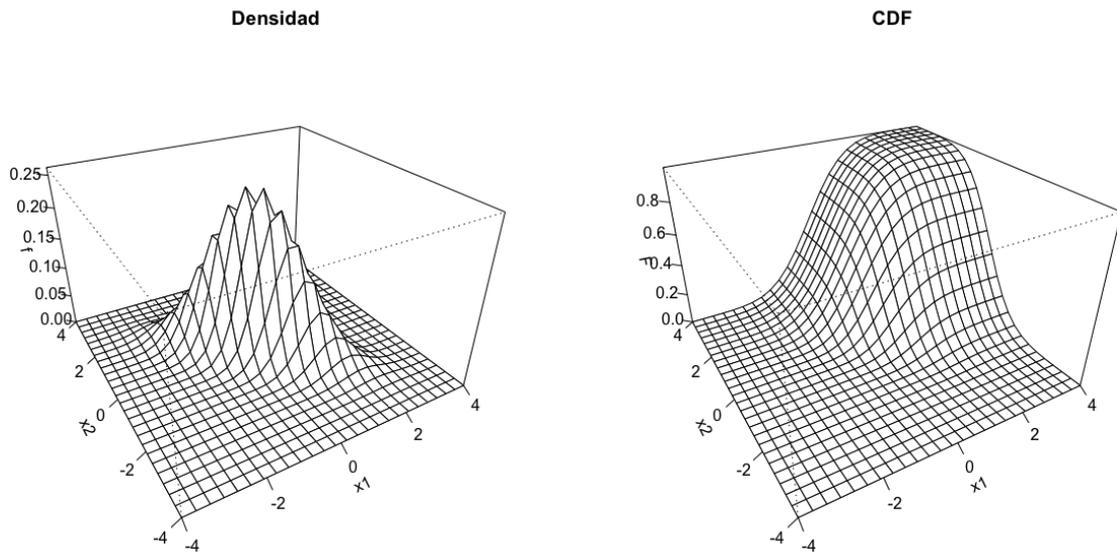


Figura 2.1: Gráfica de la función de densidad de la cópula normal bivariada con $\rho = -0.8$.

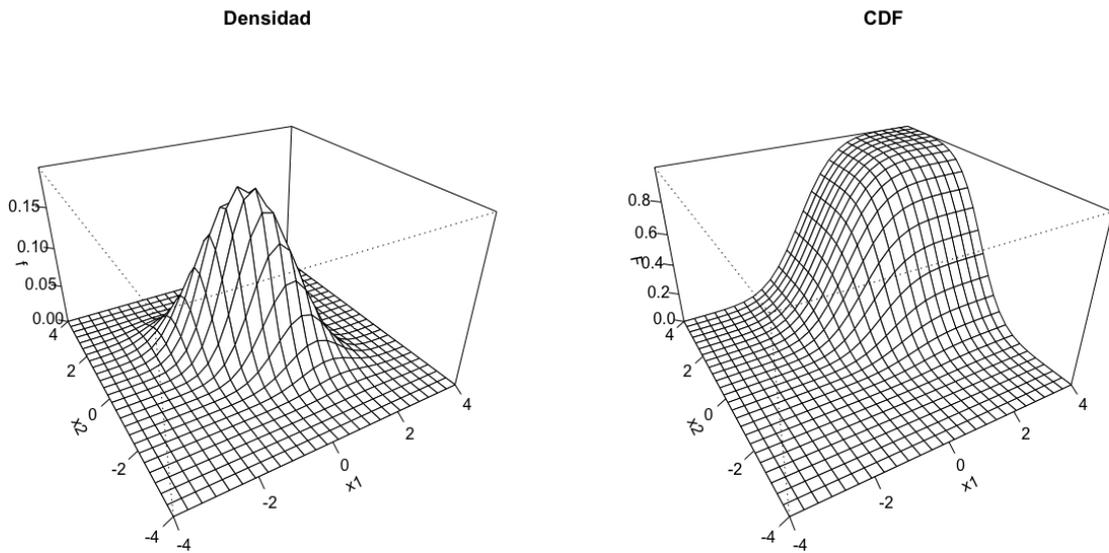


Figura 2.2: Gráfica de la función de densidad de la cópula normal bivariada con $\rho = -0.6$.

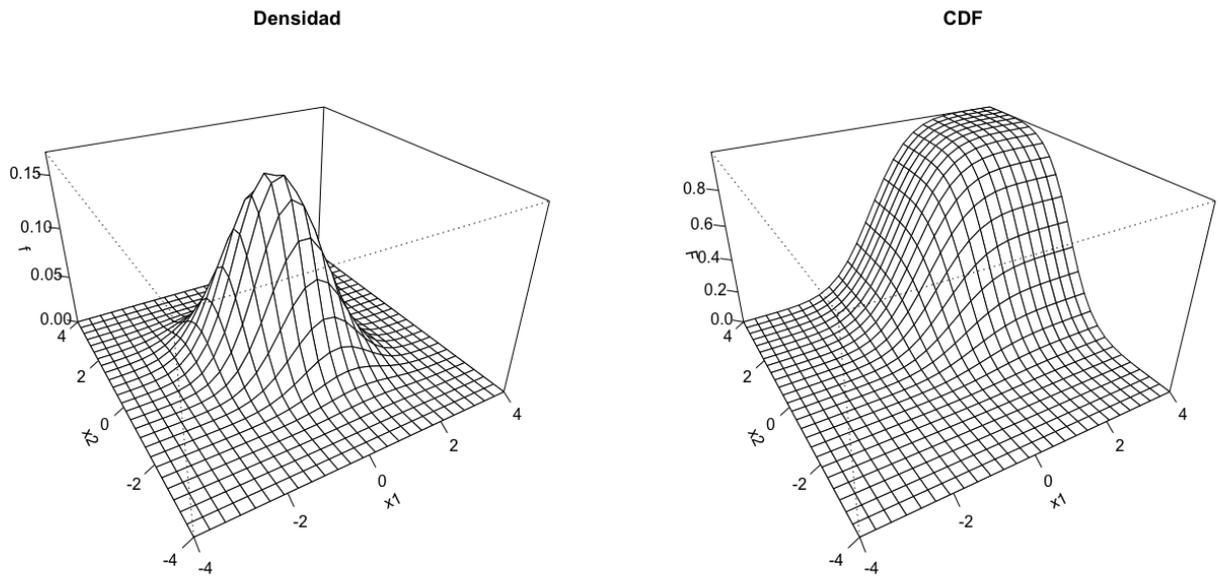


Figura 2.3: Gráfica de la función de densidad de la cópula normal bivariada con $\rho = -0.4$.

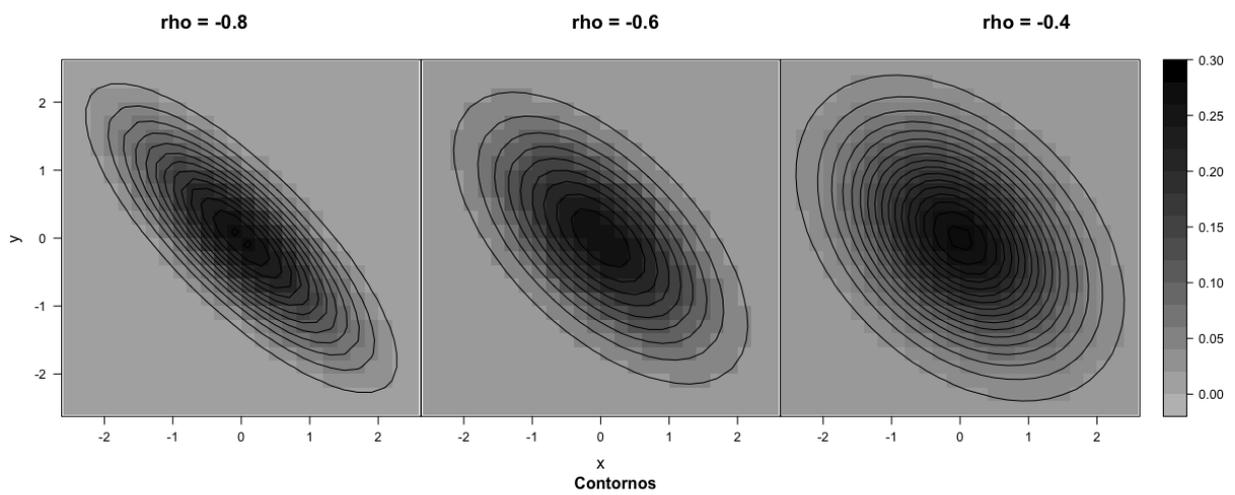


Figura 2.4: Gráfica de los contornos de la función de densidad de la cópula normal bivariada con los parámetros de correlación $\rho = -0.8$, $\rho = -0.6$ y $\rho = -0.4$.

2.4. Medida de Asociación Tau de Kendall ($\tau_{X,Y}$)

El coeficiente de correlación ρ de las variables aleatorias X y Y , es decir $\rho(X, Y)$, nos proporciona un indicador con el cuál se puede valorar la dependencia lineal que existe entre las variables aleatorias X y Y . Este valor oscila entre 1 y -1 . Cuando ρ está más próximo a 1 significa que la relación lineal que existe entre las variables X y Y es mayor en sentido positivo, lo que significa que X crece conforme Y crece. Cuando ρ está más próximo a -1 significa que la relación lineal que existe entre las variables X y Y es mayor en sentido negativo, es decir, las variables se mueven de forma lineal pero en sentido opuesto.

Un caso muy importante es cuando $\rho(X, Y)$ vale 0, esto significa que existe ausencia de una relación lineal entre las variables X y Y , pero esto no implica que no exista algún otro tipo de relación que sea no lineal. Para ello existen las denominadas medidas de asociación, en este trabajo de tesis ocuparemos la Tau de Kendall, la cual denotaremos con la letra τ , que cuantifica las relaciones no necesariamente lineales.

La medida de asociación Tau de Kendall oscila entre los valores 1 y -1 . Cuando Tau de Kendall toma alguno de estos valores extremos refleja una relación de dependencia. La dependencia es llamada negativa perfecta cuando se alcanza el valor -1 y es llamada positiva perfecta cuando se alcanza el valor 1. Conforme el valor de la medida de asociación se va alejando de estos valores, implica que existe falta de dependencia entre las variables, lo cuál significa que la medida de asociación Tau de Kendall determina cómo se relacionan los valores de la variable aleatoria X con los valores de la variable Y .

La fórmula de la medida de asociación o también denominada medida de concordancia, Tau de Kendall, para la cópula normal bivariada con parámetro de correlación ρ viene dada por la siguiente fórmula

$$\tau_{X,Y}(\rho) = \frac{2}{\pi} \arcsin \rho. \quad (2.39)$$

Para ilustrar el efecto de la tau de Kendall sobre la cópula normal bivariada mostramos a continuación los contornos bajo esta función.

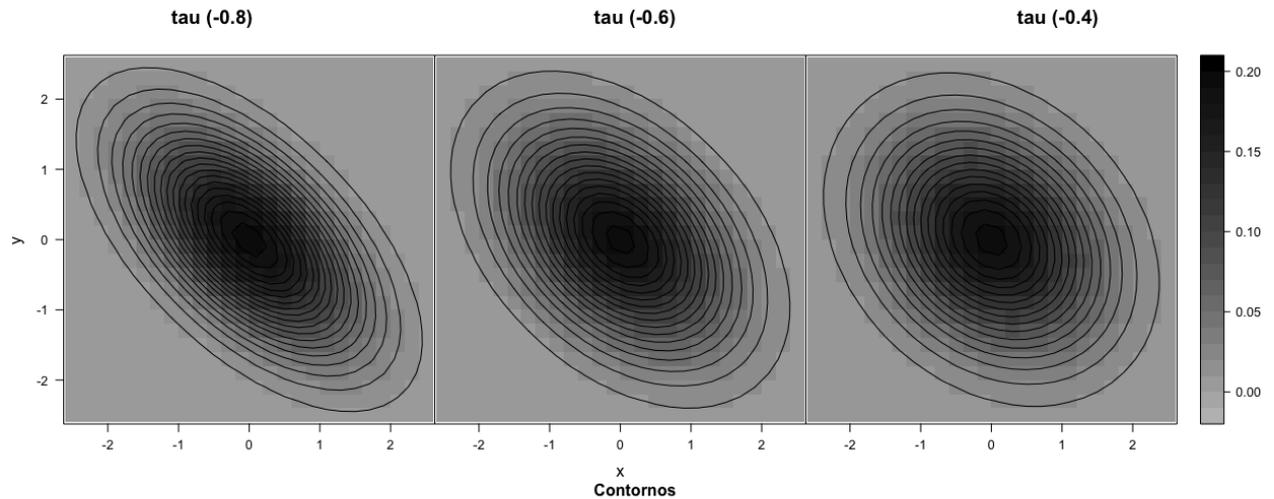


Figura 2.5: Gráfica de los contornos de la función de densidad de la cópula normal bivariada con $\tau(-0.8)$, $\tau(-0.6)$ y $\tau(-0.4)$.

2.5. Cópula Normal Bivariada con Marginales Normal Asimétrica y Weibull

En este trabajo de tesis vamos a ocupar las funciones de distribución Normal Asimétrica y Weibull.

Las tres funciones Normales Asimétricas que utilizamos tienen los siguientes parámetros:

1. $Sn(x_i = 3, \omega = 2, \alpha = -5)$
2. $Sn(x_i = 3, \omega = 2, \alpha = 5)$
3. $Sn(x_i = 3, \omega = 3, \alpha = 0)$

las gráficas de las funciones de densidad se muestran en la Figura 2.6.

Las dos funciones Weibull que utilizamos tienen los siguientes parámetros:

1. $W(shape = 1.5, scale = 1)$
2. $W(shape = 5, scale = 1)$

las gráficas de las funciones de densidad se muestran en la Figura 2.7.

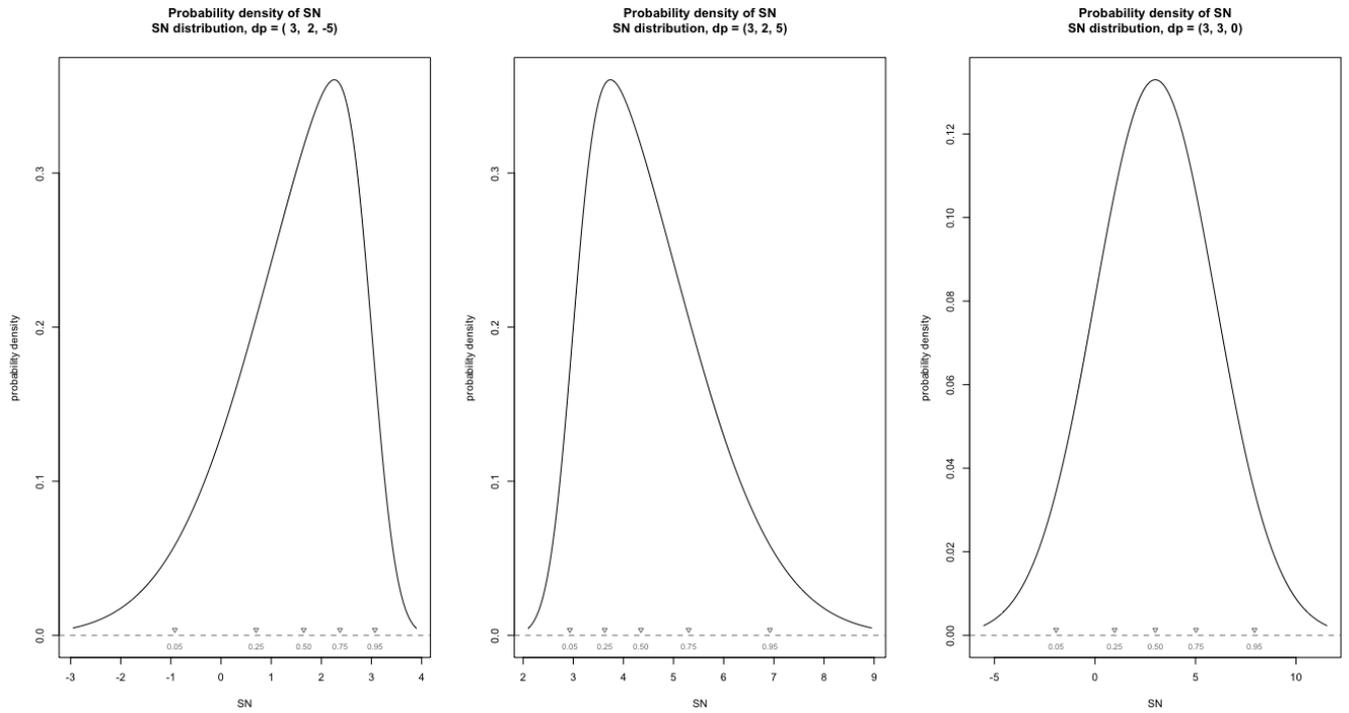


Figura 2.6: Gráficas de las funciones de densidad de la distribución normal asimétrica.

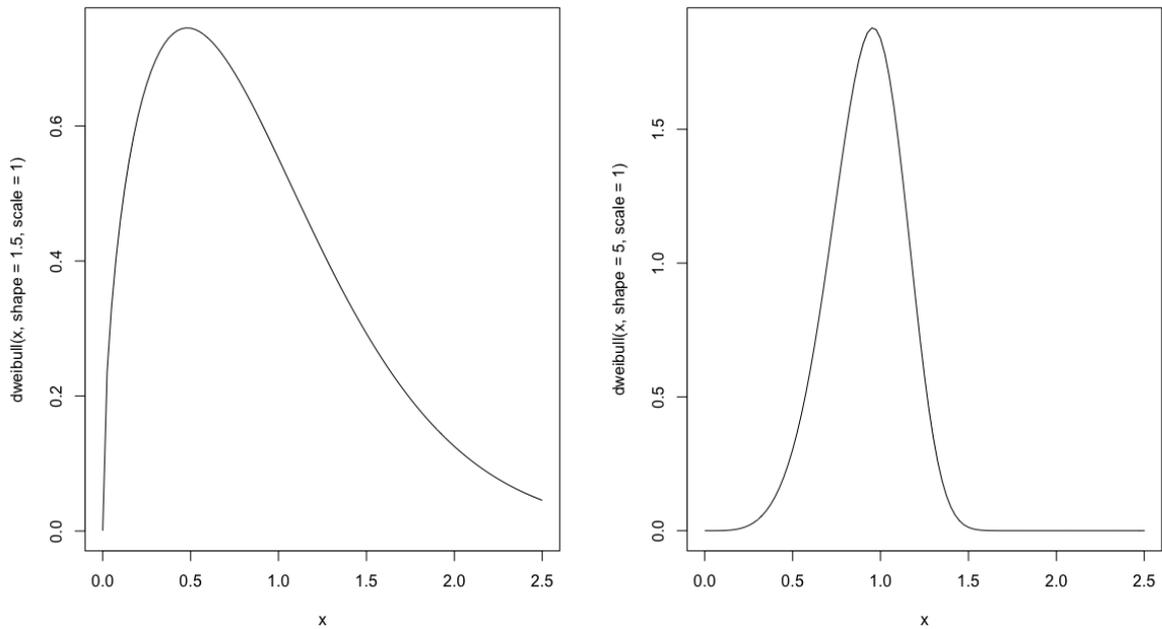


Figura 2.7: Gráficas de las funciones de densidad de la distribución Weibull.

Los parámetros de correlación que utilizamos en este trabajo de tesis son los siguientes: $\rho = -0.8$, $\rho = -0.6$ y $\rho = -0.4$. A estos parámetros de correlación se les va a aplicar la Ecuación (2.39), es decir, vamos a ocupar la medida de asociación Tau de Kendall en los siguientes ejemplos.

2.5.1. Representación Gráfica de la Cópula Normal Bivariada con Marginales Normal Asimétrica y Weibull

A continuación mostramos 3 ejemplos de la cópula normal bivariada con la primera marginal normal asimétrica y la segunda marginal Weibull, especificando sus parámetros correspondientes y su medida de asociación, así mismo, se muestran los contornos de cada una de las cópulas usando las 3 medidas de asociación $\tau(-0.8)$, $\tau(-0.6)$ y $\tau(-0.4)$.

Ejemplo 1

La gráfica de la función de densidad de la cópula normal bivariada con medida de asociación $\tau(-0.8)$ con la primera marginal normal asimétrica con parámetros $Sn(xi = 3, omega = 2, alpha = -5)$ y segunda marginal Weibull con parámetros $W(shape = 1.5, scale = 1)$ se muestra en la Figura 2.8.

En la Figura 2.9 se muestran los contornos de esta cópula aplicando las 3 medidas de asociación $\tau(-0.8)$, $\tau(-0.6)$ y $\tau(-0.4)$.

Ejemplo 2

La gráfica de la función de densidad de la cópula normal bivariada con medida de asociación $\tau(-0.6)$ con la primera marginal normal asimétrica con parámetros $Sn(xi = 3, omega = 2, alpha = 5)$ y segunda marginal Weibull con parámetros $W(shape = 5, scale = 1)$ se muestra en la Figura 2.10.

En la Figura 2.11 se muestran los contornos de esta cópula aplicando las 3 medidas de asociación $\tau(-0.8)$, $\tau(-0.6)$ y $\tau(-0.4)$.

Ejemplo 3

La gráfica de la función de densidad de la cópula normal bivariada con medida de asociación $\tau(-0.4)$ con la primera marginal normal asimétrica con parámetros $Sn(xi = 3, omega = 3, alpha = 0)$ y segunda marginal Weibull con parámetros $W(shape = 1.5, scale = 1)$ se muestra en la Figura 2.12.

En la Figura 2.13 se muestran los contornos de esta cópula aplicando las 3 medidas de asociación $\tau(-0.8)$, $\tau(-0.6)$ y $\tau(-0.4)$.

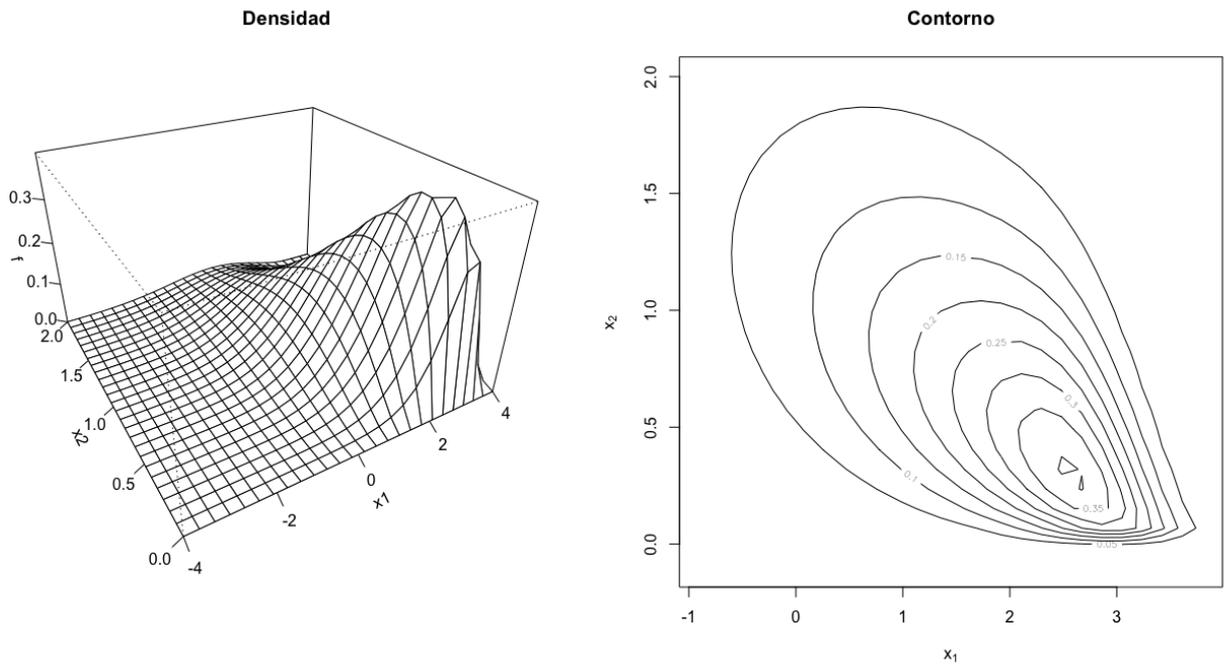


Figura 2.8: Gráfica de la función de densidad de la cópula normal biviariada con marginales normal asimétrica y Weibull con parámetros $Sn(x_i = 3, \omega = 2, \alpha = -5)$ y $W(shape = 1.5, scale = 1)$ respectivamente, con medida de asociación $\tau(-0.8)$ y su respectivo contorno.

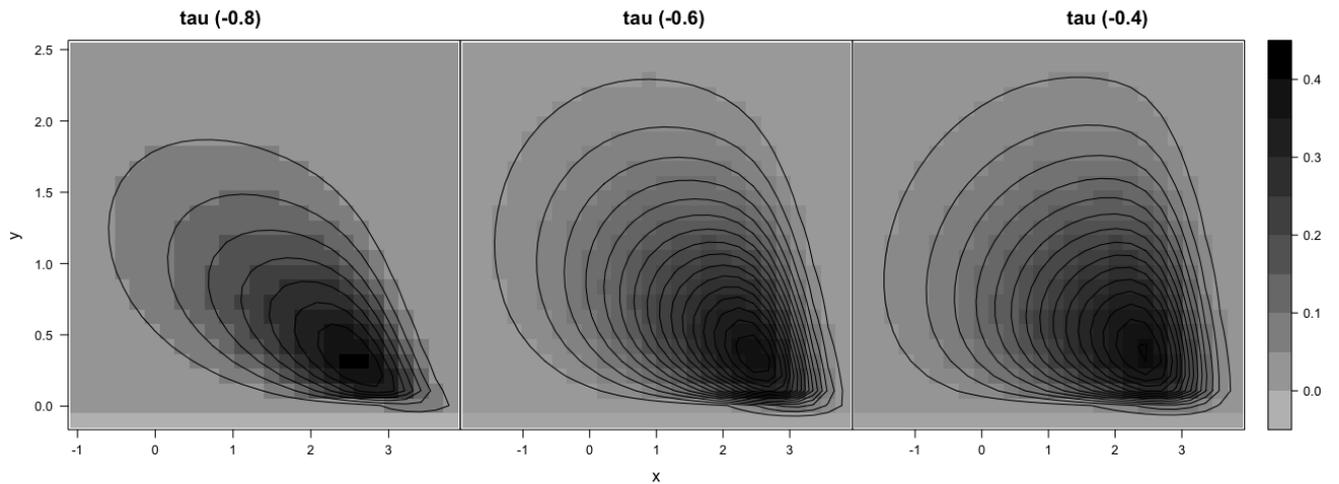


Figura 2.9: Gráfica de los contornos de la función de densidad de la cópula normal biviariada con marginales normal asimétrica y Weibull con parámetros $Sn(x_i = 3, \omega = 2, \alpha = -5)$ y $W(shape = 1.5, scale = 1)$ respectivamente y con medidas de asociación $\tau(-0.8)$, $\tau(-0.6)$ y $\tau(-0.4)$.

