



CENTRO DE INVESTIGACIÓN EN MATEMÁTICAS A.C

---

---

**MODELOS GRÁFICOS NO DIRIGIDOS VÍA CÓPULAS CONDICIONALES  
NO PARAMÉTRICAS**

**T E S I S**

QUE PARA OBTENER EL GRADO DE:  
**MAESTRO EN CIENCIAS CON ESPECIALIDAD EN PROBABILIDAD Y ESTADÍSTICA**

PRESENTA:  
**ISRAEL EMMANUEL AMBRIZ LOBATO**

CODIRECTORES DE TESIS:  
**DR. ROGELIO RAMOS QUIROGA  
DR. ROLANDO JOSÉ BISCAY LIRIO**

2016



Sí, me contradigo ¿y qué?  
Yo soy inmenso y contengo multitudes.

Walt Whitman



Para las mujeres que más amo y admiro:

Ángeles Lobato

Fer Ambriz

Estefanía Lugo.



## Agradecimientos

A mis padres, Ángeles Lobato y José Luis Ambriz: son el pilar de mi vida.

A toda mi familia: son seres maravillosos, fuente de amor y sabiduría.

A mis asesores de tesis, Dres. Rogelio Ramos y Rolando Biscay: agradezco mucho su tiempo, compromiso y paciencia, pero sobre todo su calidez humana y gran apoyo.

A mis sinodales, Dres. Arturo Erdely y Miguel Nakamura: además de agradecerles por su tiempo y las aportaciones que le hicieron a mi trabajo, agradezco mucho su amplia disposición para orientar mi carrera profesional.

Al CIMAT: por la oportunidad de obtener los estudios de maestría.

Al cuerpo de investigadores de Probabilidad y Estadística: además de todo el conocimiento que me compartieron, la calidad humana con la que cuentan es invaluable. Quiero agradecer especialmente a los Dres. Víctor Pérez Abreu y Enrique Villa, además a la Dra. Graciela González; personas que me brindaron su apoyo y me orientaron cuando lo necesité.

A mis grandes amigos, Manuel Pedraza, Lilia Rivera, Lilia López e Ilce Medina: por todos esos momentos de buena vibra y por todo el apoyo; son excelentes personas que siempre llevo en mi corazón.

A Estefanía Lugo: muchas gracias por estos ya más de tres años de locuras maravillosas, el tren ha llegado mucho más lejos de lo que me alcancé a imaginar en aquel momento y ha visitado lugares increíbles.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT): por la beca para completar mis estudios de maestría.





# Índice general

<b>Glosario</b>	<b>III</b>
<b>Notación</b>	<b>V</b>
<b>Introducción</b>	<b>VII</b>
<b>1. Modelos gráficos no dirigidos</b>	<b>1</b>
1.1. Organización del capítulo	1
1.2. Modelos gráficos	1
1.3. Modelos gráficos no dirigidos	3
Propiedades de Markov para gráficas no dirigidas	3
1.4. Aplicaciones	4
Análisis descriptivo	4
Selección de variables	4
Estimación	5
Construcción de distribuciones conjuntas	5
1.5. Casos de aplicación	6
1.6. Modelos gráficos no dirigidos triviales	6
1.7. Modelo gráfico no dirigido gaussiano	7
GIC gaussiana	7
1.8. Discusión	8
<b>2. Cópulas: multivariadas y condicionales</b>	<b>11</b>
2.1. Organización del capítulo	11
2.2. Preliminares: vectores aleatorios y funciones de distribución multivariadas	12
2.3. Cópulas: definición y propiedades elementales	14
2.4. Teorema de Sklar	15
2.5. Cópulas y vectores aleatorios	17
2.6. Dependencia vía cópulas	17
Propiedades de una medida de dependencia	17
Medidas de dependencia basadas en cópulas	18
Dependencia en colas	19
2.7. Cópulas condicionales	20
Panorama general	20
Extensión del Teorema de Sklar	21
Supuesto simplificador	21
2.8. Discusión	22
<b>3. Modelos gráficos no dirigidos de cópulas simplificadas</b>	<b>23</b>
3.1. Organización del capítulo	23
3.2. Modelos gráficos no dirigidos de cópulas simplificadas	23
3.3. Cópulas elípticas simplificadas	24
Independencia en cópulas elípticas	25
Cópula gaussiana	25