2.5. CÓPULA NORMAL BIVARIADA CON MARGINALES NORMAL ASIMÉTRICA Y WEIBULL 31

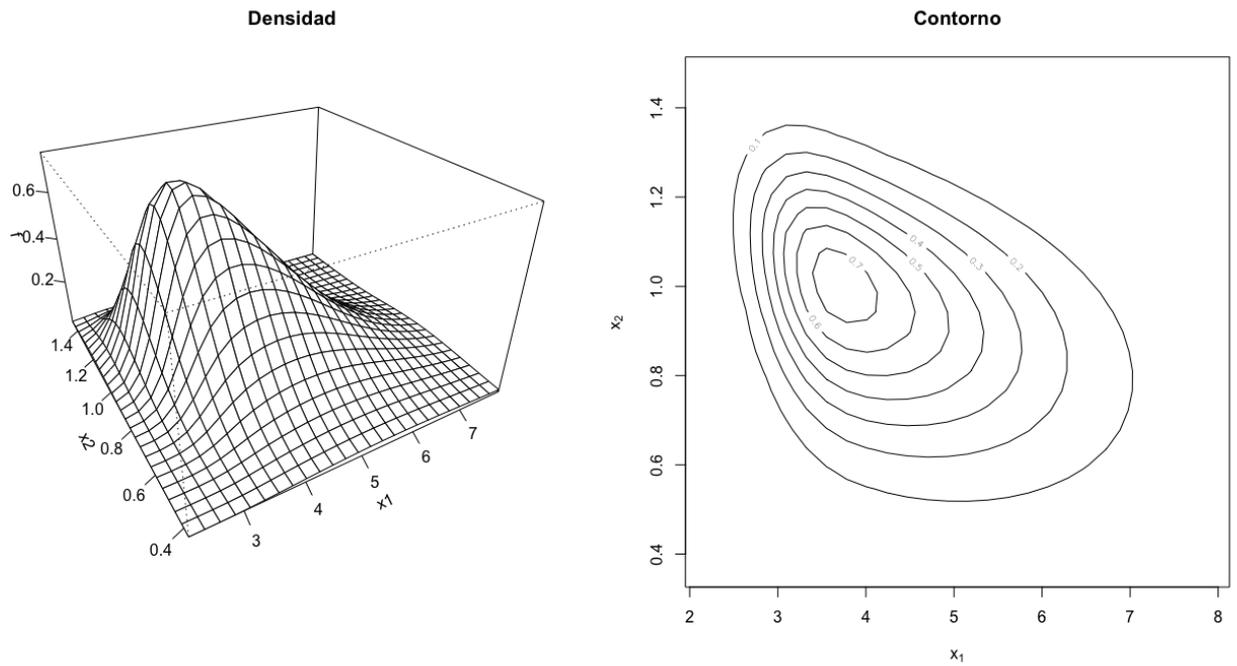


Figura 2.10: Gráfica de la función de densidad de la cópula normal bivariada con marginales normal asimétrica y Weibull con parámetros Sn ($xi = 3$, $omega = 2$, $alpha = 5$) y W ($shape = 5$, $scale = 1$) respectivamente, con medida de asociación τ (-0.6) y su respectivo contorno.

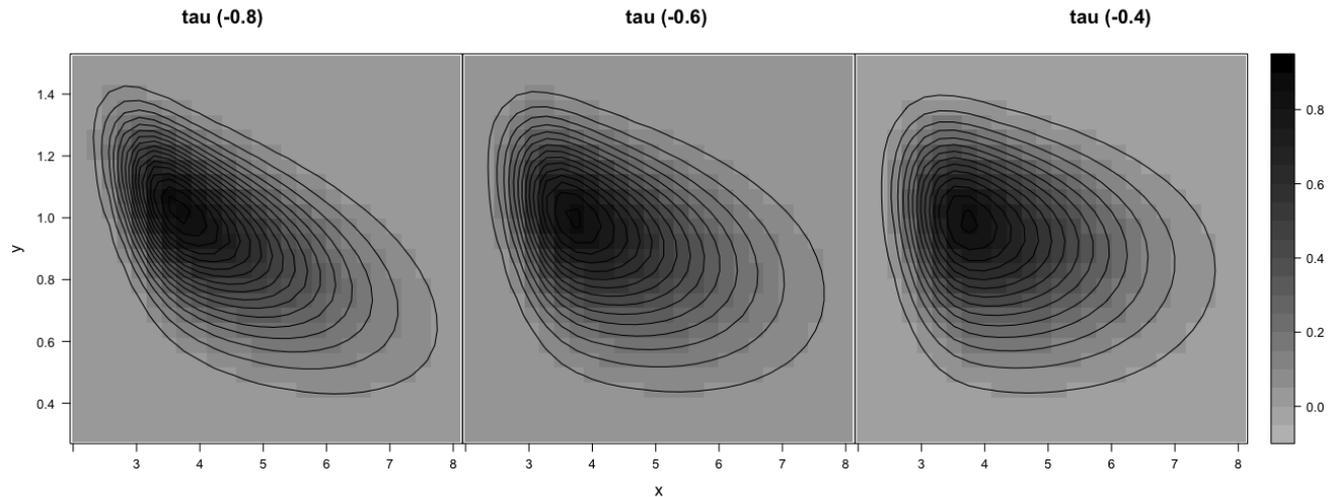


Figura 2.11: Gráfica de los contornos de la función de densidad de la cópula normal bivariada con marginales normal asimétrica y Weibull con parámetros Sn ($xi = 3$, $omega = 2$, $alpha = 5$) y W ($shape = 5$, $scale = 1$) respectivamente y con medidas de asociación τ (-0.8), τ (-0.6) y τ (-0.4).

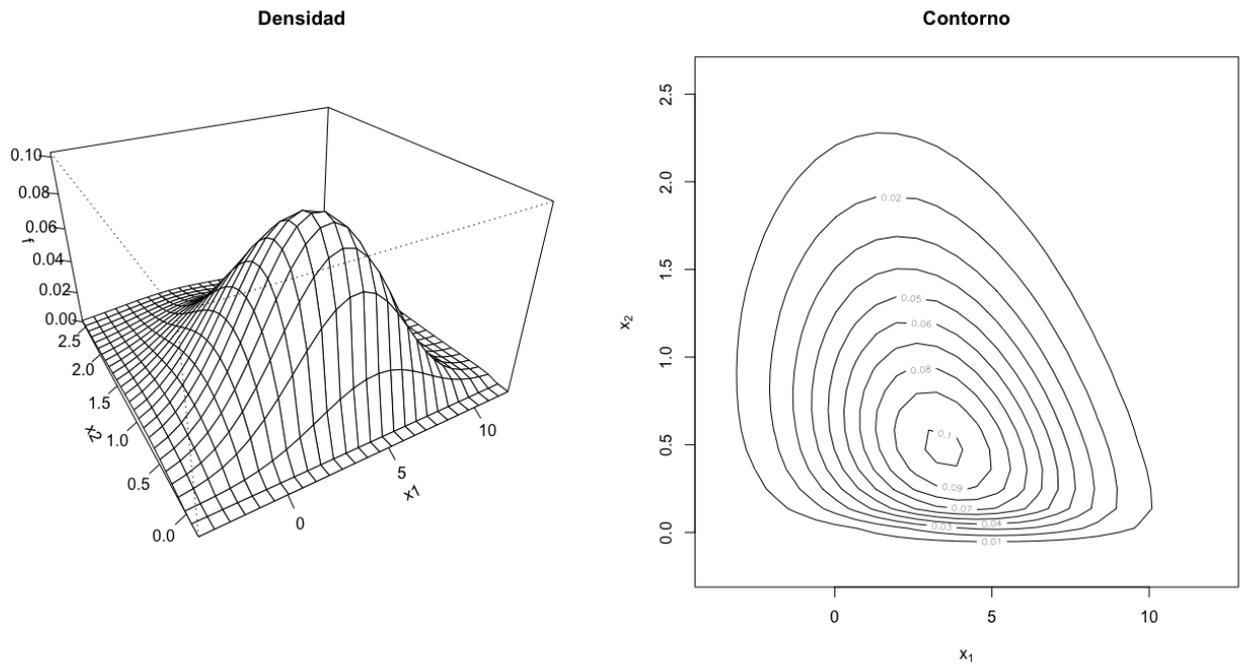


Figura 2.12: Gráfica de la función de densidad de la cópula normal biviada con marginales normal asimétrica y Weibull con parámetros Sn ($\xi = 3$, $\omega = 3$, $\alpha = 0$) y W ($shape = 1.5$, $scale = 1$) respectivamente, con medida de asociación τ (-0.4) y su respectivo contorno.

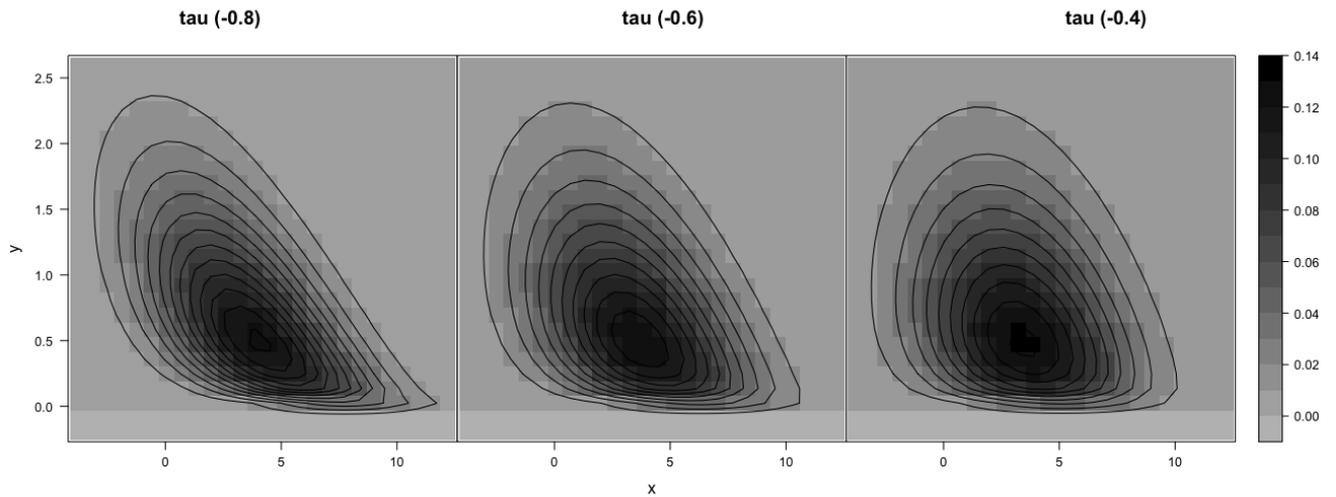


Figura 2.13: Gráfica de los contornos de la función de densidad de la cópula normal biviada con marginales normal asimétrica y Weibull con parámetros Sn ($\xi = 3$, $\omega = 3$, $\alpha = 0$) y W ($shape = 1.5$, $scale = 1$) respectivamente y con medidas de asociación τ (-0.8), τ (-0.6) y τ (-0.4).

Capítulo 3

Inferencia Estadística sobre Cópulas Bivariadas

3.1. Introducción

La inferencia estadística es una parte de la estadística que estudia técnicas y procedimientos con el fin de extender o generalizar la información de una muestra a la población completa, es decir, consiste en deducir propiedades de una población basándose en una parte de esta.

Con determinadas técnicas se puede realizar una predicción sobre cuáles podrían ser los parámetros de la población. El método de máxima verosimilitud es de los métodos más importantes para estimar parámetros de una distribución de probabilidad y es el método que ocuparemos en este trabajo de tesis. A manera de introducción mencionaremos en lo general de qué se trata este método.

Supongamos que tenemos un vector aleatorio (X_1, X_2, \dots, X_n) cuya distribución depende de un parámetro θ , entonces, la función de verosimilitud, $L(\theta, \mathbf{x})$, de este vector aleatorio es la función de densidad evaluada en el vector numérico fijo (x_1, x_2, \dots, x_n) que depende del parámetro θ

$$L(\theta, \mathbf{x}) = f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n; \theta)$$

si X_1, X_2, \dots, X_n son independientes se obtiene que

$$L(\theta, \mathbf{x}) = f_{X_1}(x_1; \theta) \cdots f_{X_n}(x_n; \theta)$$

aún más, cuando las variables son idénticamente distribuidas la función $L(\theta, \mathbf{x})$ está dada por

$$\begin{aligned} L(\theta, \mathbf{x}) &= f(x_1; \theta) \cdots f(x_n; \theta) \\ &= \prod_{i=1}^n f(x_i; \theta) \end{aligned} \tag{3.1}$$

esto último se da cuando la muestra es aleatoria.

El método de máxima verosimilitud consiste en obtener el valor de θ que maximiza la Función de Verosimilitud 3.1. Al valor de $\hat{\theta}$ en donde $L(\theta, \mathbf{x})$ alcanza su máximo se le llama estimación de máxima

verosimilitud o estimación máxima verosímil. Esto significa que θ es tal que el valor numérico observado (x_1, x_2, \dots, x_n) de la muestra aleatoria tenga probabilidad máxima. En la práctica muchas veces resulta más fácil resolver el problema de optimización de $\ln(L(\theta, \mathbf{x}))$ que $L(\theta, \mathbf{x})$, observamos que ambas funciones alcanza el máximo en el mismo punto debido a las propiedades de continuidad y monotonía de la función \ln . A la función $\ln(L(\theta, \mathbf{x}))$ se le denomina la función de log-verosimilitud y está dada de la siguiente manera

$$\begin{aligned}\ell(\theta, \mathbf{x}) &= \ln(L(\theta, \mathbf{x})) \\ &= \ln \left\{ \prod_{i=1}^n f(x_i; \theta) \right\} \\ &= \sum_{i=1}^n \ln \{f(x_i; \theta)\}.\end{aligned}\tag{3.2}$$

Pueden existir varios estimadores para un parámetro dado, lo que nos lleva al problema de cuál parámetro elegir. Es conveniente, entonces, que veamos algunas propiedades de los estimadores.

Definición 3.1 Un estimador $\hat{\theta}$ es insesgado para un parámetro θ si $E[\hat{\theta}] = \theta$, de otra manera es llamado sesgado. El sesgo de $\hat{\theta}$ es definido como $bias(\hat{\theta}) = E[\hat{\theta}] - \theta$.

Procedemos a mostrar algunas definiciones referentes a la varianza de un estimador insesgado.

Definición 3.2 Un estimador insesgado $\hat{\theta}_1$ es relativamente más eficiente que otro estimador insesgado $\hat{\theta}_2$ si se cumple que $Var(\hat{\theta}_1) \leq Var(\hat{\theta}_2)$.

Por otra parte, supongamos que X_1, X_2, \dots, X_n forma una muestra aleatoria de la distribución con función de densidad de probabilidad $f(x; \theta)$. Sujeta a ciertas condiciones de regularidad sobre $f(x; \theta)$, tenemos que para cualquier estimador insesgado $\hat{\theta}$ para θ

$$Var[\hat{\theta}] \geq I_{\theta}^{-1}\tag{3.3}$$

donde

$$\begin{aligned}I_{\theta} &= E \left[\left(\frac{\partial \ln[L(\theta, \mathbf{x})]}{\partial \theta} \right)^2 \right] \\ &= E \left[\left(\frac{\partial \ell(\theta, \mathbf{x})}{\partial \theta} \right)^2 \right].\end{aligned}\tag{3.4}$$

A I_{θ}^{-1} se le conoce como la cota inferior de Cramér-Rao y a la Ecuación (3.3) se le llama desigualdad de Cramér-Rao. Esto motiva a dar las siguientes definiciones.

Definición 3.3 Un estimador insesgado es eficiente si su varianza alcanza la cota inferior de Cramér-Rao.

La Ecuación (3.3) y la definición anterior motiva a lo siguiente

Definición 3.4 La eficiencia de un estimador insesgado $\hat{\theta}$ es

$$Eff(\hat{\theta}) = \frac{I_{\theta}^{-1}}{Var[\hat{\theta}]} \leq 1 \quad (3.5)$$

por lo tanto, un estimador insesgado es eficiente si $Eff(\hat{\theta}) = 1$.

Existen también estimadores insesgados que alcanzan la eficiencia cuando la muestra aleatoria tiende a infinito, esto motiva una nueva definición.

Definición 3.5 Un estimador insesgado $\hat{\theta}_n = \hat{\theta}(X_1, \dots, X_n)$ es asintóticamente eficiente si

$$\lim_{n \rightarrow \infty} Eff(\hat{\theta}_n) = 1.$$

Como observación tenemos que la eficiencia y la eficiencia relativa dependen teóricamente del tamaño de la muestra disponible, pero a menudo es posible usar la eficiencia relativa asintótica, la cual se define como el límite de las eficiencias relativas a medida que el tamaño de la muestra crece, como principal medida de comparación.

En este trabajo de tesis abordamos cópulas bivariadas las cuales dependen de un parámetro al cual denominaremos θ , el cual es llamado parámetro de dependencia.

Para la estimación de parámetros de una cópula por el método de máxima verosimilitud se presentarán dos métodos. El primer método consiste en estimar todos los parámetros, tanto de las distribuciones marginales como de la distribución de la cópula, todo esto conjuntamente, a esto se le conoce como el método exacto de máxima verosimilitud. El segundo método consiste en estimar primero los parámetros de las distribuciones marginales para posteriormente estimar los parámetros de la cópula, a esto se le conoce como el método IFM (Inference For the Margins). Como observación, aclaramos que el primer método es más eficiente estadísticamente, pero también es más complejo computacionalmente por todos los parámetros a calcular conjuntamente.

El lector que este interesado en los temas presentados en éste capítulo le recomendamos consultar [7], [11], [12] y [21].

3.2. Máxima Verosimilitud para Cópulas Bivariadas

En la Definición 2.3 se muestra la función de densidad conjunta h de una cópula bivariada, la cual esta dada por la siguiente ecuación

$$h(x, y) = c(F(x), G(y)) f(x) g(y). \quad (3.6)$$

Sea $\beta = (\beta_1, \beta_2, \dots, \beta_\gamma)$ el vector de parámetros de la cópula $c(F(x), G(y))$. Los parámetros de las distribuciones marginales están dados de la siguiente manera: para la distribución marginal F están dados por $\alpha_{11}, \alpha_{12}, \dots, \alpha_{1\epsilon}$ y para la distribución marginal G están dados por $\alpha_{21}, \alpha_{22}, \dots, \alpha_{2\delta}$. Sea α el vector

de parámetros de las dos distribuciones marginales, es decir, $\boldsymbol{\alpha} = (\alpha_{11}, \alpha_{12}, \dots, \alpha_{1\epsilon}, \alpha_{21}, \alpha_{22}, \dots, \alpha_{2\delta})$. Por lo tanto, la Ecuación (3.6) puede ser vista como

$$h(x, y, \boldsymbol{\alpha}, \boldsymbol{\beta}) = c(F(x; \alpha_{11}, \alpha_{12}, \dots, \alpha_{1\epsilon}), G(y; \alpha_{21}, \alpha_{22}, \dots, \alpha_{2\delta}); \boldsymbol{\beta}) f(x; \alpha_{11}, \alpha_{12}, \dots, \alpha_{1\epsilon}) \cdot g(y; \alpha_{21}, \alpha_{22}, \dots, \alpha_{2\delta}).$$

Para simplificar la Ecuación anterior definamos a la marginal F como F_1 , a la variable x como x_1 , a los parámetros de la marginal F por $\boldsymbol{\alpha}_1 = \alpha_{11}, \alpha_{12}, \dots, \alpha_{1\epsilon}$. Por otra parte, se define a la marginal G como F_2 , a la variable y como x_2 y a los parámetros de la marginal G por $\boldsymbol{\alpha}_2 = \alpha_{21}, \alpha_{22}, \dots, \alpha_{2\delta}$. Obteniendo como resultado la siguiente ecuación

$$h(x_1, x_2, \boldsymbol{\alpha}, \boldsymbol{\beta}) = c(F_1(x_1; \boldsymbol{\alpha}_1), F_2(x_2; \boldsymbol{\alpha}_2); \boldsymbol{\beta}) f_1(x_1; \boldsymbol{\alpha}_1) f_2(x_2; \boldsymbol{\alpha}_2). \quad (3.7)$$

Supongamos que realizamos una muestra aleatoria e independiente de tamaño n de la distribución bivariada, tal muestra esta representada por el conjunto $\{(x_{1i}, x_{2i}) \mid i = 1, 2, \dots, n\}$. Entonces, la función de verosimilitud está dada por

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{x}) = \prod_{i=1}^n c(F_1(x_{1i}; \boldsymbol{\alpha}_1), F_2(x_{2i}; \boldsymbol{\alpha}_2); \boldsymbol{\beta}) f_1(x_{1i}; \boldsymbol{\alpha}_1) f_2(x_{2i}; \boldsymbol{\alpha}_2). \quad (3.8)$$

Por lo tanto, podemos obtener la función log –verosimilitud como a continuación se indica

$$\begin{aligned} \ell(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{x}) &= \ln L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{x}) \\ &= \sum_{i=1}^n \left(\ln c(F_1(x_{1i}; \boldsymbol{\alpha}_1), F_2(x_{2i}; \boldsymbol{\alpha}_2); \boldsymbol{\beta}) + \sum_{j=1}^2 \ln f_j(x_{ji}; \boldsymbol{\alpha}_j) \right) \\ &= \sum_{i=1}^n \ln c(F_1(x_{1i}; \boldsymbol{\alpha}_1), F_2(x_{2i}; \boldsymbol{\alpha}_2); \boldsymbol{\beta}) + \sum_{i=1}^n \sum_{j=1}^2 \ln f_j(x_{ji}; \boldsymbol{\alpha}_j). \end{aligned} \quad (3.9)$$

La función log –verosimilitud debe de ser maximizada.

Para maximizar la función log –verosimilitud debemos notar que el vector a ser estimado es $\boldsymbol{\theta}$, el cual está definido como

$$\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}) \quad (3.10)$$

donde $\boldsymbol{\alpha}$ el vector de parámetros de las distribuciones marginales y $\boldsymbol{\beta}$ es el vector de parámetros de la distribución de la cópula.

Entonces, la función log –verosimilitud alcanza su máximo en $\hat{\boldsymbol{\theta}}$, tal que

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \arg \max_{(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \Theta} \ell(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{x}) \\ &= \arg \max_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta}, \mathbf{x}). \end{aligned} \quad (3.11)$$

Donde Θ es el espacio parametral.

3.3. Método Exacto de Máxima Verosimilitud para Cópulas Bivariadas

El método exacto de máxima verosimilitud radica en maximizar la función

$$\ell(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{x}) = \sum_{i=1}^n \ln c(F_1(x_{1i}; \boldsymbol{\alpha}_1), F_2(x_{2i}; \boldsymbol{\alpha}_2); \boldsymbol{\beta}) + \sum_{i=1}^n \sum_{j=1}^2 \ln f_j(x_{ji}; \boldsymbol{\alpha}_j) \quad (3.12)$$

pero esto implica maximizar una función multidimensional, ya que

$$\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_\gamma)$$

y

$$\boldsymbol{\alpha} = (\alpha_{11}, \alpha_{12}, \dots, \alpha_{1\varepsilon}, \alpha_{21}, \alpha_{22}, \dots, \alpha_{2\delta}).$$

Entonces se debe de maximizar una función $(\gamma + (\varepsilon + \delta))$ -dimensional, donde γ es el número de parámetros de la cópula, ε y δ son los números de parámetros de las marginales F_1 y F_2 respectivamente.

El método exacto de máxima verosimilitud radica en que se deben de estimar todos los parámetros de las distribuciones marginales y de la distribución de la cópula conjuntamente, esto implica un gran esfuerzo computacional.

3.4. Método IFM

Como hemos visto, a medida que aumentan los parámetros aumenta la dimensión de la función que se desea maximizar y por lo tanto, el problema de optimización se hace cada vez más difícil de resolver. Para ello Joe y Xu [12], propusieron un método de estimación de dos etapas llamado el método IFM (Inference For the Margins).

Observando la Ecuación (3.9)

$$\ell(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{x}) = \sum_{i=1}^n \ln c(F_1(x_{1i}; \boldsymbol{\alpha}_1), F_2(x_{2i}; \boldsymbol{\alpha}_2); \boldsymbol{\beta}) + \sum_{i=1}^n \sum_{j=1}^2 \ln f_j(x_{ji}; \boldsymbol{\alpha}_j) \quad (3.13)$$

podemos notar que está compuesta por la suma de dos expresiones. En cuanto a la primera expresión, observamos que involucra los parámetros de las distribuciones marginales y de la distribución de la cópula. Por otra parte, la segunda expresión depende únicamente de los parámetros de las funciones marginales. El método IFM consiste en maximizar cada una de las expresiones por separado, es decir, la Ecuación (3.9) va a ser maximizada en dos etapas.

Primera Etapa

Utilizando el método de máxima verosimilitud estimamos los parámetros de las marginales ($\boldsymbol{\alpha}$), es decir

$$\hat{\boldsymbol{\alpha}}_{IFM} = \arg \max_{\boldsymbol{\alpha}} \sum_{i=1}^n \sum_{j=1}^2 \ln f_j(x_{ji}; \boldsymbol{\alpha}_j). \quad (3.14)$$

Notamos que

$$\begin{aligned}
\hat{\boldsymbol{\alpha}}_{IFM} &= \arg \max_{\boldsymbol{\alpha}} \sum_{i=1}^n \sum_{j=1}^2 \ln f_j(x_{ji}; \boldsymbol{\alpha}_j) \\
&= \arg \max_{\boldsymbol{\alpha}} \sum_{i=1}^n (\ln f_1(x_{1i}; \boldsymbol{\alpha}_1) + \ln f_2(x_{2i}; \boldsymbol{\alpha}_2)) \\
&= \arg \max_{\boldsymbol{\alpha}} \left\{ \sum_{i=1}^n f_1(x_{1i}; \boldsymbol{\alpha}_1) + \sum_{i=1}^n f_2(x_{2i}; \boldsymbol{\alpha}_2) \right\} \\
&= \arg \max_{\boldsymbol{\alpha}_1} \sum_{i=1}^n f_1(x_{1i}; \boldsymbol{\alpha}_1) + \arg \max_{\boldsymbol{\alpha}_2} \sum_{i=1}^n f_2(x_{2i}; \boldsymbol{\alpha}_2). \tag{3.15}
\end{aligned}$$

Sean

$$\hat{\boldsymbol{\alpha}}_{1IFM} = \arg \max_{\boldsymbol{\alpha}_1} \sum_{i=1}^n f_1(x_{1i}; \boldsymbol{\alpha}_1)$$

y

$$\hat{\boldsymbol{\alpha}}_{2IFM} = \arg \max_{\boldsymbol{\alpha}_2} \sum_{i=1}^n f_2(x_{2i}; \boldsymbol{\alpha}_2).$$

Donde $\hat{\boldsymbol{\alpha}}_{IFM} = (\hat{\boldsymbol{\alpha}}_{1IFM}, \hat{\boldsymbol{\alpha}}_{2IFM})$.

Lo cual es equivalente a estimar los parámetros de las marginales individuales utilizando máxima verosimilitud del modo habitual [21].

Segunda Etapa

Habiendo encontrado $\hat{\boldsymbol{\alpha}}_{IFM}$ procedemos a estimar los parámetros $\boldsymbol{\beta}$ de la cópula, de la siguiente manera

$$\hat{\boldsymbol{\beta}}_{IFM} = \arg \max_{\boldsymbol{\beta}} \sum_{i=1}^n \ln c(F_1(x_{1i}; \hat{\boldsymbol{\alpha}}_{1IFM}), F_2(x_{2i}; \hat{\boldsymbol{\alpha}}_{2IFM}); \boldsymbol{\beta}). \tag{3.16}$$

Juntando los resultados de cada etapa se obtiene el estimador $\hat{\boldsymbol{\theta}}_{IFM}$, el cual está definido como

$$\hat{\boldsymbol{\theta}}_{IFM} = (\hat{\boldsymbol{\alpha}}_{IFM}, \hat{\boldsymbol{\beta}}_{IFM}). \tag{3.17}$$

Debemos de observar que el método IFM baja la dimensión del problema original, de optimizar la Ecuación (3.9), y lo transforma a su vez a varios problemas mas pequeños que tienen un menor número de parámetros a ser estimados. Esto implica que, en vez de resolver un problema computacionalmente complejo se resuelven varios problemas computacionalmente más simples, de menor dimensión.

3.5. Ejemplo de Máxima verosimilitud en Cópulas Bivariadas

3.5.1. Cópula Normal Bivariada

En la Ecuación (2.31) se obtuvo que

$$\begin{aligned} h(x, y) &= c(F(x), G(y)) f(x) g(y) \\ &= \phi_2(x, y; \rho) \end{aligned}$$

y por la Ecuación (3.8) se obtiene la función de verosimilitud

$$\begin{aligned} L(\boldsymbol{\theta}, \mathbf{x}) &= \prod_{i=1}^n c(F_1(x_{1i}; \boldsymbol{\alpha}_1), F_2(x_{2i}; \boldsymbol{\alpha}_2); \boldsymbol{\beta}) f_1(x_{1i}; \boldsymbol{\alpha}_1) f_2(x_{2i}; \boldsymbol{\alpha}_2) \\ &= \prod_{i=1}^n h(x_i, y_i, \boldsymbol{\theta}) \\ &= \prod_{i=1}^n \phi_2(x_i, y_i; \boldsymbol{\theta}) \end{aligned} \tag{3.18}$$

con $x_i := \Phi^{-1}(u_i)$ y $y_i := \Phi^{-1}(v_i)$.

Por lo tanto

$$\begin{aligned} L(\boldsymbol{\theta}, \mathbf{x}) &= \prod_{i=1}^n \phi_2(x_i, y_i; \boldsymbol{\theta}) \\ &= \prod_{i=1}^n \frac{1}{2\pi\sqrt{1-\theta^2}} \exp\left(-\frac{x_i^2 - 2\theta x_i y_i + y_i^2}{2(1-\theta^2)}\right) \\ &= \left(\frac{1}{2\pi\sqrt{1-\theta^2}}\right)^n \exp\left(\frac{-1}{2(1-\theta^2)} \sum_{i=1}^n (x_i^2 - 2\theta x_i y_i + y_i^2)\right). \end{aligned} \tag{3.19}$$

De la Ecuación (3.19) se obtiene la función log –verosimilitud

$$\begin{aligned}
\ell(\boldsymbol{\theta}, \mathbf{x}) &= \log \left\{ \left(\frac{1}{2\pi\sqrt{1-\theta^2}} \right)^n \exp \left(\frac{-1}{2(1-\theta^2)} \sum_{i=1}^n (x_i^2 - 2\theta x_i y_i + y_i^2) \right) \right\} \\
&= \log \left\{ \left(\frac{1}{2\pi\sqrt{1-\theta^2}} \right)^n \right\} + \log \left\{ \exp \left(\frac{-1}{2(1-\theta^2)} \sum_{i=1}^n (x_i^2 - 2\theta x_i y_i + y_i^2) \right) \right\} \\
&= \log \left\{ \left((2\pi)^2 (1-\theta^2) \right)^{-\frac{n}{2}} \right\} - \frac{1}{2(1-\theta^2)} \sum_{i=1}^n (x_i^2 - 2\theta x_i y_i + y_i^2) \\
&= -\frac{n}{2} \log \left((2\pi)^2 (1-\theta^2) \right) - \frac{1}{2(1-\theta^2)} \sum_{i=1}^n (x_i^2 - 2\theta x_i y_i + y_i^2).
\end{aligned} \tag{3.20}$$

Para ejemplificar de manera gráfica cómo se van estimando los parámetros de una cópula tomemos el siguiente ejemplo: supongamos que tenemos una cópula normal bivariada con medida de asociación $\tau(-0.8)$, cuya primera marginal se distribuye de manera normal asimétrica con parámetros $Sn(x_i = 3, \text{omega} = 2, \text{alpha} = -5)$ y la segunda marginal tiene una distribución Weibull con parámetros $W(\text{shape} = 1.5, \text{scale} = 1)$, para estimar los parámetros de esta cópula tomemos una muestra aleatoria e independiente de tamaño $n = 10$, por el método exacto de máxima verosimilitud, estimamos los parámetros de las marginales y de la cópula, mediante los parámetros estimados podemos graficar una función de densidad, ahora procedemos a graficar los contornos de esta función, los cuales se muestran en la Figura 3.1. Se procede de manera análoga para los casos cuando $n = 50$ y $n = 100$. Observamos que mientras más grande se va haciendo n , más exacta va siendo la estimación de parámetros.

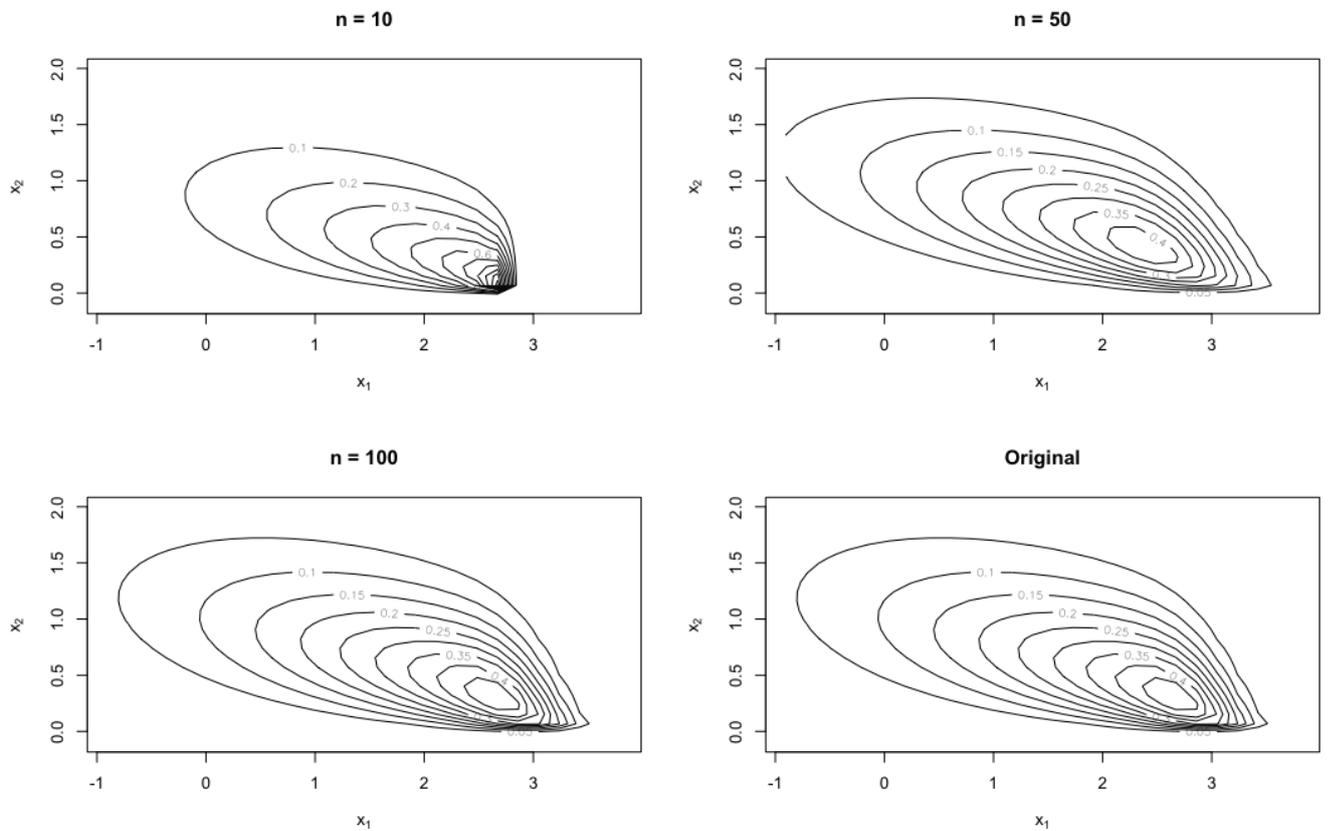


Figura 3.1: Contornos de las funciones de densidad estimadas con muestras aleatorias e independientes de tamaño 10, 50 y 100 de la cópula normal bivariada con marginales normal asimétrica y Weibull con parámetros $Sn(x_i = 3, \omega = 2, \alpha = -5)$ y $W(\text{shape} = 1.5, \text{scale} = 1)$ respectivamente, con medida de asociación $\tau(-0.8)$.

Existe un equilibrio entre la capacidad de cálculo y la eficiencia relativa asintótica de los estimadores. En general, el método IFM tiene una buena eficiencia, excepto posiblemente por una dependencia extrema cerca de los límites de Fréchet-Hoeffding, los cuales se dieron en la Definición 2.2. La estimación computacional del método IFM es mucho más sencilla, especialmente cuando el número total de parámetros excede el rango de 15 a 20. Cuando la estimación de máxima verosimilitud es factible computacionalmente, el procedimiento de dos etapas proporciona un buen punto de partida [11].

En el siguiente capítulo abordaremos el tema de la Curva ROC para posteriormente aplicar la teoría ya desarrollada sobre la cópula normal bivariada.

Capítulo 4

Curva ROC Acumulativa/Dinámica

4.1. Introducción

Este capítulo se inicia con una introducción histórica de la curva ROC, la cual hace referencia a que se desarrolló en la segunda guerra mundial, posteriormente se da una breve introducción al análisis de supervivencia en el cual se definen las funciones de supervivencia y de riesgo, y finalmente se muestran varios modelos paramétricos que son de gran importancia, esto se hace con el único propósito de introducir a manera de ejemplo la aplicación de la curva ROC al análisis de supervivencia.

La curva ROC en el presente capítulo se da en términos de la Sensibilidad Acumulativa y la Especificidad Dinámica, bajo estas definiciones se demuestra que la curva es monótona, posteriormente se estudian a detalle los 3 casos extremos de la curva ROC, para ello es necesario saber las condiciones necesarias y suficientes para que una cópula alcance la cota inferior y superior de Frechét-Hoeffding. Finalmente, se define matemáticamente el área debajo de la curva o AUC, por sus siglas en inglés (Area Under the Curve), de la curva ROC y se da su interpretación.

El lector que este interesado en los temas relacionados con la curva ROC le recomendamos consultar [9],[13], [14] y [24]. Para el análisis de supervivencia recomendamos [4].

4.2. Introducción Histórica a la Curva ROC

A principios de la década de 1940, el radar seguía siendo una tecnología secreta que se estaba desarrollando. El comando de combate británico logró interceptar bombarderos alemanes, en parte, basado en señales de radar muy primitivas. Un borrón en una pantalla de radar podrían ser bombarderos dirigidos a través del Canal de la Mancha o podría ser ruido aleatorio, tal vez una bandada de gaviotas o simplemente nada.

Aquellos encargados de tomar decisiones basadas en datos de radares tempranos enfrentaron un problema difícil en la clasificación binaria, las cuales se detallarán a continuación.

Dar una orden positiva para que la Real Fuerza Aérea Británica ataque podía consumir recursos preciosos. Los pilotos se agotaban y el combustible de aviación no podía ser substituido mientras que los

barcos alemanes bloqueaban el país. Entonces, enviando combatientes cuando no había bombarderos, es decir, un falso positivo, tuvo costos evidentemente significativos, por lo que era imposible responder a cada desenfoque en un radar con una orden positiva.

Por otro lado, una orden negativa era hacer nada, permanecer en el suelo, esto podía tener consecuencias catastróficas. Si una imagen de radar fuera realmente un escuadrón de bombarderos alemanes, dar un falso negativo, les permitiría atacar a Londres sin oposición alguna. Para evaluar estos sistemas de radares primitivos alguien bastante brillante inventó una metodología muy inteligente conocida como la Curva Característica Operativa del Receptor o Curva ROC, por sus siglas en inglés (Receiver Operating Characteristic). La Curva en sí misma permitió a los tomadores de decisiones elegir cuándo preparar sus aviones o cuando dejarlos en el suelo basándose en su mejor estimación del costo relativo de los dos tipos de errores. La falsa alarma o falso positivo y la falta de reacción ante un ataque real, el falso negativo.

Tenemos, por lo tanto, dos condiciones, estas son: Bombarderos y Ruido, este último podrían ser gaviotas. Así mismo, tenemos dos clasificaciones, a saber, Positivo, lo cual significaba enviar a los combatientes y Negativo, lo cual significaba hacer nada.

El método mediante el cual se calculaba la Curva ROC era que, a todas las imágenes de radar se les asignaba una puntuación numérica. En este caso, se asignaba el número correspondiente al área máxima del desenfoque que aparecía en la pantalla del radar. La condición final, el Bombardero o el Ruido (gaviotas), también se calculaba después de recopilarse los datos. Ahora, el comando de combate podía decidir si lo aviones permanecerían en el suelo o atacaban.

El comando de combate podría establecer un umbral que esté por arriba de cualquier puntuación anteriormente obtenida para dar una clasificación positiva a toda puntuación arriba de dicho umbral, por lo tanto, cada clasificación sería clasificada como negativa. El caso inverso a este era decidir que cada imagen justificaba una clasificación positiva. Esto implicaría establecer el umbral, para dar una clasificación positiva, por debajo de cada puntaje, es decir, por debajo del mínimo de los puntajes obtenidos.

En realidad, el umbral que divide a las clasificaciones en positivas y negativas está situado en medio, aún más, dependiendo del posicionamiento de dicho umbral dependen también los costos relativos de falsos negativos versus falsos positivos. También notamos que si mantenemos el método de puntuación constante pero cambiando el posicionamiento del umbral conduce a clasificaciones diferentes. Es ahí donde surge la Curva ROC, la cual se va dibujando mediante la identificación de un umbral dado.

La suma del área debajo de la Curva ROC sobre todos los posibles umbrales da el área debajo de la curva, a la cual se le denomina AUC, por sus siglas en inglés (Area Under the Curve). Maximizar dicha área sigue siendo actualmente una medida muy buena para medir la potencia de un modelo de clasificación binaria. El área debajo la Curva ROC ayudó a ganar la batalla a Gran Bretaña y ha seguido siendo la forma más utilizada para optimizar los sistemas de clasificación binaria, pero su inventor permanece anónimo.

En el momento en que las referencias a la Curva ROC y al AUC comenzaron a aparecer en escritos sin clasificar a principios de los años 1950, nadie se molestó en darle crédito, sin embargo, los primeros documentos donde aparecen la Curva ROC y el AUC permanecen clasificados.

A continuación damos una breve introducción al análisis de supervivencia, esto se hace con el único propósito de introducir, a manera de ejemplo, la aplicación de la curva ROC al análisis de supervivencia.

4.3. Análisis de supervivencia

El análisis de supervivencia estaba muy relacionado anteriormente con la mortalidad pero la supervivencia puede ser cualquier medida que se evalúe hasta que se cumpla un evento, por ejemplo, el tiempo que ocurre hasta que se tenga un evento. Para ilustrar esto pongamos el siguiente ejemplo: supongamos que dejáramos pasar cierto tiempo t hasta que un paciente pueda desarrollar el evento cáncer, esto nos lleva de manera natural a preguntarnos ¿cuál será el tiempo hasta que el evento cáncer ocurra?, esta pregunta nos dirige a otra, ¿cuál es la probabilidad de que una persona sobreviva al evento cáncer, o en su caso, que no presente dicho evento?. Visto de este modo, existe un evento llamado mortalidad, el cual puede afectar a la prevalencia de la incidencia, donde el término prevalencia se refiere a cuantas personas hay que tienen cierta enfermedad en el total de la población y la incidencia se refiere a la razón de los casos de la enfermedad frente a los no casos, esto quiere decir que si una persona no muere hasta cierto tiempo t , sobrevive.

En el análisis de supervivencia es de gran interés ver cómo se comporta el evento, por lo cual se tendrá que hacer un seguimiento de este. Al hacer el seguimiento existen varias cuestiones a tomar en cuenta, por ejemplo la censura, la cual menciona que se puede hacer un seguimiento pero existen diversas razones por las cuales el seguimiento no se completa, como ejemplos podrían ser las siguientes: se pierde el sujeto de estudio, lo que significa que ya no se puede estudiar, otro ejemplo es que el sujeto se pudo haber muerto y se quería saber si se desarrollaba cierta enfermedad, etc. Entonces existe una censura siempre y cuando el evento no fuera la mortalidad del sujeto, ya que entonces no sería censura. Otro tipo de censura es la administrativa, la cual se refiere a que durante todo el tiempo de seguimiento al sujeto, no ocurrió el evento, como por ejemplo, sobrevivió a no tener el evento. Hacemos la mención de que en este trabajo de tesis no se trabajará con la censura.

Es de interés calcular la probabilidad de que un sujeto sobreviva hasta tener cierto evento. Esto se puede hacer mediante las siguientes dos funciones:

- 1.- **Función de supervivencia:** Probabilidad de tener el evento dado que no se ha tenido hasta cierto tiempo de seguimiento.
- 2.- **Función Riesgo:** Riesgo instantáneo de tener un evento, es decir, dado que no se ha tenido hasta un tiempo t .

A continuación se analizarán a detalle estas funciones.

4.3.1. Función de Supervivencia

Antes de dar la definición de función de supervivencia es necesario definir la variable aleatoria de supervivencia, la cual se muestra a continuación.

Definición 4.1 Una variable aleatoria T es una variable aleatoria de supervivencia si el resultado observado t de T se encuentra en el intervalo $[0, \infty)$.

Los datos de supervivencia se denotan por t_1, t_2, \dots, t_n y son observaciones de una variable aleatoria de supervivencia.

Dado que T es una variable aleatoria, se denotará por f a su función de densidad de probabilidad y por F a su función de distribución acumulativa, así obtenemos:

$$F(t) = \Pr [T \leq t] = \int_0^t f(u) du.$$

Se procede a dar entonces una de las definiciones más importantes de este capítulo.

Definición 4.2 La función de supervivencia o función de fiabilidad, la cual se denotará por la letra S , se define para todos los valores t por $S(t) = 1 - F(t)$.

Lo cual significa que

$$S(t) = \Pr [T > t] = \int_t^{\infty} f(u) du.$$

La función de supervivencia representa la probabilidad de que un individuo (o individuos) sobreviva desde el punto de origen hasta algún punto mayor que t . Es decir, esta función cuantifica la capacidad que tiene un individuo o grupo de individuos para cumplir con éxito su función, en un intervalo de tiempo $(0, t]$.

La función de supervivencia tiene las siguientes propiedades:

- 1) Monótona, decreciente y continua.
- 2) $S(0) = 1$ y $\lim_{t \rightarrow \infty} S(t) = 0$.
- 3) $S(t) = \Pr \{T > t\} = \int_t^{\infty} f(u) du \quad 0 \leq t \leq \infty$ donde f es la función de densidad asociada a T .

4.3.2. Función de Riesgo

La función de riesgo, $h(t)$, es utilizada para expresar el riesgo de muerte al tiempo t y es obtenida de la probabilidad de que un individuo muera en el tiempo t dado que ha sobrevivido hasta ese tiempo. La fuerza de mortalidad representa la tasa instantánea de mortalidad para un individuo que sobrevive al tiempo t , es decir, considere la probabilidad de que la variable aleatoria asociada con el tiempo de supervivencia del individuo, T , está entre t y $t + \delta t$ condicionada a que T sea mayor o igual a t , por lo que tenemos $\Pr [t \leq T < t + \delta t \mid T \geq t]$. Esta probabilidad condicional es entonces expresada como una probabilidad por unidad de tiempo dividida por el intervalo de tiempo δt , por lo tanto, la función de riesgo es el valor del límite de esta cantidad cuando el valor δt tiende a cero, es decir

$$h(t) = \lim_{\delta t \rightarrow 0} \left\{ \frac{\Pr [t \leq T < t + \delta t \mid T \geq t]}{\delta t} \right\}. \quad (4.1)$$

De la ecuación anterior se obtiene que $h(t) \delta t$ es aproximadamente la probabilidad de que un individuo muera en el intervalo $(t, t + \delta t)$ condicionada a que el individuo haya sobrevivido al tiempo t . También notamos que de la definición de la función de riesgo se obtiene

$$\Pr [t \leq T < t + \delta t \mid T \geq t] = \frac{\Pr [t \leq T < t + \delta t]}{\Pr [T \geq t]}.$$

Lo cual es equivalente a

$$\frac{F(t + \delta t) - F(t)}{S(t)}$$

donde $F(t)$ es la función de distribución de T .

Sustituyendo estos resultados en la Ecuación (4.1), obtenemos

$$h(t) = \lim_{\delta t \rightarrow 0} \left\{ \frac{F(t + \delta t) - F(t)}{\delta t} \right\} \frac{1}{S(t)}$$

donde

$$\lim_{\delta t \rightarrow 0} \left\{ \frac{F(t + \delta t) - F(t)}{\delta t} \right\}$$

es la definición de la derivada de $F(t)$ con respecto a t , la cual es $f(t)$, dado que $F(t) = \int_0^t f(u) du$, por lo tanto

$$h(t) = \frac{f(t)}{S(t)}$$

y dado a que $S(t) = \Pr[T > t] = \int_t^\infty f(u) du$, se tiene que

$$h(t) = -\frac{d}{dt} \{\log S(t)\} \quad (4.2)$$

y por lo tanto

$$S(t) = \exp\{-H(t)\} \quad (4.3)$$

donde

$$H(t) = \int_0^t h(u) du. \quad (4.4)$$

La función $H(t)$ es llamada la función acumulada de riesgo.

Observamos que, por la Ecuación (4.4), la función acumulada de riesgo puede ser obtenida de la función de supervivencia, a saber

$$H(t) = -\log S(t).$$

Por último, una función $(h(x))$, es de riesgo sí y solo sí satisface las siguientes propiedades:

- 1) $h(x) \geq 0$ para toda x .
- 2) $\int_0^\infty h(x) dx = \infty$.

4.3.3. Modelos

En el análisis de supervivencia, los métodos tradicionalmente usados pueden dividirse en tres grupos, los cuales son: paramétricos, semiparamétricos y no paramétricos. En éste trajo de tesis utilizaremos únicamente los métodos paramétricos.

Existen numerosos modelos paramétricos que son utilizados en el análisis de supervivencia. Entre estos modelos, existen varias distribuciones que toman un papel primordial dada su utilidad. Ejemplos de estas distribuciones, las cuales detallaremos a continuación son: la exponencial, la Weibull y la log-logística.

Modelo Exponencial

La función de densidad de una distribución exponencial, con parámetro λ , es

$$f(t) = \lambda \exp\{-\lambda t\} \quad t \geq 0, \lambda > 0$$

su función de distribución es

$$F(t) = 1 - \exp\{-\lambda t\} \quad t \geq 0$$

entonces, la función de supervivencia esta dada por

$$S(t) = 1 - F(t) = \exp\{-\lambda t\} \quad t \geq 0$$

por lo tanto, podemos obtener su función de riesgo, la cual es

$$h(t) = \frac{f(t)}{S(t)} = \frac{\lambda \exp\{-\lambda t\}}{\exp\{-\lambda t\}} = \lambda \quad t \geq 0$$

finalmente, la función acumulada de riesgo es

$$H(t) = \lambda t \quad t \geq 0.$$

Propiedad de no Memoria

Una característica importante de la distribución exponencial es la propiedad de no memoria, la cual está dada de la siguiente manera:

Sea T una variable aleatoria con distribución exponencial, entonces

$$\Pr [T > a + t | T > a] = \Pr [T > t] \quad \text{para toda } a > 0 \text{ y } t > 0.$$

La interpretación de la propiedad de no memoria es la siguiente: Si T es el tiempo de vida de un individuo, la anterior igualdad establece que la probabilidad de que el individuo sobreviva al tiempo $a + t$ dado que ha sobrevivido al tiempo a , es igual a la probabilidad de que sobreviva al tiempo t .

La condición anterior puede escribirse también en términos de la función de supervivencia, es decir

$$\begin{aligned} \Pr [T > a + t \mid T > a] &= \frac{\Pr [T > a + t]}{\Pr [T > a]} \\ &= \frac{S(a + t)}{S(a)} \end{aligned} \quad (4.5)$$

$$\begin{aligned} &= \frac{\exp \{-\lambda (a + t)\}}{\exp \{-\lambda (a)\}} \\ &= \exp \{-\lambda (t)\} \\ &= \Pr [T > t] \\ &= S(t). \end{aligned} \quad (4.6)$$

Por lo tanto, de las Ecuaciones (4.5) y (4.6) obtenemos la siguiente expresión

$$S(a + t) = S(a) S(t).$$

Esta propiedad caracteriza a la función exponencial.

Modelo Weibull

Una desventaja que presenta la distribución exponencial es que no sirve como modelo para tiempos de vida en los que la función de riesgo no es una función constante, es decir, que varía con el tiempo. En cambio, las funciones de riesgo, de la familia de distribuciones Weibull, pueden ser crecientes y decrecientes. Dado que en la práctica se pueden encontrar tendencias crecientes y decrecientes, como ejemplo sería el riesgo de mortalidad por parto: a temprana edad es alto, a mediana edad es bajo porque la mujer ya está preparada y a edad avanzada es alto.

Supongamos que T tiene una distribución Weibull con parámetro de escala λ y parámetro de forma α , es decir, $T \sim W\left(\frac{1}{\lambda}, \alpha\right)$.

La función de densidad de una distribución Weibull es

$$f(t) = \alpha \lambda^\alpha t^{\alpha-1} \exp \{-(\lambda t)^\alpha\} \quad \text{con } \alpha, \lambda > 0.$$

Se deduce que la distribución exponencial es un caso particular de la distribución Weibull para el caso donde $\alpha = 1$.

Su función de supervivencia es

$$S(t) = \exp \{-(\lambda t)^\alpha\}.$$

Notamos que si $\alpha \rightarrow \infty$, se tiene un tiempo de vida constante, pues $S(t) \rightarrow 0$.

Su función de riesgo es

$$h(t) = \frac{f(t)}{S(t)} = \alpha\lambda(\lambda t)^{\alpha-1}.$$

Notamos que la función de riesgo es creciente para $\alpha > 1$ y $h(t) \rightarrow \infty$ cuando $t \rightarrow \infty$. Es decreciente para el caso en que $\alpha < 1$ y $h(t)$ tiende asintóticamente a 0 cuando $t \rightarrow \infty$.

Modelo Log-Logístico

Cualquier distribución definida para $t \in [0, \infty)$ puede servir como una distribución de supervivencia. De manera más general, podemos partir de una variable aleatoria, W , con una distribución estándar en $(-\infty, \infty)$ y generar una familia de distribuciones de supervivencia mediante la introducción de cambios de forma y escala de la siguiente manera

$$Y = \log T = \alpha + \sigma W.$$

Una distribución de esta forma es la Log-Logística.

T tiene distribución Log-logística sí y solo sí

$$Y = \log T = \alpha + \sigma W.$$

Donde W tiene una distribución Logística estándar, con función de densidad

$$f_W(w) = \frac{e^w}{(1+e^w)^2}$$

su función de distribución es

$$F_W(w) = \frac{e^w}{1+e^w}$$

con función de supervivencia

$$S_W(w) = \frac{1}{1+e^w}.$$

Cambiando las variables a T encontramos que la función de supervivencia Log-logística es

$$S(t) = \frac{1}{1+(\lambda t)^p}.$$

Donde $\alpha = -\log \lambda$ y $p = 1/\sigma$.

Ahora ocupamos la Ecuación (4.2) para obtener la función de riesgo de la siguiente manera

$$h(t) = -\frac{d}{dt} \{\log S(t)\} = \frac{\lambda p (\lambda t)^{p-1}}{1+(\lambda t)^p}.$$

Por último, observamos que la función de riesgo de la distribución Log-logística cumple con las siguientes 2 propiedades:

- 1) Monótona decreciente desde ∞ si $p < 1$.
- 2) Monótona decreciente de λ si $p = 1$.

A continuación mencionamos una de las aplicaciones que tiene la curva ROC al análisis de supervivencia.

4.4. Aplicación de la Curva ROC al Análisis de Supervivencia

La curva ROC, por sus siglas en inglés (Receiver Operating Characteristic), cuantifica la capacidad de un indicador diagnóstico para discriminar entre enfermos y sanos. Para ejemplificar la curva ROC tomemos en cuenta la Figura 4.1, en donde se muestra la distribución del indicador en los sanos y la distribución del indicador en los enfermos.

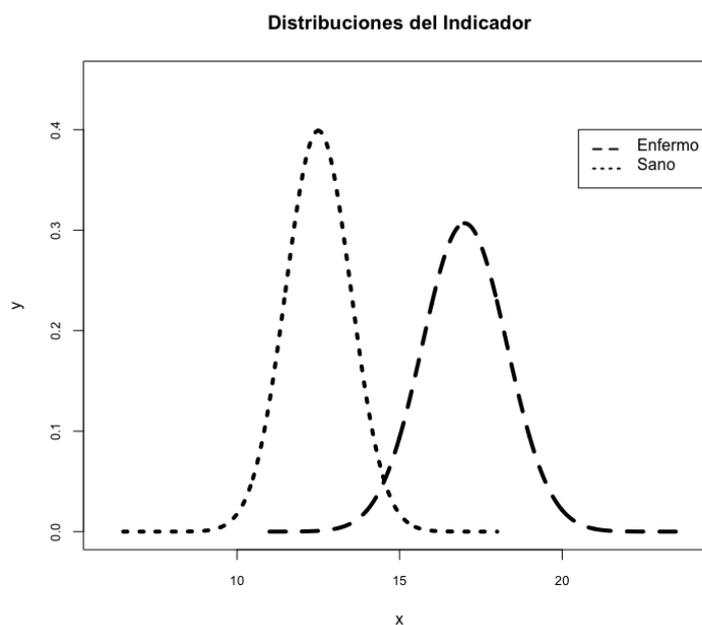


Figura 4.1: Distribuciones del indicador en los sanos y en los enfermos.

Como ambas poblaciones comparten ciertos valores del indicador, la clasificación no puede ser perfecta. Cualquier umbral o punto de corte clasificará erróneamente algunos de los casos de la zona traslapada.

La curva ROC indica para cada posible umbral la proporción de positivos en ambas poblaciones, en ordenadas para los enfermos y en abscisas para los sanos, de esta forma, la línea vertical en el cuadrado unitario $[0, 1] \times [0, 1]$ indica la Sensibilidad o proporción de positivos en los enfermos, por lo tanto nos interesa que sea alta, por otra parte, la línea horizontal inferior en el cuadrado unitario $[0, 1] \times [0, 1]$ indica la proporción de positivos en los sanos, por lo tanto nos conviene que sea baja, al complemento

de esto se le denomina Especificidad. Notamos que la Curva ROC se encuentra ubicada en el cuadrado unitario $[0, 1] \times [0, 1]$.

Si miramos la Figura 4.1 notamos que diferentes umbrales o puntos de corte implicarán diferentes proporciones de errores tanto para sanos como para enfermos, por ejemplo, situemos el umbral diagnóstico en el extremo izquierdo, tal y como lo muestra la Figura 4.2, y si declaramos como positivos a todos los casos por encima de este punto tendremos un cien por ciento de positivos tanto para los enfermos como para los sanos. En la gráfica de la Curva ROC, este valor se sitúa en el extremo superior derecho tal y como lo muestra la Figura 4.4, observamos también que la línea vertical de esta Figura se corresponde con la Figura 4.2 con el área debajo de la curva de los enfermos a la derecha del umbral e indica que un cien por ciento de los enfermos son correctamente clasificados como positivos, por otra parte, esto también indica que el cien por ciento de los sanos también están a la derecha del umbral y también son clasificados como positivos pero ahora erróneamente, en resumen, este punto de corte clasifica bien a todos los enfermos pero mal a todos los sanos. Notamos también que cuando el umbral se va desplazando hacia la derecha, como en la Figura 4.3, la Curva ROC mostrará como varían estas proporciones, disminuirá el error en los sanos a cambio de empeorar el acierto en los enfermos, es decir, baja un error pero sube otro, en concreto, veremos que primero sube la Especificidad y luego más tarde bajará la Sensibilidad.

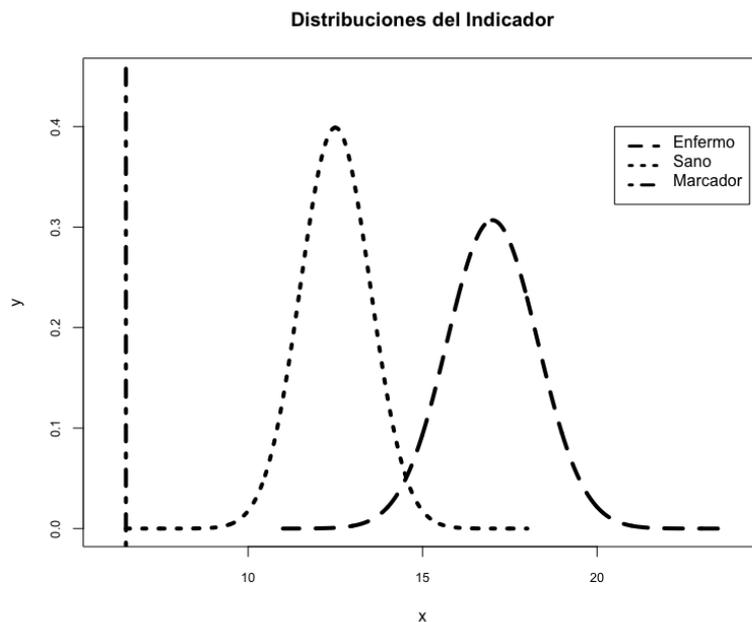


Figura 4.2: Umbral del diagnóstico en el extremo izquierdo.

Esto motiva a dar la definición de Sensibilidad y Especificidad. Definimos a la Sensibilidad como a la probabilidad de que los pacientes se clasifiquen como positivos dado que están enfermos y a la Especificidad como a la probabilidad de que los pacientes se clasifiquen como negativos dado que están sanos.

Al desplazar el umbral hacia la derecha vemos como baja la proporción de positivos en los sanos, al inicio, el acierto en los enfermos sigue al cien por ciento pero luego empieza a cambiar, de esta forma, vamos consiguiendo tener menos positivos en los sanos pero al precio de tener algunos negativos en

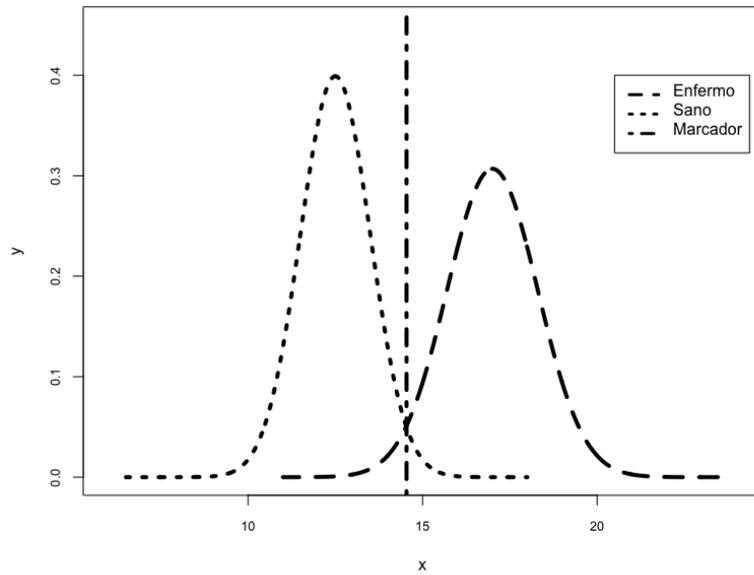


Figura 4.3: Umbral del diagnóstico en el cruce de las distribuciones.

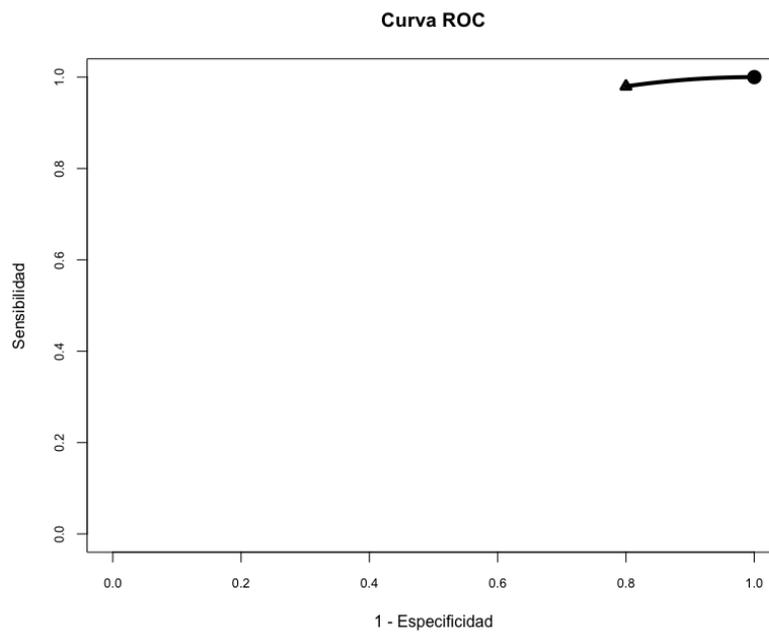


Figura 4.4: Inicio de la gráfica de la curva ROC.

los enfermos. Así, llegaremos al punto donde la Especificidad sea igual a la Sensibilidad, es decir, donde se clasifica bien a la misma proporción de enfermos que de sanos. Posterior a ese punto, si se sigue desplazando el umbral hacia la derecha lo que descende más rápido es la sensibilidad o proporción de positivos en los enfermos y se va aproximando al punto en el que todos los casos serían declarados negativos, por lo que ambas proporciones de positivos serían cero, magnífico para los sanos pero pésimos para los enfermos. Así, la Curva ROC muestra simultáneamente la Sensibilidad y Especificidad, o dicho de otra forma, dibuja para todos los posibles umbrales los pares de proporciones de positivos en ambas poblaciones. En la Figura 4.5 se muestra la imagen de la Curva ROC completa.

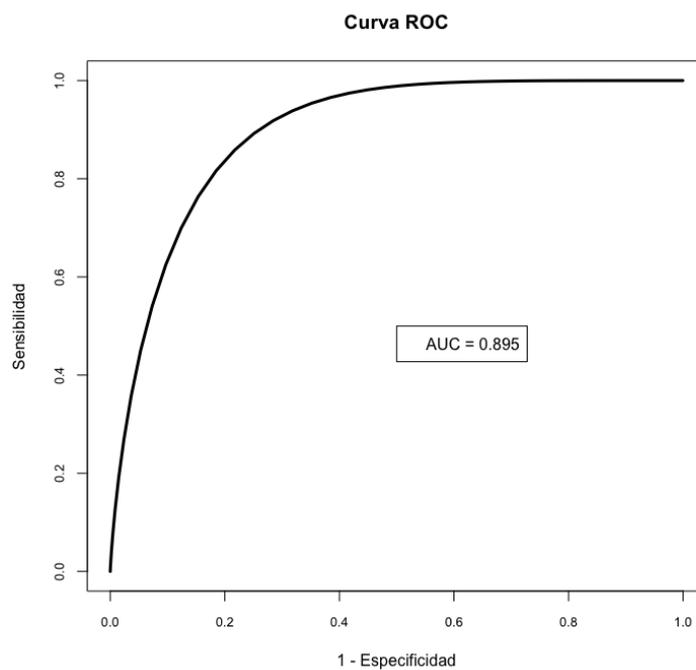


Figura 4.5: Curva ROC.

Si el indicador diagnóstico discriminara muy bien a los sanos de los enfermos, las dos distribuciones casi no se superpondrían y la Curva ROC correspondiente cubriría un área muy grande, esta área bajo la Curva ROC resume la capacidad de un indicador y es conocida como AUC, por sus siglas en inglés (Area Under the Curve). A medida que más se solapan las curvas más se parecen los resultados de sanos y enfermos, por lo tanto, la capacidad discriminativa del indicador es muy baja, lo cual implica que el valor del AUC es cada vez más bajo. Como observación notamos que el caso cuando las gráficas se traslapan completamente, el valor del indicador es independiente del estado del paciente y esto implica que el valor del AUC es de 0.5, lo cual significa que el indicador no aporta información, por otra parte, también tenemos los casos donde la Curva ROC alcanza su valor máximo y mínimo. Estos tres casos serán estudiados detalladamente en el presente capítulo.

4.5. Curva ROC Acumulativa/Dinámica

Antes de definir formalmente la Curva ROC es necesario que veamos algunos conceptos, definiciones y notaciones.

Comencemos definiendo a T como el tiempo de supervivencia, sea M la variable que denota la prueba del diagnóstico, es decir, M es el marcador de valor continuo y sea c un umbral tal que $c \in (-\infty, \infty)$. Usamos la notación $T \leq t$ para indicar que el paciente ha tenido el evento antes del tiempo t , es decir, ha presentado la enfermedad (o falleció) antes del tiempo t y $T > t$ si el paciente ha presentado la enfermedad (o murió) después del tiempo t , por lo tanto, denotamos a los sanos (o vivos) con $T > t$ y a los enfermos (o muertos) con $T \leq t$.

Con base en lo anterior se darán ahora las Definiciones de la Sensibilidad Acumulativa y la Especificidad Dinámica.

Definición 4.3 *La Sensibilidad Acumulativa y Especificidad Dinámica son funciones que dependen del tiempo t y están dadas de la siguiente manera*

$$\text{Sensibilidad}^{\mathbf{C}}(c, t) = \Pr [M > c | T \leq t] \quad (4.7)$$

$$\text{Especificidad}^{\mathbf{D}}(c, t) = \Pr [M \leq c | T > t]. \quad (4.8)$$

Donde M denota un marcador de valor escalar y c es un valor dado, con $c \in (-\infty, \infty)$.

Usaremos los superíndices \mathbf{C} y \mathbf{D} para denotar a la Sensibilidad Acumulativa y a la Especificidad Dinámica respectivamente.

A los pacientes que hayan presentando cierta enfermedad se les llamará Casos y aquellos pacientes sanos se les denominará como Controles. Notamos que para cualquier tiempo fijo t la población entera se clasifica ya sea como un caso o como un control. El interés científico radica en discriminar entre sujetos que mueren antes de un tiempo dado t y los que sobreviven más allá de t . Es por esta razón que se le denomina Sensibilidad Acumulativa y Especificidad Dinámica.

Analizando a detalle la Definición 4.3 observamos que la Sensibilidad Acumulativa

$$\text{Sensibilidad}^{\mathbf{C}}(c, t) = \Pr [M > c | T \leq t]$$

mide la fracción esperada de sujetos con un marcador mayor que c entre la subpoblación de individuos que mueren antes del tiempo t , mientras que la Especificidad Dinámica

$$\text{Especificidad}^{\mathbf{D}}(c, t) = \Pr [M \leq c | T > t]$$

mide la fracción esperada de sujetos con un marcador menor o igual que c entre los que sobreviven más allá del tiempo t .

Las funciones de Sensibilidad y Especificidad reciben otro nombre, a continuación veremos cuales son estos.

Dado un umbral $c \in (-\infty, \infty)$, un paciente se clasifica como positivo si $M > c$ y negativo en caso contrario. Se sigue que la tasa de verdaderos positivos, TP , por sus siglas en inglés, o la *Sensibilidad*, se define como

$$TP_t^{\mathbf{C}}(c) = \text{Sensibilidad}^{\mathbf{C}}(c, t) = \Pr[M > c | T \leq t] \quad (4.9)$$

y la tasa de falsos positivos, FP , por sus siglas en inglés, o $1 - \text{Especificidad}$, se define como

$$FP_t^{\mathbf{D}}(c) = 1 - \text{Especificidad}^{\mathbf{D}}(c, t) = 1 - \Pr[M \leq c | T > t] \quad (4.10)$$

esta última función surge por lo siguiente:

$$\begin{aligned} FP_t^{\mathbf{D}}(c) &= 1 - \text{Especificidad}^{\mathbf{D}}(c, t) \\ &= 1 - \Pr[M \leq c | T > t] \\ &= 1 - \frac{\Pr[M \leq c, T > t]}{\Pr[T > t]} \\ &= \frac{\Pr[T > t] - \Pr[M \leq c, T > t]}{\Pr[T > t]} \\ &= \frac{\Pr[M > c, T > t]}{\Pr[T > t]} \\ &= \Pr[M > c | T > t] \\ &= p \end{aligned} \quad (4.11)$$

con $p \in [0, 1]$.

Es importante notar que hemos usado la Ley de Probabilidad Total, la cual nos dice lo siguiente

$$\Pr[Y > y] = \Pr[X \leq x, Y > y] + \Pr[X > x, Y > y].$$

Con los resultados ya obtenidos podemos definir a la Curva $ROC_t^{\mathbf{C/D}}(p)$, la cual evalúa el poder discriminatorio y diagnóstico del marcador a través de la clasificación correcta utilizando la siguiente función

$$ROC_t^{\mathbf{C/D}}(p) = TP_t^{\mathbf{C}}\left(\left[FP_t^{\mathbf{D}}\right]^{-1}(p)\right) \quad (4.12)$$

para $p \in [0, 1]$ y

$$\left[FP_t^{\mathbf{D}}\right]^{-1}(p) = \inf_c \{c : FP_t^{\mathbf{D}}(c) \leq p\}. \quad (4.13)$$

Analizando a detalle la Función (4.12) obtenemos lo siguiente

$$\begin{aligned}
 ROC_t^{C/D}(p) &= TP_t^C \left([FP_t^D]^{-1}(p) \right) \\
 &= TP_t^C(c^*) \\
 &= Sensibilidad^C(c^*, t)
 \end{aligned} \tag{4.14}$$

para $p \in [0, 1]$ y $c^* = \inf_c \{c : FP_t^D(c) \leq p\}$.

La Ecuación (4.11) nos permite reescribir a la Ecuación (4.13) de la siguiente manera

$$\begin{aligned}
 [FP_t^D]^{-1}(p) &= \inf_c \{c : FP_t^D(c) \leq p\} \\
 &= \inf_c \{c : \Pr[M > c | T > t] \leq p\}
 \end{aligned} \tag{4.15}$$

donde $p \in [0, 1]$.

Esto nos permite ver más claramente el funcionamiento de la función $[FP_t^D]^{-1}(p)$ y nos ayuda a comprender mejor la Ecuación (4.12). Por lo tanto, podemos ver a la Curva $ROC_t^{C/D}$ como la relación TP_t^C vs FP_t^D , la cual puede ser evaluada a través del conjunto de parejas ordenadas

$$\begin{aligned}
 \{(p, ROC_t^{C/D}(p)) ; p \in [0, 1]\} &= \left\{ \left(FP_t^D(c), TP_t^C \left([FP_t^D]^{-1}(p) \right) \right) ; p \in [0, 1], c \in (-\infty, \infty) \right\} \\
 &= \left\{ \left(1 - Especificidad^D(c, t), Sensibilidad^C(c, t) \right) ; t, c \in (-\infty, \infty) \right\}.
 \end{aligned} \tag{4.16}$$

Hacemos la aclaración de que la curva $ROC_t^{C/D}(p)$ esta definida para un t fijo.

Por último, usamos la notación

$$AUC(t) = \int_0^1 ROC_t^{C/D}(p) dp \tag{4.17}$$

para denotar el área bajo la Curva ROC Acumulativa-Dinámica para el tiempo t . La interpretación de esta ecuación será detallada en la Sección 4.6.

De la Definición 4.3 y del análisis realizado anteriormente se tiene que

$$Sensibilidad^C(c, t) = \Pr[M > c | T \leq t] \tag{4.18}$$

$$\text{Especificidad}^D(c, t) = \Pr[M \leq c | T > t]. \quad (4.19)$$

De la Ley de Probabilidad Total se obtienen expresiones para $\Pr[T \leq t]$ y $\Pr[M \leq c]$ como a continuación se indica.

Para $\Pr[T \leq t]$ se tiene lo siguiente

$$\Pr[T \leq t] = \Pr[M \leq c, T \leq t] + \Pr[M > c, T \leq t]$$

por lo tanto

$$\Pr[M > c, T \leq t] = \Pr[T \leq t] - \Pr[M \leq c, T \leq t]. \quad (4.20)$$

Análogamente para $\Pr[M \leq c]$ se obtiene

$$\Pr[M \leq c] = \Pr[M \leq c, T \leq t] + \Pr[M \leq c, T > t]$$

por lo tanto

$$\Pr[M \leq c, T > t] = \Pr[M \leq c] - \Pr[M \leq c, T \leq t]. \quad (4.21)$$

En base a la Ley de Probabilidad Total se obtienen las siguientes expresiones para la Sensibilidad y la Especificidad

$$\begin{aligned} \text{Sensibilidad}^C(c, t) &= \Pr[M > c | T \leq t] \\ &= \frac{\Pr[M > c, T \leq t]}{\Pr[T \leq t]} \\ &= \frac{\Pr[T \leq t] - \Pr[M \leq c, T \leq t]}{\Pr[T \leq t]} \\ &= 1 - \frac{\Pr[M \leq c, T \leq t]}{\Pr[T \leq t]}. \end{aligned} \quad (4.22)$$

$$\begin{aligned}
\text{Especificidad}^D(c, t) &= \Pr [M \leq c | T > t] \\
&= \frac{\Pr [M \leq c, T > t]}{\Pr [T > t]} \\
&= \frac{\Pr [M \leq c] - \Pr [M \leq c, T \leq t]}{\Pr [T > t]}. \tag{4.23}
\end{aligned}$$

A continuación se demostrará la monotonicidad de las funciones de Sensibilidad y Especificidad.

Monotonicidad de la curva ROC Acumulativa/Dinámica

Antes de demostrar que las funciones de Sensibilidad y Especificidad son funciones monótonas con respecto a c y t se recordarán algunos conceptos acerca de las funciones monótonas.

Sea $\Pi \subseteq \mathbb{R}$, entonces la función $f : \Pi \rightarrow \mathbb{R}$ se dice que es creciente en Π cuando $\pi_1, \pi_2 \in \Pi$, con $\pi_1 \leq \pi_2$, entonces $f(\pi_1) \leq f(\pi_2)$. La función f es llamada estrictamente creciente en Π si $\pi_1, \pi_2 \in \Pi$, con $\pi_1 < \pi_2$, entonces $f(\pi_1) < f(\pi_2)$. Similarmente, la función $g : \Pi \rightarrow \mathbb{R}$ es llamada decreciente en Π cuando $\pi_1, \pi_2 \in \Pi$, con $\pi_1 \leq \pi_2$, entonces $g(\pi_2) \leq g(\pi_1)$. La función g es llamada estrictamente decreciente en Π si $\pi_1, \pi_2 \in \Pi$, con $\pi_1 < \pi_2$, entonces $g(\pi_2) < g(\pi_1)$.

Si una función es creciente o decreciente en Π se dice que es monótona en Π . Análogamente, si una función es estrictamente creciente o estrictamente decreciente en Π se dice que es estrictamente monótona en Π .

Sea la función $f : \Pi \rightarrow \mathbb{R}$ creciente en Π , definamos ahora a la función ϕ como $\phi := -f$, dado a que f es creciente, es decir, si $\pi_1, \pi_2 \in \Pi$, con $\pi_1 \leq \pi_2$, entonces $f(\pi_1) \leq f(\pi_2)$, lo cual implica que $-f(\pi_2) \leq -f(\pi_1)$ o equivalentemente $\phi(\pi_2) \leq \phi(\pi_1)$, concluyendo de esta manera que si una función f es creciente en Π , entonces la función $\phi := -f$ es decreciente en Π . Similarmente, si una función g es decreciente, entonces la función $\psi = -g$ es creciente.

Es posible generalizar el caso anterior tal y como se muestra a continuación.

Sea f una función creciente en Π , definamos ahora a la función φ como $\varphi := k \cdot f$, donde $k \in \mathbb{R}$. Dado a que f es creciente, es decir, si $\pi_1, \pi_2 \in \Pi$, con $\pi_1 \leq \pi_2$, entonces $f(\pi_1) \leq f(\pi_2)$. Para poder determinar el comportamiento de la función φ es necesario analizar dos casos.

1. **Caso 1** ($k \geq 0$)
 $k \cdot f(\pi_1) \leq k \cdot f(\pi_2)$. Por lo tanto, la función es creciente.
2. **Caso 2** ($k < 0$)
 $k \cdot f(\pi_1) > k \cdot f(\pi_2)$. Por lo tanto, la función es decreciente.

Un resultado análogo al anterior es que si f es una función monótona, entonces la función $f + k$, donde $k \in \mathbb{R}$, es monótona en el mismo sentido.

Teniendo estos resultados se procede a demostrar la monotonicidad de las funciones de Sensibilidad y Especificidad.

De la Definición de la función de Sensibilidad (4.22) se obtiene la siguiente expresión

$$\begin{aligned} \text{Sensibilidad}^{\mathcal{C}}(c, t) &= \Pr[M > c | T \leq t] \\ &= 1 - \frac{\Pr[M \leq c, T \leq t]}{\Pr[T \leq t]}. \end{aligned} \quad (4.24)$$

Supongamos ahora que t es un valor dado fijo y c varía en el intervalo $(\infty, -\infty)$. Dado que la ecuación $\Pr[M \leq c, T \leq t]$ representa la probabilidad de que la variable aleatoria M toma un valor menor o igual que c y que la variable aleatoria T toma un valor menor o igual a t , se tiene por lo tanto que esta ecuación es una función monótonamente creciente cuando la variable c toma valores cada vez más grandes, y dado a que t es fijo se obtiene que la función $\Pr[T \leq t] = k$, donde k es una constante tal que $0 < k \leq 1$. Entonces la Ecuación (4.24) se puede escribir de la siguiente manera

$$\begin{aligned} \text{Sensibilidad}^{\mathcal{C}}(c, t) &= \Pr[M > c | T \leq t] \\ &= 1 - \frac{\Pr[M \leq c, T \leq t]}{\Pr[T \leq t]} \\ &= 1 - \frac{1}{k} \Pr[M \leq c, T \leq t]. \end{aligned} \quad (4.25)$$

Como se mencionó anteriormente, la función $\Pr[M \leq c, T \leq t]$ es monótonamente creciente cuando la variable c toma valores cada vez más grandes, de igual manera la función $\frac{1}{k} \Pr[M \leq c, T \leq t]$ es monótonamente creciente, por lo tanto, $-\frac{1}{k} \Pr[M \leq c, T \leq t]$ es monótonamente decreciente, dado a que $-\frac{1}{k} < 0$. Obteniendo como conclusión que la función

$$\text{Sensibilidad}^{\mathcal{C}}(c, t) = 1 - \frac{1}{k} \Pr[M \leq c, T \leq t]$$

es monótonamente decreciente con respecto a c .

Ahora, de la definición de Especificidad (4.23), se obtiene la siguiente expresión

$$\begin{aligned}
\text{Especificidad}^{\mathbf{D}}(c, t) &= \Pr [M \leq c | T > t] \\
&= \frac{\Pr [M \leq c, T > t]}{\Pr [T > t]}.
\end{aligned} \tag{4.26}$$

Supongamos ahora que t es un valor dado fijo y c varía en el intervalo $(\infty, -\infty)$, entonces podemos escribir la ecuación anterior como

$$\begin{aligned}
\text{Especificidad}^{\mathbf{D}}(c, t) &= \Pr [M \leq c | T > t] \\
&= \frac{\Pr [M \leq c, T > t]}{\Pr [T > t]} \\
&= \frac{1}{k} \Pr [M \leq c, T > t].
\end{aligned} \tag{4.27}$$

Notamos que $\Pr [M \leq c, T > t]$ es monótonamente creciente cuando la variable c toma valores cada vez más grandes, por lo tanto, la $\text{Especificidad}^{\mathbf{D}}(c, t)$ es monótonamente creciente con respecto a c .

Ahora procedemos a estudiar a detalle los 3 casos extremos de la curva ROC Acumulativa/Dinámica, para ello es necesario saber las condiciones necesarias y suficientes para que una cópula alcance la cota inferior y superior de Frechét-Hoeffding, así mismo, daremos la representación gráfica de estos 3 casos extremos.

Independencia

Si M y T son variables aleatorias independientes, es decir

$$\Pr [M \leq c, T \leq t] = \Pr [M \leq c] \cdot \Pr [T \leq t].$$

Podemos obtener, entonces, los siguientes resultados:

De la Definición 4.3 y de la Ecuación (4.18) se obtiene

$$\begin{aligned} \text{Sensibilidad}^{\mathbf{C}}(c, t) &= \Pr [M > c | T \leq t] \\ &= \frac{\Pr [M > c, T \leq t]}{\Pr [T \leq t]} \\ &= \Pr [M > c]. \end{aligned} \tag{4.28}$$

De la Definición 4.3 y de la Ecuación (4.19) se obtiene

$$\begin{aligned} \text{Especificidad}^{\mathbf{D}}(c, t) &= \Pr [M \leq c | T > t] \\ &= \frac{\Pr [M \leq c, T > t]}{\Pr [T > t]} \\ &= \Pr [M \leq c]. \end{aligned} \tag{4.29}$$

Entonces, por la Ecuación (4.16), obtenemos la Curva ROC Acumulativa-Dinámica dada por el siguiente conjunto de parejas ordenadas:

$$\begin{aligned} \left\{ \left(p, \text{ROC}_t^{\mathbf{C}/\mathbf{D}}(p) \right) ; p \in [0, 1] \right\} &= \left\{ \left(1 - \text{Especificidad}^{\mathbf{D}}(c, t), \text{Sensibilidad}^{\mathbf{C}}(c, t) \right) ; t, c \in (-\infty, \infty) \right\} \\ &= \left\{ (\Pr [M > c], \Pr [M > c]) , c \in (-\infty, \infty) \right\}. \end{aligned} \tag{4.30}$$

Lo cual sucede sí y sólo sí las variables aleatorias M y T son independientes.

Lo que implica que, cuando las variables son independientes, al dibujar la Curva $\text{ROC}_t^{\mathbf{C}/\mathbf{D}}(p)$ se obtiene la función identidad en el cuadrado unitario.

El dibujo de la Curva $\text{ROC}_t^{\mathbf{C}/\mathbf{D}}(p)$, cuando las variables son independientes, se muestra en la Figura 4.6.

En la Sección 4.4 mostramos un ejemplo con el cual obtuvimos una curva ROC. Basándonos en esa misma dinámica procedemos de manera análoga y damos el siguiente ejemplo:

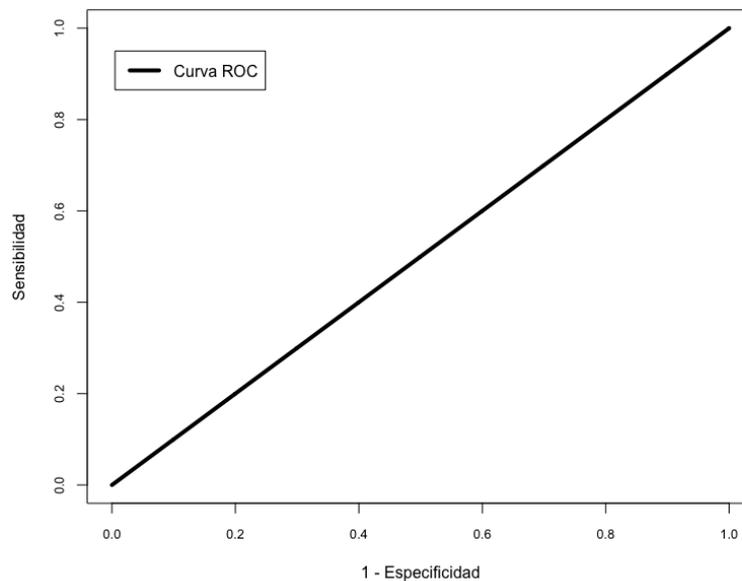


Figura 4.6: Curva $ROC_t^{C/D}(p)$ cuando las variables M y T son independientes.

Si el umbral diagnóstico se sitúa en el extremo izquierdo, tal y como lo muestra la Figura 4.7 y si declaramos como positivos a todos los casos por encima de este punto tendremos un cien por ciento de positivos tanto para los enfermos como para los sanos, ya que ambos comparten la misma distribución. Este valor se sitúa en el extremo superior derecho, ya que es ahí donde inicia su curva $ROC_t^{C/D}(p)$ correspondiente a esta gráfica, tal y como lo muestra la Figura 4.6, observamos también que la línea vertical de esta figura se corresponde con la Figura 4.7 con el área debajo de la curva de los enfermos a la derecha del umbral e indica que un cien por ciento de los enfermos son correctamente clasificados como positivos, por otra parte, esto también indica que el cien por ciento de los sanos también están a la derecha del umbral y también son clasificados como positivos pero ahora erróneamente, en resumen, este punto de corte clasifica bien a todos los enfermos pero mal a todos los sanos. Notamos también que cuando el umbral se va desplazando hacia la derecha, la curva $ROC_t^{C/D}(p)$ mostrará como varían estas proporciones, disminuirá el error en los sanos a cambio de empeorar el acierto en los enfermos pero lo hacen a la misma tasa, es decir, baja un error pero sube otro en la misma proporción, formando así la Curva $ROC_t^{C/D}(p)$, la cual es la función identidad en el cuadrado unitario.

Para obtener el segundo caso extremo de la curva $ROC_t^{C/D}(p)$ es necesario saber cuando se alcanza la cota superior de Fréchet-Hoeffding.

Cota Superior de Fréchet-Hoeffding

Podemos ahora aplicar el Teorema 2.3 para dibujar la curva $ROC_t^{C/D}(p)$ dadas las variables aleatorias M y T . Es importante recordar que este Teorema nos da las condiciones necesarias y suficientes para que

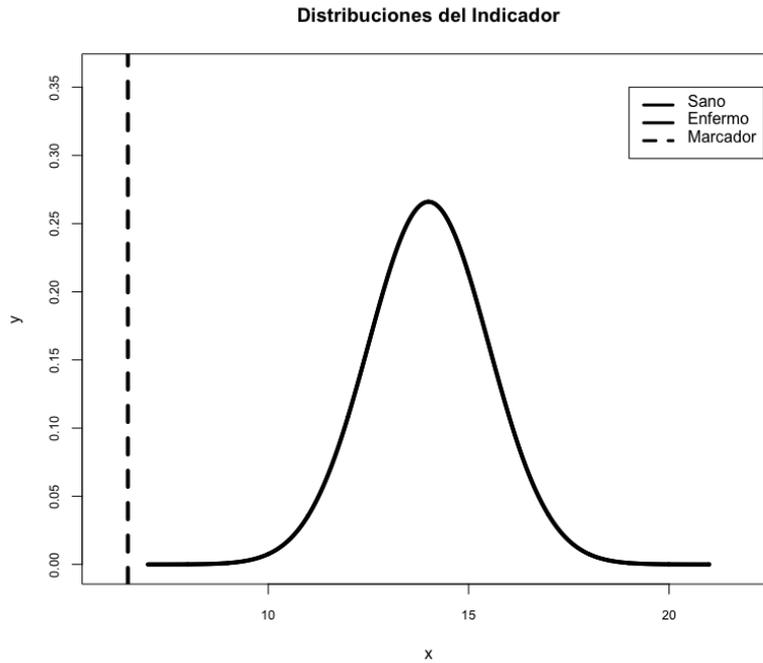


Figura 4.7: Umbral del diagnóstico en el extremo izquierdo.

la Cota Superior de Fréchet-Hoeffding sea alcanzada, por lo tanto, tenemos 2 casos.

Caso 1

$$\Pr [M > c, T \leq t] = 0 \quad (4.31)$$

Caso 2

$$\Pr [M \leq c, T > t] = 0 \quad (4.32)$$

A continuación se estudiarán a detalle estos dos casos para poder, de esta manera, dibujar la Curva $ROC_t^{C/D}(p)$ en cada uno de ellos.

Análisis del Caso 1

Por la Ecuación (4.18) se obtiene

$$\begin{aligned} \text{Sensibilidad}^C(c, t) &= \Pr [M > c | T \leq t] \\ &= \frac{\Pr [M > c, T \leq t]}{\Pr [T \leq t]}. \end{aligned}$$

Ahora, si se cumple la Ecuación (4.31) se tiene que

$$\text{Sensibilidad}^C(c, t) = 0.$$

Por lo tanto, el valor de la Sensibilidad es igual a cero en todos los puntos.

Concluyendo de esta manera que, la primera parte de la Curva ROC Acumulativa-Dinámica, está dada por el siguiente conjunto de parejas ordenadas:

$$\begin{aligned} \left\{ \left(p, ROC_t^{C/D}(p) \right) ; p \in [0,1] \right\} &= \left\{ \left(1 - Especificidad^D(c,t), Sensibilidad^C(c,t) \right) ; t, c \in (-\infty, \infty) \right\} \\ &= \left\{ (1 - \Pr[M \leq c | T > t], 0) ; t, c \in (-\infty, \infty) \right\}. \end{aligned} \quad (4.33)$$

Hacemos la aclaración de que la curva $ROC_t^{C/D}(p)$ esta definida para un t fijo. Lo anterior implica que la primera parte de la Curva $ROC_t^{C/D}(p)$, derivada del Caso 1, esta dada por las parejas ordenadas del conjunto $[0,1] \times \{0\}$, tal y como lo muestra la Figura 4.8.

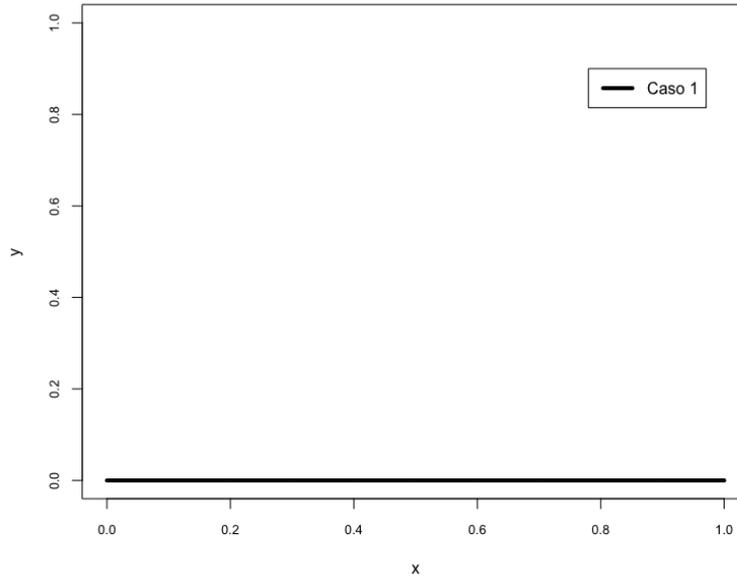


Figura 4.8: Parte de la Curva $ROC_t^{C/D}(p)$ derivada del análisis del Caso 1.

Análisis del Caso 2

Por la Ecuación (4.19) se obtiene

$$\begin{aligned} Especificidad^D(c,t) &= \Pr[M \leq c | T > t] \\ &= \frac{\Pr[M \leq c, T > t]}{\Pr[T > t]}. \end{aligned}$$

Ahora, si se cumple la Ecuación (4.32) se tiene que

$$\text{Especificidad}^D(c, t) = 0.$$

Por lo tanto, el valor de la Especificidad es igual a cero en todos los puntos.

Concluyendo de esta manera que la segunda parte de la Curva $ROC_t^{C/D}(p)$ está dada por el siguiente conjunto de parejas ordenadas:

$$\begin{aligned} \left\{ (p, ROC_t^{C/D}(p)) ; p \in [0, 1] \right\} &= \left\{ (1 - \text{Especificidad}^D(c, t), \text{Sensibilidad}^C(c, t)) ; t, c \in (-\infty, \infty) \right\} \\ &= \left\{ (1, \Pr[M > c | T \leq t]) ; t, c \in (-\infty, \infty) \right\}. \end{aligned} \quad (4.34)$$

Lo que implica que la segunda parte de la Curva $ROC_t^{C/D}(p)$, derivada del Caso 2, son las parejas ordenadas del conjunto $\{1\} \times [0, 1]$, tal y como lo muestra la Figura 4.9.

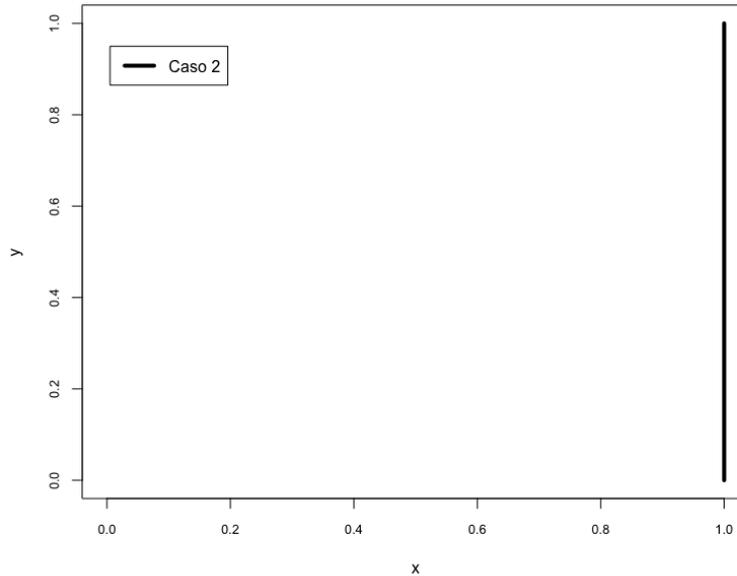


Figura 4.9: Parte de la Curva $ROC_t^{C/D}(p)$ derivada del análisis del Caso 2.

Juntando los Casos 1 y 2 obtenemos la curva $ROC_t^{C/D}(p)$ completa, la cual se muestra en la Figura 4.10.

En la Sección 4.4 mostramos un ejemplo con el cual obtuvimos la curva ROC. Basándonos en esa misma dinámica procedemos de manera análoga y damos el siguiente ejemplo:

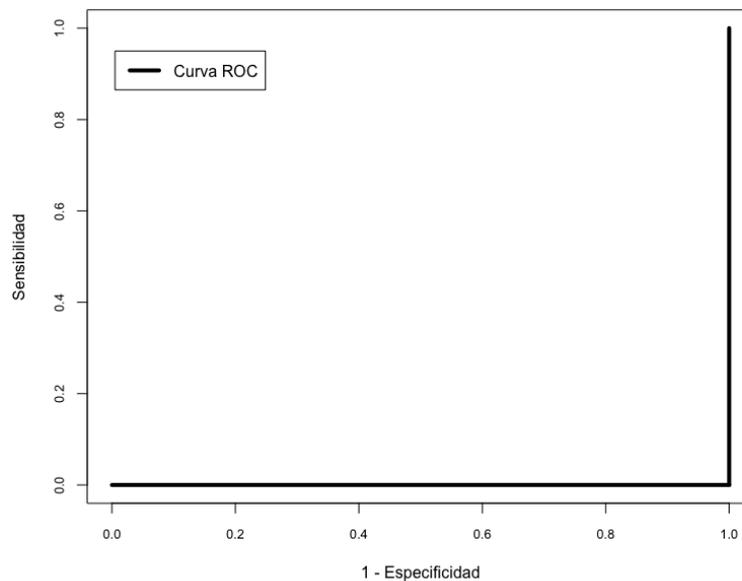


Figura 4.10: Curva $ROC_t^{C/D}(p)$.

Si el umbral diagnóstico se sitúa en el extremo izquierdo, tal y como lo muestra la Figura 4.11, y si declaramos como positivos a todos los casos por encima de este punto tendremos un cien por ciento de positivos tanto para los enfermos como para los sanos. Este valor se sitúa en el extremo superior derecho, ya que es ahí donde inicia su curva $ROC_t^{C/D}(p)$ correspondiente a esta gráfica, tal y como lo muestra la Figura 4.10, observamos también que la línea vertical de esta figura se corresponde con la Figura 4.11 con el área debajo de la curva de los enfermos a la derecha del umbral e indica que un cien por ciento de los enfermos son correctamente clasificados como positivos, por otra parte, esto también indica que el cien por ciento de los sanos también están a la derecha del umbral y también son clasificados como positivos pero ahora erróneamente, en resumen, este punto de corte clasifica bien a todos los enfermos pero mal a todos los sanos. Notamos también que cuando el umbral se va desplazando hacia la derecha, la Curva $ROC_t^{C/D}(p)$ mostrará como varían estas proporciones, disminuirá el acierto en los enfermos sin variar el error en los sanos, en concreto, veremos que primero bajará la Sensibilidad y luego más tarde subirá la Especificidad, formando así la curva $ROC_t^{C/D}(p)$ referente a la Figura 4.10.

Para obtener finalmente el tercer caso extremo de la curva $ROC_t^{C/D}(p)$ es necesario saber las condiciones necesarias y suficientes bajo las cuales una cópula alcanza la cota inferior de Fréchet-Hoeffding.

Cota Inferior de Fréchet-Hoeffding

Para obtener la Curva $ROC_t^{C/D}(p)$ de la cota Inferior W de Fréchet-Hoeffding es necesario un resultado análogo al Teorema 2.3, el cual nos ayudará a realizar nuestro análisis correspondiente. A continuación se enuncia y se demuestra dicho Teorema.

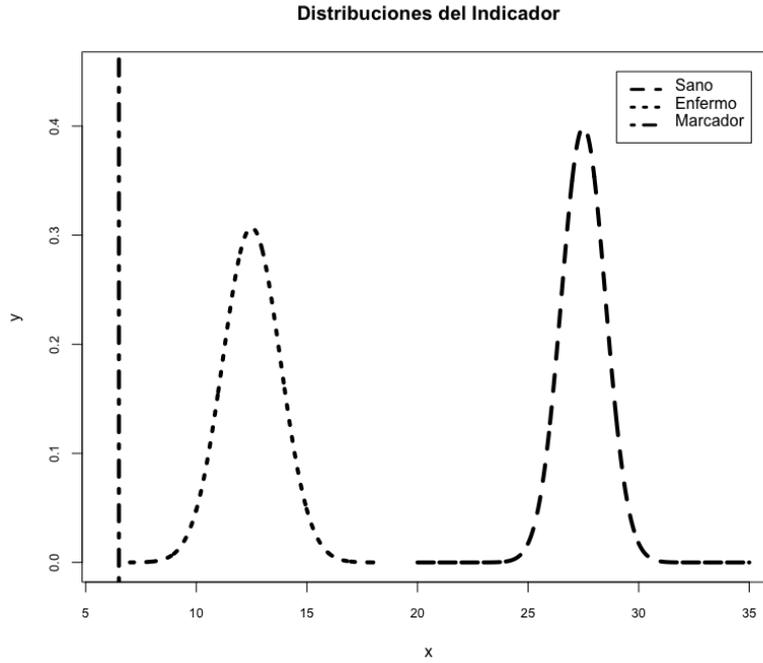


Figura 4.11: Umbral del diagnóstico en el extremo izquierdo.

Teorema 4.1 Sean X y Y variables aleatorias con función de distribución conjunta H , entonces H es igual a su cota inferior de Fréchet-Hoeffding sí y solo sí para todo (x, y) en \mathbb{R}^2 se cumple que $\Pr[X > x, Y > y] = 0$ o $\Pr[X \leq x, Y \leq y] = 0$.

Demostración 4.1.1 De la Ley de Probabilidad Total se tienen las siguientes ecuaciones

$$\Pr[X \leq x] = \Pr[X \leq x, Y \leq y] + \Pr[X \leq x, Y > y]$$

$$\Pr[Y \leq y] = \Pr[X \leq x, Y \leq y] + \Pr[X > x, Y \leq y].$$

Por lo tanto, se tiene lo siguiente

$$\begin{aligned} F(x) &= \Pr[X \leq x] = \Pr[X \leq x, Y \leq y] + \Pr[X \leq x, Y > y] \\ &= H(x, y) + \Pr[X \leq x, Y > y] \end{aligned} \tag{4.35}$$

y

$$\begin{aligned} G(y) &= \Pr[Y \leq y] = \Pr[X \leq x, Y \leq y] + \Pr[X > x, Y \leq y] \\ &= H(x, y) + \Pr[X > x, Y \leq y]. \end{aligned} \tag{4.36}$$

Dado que $H(x, y)$ alcanza su cota inferior, las Ecuaciones (4.35) y (4.36) pueden escribirse de la siguiente manera

$$F(x) = W(F(x), G(y)) + \Pr[X \leq x, Y > y] \quad (4.37)$$

$$G(y) = W(F(x), G(y)) + \Pr[X > x, Y \leq y]. \quad (4.38)$$

Analicemos, entonces, los dos casos existentes dados por el Teorema 2.2.

Caso 1)

$$W(F(x), G(y)) = F(x) + G(y) - 1.$$

Sustituyendo el valor de $W(F(x), G(y))$ en la Ecuación (4.37) se obtiene

$$F(x) = F(x) + G(y) - 1 + \Pr[X \leq x, Y > y]$$

reduciendo términos

$$1 - G(y) = \Pr[X \leq x, Y > y]$$

por lo tanto

$$1 = \Pr[X \leq x, Y > y] + \Pr[Y \leq y]$$

lo que implica que $\Pr[X > x, Y > y] = 0$.

Análogamente, sustituyendo el valor de $W(F(x), G(y))$ en la Ecuación (4.38) se obtiene

$$G(y) = F(x) + G(y) - 1 + \Pr[X > x, Y \leq y]$$

reduciendo términos

$$1 - F(x) = \Pr[X > x, Y \leq y]$$

por lo tanto

$$1 = \Pr[X > x, Y \leq y] + \Pr[X \leq x]$$

de manera similar implica que $\Pr[X > x, Y > y] = 0$.

Caso 2)

$$W(F(x), G(y)) = 0$$

Sustituyendo el valor de $W(F(x), G(y))$ en la Ecuación (4.37) se obtiene

$$F(x) = \Pr[X \leq x] = \Pr[X \leq x, Y > y]$$

por lo tanto

$$\Pr[X \leq x, Y \leq y] = 0.$$

Análogamente, sustituyendo el valor de $W(F(x), G(y))$ en la Ecuación (4.38) se obtiene

$$G(y) = \Pr[Y \leq y] = \Pr[X > x, Y \leq y]$$

por lo tanto

$$\Pr[X \leq x, Y \leq y] = 0.$$

De los Casos 1 y 2 se desprende que la cota inferior de Fréchet-Hoeffding $W(F(x), G(y))$ se alcanza sí y solo sí $\Pr[X > x, Y > y] = 0$ ó $\Pr[X \leq x, Y \leq y] = 0$.

A continuación se estudiarán a detalle estos dos casos para poder, de esta manera, dibujar la Curva $ROC_t^{C/D}(p)$.

Análisis del Caso 1

De la Ecuación (4.23) se tiene que

$$\begin{aligned} \text{Especificidad}^D(c, t) &= \Pr[M \leq c | T > t] \\ &= \frac{\Pr[M \leq c, T > t]}{\Pr[T > t]} \\ &= \frac{\Pr[M \leq c] - \Pr[M \leq c, T \leq t]}{\Pr[T > t]}. \end{aligned} \quad (4.39)$$

Aplicando el Caso 1, es decir, cuando $W(F(x), G(y)) = F(x) + G(y) - 1$, a la Ecuación (4.39) obtenemos

$$\begin{aligned} \text{Especificidad}^D(c, t) &= \frac{\Pr[M \leq c] - \Pr[M \leq c, T \leq t]}{\Pr[T > t]} \\ &= \frac{\Pr[M \leq c] - \Pr[M \leq c] - \Pr[T \leq t] + 1}{\Pr[T > t]} \\ &= \frac{\Pr[T > t]}{\Pr[T > t]} \\ &= 1. \end{aligned} \quad (4.40)$$

Concluyendo de esta manera que, la primera parte de la Curva $ROC_t^{C/D}(p)$ está dada por el siguiente conjunto de parejas ordenadas:

$$\begin{aligned} \left\{ \left(p, ROC_t^{C/D}(p) \right) ; p \in [0,1] \right\} &= \left\{ \left(1 - Especificidad^D(c,t), Sensibilidad^C(c,t) \right) ; t, c \in (-\infty, \infty) \right\} \\ &= \left\{ (0, \Pr[M > c | T \leq t]) ; t, c \in (-\infty, \infty) \right\}. \end{aligned} \quad (4.41)$$

Hacemos la aclaración de que la curva $ROC_t^{C/D}(p)$ esta definida para un t fijo. Lo anterior implica que la primera parte de la Curva $ROC_t^{C/D}(p)$, derivada por el Caso 1, está dada por las parejas ordenadas del conjunto $\{0\} \times [0,1]$, tal y como lo muestra la Figura 4.12.

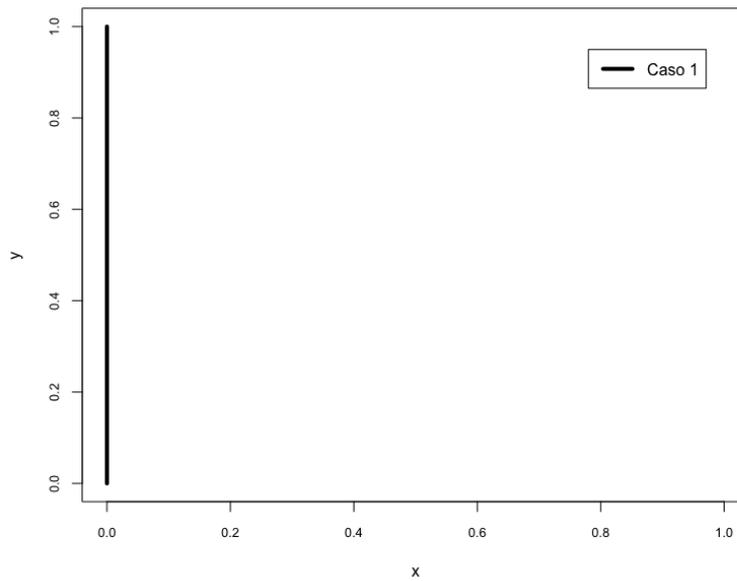


Figura 4.12: Parte de la Curva $ROC_t^{C/D}(p)$ derivada del análisis del Caso 1.

Análisis del Caso 2

De la Ecuación (4.22) se tiene que

$$\begin{aligned}
\text{Sensibilidad}^{\mathbf{C}}(c, t) &= \Pr [M > c | T \leq t] \\
&= \frac{\Pr [M > c, T \leq t]}{\Pr [T \leq t]} \\
&= \frac{\Pr [T \leq t] - \Pr [M \leq c, T \leq t]}{\Pr [T \leq t]} \\
&= 1 - \frac{\Pr [M \leq c, T \leq t]}{\Pr [T \leq t]}. \tag{4.42}
\end{aligned}$$

Aplicando el Caso 2, es decir, cuando $W(F(x), G(y)) = 0$ a la Ecuación (4.42) obtenemos

$$\text{Sensibilidad}^{\mathbf{C}}(c, t) = 1.$$

Concluyendo de esta manera que, la segunda parte de la Curva $ROC_t^{\mathbf{C}/\mathbf{D}}(p)$, está dada por el siguiente conjunto de parejas ordenadas:

$$\begin{aligned}
\left\{ (p, ROC_t^{\mathbf{C}/\mathbf{D}}(p)) ; p \in [0, 1] \right\} &= \left\{ (1 - \text{Especificidad}^{\mathbf{D}}(c, t), \text{Sensibilidad}^{\mathbf{C}}(c, t)) ; t, c \in (-\infty, \infty) \right\} \\
&= \left\{ (1 - \Pr [M \leq c | T > t], 1) ; t, c \in (-\infty, \infty) \right\}. \tag{4.43}
\end{aligned}$$

Lo que implica que la segunda parte de la Curva $ROC_t^{\mathbf{C}/\mathbf{D}}(p)$, derivada por el Caso 2, son las parejas ordenadas del conjunto $[0, 1] \times \{1\}$, tal y como lo muestra la Figura 4.13.

Juntando los Casos 1 y 2 obtenemos la curva $ROC_t^{\mathbf{C}/\mathbf{D}}(p)$ completa, la cual se muestra en la Figura 4.14.

En la Sección 4.4 mostramos un ejemplo con el cual obtuvimos la curva ROC. Basándonos en esa misma dinámica procedemos de manera análoga y damos el siguiente ejemplo:

Si el umbral diagnóstico se sitúa en el extremo izquierdo, tal y como lo muestra la Figura 4.15, y si declaramos como positivos a todos los casos por encima de este punto tendremos un cien por ciento de positivos tanto para los enfermos como para los sanos. Este valor se sitúa en el extremo superior derecho, ya que es ahí donde inicia su curva $ROC_t^{\mathbf{C}/\mathbf{D}}(p)$ correspondiente a esta gráfica, tal y como lo muestra la Figura 4.14, observamos también que la línea vertical de esta figura se corresponde con la Figura 4.15 con el área debajo de la curva de los enfermos a la derecha del umbral e indica que un cien por ciento de los enfermos son correctamente clasificados como positivos, por otra parte, esto también indica que el cien por ciento de los sanos también están a la derecha del umbral y también son clasificados como positivos pero ahora erróneamente, en resumen, este punto de corte clasifica bien a todos los enfermos pero mal a todos los sanos. Notamos también que cuando el umbral se va desplazando hacia la derecha, la Curva $ROC_t^{\mathbf{C}/\mathbf{D}}(p)$ mostrará como varían estas proporciones, disminuirá el error en los sanos sin variar el acierto en los enfermos, en concreto, veremos que primero sube la Especificidad y luego más tarde bajará la Sensibilidad, formando así la curva $ROC_t^{\mathbf{C}/\mathbf{D}}(p)$ referente a la Figura 4.14.

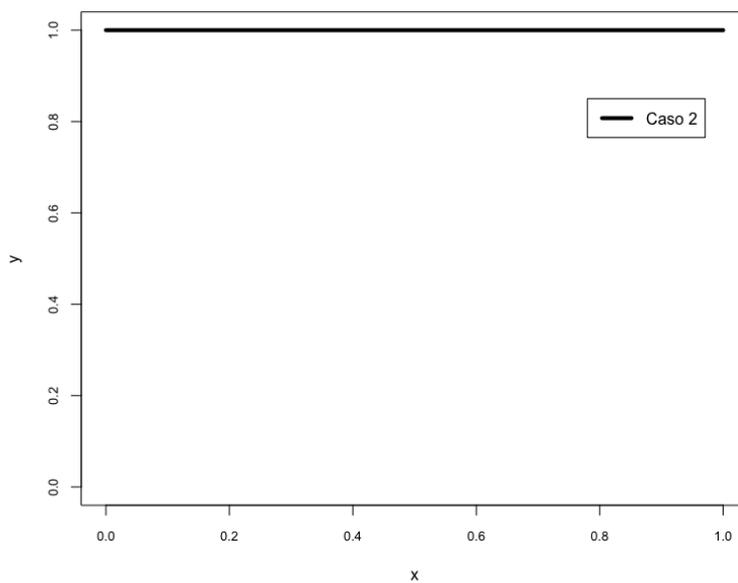


Figura 4.13: Parte de la Curva $ROC_t^{C/D}(p)$ derivada del análisis del Caso 2.

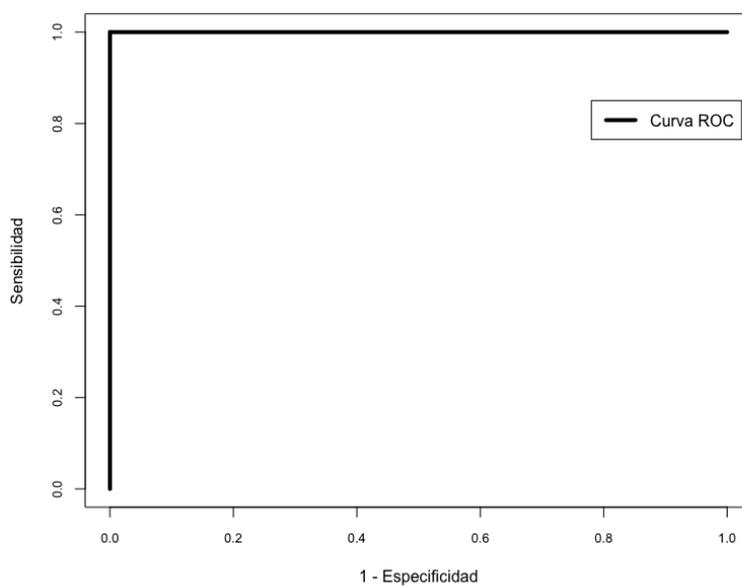


Figura 4.14: Curva $ROC_t^{C/D}(p)$.

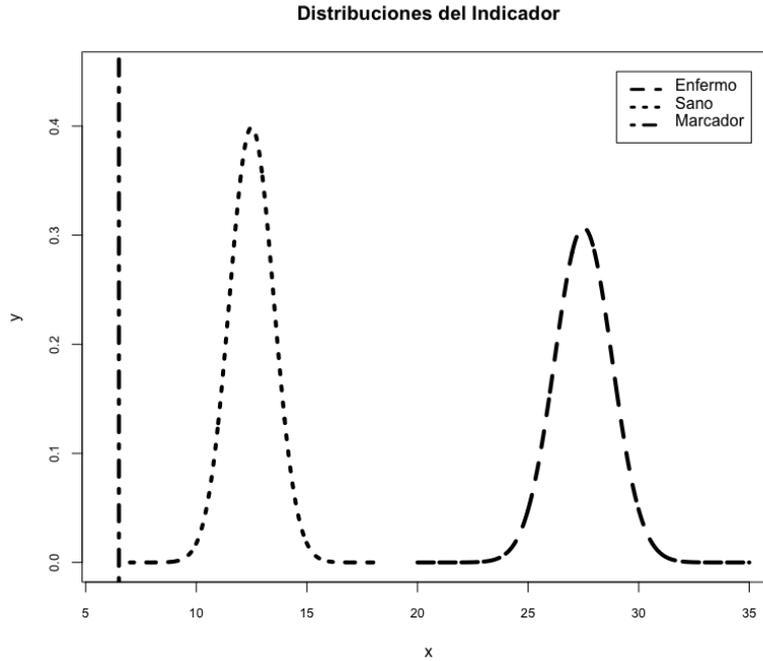


Figura 4.15: Umbral del diagnóstico en el extremo izquierdo.

Concluyendo de esta manera el análisis de los 3 casos extremos de la Curva $ROC_t^{C/D}(p)$. Prodecemos ahora a dar la interpretación del área debajo de la curva (AUC) de la curva $ROC_t^{C/D}(p)$.

4.6. Interpretación del Área Debajo de la Curva (AUC)

La Ecuación (4.17) muestra la definición matemática del AUC, la cual denota el área bajo la Curva ROC Acumulativa-Dinámica para el tiempo t .

$$AUC(t) = \int_0^1 ROC_t^{C/D}(p) dp.$$

Recordemos que la curva $ROC_t^{C/D}(p)$ cuantifica la capacidad de un indicador diagnóstico para discriminar entre enfermos y sanos (vivos y muertos) y nos indica para cada posible umbral la proporción de positivos en ambas poblaciones, en ordenadas para los enfermos y en abscisas para los sanos. Si el indicador diagnóstico discriminara muy bien a los sanos de los enfermos, la Curva $ROC_t^{C/D}(p)$ correspondiente cubriría un área muy grande, recordando que la Curva $ROC_t^{C/D}(p)$ se encuentra ubicada en el cuadrado unitario $[0, 1] \times [0, 1]$, esta área bajo la Curva $ROC_t^{C/D}(p)$ resume la capacidad de un indicador, lo cual indica que para valores de AUC cercanos a 1, como por ejemplo la Figura 4.16, sugieren una alta precisión diagnóstica del indicador, por otra parte, si los valores del AUC son cada vez más pequeños nos indica que la precisión diagnóstica del indicador es deficiente, es decir, no nos proporciona información suficiente para discriminar entre enfermos y sanos.

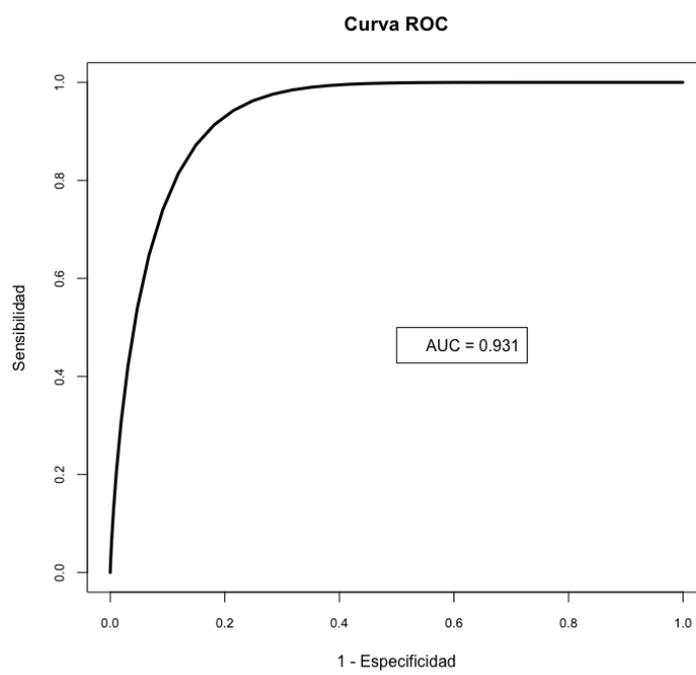


Figura 4.16: Área debajo de la curva $ROC_t^{C/D}(p)$, $AUC=0.931$.

Capítulo 5

Sensibilidad Incidente y Especificidad Dinámica

5.1. Introducción

En el estudio de la Curva ROC se han ido desarrollado varias definiciones de Sensibilidad y Especificidad, una de estas definiciones surge del siguiente planteamiento: supongamos que los casos estudiados para definir a la Sensibilidad la cual es dependiente del tiempo, son casos incidentes, es decir, en los que $T = t$, a esto se le denomina Sensibilidad Incidente, aún más, vamos a suponer que los controles de los marcadores sean dinámicos y definidos para el tiempo t , es decir, como aquellos sujetos con $T > t$. Es importante hacer la aclaración que no estamos tomando un valor fijo t , más bien, hacemos variar t hasta que $t = T$, a esto se le denomina Especificidad Dinámica.

Dada las nuevas definiciones de Sensibilidad y Especificidad se muestra un ejemplo en el cual vamos a suponer que la función de distribución del marcador es $F(c)$ y la función de distribución del tiempo es $G(t)$, y ocupando la cópula normal bivariada obtenemos las expresiones explícitas de la Sensibilidad Incidente y la Especificidad Dinámica. Finalmente mostramos un método para obtener la función de densidad conjunta de una cópula normal bivariada cuando una de las variables es discreta.

5.2. Sensibilidad Incidente y Especificidad Dinámica.

Como hemos visto anteriormente, a T se le ha definido como el tiempo de supervivencia, a M la variable que denota la prueba del diagnóstico y c un umbral tal que $c \in (-\infty, \infty)$. También usamos la notación $T \leq t$ para indicar que el paciente ha tenido el evento antes del tiempo t , es decir, ha presentado la enfermedad antes del tiempo t y $T > t$ si el paciente ha presentado la enfermedad después del tiempo t .

En este capítulo, vamos a suponer que los casos estudiados para definir a la Sensibilidad, son casos incidentes, es decir, en los que $T = t$, vamos a representar este caso mediante la notación $\partial(t)$, a esto se le denominará como Sensibilidad Incidente, aún más, vamos a suponer que los controles de los marcadores sean dinámicos y definidos para el tiempo t , es decir, como aquellos sujetos con $T > t$. Es importante hacer la aclaración que nosotros hacemos variar t hasta que $t = T$ y es debido a ello que se le denomina

Especificidad Dinámica [14].

Como notación usaremos los superíndices **I** y **D** para denotar a las nuevas Definiciones de Sensibilidad Incidente y Especificidad Dinámica respectivamente, las cuales se darán a continuación.

Definición 5.1 *La Sensibilidad Incidente y la Especificidad Dinámica son funciones que dependen del tiempo t y están dadas de la siguiente manera*

$$\begin{aligned} \text{Sensibilidad}^I(c, t) &= \Pr [M > c | T = t] \\ &= \Pr [M > c | \partial(t)]. \end{aligned} \quad (5.1)$$

$$\text{Especificidad}^D(c, t) = \Pr [M \leq c | T > t]. \quad (5.2)$$

Donde M denota un marcador de valor escalar y c es un valor dado, con $c \in (-\infty, \infty)$.

Utilizando estas nuevas definiciones podemos notar que para tiempos tempranos, es decir, para $t < T$, un sujeto puede jugar el papel de control, pero luego juega el papel del caso cuando $t = T$.

Analizando a detalle la definición pasada, observamos que la Sensibilidad Incidente

$$\text{Sensibilidad}^I(c, t) = \Pr [M > c | \partial(t)]$$

mide la fracción esperada de sujetos con un marcador mayor que c entre la subpoblación de individuos que mueren al tiempo t , mientras que la Especificidad Dinámica

$$\text{Especificidad}^D(c, t) = \Pr [M \leq c | T > t]$$

mide la fracción esperada de sujetos con un marcador menor o igual que c entre los que sobreviven más allá del tiempo t .

5.2.1. Distribución del Marcador $F(c)$ y del Tiempo $G(t)$

Con base en las nuevas definiciones de Sensibilidad Incidente y Especificidad Dinámica, ilustraremos un ejemplo en el cual vamos a suponer que $F(c)$ y $G(t)$ son funciones de distribución del marcador y del tiempo respectivamente, entonces, ocupando la Cópula Normal Bivariada obtenemos una serie de interesantes resultados los cuales serán detallados a continuación.

De la Ley de Probabilidad Total, obtenemos

$$\Pr [\partial(t)] = \Pr [M \leq c, \partial(t)] + \Pr [M > c, \partial(t)].$$

Ocupando la ecuación anterior y la Ecuación (5.1) se obtiene que

$$\begin{aligned}
\text{Sensibilidad}^I(c, t) &= \Pr [M > c | \partial(t)] \\
&= \frac{\Pr [M > c, \partial(t)]}{\Pr [\partial(t)]} \\
&= \frac{\Pr [\partial(t)] - \Pr [M \leq c, \partial(t)]}{\Pr [\partial(t)]} \\
&= 1 - \frac{\Pr [M \leq c, \partial(t)]}{\Pr [\partial(t)]} \\
&= 1 - \Pr [M \leq c | \partial(t)].
\end{aligned} \tag{5.3}$$

Ocupando la Definición 2.6 y la Ecuación (2.34) obtenemos

$$\begin{aligned}
\text{Sensibilidad}^I(c, t) &= 1 - \Pr [M \leq c | \partial(t)] \\
&= 1 - \Phi \left(\frac{\Phi^{-1}(F(c)) - \rho \Phi^{-1}(G(t))}{\sqrt{1 - \rho^2}} \right) \\
&= \Phi \left(\frac{\rho \Phi^{-1}(G(t)) - \Phi^{-1}(F(c))}{\sqrt{1 - \rho^2}} \right).
\end{aligned} \tag{5.4}$$

En la Ecuación (5.4) se ocupó la siguiente propiedad: $\Phi(-x) = 1 - \Phi(x)$, donde $\Phi(x) = \Pr [X < x]$ para $X \sim N(0, 1)$.

Ocupando nuevamente la Ley de Probabilidad Total obtenemos que

$$\Pr [T > t] = \Pr [M \leq c, T > t] + \Pr [M > c, T > t].$$

Por lo tanto, la Especificidad Dinámica, Ecuación (5.2), está dada de la siguiente manera

$$\begin{aligned}
\text{Especificidad}^D(c, t) &= \Pr[M \leq c | T > t] \\
&= \frac{\Pr[M \leq c, T > t]}{\Pr[T > t]} \\
&= \frac{\Pr[T > t] - \Pr[M > c, T > t]}{\Pr[T > t]} \\
&= 1 - \frac{\Pr[M > c, T > t]}{\Pr[T > t]} \\
&= 1 - \frac{S_2^N[c, t; \rho]}{1 - \Pr[T \leq t]} \\
&= 1 - \frac{S_2^N[c, t; \rho]}{1 - G(t)}. \tag{5.5}
\end{aligned}$$

Ocupamos la notación $S_2^N[x, y; \rho]$ para denotar a $\Pr[X > x, Y > y]$ cuando $[X, Y]$ tienen una distribución normal bivariada unitaria con correlación ρ .

En resumen, si $F(c)$ y $G(t)$ son funciones de distribución del marcador y del tiempo respectivamente y ocupando la cópula normal bivariada obtenemos los siguientes resultados:

$$\text{Sensibilidad}^I(c, t) = \Phi\left(\frac{\rho \Phi^{-1}(G(t)) - \Phi^{-1}(F(c))}{\sqrt{1 - \rho^2}}\right) \tag{5.6}$$

y

$$\text{Especificidad}^D(c, t) = 1 - \frac{S_2^N[c, t; \rho]}{1 - G(t)}. \tag{5.7}$$

En el estudio de la Sensibilidad Incidente y la Especificidad Dinámica ocupando la cópula normal bivariada, surge una pregunta interesante, ¿qué pasa cuando una de las variables aleatorias es discreta?. La respuesta a esta pregunta se da a continuación, mostrando como obtener la densidad conjunta cuando una de las variables es discreta [16].

5.3. Función de Densidad Conjunta Cuando una de las Variables es Discreta.

5.3.1. Asignación de una Distribución al Indicador Binario

Supongamos que tenemos dos variables que se correlacionan X y Y , donde X es discreta y Y es continua. El resultado discreto representa la muerte o la supervivencia del paciente y el resultado continuo

corresponde al tiempo t . Supongamos que X tiene una distribución F_X , es decir, $X \sim F_X$ y que Y tiene una distribución G_Y , es decir, $Y \sim G_Y$. Supongamos ahora que X tiene 3 valores distintos, los cuales son: 0, p y 1 (que posiblemente representan puntajes ordinales o estados nominales). Para modelar la distribución conjunta $H_{X,Y}$ de X y Y , necesitamos de la variable Y^* , la cual tiene una distribución G_{Y^*} , es decir, $Y^* \sim G_{Y^*}$, donde Y^* es la variable latente continua no observada subyacente a X , tal que

$$X = \begin{cases} 0, & \text{si } Y^* \in (-\infty, 0) \\ p, & \text{si } Y^* \in [0, 1) \\ 1, & \text{si } Y^* \in [1, \infty) . \end{cases} \quad (5.8)$$

Donde $0 < p < 1$ son 3 umbrales. Ahora, asumiendo que la distribución conjunta de $H_{Y^*,Y}$ de Y^* y Y está determinada por una cópula normal, obtenemos lo siguiente

$$H_{Y^*,Y}(y^*, y) = \Phi_2 \left(\Phi^{-1}(F_{Y^*}(y^*)), \Phi^{-1}(F_Y(y)) \right).$$

Lo anterior se deriva de la Definición 2.4. Cabe mencionar que la elección de la cópula es arbitraria pero en éste trabajo de Tesis adoptamos la cópula normal debido a sus propiedades ya estudiadas.

Para obtener la distribución conjunta de X y Y es necesario hacer un pequeño análisis previamente, el cual comienza notando que cuando la variable $X = 0$, la expresión $\Pr[X = 0, Y \leq y]$ es igual, por la Ecuación (5.8), a la expresión $\Pr[Y^* \leq 0, Y \leq y]$, por lo tanto, tenemos lo siguiente

$$\begin{aligned} \Pr[X = 0, Y \leq y] &= \Pr[Y^* \leq 0, Y \leq y] \\ &= H_{Y^*,Y}(0, y) . \end{aligned}$$

Cuando la variable $X = p$, la expresión $\Pr[X = p, Y \leq y]$ es igual, por la Ecuación (5.8), a la expresión $\Pr[Y^* \leq 1, Y \leq y] - \Pr[Y^* \leq 0, Y \leq y]$, por lo tanto, tenemos lo siguiente

$$\begin{aligned} \Pr[X = p, Y \leq y] &= \Pr[Y^* \leq 1, Y \leq y] - \Pr[Y^* \leq 0, Y \leq y] \\ &= H_{Y^*,Y}(1, y) - H_{Y^*,Y}(0, y) . \end{aligned}$$

Cuando la variable $X = 1$, la expresión $\Pr[X = 1, Y \leq y]$ es igual, por la Ecuación (5.8), a la expresión $\Pr[Y^* \leq \infty, Y \leq y] - \Pr[Y^* \leq 1, Y \leq y]$, por lo tanto, tenemos lo siguiente

$$\begin{aligned} \Pr[X = 1, Y \leq y] &= \Pr[Y^* \leq \infty, Y \leq y] - \Pr[Y^* \leq 1, Y \leq y] \\ &= H_{Y^*,Y}(\infty, y) - H_{Y^*,Y}(1, y) \\ &= G_Y(y) - H_{Y^*,Y}(1, y) . \end{aligned}$$

Lo último se obtiene de la Definición 8.5, la cual se encuentra en el Apéndice de este trabajo de Tesis.

De manera general, se obtiene que la distribución conjunta de X y Y está dada por la siguiente ecuación

$$\Pr[X = x, Y \leq y] = \begin{cases} H_{Y^*,Y}(0, y) & \text{si } x = 0 \\ H_{Y^*,Y}(1, y) - H_{Y^*,Y}(0, y) & \text{si } x = p \\ G_Y(y) - H_{Y^*,Y}(1, y) & \text{si } x = 1. \end{cases} \quad (5.9)$$

La función de densidad conjunta de X y Y está dada por

$$h_{X,Y}(x, y) = \frac{\partial}{\partial y} \Pr[X = x, Y \leq y] \quad (5.10)$$

y por las Ecuaciones (2.37) y (2.38) se puede ver de la siguiente manera

$$h_{X,Y}(x, y) = \Phi \left(\frac{\Phi^{-1}(v) - \rho \Phi^{-1}(u)}{\sqrt{1 - \rho^2}} \right) \cdot g_Y(y) \quad (5.11)$$

donde $u = G_{Y^*}(y)$ y $v = G_Y(y)$. Aplicando esto último a la función de densidad conjunta (5.9), la cual es una función definida por partes, cuando $X = 0$, obtenemos lo siguiente

$$\begin{aligned} h_{X,Y}(x, y) &= \left\{ \Phi \left(\frac{\Phi^{-1}(G_{Y^*}(0)) - \rho \Phi^{-1}(G_Y(y))}{\sqrt{1 - \rho^2}} \right) - \Phi \left(\frac{\Phi^{-1}(G_{Y^*}(-\infty)) - \rho \Phi^{-1}(G_Y(y))}{\sqrt{1 - \rho^2}} \right) \right\} g_Y(y) \\ &= \left\{ \Phi \left(\frac{\Phi^{-1}(G_{Y^*}(0)) - \rho \Phi^{-1}(G_Y(y))}{\sqrt{1 - \rho^2}} \right) \right\} g_Y(y) \end{aligned}$$

cuando $X = p$ obtenemos

$$h_{X,Y}(x, y) = \left\{ \Phi \left(\frac{\Phi^{-1}(G_{Y^*}(1)) - \rho \Phi^{-1}(G_Y(y))}{\sqrt{1 - \rho^2}} \right) - \Phi \left(\frac{\Phi^{-1}(G_{Y^*}(0)) - \rho \Phi^{-1}(G_Y(y))}{\sqrt{1 - \rho^2}} \right) \right\} g_Y(y)$$

cuando $X = 1$ obtenemos

$$\begin{aligned} h_{X,Y}(x, y) &= \left\{ \Phi \left(\frac{\Phi^{-1}(G_{Y^*}(\infty)) - \rho \Phi^{-1}(G_Y(y))}{\sqrt{1 - \rho^2}} \right) - \Phi \left(\frac{\Phi^{-1}(G_{Y^*}(1)) - \rho \Phi^{-1}(G_Y(y))}{\sqrt{1 - \rho^2}} \right) \right\} g_Y(y) \\ &= \left\{ 1 - \Phi \left(\frac{\Phi^{-1}(G_{Y^*}(1)) - \rho \Phi^{-1}(G_Y(y))}{\sqrt{1 - \rho^2}} \right) \right\} g_Y(y) \\ &= \left\{ \Phi \left(- \frac{\Phi^{-1}(G_{Y^*}(1)) - \rho \Phi^{-1}(G_Y(y))}{\sqrt{1 - \rho^2}} \right) \right\} g_Y(y). \end{aligned}$$

5.3. FUNCIÓN DE DENSIDAD CONJUNTA CUANDO UNA DE LAS VARIABLES ES DISCRETA. 83

Es importante notar que ocupamos la siguiente propiedad: $\Phi(-x) = 1 - \Phi(x)$, donde $\Phi(x) = \Pr[X < x]$.

Con el análisis realizado previamente obtenemos la función de densidad conjunta completa, la cual está definida de la siguiente manera

$$h_{X,Y}(x,y) = \begin{cases} \left\{ \Phi \left(\frac{\Phi^{-1}(G_{Y^*}(0)) - \rho \Phi^{-1}(G_Y(y))}{\sqrt{1-\rho^2}} \right) \right\} g_Y(y) \\ \text{si } x = 0 \\ \\ \left\{ \Phi \left(\frac{\Phi^{-1}(G_{Y^*}(1)) - \rho \Phi^{-1}(G_Y(y))}{\sqrt{1-\rho^2}} \right) - \Phi \left(\frac{\Phi^{-1}(G_{Y^*}(0)) - \rho \Phi^{-1}(G_Y(y))}{\sqrt{1-\rho^2}} \right) \right\} g_Y(y) \\ \text{si } x = p \\ \\ \left\{ \Phi \left(- \frac{\Phi^{-1}(G_{Y^*}(1)) - \rho \Phi^{-1}(G_Y(y))}{\sqrt{1-\rho^2}} \right) \right\} g_Y(y) \\ \text{si } x = 1. \end{cases}$$

Capítulo 6

Curvas ROC Dependientes del Tiempo

6.1. Introducción

Con base en las nuevas definiciones de Sensibilidad y Especificidad, se define la curva ROC Incidente/Dinámica, la cual denotaremos por $ROC_t^{I/D}(p)$. A partir de ello se muestran una serie de ejemplos utilizando la cópula normal bivariada. El primero de ellos supone que el marcador M y el logaritmo del tiempo de supervivencia $\log(T)$ siguen una distribución normal bivariada. Para diferentes valores del tiempo $\log(T)$ se grafica su respectiva curva $ROC_t^{I/D}(p)$ y se calcula su área debajo de la curva (AUC).

Ocupando nuevamente la cópula normal bivariada y suponiendo que el marcador M tiene una distribución Normal Asimétrica y que el tiempo T tiene una distribución Weibull, las cuales siguen una distribución normal bivariada, entonces procedemos, mediante la teoría ya desarrollada, a graficar varias curvas $ROC_t^{I/D}(p)$, las cuales están dadas por los tiempos utilizados y para cada una de estas curvas $ROC_t^{I/D}(p)$ calculamos su AUC. Esta información es puesta en tablas y se observa en que parte del tiempo T se obtiene la mejor y la peor curva $ROC_t^{I/D}(p)$.

6.2. Curvas ROC Incidente-Dinámica

En la sección pasada se definió la Sensibilidad Incidente y la Especificidad Dinámica, las cuales dependen del tiempo, por lo tanto, es posible definir, calcular e interpretar las Curvas ROC en base a esta nueva definición. En esta sección nos centramos únicamente en la definición de las Curvas ROC Incidente/Dinámica (**I/D**), definidas como la función $ROC_t^{I/D}(p)$, donde p denota la tasa de falsos positivos dinámicos y $ROC_t^{I/D}(p)$ denota la correspondiente tasa de verdaderos positivos incidentes. Específicamente, definamos a c^p como el umbral que produce una tasa de falsos positivos p , es decir,

$$\begin{aligned}
\Pr [M > c^p | T > t] &= \Pr [M > c^p | T > t] \\
&= \frac{\Pr [M > c^p, T > t]}{\Pr [T > t]} \\
&= \frac{\Pr [T > t] - \Pr [M \leq c^p, T > t]}{\Pr [T > t]} \\
&= 1 - \frac{\Pr [M \leq c^p, T > t]}{\Pr [T > t]} \\
&= 1 - \Pr [M \leq c^p | T > t] \\
&= 1 - \text{Especificidad}^{\mathbf{D}}(c^p, t) \\
&= p
\end{aligned} \tag{6.1}$$

con $p \in [0, 1]$.

Es importante notar que hemos usado la Ley de Probabilidad Total, la cual nos dice que

$$\Pr [Y > y] = \Pr [X \leq x, Y > y] + \Pr [X > x, Y > y].$$

Observamos que la tasa verdadera positiva, $ROC_t^{I/D}(p)$, es la Sensibilidad que se obtiene usando este umbral, es decir,

$$ROC_t^{I/D}(p) = \text{Sensibilidad}^{\mathbf{I}}(c^p, t) = \Pr [M > c^p | T = t]. \tag{6.2}$$

Notamos que si utilizamos las funciones de verdadero-positivo y falso-positivo

$$TP_t^{\mathbf{I}}(c) = \text{Sensibilidad}^{\mathbf{I}}(c, t)$$

$$FP_t^{\mathbf{D}}(c) = 1 - \text{Especificidad}^{\mathbf{D}}(c, t)$$

y recordando que se definió a c^p como el umbral que produce una tasa de falsos-positivos, obtenemos

$$FP_t^{\mathbf{D}}(c^p) = 1 - \text{Especificidad}^{\mathbf{D}}(c^p, t)$$

y por la Ecuación 6.1 obtenemos

$$FP_t^{\mathbf{D}}(c^p) = 1 - \text{Especificidad}^{\mathbf{D}}(c^p, t) = p. \tag{6.3}$$

Lo que nos quiere decir esta última ecuación es que, por cada valor que se le asigne a c^p obtenemos un valor p distinto. Por lo tanto, podemos escribir a la Curva $ROC_t^{I/D}(p)$ como la composición de la función $TP_t^{\mathbf{I}}(c)$ y la función inversa $[FP_t^{\mathbf{D}}]^{-1}(p) = c^p$. Así obtenemos lo siguiente

$$\begin{aligned}
ROC_t^{I/D}(p) &= TP_t^I \left([FP_t^D]^{-1}(p) \right) \\
&= TP_t^I(c^p) \\
&= Sensibilidad^I(c^p, t)
\end{aligned} \tag{6.4}$$

para $p \in [0, 1]$.

Por lo tanto, podemos ver a la Curva $ROC_t^{I/D}(p)$ como el conjunto de parejas ordenadas $\{(p, ROC_t^{I/D}(p)) ; p \in [0, 1]\}$.

Es importante notar que hemos usado la función $[FP_t^D]^{-1}(p)$, sin embargo, no se ha definido de manera formal matemáticamente hablando, por lo tanto, vamos a definirla explícitamente

$$[FP_t^D]^{-1}(p) = \inf_c \{c : FP_t^D(c) \leq p\} \tag{6.5}$$

y por las Ecuaciones (6.1) y (6.3) se tiene que

$$FP_t^D(c) = 1 - Especificidad^D(c, t) = \Pr[M > c | T > t] = p. \tag{6.6}$$

Sustituyendo la Ecuación (6.6) en la Ecuación (6.5) obtenemos

$$[FP_t^D]^{-1}(p) = \inf_c \{c : \Pr[M > c | T > t] \leq p\} \tag{6.7}$$

donde $p \in [0, 1]$.

Esto nos permite ver más claramente el funcionamiento de la función $[FP_t^D]^{-1}(p)$ y nos ayuda a comprender mejor la Ecuación (6.4).

$$ROC_t^{I/D}(p) = TP_t^I \left([FP_t^D]^{-1}(p) \right). \tag{6.8}$$

Por último, usamos la notación

$$AUC(t) = \int_0^1 ROC_t^{I/D}(p) dp \tag{6.9}$$

para denotar el área bajo la curva ROC Incidente-Dinámica para el tiempo t .

A continuación mostramos un ejemplo en el cual ocupamos la cópula normal bivariada con marginales normal y log-normal.

6.2.1. Curva $ROC_t^{I/D}(p)$ con Marcador Normal y Tiempo Log-Normal

Después de desarrollar la teoría con respecto a las nuevas definiciones de Sensibilidad Incidente y Especificidad Dinámica, las cuales reciben también el nombre de conceptos de precisión dependientes del tiempo, ilustraremos un ejemplo en el cual vamos a suponer que el marcador M y el logaritmo del tiempo de supervivencia $\log(T)$ siguen una distribución normal bivariada. Por convención consideraremos un valor del marcador alto como indicativo del inicio de la enfermedad temprana y examinaremos las distribuciones bivariadas con una correlación negativa entre el marcador M y $\log(T)$.

Si $[M, \log(T)]$ tiene una distribución normal bivariada con media $(0, 0)$ y desviaciones estándar unitarias, entonces la Sensibilidad Incidente dependiente del tiempo es

$$\text{Sensibilidad}^I(c, \log(t)) = \Pr[M > c | \partial(\log(t))].$$

Ocupando la Ley de Probabilidad Total, obtenemos

$$\Pr[\partial(\log(t))] = \Pr[M \leq c, \partial(\log(t))] + \Pr[M > c, \partial(\log(t))].$$

Por lo tanto

$$\begin{aligned} \text{Sensibilidad}^I(c, \log(t)) &= \Pr[M > c | \partial(\log(t))] \\ &= \frac{\Pr[M > c, \partial(\log(t))]}{\Pr[\partial(\log(t))]} \\ &= \frac{\Pr[\partial(\log(t))] - \Pr[M \leq c, \partial(\log(t))]}{\Pr[\partial(\log(t))]} \\ &= 1 - \frac{\Pr[M \leq c, \partial(\log(t))]}{\Pr[\partial(\log(t))]} \\ &= 1 - \Pr[M \leq c | \partial(\log(t))]. \end{aligned} \tag{6.10}$$

Por la Definición 2.6 y por la Ecuación (2.34) se obtiene que

$$\begin{aligned}
\text{Sensibilidad}^I(c, \log(t)) &= TP_{\log(t)}^I(c) \\
&= 1 - \Pr[M \leq c \mid \partial(\log(t))] \\
&= 1 - \Phi\left(\frac{\Phi^{-1}(\Phi(c)) - \rho \Phi^{-1}(\Phi(\log(t)))}{\sqrt{1 - \rho^2}}\right) \\
&= 1 - \Phi\left(\frac{c - \rho \log(t)}{\sqrt{1 - \rho^2}}\right) \\
&= \Phi\left(\frac{\rho \log(t) - c}{\sqrt{1 - \rho^2}}\right).
\end{aligned} \tag{6.11}$$

Ocupando nuevamente la Ley de Probabilidad Total obtenemos que

$$\Pr[\log(T) > \log(t)] = \Pr[M \leq c, \log(T) > \log(t)] + \Pr[M > c, \log(T) > \log(t)].$$

Donde

$$\Pr[M \leq c, \log(T) > \log(t)] = \Pr[\log(T) > \log(t)] - \Pr[M > c, \log(T) > \log(t)].$$

Por lo tanto, la Especificidad Dinámica está dada por

$$\begin{aligned}
\text{Especificidad}^D(c, \log(t)) &= \Pr[M \leq c \mid \log(T) > \log(t)] \\
&= \frac{\Pr[M \leq c, \log(T) > \log(t)]}{\Pr[\log(T) > \log(t)]} \\
&= \frac{\Pr[\log(T) > \log(t)] - \Pr[M > c, \log(T) > \log(t)]}{\Pr[\log(T) > \log(t)]} \\
&= 1 - \frac{\Pr[M > c, \log(T) > \log(t)]}{\Pr[\log(T) > \log(t)]} \\
&= 1 - \frac{S_2^N[c, \log(t); \rho]}{1 - \Pr[\log(T) \leq \log(t)]} \\
&= 1 - \frac{S_2^N[c, \log(t); \rho]}{1 - \Phi(\log(t))} \\
&= 1 - \frac{S_2^N[c, \log(t); \rho]}{\Phi(-\log(t))}.
\end{aligned} \tag{6.12}$$

Concluyendo que

$$\begin{aligned}
 FP_{\log(t)}^{\mathbf{D}}(c) &= 1 - \text{Especificidad}^{\mathbf{D}}(c, \log(t)) \\
 &= 1 - \left(1 - \frac{S_2^N[c, \log(t); \rho]}{\Phi(-\log(t))} \right) \\
 &= \frac{S_2^N[c, \log(t); \rho]}{\Phi(-\log(t))}.
 \end{aligned} \tag{6.13}$$

Es importante notar que en la Ecuación (6.11) se ocupó la siguiente propiedad: $\Phi(-x) = 1 - \Phi(x)$, donde $\Phi(x) = \Pr[X < x]$ para $X \sim N(0,1)$ y en la Ecuación (6.12) se ocupó la notación $S_2^N[x, y; \rho]$ para denotar a $\Pr[X > x, Y > y]$ cuando $[X, Y]$ tienen una distribución normal bivariada unitaria con correlación ρ .

Finalmente, resumiendo nuestros resultados, obtenemos las siguientes Ecuaciones:

Por la Ecuación (6.11) conseguimos definir a $TP_{\log(t)}^{\mathbf{I}}(c)$

$$\begin{aligned}
 TP_{\log(t)}^{\mathbf{I}}(c) &= \text{Sensibilidad}^{\mathbf{I}}(c, \log(t)) \\
 &= \Pr[M > c \mid \partial(\log(t))] \\
 &= \Phi\left(\frac{\rho \log(t) - c}{\sqrt{1 - \rho^2}}\right)
 \end{aligned} \tag{6.14}$$

y con ayuda de las Ecuaciones (6.1) y (6.13) conseguimos definir a $FP_{\log(t)}^{\mathbf{D}}(c)$

$$\begin{aligned}
 FP_{\log(t)}^{\mathbf{D}}(c) &= 1 - \text{Especificidad}^{\mathbf{D}}(c, \log(t)) \\
 &= \Pr[M > c \mid \log(T) > \log(t)] \\
 &= \frac{S_2^N[c, \log(t); \rho]}{\Phi(-\log(t))}.
 \end{aligned} \tag{6.15}$$

De manera general, llegamos a la conclusión de que si $F(c)$ y $G(t)$ son funciones de distribución del marcador y del tiempo respectivamente, entonces, ocupando la Cópula Normal Bivariada obtenemos que las Ecuaciones de $TP_{(t)}^{\mathbf{I}}(c)$ y $FP_{(t)}^{\mathbf{D}}(c)$ están dadas de la siguiente manera

$$TP_{(t)}^{\mathbf{I}}(c) = \Phi\left(\frac{\rho \Phi^{-1}(G(t)) - \Phi^{-1}(F(c))}{\sqrt{1 - \rho^2}}\right) \tag{6.16}$$

$$FP_{(t)}^D(c) = \frac{S_2^N[c, t; \rho]}{1 - G(t)}. \quad (6.17)$$

Las Ecuaciones (6.16) y (6.17) son en esencia las mismas que las Ecuaciones (5.6) y (5.7) previamente encontradas. Estas ecuaciones nos van a ser de gran utilidad en la siguiente sección, pero antes de eso vamos a mostrar diferentes curvas $ROC_t^{I/D}(p)$ referentes al marcador normal y al tiempo Log-Normal, recordando que $[M, \log(T)]$ tiene una distribución normal bivariada con media $(0, 0)$, con desviaciones estándar unitarias y vamos a suponer que el parámetro de correlación es $\rho = -0.8$. Dichas curvas $ROC_t^{I/D}(p)$ se muestran en la Figura 6.1.

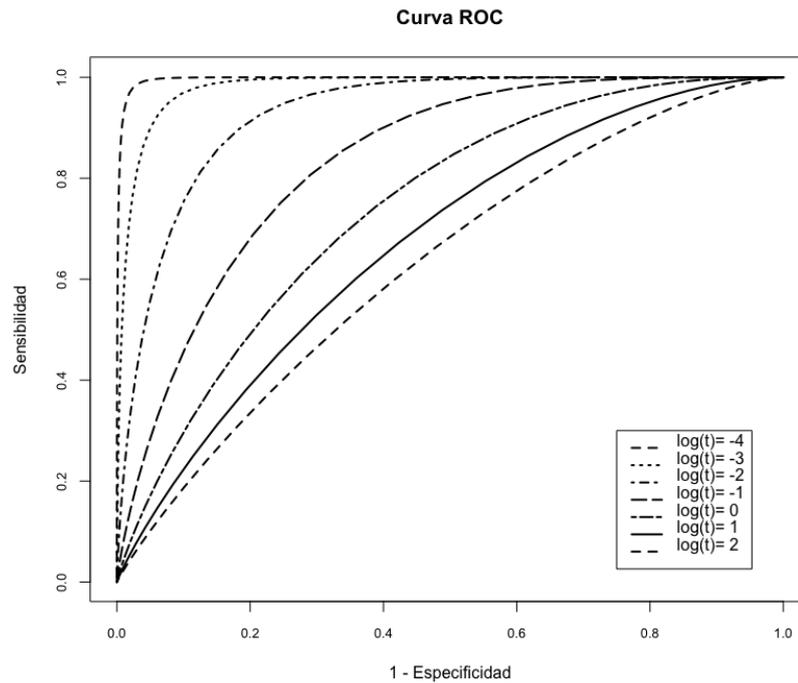


Figura 6.1: Curvas $ROC_t^{I/D}(p)$ derivadas de la cópula normal bivariada con marginales normal y log-normal.

Para construir estas curvas $ROC_t^{I/D}(p)$ tomamos valores enteros fijos del tiempo que van desde $\log(t) = -4$ hasta $\log(t) = 2$, obteniendo un total de 7 curvas. Notamos que la mejor curva $ROC_t^{I/D}(p)$ se alcanza cuando $\log(t) = -4$, ya que el $AUC = 0.996$, tal y como lo muestra el Cuadro 6.1.

A continuación mostramos otro ejemplo, en el cual asignamos distribuciones distintas al marcador y al tiempo.

6.2.2. Curvas $ROC_t^{I/D}(p)$ con Marcador Normal Asimétrica y Tiempo Weibull

En la Sección 2.5 se mostraron 3 ejemplos de la cópula normal bivariada, estos fueron los siguientes:

Tiempo	Área
$\log(t) = -4$	0.996
$\log(t) = -3$	0.980
$\log(t) = -2$	0.927
$\log(t) = -1$	0.833
$\log(t) = 0$	0.740
$\log(t) = 1$	0.672
$\log(t) = 2$	0.628

Cuadro 6.1: Áreas bajo la curva $ROC_t^{I/D}(p)$ al tiempo $\log(t)$.

1. Cópula normal bivariada con medida de asociación $\tau(-0.8)$ con la primera marginal normal asimétrica con parámetros $Sn(xi = 3, \omega = 2, \alpha = -5)$ y segunda marginal Weibull con parámetros $W(shape = 1.5, scale = 1)$.
2. Cópula normal bivariada con medida de asociación $\tau(-0.6)$ con la primera marginal normal asimétrica con parámetros $Sn(xi = 3, \omega = 2, \alpha = 5)$ y segunda marginal Weibull con parámetros $W(shape = 5, scale = 1)$.
3. Cópula normal bivariada con medida de asociación $\tau(-0.4)$ con la primera marginal normal asimétrica con parámetros $Sn(xi = 3, \omega = 3, \alpha = 0)$ y segunda marginal Weibull con parámetros $W(shape = 1.5, scale = 1)$.

Con base en el ejemplo de la Curva $ROC_t^{I/D}(p)$ con Marcador normal y Tiempo Log-Normal y con ayuda de las Ecuaciones (6.16) y (6.17) vamos a construir para cada uno de estos ejemplos 5 curvas $ROC_t^{I/D}(p)$, haciendo énfasis en que el marcador sigue una distribución Normal Asimétrica y el tiempo sigue una distribución Weibull.

Los cuantiles q tal que $q \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$ de la distribución Weibull son los tiempos fijos que vamos a utilizar para la construcción de las 5 curvas $ROC_t^{I/D}(p)$ para cada ejemplo.

A continuación mostramos las 5 curvas $ROC_t^{I/D}(p)$ para cada uno de los ejemplos.

Ejemplo 1

Las curvas $ROC_t^{I/D}(p)$ de la cópula normal bivariada con medida de asociación $\tau(-0.8)$ con la primera marginal normal asimétrica con parámetros $Sn(xi = 3, \omega = 2, \alpha = -5)$, la segunda marginal Weibull con parámetros $W(shape = 1.5, scale = 1)$, cuyos tiempos fijos utilizados son los cuantiles q tal que $q \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$ de la distribución Weibull, se muestran en la Figura 6.2.

La mejor curva $ROC_t^{I/D}(p)$ se alcanza cuando $q = 0.05$ ya que se alcanza un $AUC = 0.793$, tal y como lo muestra el Cuadro 6.2.

Ejemplo 2

Las curvas $ROC_t^{I/D}(p)$ de la cópula normal bivariada con medida de asociación $\tau(-0.6)$ con la primera marginal normal asimétrica con parámetros $Sn(xi = 3, \omega = 2, \alpha = 5)$, la segunda marginal

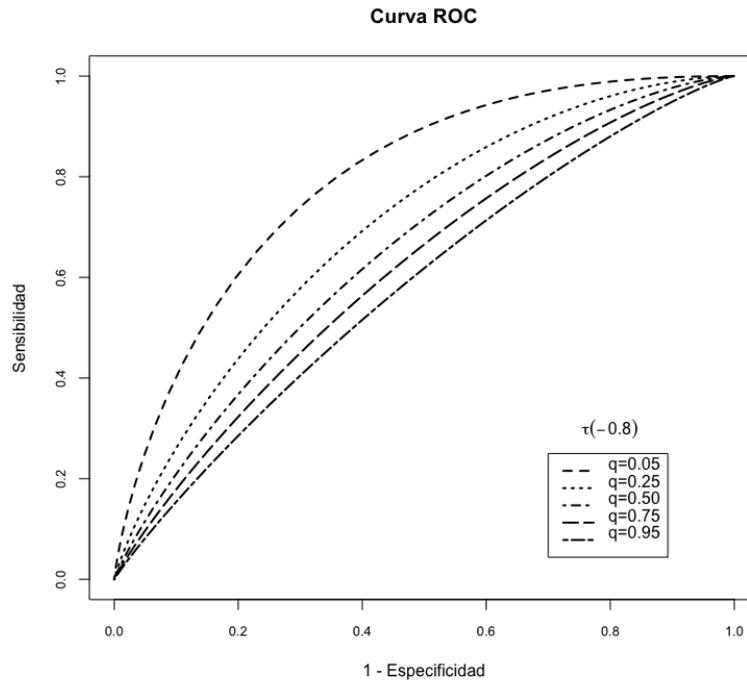


Figura 6.2: Curvas $ROC_t^{I/D}(p)$ derivadas de la cópula normal bivariada con marginales normal asimétrica y Weibull con parámetros $Sn(x_i = 3, \omega = 2, \alpha = -5)$ y $W(shape = 1.5, scale = 1)$ respectivamente, con medida de asociación $\tau(-0.8)$.

Tiempo	Área
$q = 0.05$	0.793
$q = 0.25$	0.700
$q = 0.50$	0.650
$q = 0.75$	0.614
$q = 0.95$	0.576

Cuadro 6.2: Áreas bajo la curva $ROC_t^{I/D}(p)$ al tiempo q .

Weibull con parámetros $W(shape = 5, scale = 1)$, cuyos tiempos fijos utilizados son los cuantiles q tal que $q \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$ de la distribución Weibull, se muestran en la Figura 6.3.

La mejor curva $ROC_t^{I/D}(p)$ se alcanza cuando $q = 0.05$ ya que se alcanza un $AUC = 0.703$, tal y como lo muestra el Cuadro 6.3.

Ejemplo 3

Las curvas $ROC_t^{I/D}(p)$ de la cópula normal bivariada con medida de asociación $\tau(-0.4)$ con la primera marginal normal asimétrica con parámetros $Sn(x_i = 3, \omega = 3, \alpha = 0)$, la segunda marginal

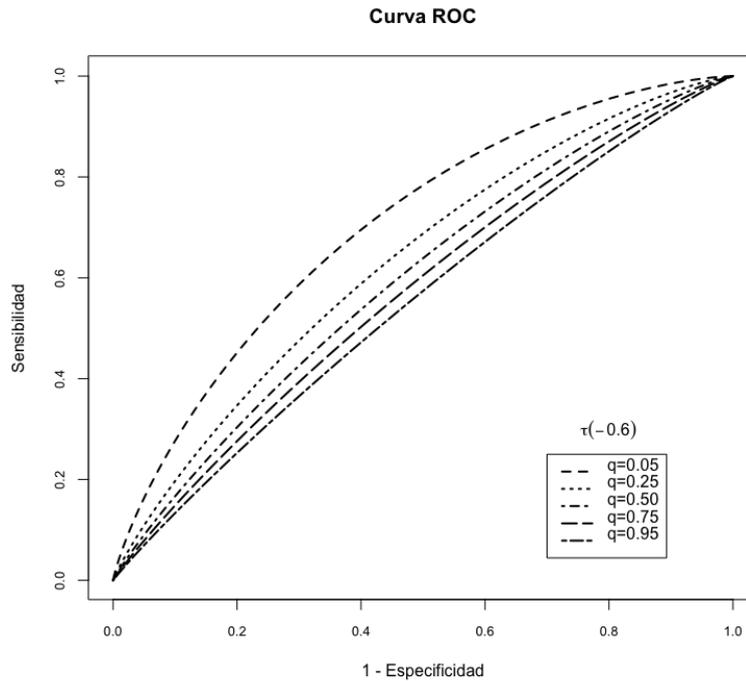


Figura 6.3: Curvas $ROC_t^{I/D}(p)$ derivadas de la cópula normal bivariada con marginales normal asimétrica y Weibull con parámetros $Sn(xi = 3, \omega = 2, \alpha = 5)$ y $W(shape = 5, scale = 1)$ respectivamente, con medida de asociación $\tau(-0.6)$.

Tiempo	Área
$q = 0.05$	0.703
$q = 0.25$	0.632
$q = 0.50$	0.597
$q = 0.75$	0.573
$q = 0.95$	0.552

Cuadro 6.3: Áreas bajo la curva $ROC_t^{I/D}(p)$ al tiempo q .

Weibull con parámetros $W(shape = 1.5, scale = 1)$, cuyos tiempos fijos utilizados son los cuantiles q tal que $q \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$ de la distribución Weibull, se muestran en la Figura 6.4.

La mejor curva $ROC_t^{I/D}(p)$ se alcanza cuando $q = 0.05$ ya que se alcanza un $AUC = 0.628$, tal y como lo muestra el Cuadro 6.4.

Una de las preguntas que nos surge es, ¿cómo hace variar la medida de asociación τ a nuestras curvas $ROC_t^{I/D}(p)$?. Para responder a esta pregunta basta con tomar uno de los ejemplos anteriormente analizados, digamos el Ejemplo 1, entonces, para cada valor de τ donde $\tau \in \{-0.8, -0.6, -0.4\}$ sacamos las 5 curvas $ROC_t^{I/D}(p)$ correspondientes a los tiempos fijos utilizados, los cuales son los cuantiles q tal que $q \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$ de la distribución Weibull. Este conjunto de curvas $ROC_t^{I/D}(p)$ se

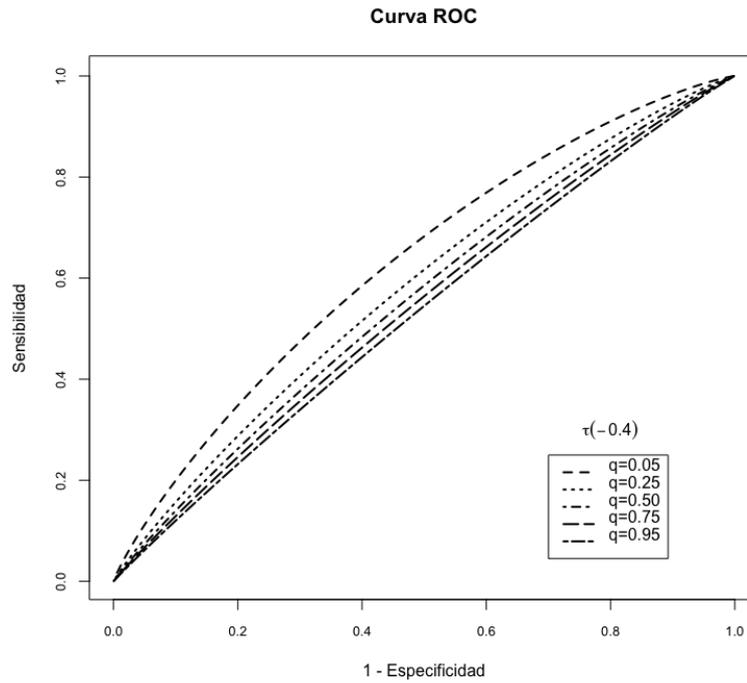


Figura 6.4: Curvas $ROC_t^{I/D}(p)$ derivadas de la cópula normal bivariada con marginales normal asimétrica y Weibull con parámetros $Sn(x_i = 3, \omega = 3, \alpha = 0)$ y $W(shape = 1.5, scale = 1)$ respectivamente, con medida de asociación $\tau(-0.4)$.

Tiempo	Área
$q = 0.05$	0.628
$q = 0.25$	0.580
$q = 0.50$	0.558
$q = 0.75$	0.542
$q = 0.95$	0.526

Cuadro 6.4: Áreas bajo la curva $ROC_t^{I/D}(p)$ al tiempo q .

muestran en la Figura 6.5.

Las curvas $ROC_t^{I/D}(p)$ que genera la cópula normal bivariada con la primera marginal normal asimétrica con parámetros $Sn(x_i = 3, \omega = 2, \alpha = -5)$, la segunda marginal Weibull con parámetros $W(shape = 1.5, scale = 1)$ alcanza el AUC máxima con $\tau(-0.8)$ en el cuantil $q = 0.05$, generando un valor de $AUC = 0.793$, por el contrario, notamos que el AUC mínimo se alcanza con $\tau(-0.4)$ en el cuantil $q = 0.95$, generando un valor de $AUC = 0.529$.

La información detallada de como la medida de asociación τ afecta a las curvas $ROC_t^{I/D}(p)$ se muestra en el Cuadro 6.5.

τ	Área máxima	Cuantil _{Área máxima}	Área mínima	Cuantil _{Área mínima}
-0.8	0.793	0.05	0.576	0.95
-0.6	0.702	0.05	0.548	0.95
-0.4	0.629	0.05	0.529	0.95

Cuadro 6.5: Áreas bajo la curva $ROC_t^{I/D}(p)$ máximas y mínimas al tiempo q utilizando tres medidas de asociación: $\tau(-0.8)$, $\tau(-0.6)$ y $\tau(-0.4)$.

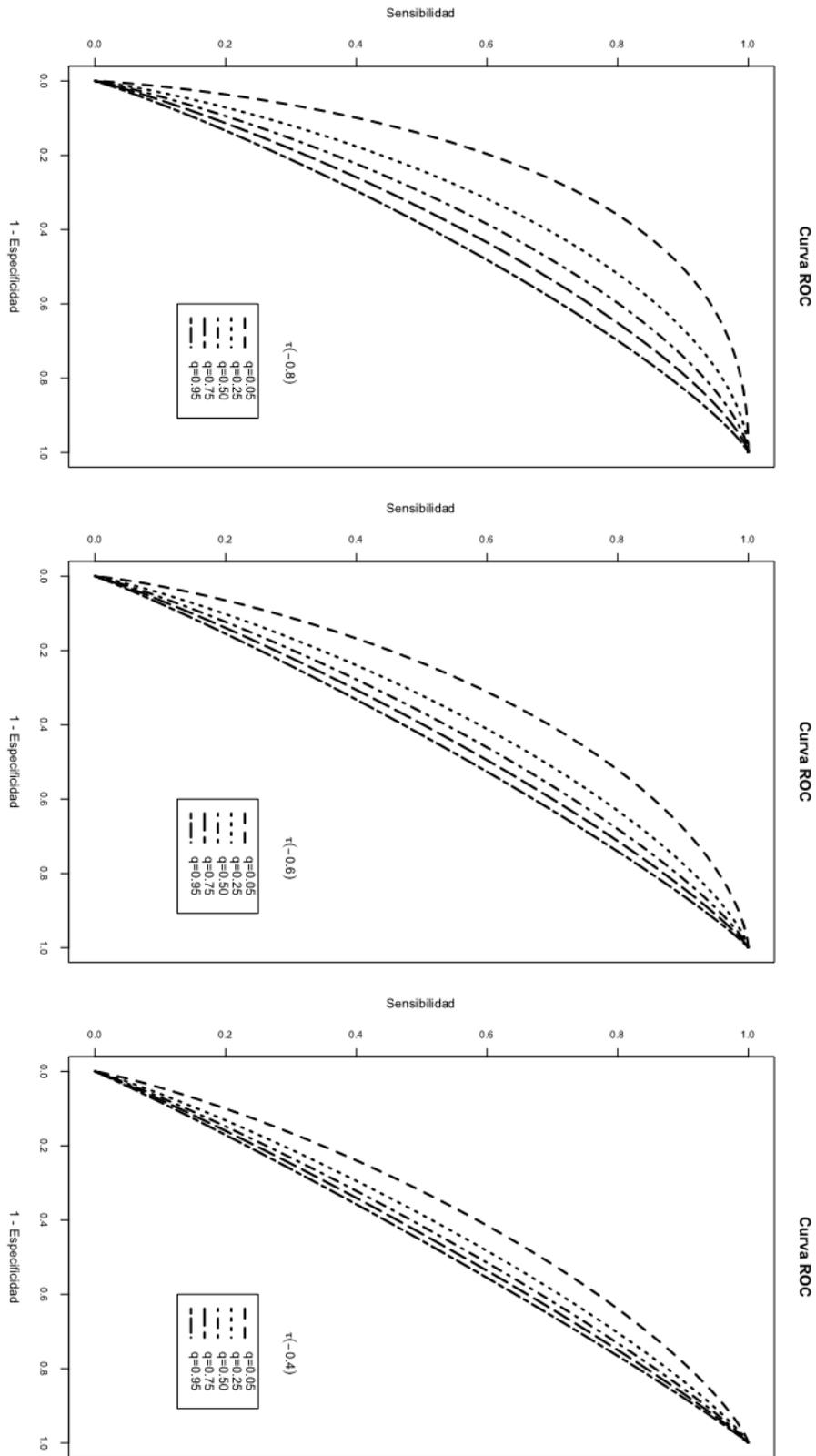


Figura 6.5: Curvas $ROC_t^{I/D}(p)$ derivadas de la cópula normal bivariada con marginales normal asimétrica y Weibull con parámetros $Sn(x_i = 3, \omega = 2, \alpha = -5)$ y $W(shape = 1.5, scale = 1)$ respectivamente, utilizando tres medidas de asociación: $\tau(-0.8)$, $\tau(-0.6)$ y $\tau(-0.4)$.

Capítulo 7

Conclusiones

En este trabajo de Tesis hemos podido definir de manera clara, concisa y detallada la curva ROC Incidente/Dinámica. Para ello fue necesario desarrollar teoría matemática que nos ayudó a comprender y visualizar todo el procedimiento que llevamos a cabo.

Uno de los basamentos de este trabajo fue el desarrollar la teoría de las cópulas bivariadas, aclaramos que este trabajo no se centra en un estudio a profundidad de las cópulas, solo se limita a abordar este tema de forma breve, de tal forma, que se desarrolle la teoría necesaria. Desarrollada la teoría, nos dedicamos al estudio detallado de la cópula normal bivariada, la cual, por sus propiedades, se utilizó como cópula base en los ejemplos de las curvas ROC.

Las conclusiones más importantes a las que llegamos en este trabajo de Tesis son la siguientes:

1. Al haber desarrollado la teoría de las cópulas bivariadas, se abordó el tema de inferencia estadística, ahí vemos áreas de oportunidad para futuras investigaciones tanto teóricas como computacionales en el proceso de la estimación de parámetros de las cópulas.
2. Al desarrollar a detalle la curva ROC Acumulativa/Dinámica, la cual se basa en los conceptos de Sensibilidad Acumulativa y la Especificidad Dinámica, notamos que es posible aplicar las cotas de Fréchet-Hoeffding para obtener 2 de los 3 casos extremos de la curva $ROC_t^{C/D}(p)$.
3. Es posible extender el concepto de la curva ROC Acumulativa/Dinámica a la curva ROC Incidente/Dinámica, la cual tiene como fundamento el suponer que los casos estudiados para definir a la Sensibilidad, son casos incidentes, es decir, en los que $T = t$, supone también que los controles de los marcadores sean dinámicos y definidos para el tiempo t , es decir, como aquellos sujetos con $T > t$.
4. Cuando se abordó el tema de la Sensibilidad Incidente y la Especificidad Dinámica nos percatamos de que era posible asignarles una función de distribución al Marcador y al Tiempo, con la condición de que ambas distribuciones fueran continuas o que una de ellas fuera continua y la otra discreta.
5. Se demuestra matemáticamente que las curvas $ROC_t^{C/D}(p)$ y $ROC_t^{I/D}(p)$ son monótonas con respecto a c , aún más, mientras más grande sea el valor del tiempo más baja será la capacidad discriminatoria de las curvas, por lo tanto, a medida que aumenta el tiempo, el valor del área debajo de la curva (AUC) es cada vez más bajo, sin embargo, no solo es el tiempo quien afecta la capacidad discriminatoria, también lo es la medida de asociación Tau de Kendall. Todo esto

se ilustra mediante varios ejemplos, el primero de ellos utiliza la cópula normal bivariada con marginales normal y log-normal y para los demás ejemplos utilizamos la cópula normal bivariada con marginales normal asimétrica y Weibull con 3 medidas de asociación Tau de Kendall.

Capítulo 8

Apéndice

8.1. Cálculo de Probabilidades

Un espacio de probabilidad es una terna $(\Omega, \mathfrak{F}, \text{Pr})$, donde Ω es un conjunto arbitrario, \mathfrak{F} es una σ -álgebra de subconjuntos de Ω y Pr una medida de probabilidad definida sobre \mathfrak{F} .

Se procede a definir un conjunto de Borel sobre \mathbb{R}^2 para poder definir un vector aleatorio.

Sea Λ la colección de los rectángulos abiertos en \mathbb{R}^2

$$\Lambda = \{(a, b) \times (c, d) : a \leq b, c \leq d\}.$$

Entonces se definen los conjuntos de Borel de \mathbb{R}^2 como los elementos de la mínima σ -álgebra generada por la colección Λ , es decir, $\mathcal{B}(\mathbb{R}^2) = \sigma(\Lambda)$. Formalmente se obtiene la siguiente definición.

Definición 8.1 La σ -álgebra de Borel de \mathbb{R}^2 está dada de la siguiente manera

$$\mathcal{B}(\mathbb{R}^2) = \sigma(\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R})).$$

Se procede a extender el concepto de variable aleatoria con valores reales a variables aleatorias con valores en \mathbb{R}^n , estas funciones se llamarán vectores aleatorios.

Definición 8.2 Un vector aleatorio es una función $X : \Omega \rightarrow \mathbb{R}^n$ tal que para cualquier conjunto Θ en $\mathcal{B}(\mathbb{R}^n)$ se cumple que la imagen inversa $X^{-1}(\Theta)$ es un elemento de \mathfrak{F} .

De la definición anterior se tiene que X es de la siguiente forma

$$X = (X_1, X_2, \dots, X_n)$$

donde cada coordenada es una variable aleatoria.

A lo largo de este trabajo únicamente se utilizarán vectores bidimensionales, es decir, vectores de la forma (X_1, X_2) , donde

$$(X_1, X_2) : \Omega \rightarrow \mathbb{R}^2.$$

Por lo cual las siguientes definiciones se harán únicamente para vectores bidimensionales.

De manera similar al caso de las variables aleatorias, un vector aleatorio bidimensional induce una medida de probabilidad sobre \mathbb{R}^2 . Esta medida se estudia mediante la función de distribución conjunta definida a continuación.

Definición 8.3 La función de distribución conjunta de un vector (X_1, X_2) , dada por $F(x_1, x_2) : \mathbb{R}^2 \rightarrow [0, 1]$, está dada de la siguiente manera:

$$F(x_1, x_2) = \Pr(X_1 \leq x_1, X_2 \leq x_2).$$

Donde $F(x_1, x_2)$ es la probabilidad de que el vector (X_1, X_2) toma en un valor del rectángulo $(-\infty, x_1] \times (-\infty, x_2]$.

A la función $F(x_1, x_2)$ se le llamará función de distribución bivariada de X_1 y X_2 , y cumple con las siguientes propiedades.

Teorema 8.1 Las funciones de distribución bivariadas $F(x_1, x_2)$ cumplen con las siguientes propiedades:

1.- Cuando ambas variables tienden a infinito se cumple lo siguiente:

$$\lim_{x_1, x_2 \rightarrow \infty} F(x_1, x_2) = 1.$$

2.- Cuando alguna de las variables tiende a cero se cumple lo siguiente:

$$\lim_{x_1, x_2 \rightarrow -\infty} F(x_1, x_2) = 0.$$

3.- $F(x_1, x_2)$ es no decreciente en cada variable.

4.- $\Pr\{(X, Y) \in (x_1, x_2] \times (y_1, y_2]\} = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) \geq 0$.

Demostración 8.1.1 La demostración de estas propiedades es análoga al caso de una dimensión.

De manera análoga al caso de variables aleatorias, los vectores aleatorios tienen asociada una función de densidad.

Definición 8.4 Sea (X_1, X_2) un vector continuo con la función de distribución $F(x_1, x_2)$, se dice que el vector (X_1, X_2) es absolutamente continuo si existe una función no negativa e integrable $f(x_1, x_2) : \mathbb{R}^2 \rightarrow [0, \infty)$ tal que para toda (x_1, x_2) en \mathbb{R}^2 , se cumple con la igualdad

$$F(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(u, v) dv du.$$

A la función $f(x_1, x_2)$ se le denota por $f_{X_1, X_2}(x_1, x_2)$ y se le llama función de densidad conjunta de X_1 y X_2 .

La función de densidad conjunta $f(x_1, x_2)$ de un vector absolutamente continuo cumple con las siguientes dos propiedades:

$$i) f(x_1, x_2) \geq 0.$$

$$ii) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) dx_1 dx_2 = 1.$$

De la función de distribución $F(x_1, x_2)$ de un vector (x_1, x_2) se puede obtener la función de distribución de cada variable aleatoria tal y como lo muestra la siguiente definición.

Definición 8.5 Sea (X_1, X_2) un vector con función de distribución $F(x_1, x_2)$, entonces a la función

$$F(x_1) = \lim_{x_2 \rightarrow \infty} F(x_1, x_2).$$

Se le llama función de distribución marginal de X_1 . Similarmente, se define la función de distribución marginal de X_2

$$F(x_2) = \lim_{x_1 \rightarrow \infty} F(x_1, x_2).$$

Es importante notar que las funciones marginales son funciones de distribución univariadas. De las funciones de densidad conjunta es posible obtener las funciones de densidad de cada una de las variables, tal como lo menciona la siguiente definición.

Definición 8.6 Dado un vector (X_1, X_2) absolutamente continuo con función de densidad $f(x_1, x_2)$, entonces a la función

$$f(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2.$$

Se le llama función de densidad marginal de X_1 . Análogamente se define la función de densidad marginal de X_2 como

$$f(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_1.$$

Ahora se procede a definir la función de densidad condicional

Definición 8.7 Sea (X_1, X_2) un vector con función de densidad $f_{X_1, X_2}(x_1, x_2)$ y sea y tal que $f_{x_2}(x_2) \neq 0$, entonces a la función

$$x \mapsto f_{X_1|X_2}(x_1 | x_2) = \frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_2}(x_2)}$$

se le llama función de densidad condicional de X dado que Y toma el valor y .

También se puede definir la función de distribución condicional de la siguiente manera:

Definición 8.8 Sea (X_1, X_2) un vector aleatorio absolutamente continuo con función de densidad $f_{X_1, X_2}(x_1, x_2)$, y sea y tal que $f_{x_2}(x_2) \neq 0$, entonces a la función

$$x \mapsto F_{X_1|X_2}(x_1 | x_2) = \int_{-\infty}^{\infty} f_{X_1|X_2}(u | x_2) du$$

se le conoce como la función de distribución condicional de X_1 dado que X_2 toma el valor x_2 .

De la función de distribución, la independencia de dos variables aleatorias se pueden expresar como indica el siguiente teorema.

Teorema 8.2 Las variables aleatorias X_1 y X_2 son independientes sí y sólo sí para cada (x_1, x_2) en \mathbb{R}^2 se cumple la siguiente igualdad

$$F_{X_1, X_2}(x_1, x_2) = F_{X_1}(x_1) \cdot F_{X_2}(x_2).$$

Bibliografía

- [1] Arnold, Helen and Shevchenko, Pavel (2006). *Dependence Modelling via the Copula Method prepared*.
- [2] Bean, M. A. (2001). *Probability: The science of Uncertainty*. United States of America: Books/Cole.
- [3] Clemen, R. T. and Reilly, T. (1999). *Correlations and Copulas for decisions and Risk Analysis*. Management Science, Vol. 45, No. 2, 208-224.
- [4] Collett, D. (2003). *Modelling Survival Data in Medical Research*. 2nd edition. Chapman and Hall/CRC.
- [5] Escarela, G. and Carriere, J. F. (2003). *Fitting competing risks with an assumed copula*. Statistical Methods in Medical Research 12, 333-349.
- [6] Fisher, R. A. (1921). *On the Probable Error of a Coefficient of Correlation Deduced from Small Sample*. Metron, V.1, 3-32.
- [7] Garthwaite, P.; Jolliffe, I. and Jones, B. (2002). *Statistical Inference*. New York: Oxford University Press.
- [8] Genest, C. and Favre, A. C. (2007). *Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask*. Journal of Hydrologic Engineering 12, 437-468.
- [9] Goncalvez, L.; Subtil, A.; Olivera, M. R.; Bermudez, P. (2014). *ROC curve estimation: an overview*. Volume 12, Number 1, 1-20.
- [10] Jaworski, P.; Durante, F.; Hardle, W. and Rychlik, T. (2010). *Copula Theory and its Applications*. London New York. Springer.
- [11] Joe, H. (2005). *Asymptotic efficiency of the two-stage estimation method for copula-based models*. Journal of Multivariate Analysis 94, 401-419.
- [12] Joe, H. and Xu, J.J. (1996). *The Estimation Method of Inference Functions for Margins for Multivariate Models*. Dept. of Statistics University of British Columbia, Tech. Rept. 166.
- [13] J. Heagerty, Patrick; Lumley Thomas. (2000). *Time-Dependent ROC Curves for Censored Survival Data and a Diagnostic Marker*. Biometrics 56, 337-344.
- [14] J. Heagerty, Patrick; Zheng Yingye. (2005). *Survival Model Predictive Accuracy and ROC Curves*. Biometrics 61, 92-105.
- [15] Kalbfleisch, J. G. (1979). *Probability and Statistical Inference*. Volume 2: Statistical Inference. 2nd edition. New York Berlin Heidelberg Tokyo: Springer-Verlag.
- [16] Leon, A. R. and Wu, B. (2009). *Copula based regression models for a bivariate mixed discrete and continuous outcome*. Statistics in Medicine 30, 175-185.

- [17] Nelsen, R. B. (2006). *An Introduction to Copulas*. 2nd edition. New York: Springer.
- [18] Pelikan M. and H. Muhlenbein (1999). *The bivariate marginal distribution algorithm*. Advances in Soft Computing-Engineering Design and Manufacturing.
- [19] Pravin K. Trivedi and David M. Zimmer (2007). *Copula Modeling: An Introduction for Practitioners*. World Scientific Publishing.
- [20] Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*. Seattle, Washington: Springer.
- [21] Shih, J. H. and Louis, T. A. (1995). *Inferences on the Association Parameter in Copula Models for Bivariate Survival Data*. Biometrics, 51, 1384-1399.
- [22] Steinbrecher, G. and Shaw W. (2008). *Quantile mechanics*. European Journal of Applied Mathematics; Cambridge Tomo 19, No 2, 87-112.
- [23] Yan, J. (2007). *Enjoy the Joy of Copulas: With a Package copula*. Journal of Statistical Software, Volume 21, Issue 4.
- [24] Zou, K. H.; Aiyi, L.; Bandos, A. I. (2012). *Statistical Evaluation of Diagnostic Performance Topics in ROC Analysis*. Chapman and Hall/CRC Biostatistics Series.



Casa abierta al tiempo
UNIVERSIDAD AUTÓNOMA METROPOLITANA

ACTA DE EXAMEN DE GRADO

No 00187
Matrícula: 2152800936

Curva ROC Incidente/Dinámica
vista desde el enfoque de la
Cópula Normal Bivariada.

En la Ciudad de México, se presentaron a las 13:00 horas del día 6 del mes de septiembre del año 2019 en la Unidad Iztapalapa de la Universidad Autónoma Metropolitana, los suscritos miembros del jurado:

DR. GABRIEL ESCARELA PEREZ
DRA. HORTENSIA JOSEFINA REYES CERVANTES
DR. GABRIEL NUÑEZ ANTONIO

Bajo la Presidencia del primero y con carácter de Secretario el último, se reunieron para proceder al Examen de Grado cuya denominación aparece al margen, para la obtención del grado de:

MAESTRO EN CIENCIAS (MATEMÁTICAS)

DE: DAVID VASQUEZ HERRERA

y de acuerdo con el artículo 78 fracción III del Reglamento de Estudios Superiores de la Universidad Autónoma Metropolitana, los miembros del jurado resolvieron:

APROBAR

Acto continuo, el presidente del jurado comunicó al interesado el resultado de la evaluación y, en caso aprobatorio, le fue tomada la protesta.



DAVID VASQUEZ HERRERA
ALUMNO

REVISÓ

MTRA. ROSALIA SERRANO DE LA PAZ
DIRECTORA DE SISTEMAS ESCOLARES

DIRECTOR DE LA DIVISIÓN DE CBI

DR. JESUS ALBERTO OCHOA TAPIA

PRESIDENTE

DR. GABRIEL ESCARELA PEREZ

VOCAL

DRA. HORTENSIA JOSEFINA REYES
CERVANTES

SECRETARIO

DR. GABRIEL NUÑEZ ANTONIO