Cópula T . . . . .	28
3.4. Cópulas Arquimedianas simplificadas . . . . .	33
Cópula Clayton . . . . .	33
3.5. Limitantes de los modelos de cópula gaussiana . . . . .	34
3.6. Discusión . . . . .	36
<b>4. Modelos gráficos no dirigidos vía cópulas condicionales</b>	<b>37</b>
4.1. Organización del capítulo . . . . .	38
4.2. Preliminares . . . . .	38
Independencia condicional vía cópulas . . . . .	38
Fórmula de Joe . . . . .	39
4.3. $\alpha$ -estimador del grafo no dirigido vía cópulas condicionales . . . . .	40
Extracción iterativa de observaciones de cópulas condicionales . . . . .	41
4.4. Control en la tasa de descubrimientos falsos (CTDF) . . . . .	42
Cotas para la TDF ante pruebas dependientes . . . . .	44
4.5. Enfoque paramétrico de cópula multivariada . . . . .	45
4.6. Estimador Bernstein del grafo no dirigido . . . . .	46
4.7. Estudio de simulación . . . . .	47
Grado de escasez y aleatoriedad en el patrón de escasez . . . . .	48
Matrices de correlación positivas definidas: cópulas gaussiana y $\mathcal{T}$ . . . . .	50
Simulación de muestras escasas de cópula gaussiana y $\mathcal{T}$ . . . . .	51
El desempeño del estimador . . . . .	51
Estructura del estudio de simulación . . . . .	52
4.8. Resultados: curvas ROC, TEG y CTDF . . . . .	53
Curvas ROC . . . . .	53
Tasa del error global (TEG) . . . . .	58
CTDF en el estudio de simulación . . . . .	60
4.9. El supuesto simplificador en la estimación del grafo no dirigido . . . . .	62
4.10. Discusión . . . . .	63
<b>5. Modelos gráficos no dirigidos en altas dimensiones</b>	<b>65</b>
5.1. Organización del capítulo . . . . .	66
5.2. La problemática en altas dimensiones . . . . .	66
5.3. Algunos enfoques destacados de estimación . . . . .	68
Modelos gráficos escasos de cópula elíptica: gaussiana y $\mathcal{T}$ . . . . .	69
5.4. Escasez vía relaciones subyacentes del grafo no dirigido . . . . .	69
Parsimonia vía el principio de escasez: la propiedad de Markov de la distribución . . . . .	70
Estimación no paramétrica vía de la matriz de dependencia condicional penalizada . . . . .	71
Una extensión adicional del principio de escasez y de parsimonia . . . . .	72
5.5. Discusión . . . . .	73
<b>Conclusiones</b>	<b>75</b>
<b>Apéndices</b>	<b>77</b>
<b>A. Distribución NonParaNormal y el modelo de cópula gaussiana</b>	<b>79</b>
A.1. Transformaciones en NPN . . . . .	79
A.2. NPN y cópula gaussiana . . . . .	79
<b>B. Condicionales de cópula gaussiana</b>	<b>81</b>
B.1. Condicionales de cópula gaussiana . . . . .	81
<b>C. Apéndice C: prueba para la fórmula de Joe</b>	<b>87</b>
<b>Bibliografía</b>	<b>89</b>

# Glosario

<b>Abreviación</b>	<b>Descripción</b>
v.a.	Vector aleatorio (por alguna generalización puede ser univariado).
FDA	función de distribución acumulada (o función de distribución).
GIC	Gráfica de independencia condicional.
NPN	Familia de distribución <i>NonParaNormal</i> .
TFP	Tasa de falsos positivos.
TFN	Tasa de falsos negativos.
TVP	Tasa de verdaderos positivos.
TVN	Tasa de verdaderos negativos.
TEG	Tasa de error global.
TDF	Tasa de descubrimientos falsos ( <i>FDR, False Discovery Rate</i> ).
CTDF	Control en la tasa de descubrimientos falsos.
ROC	<i>Receiver Operating Characteristic</i> .
LASSO	<i>Least Absolute Shrinkage and Selection Operator</i> .
p.d.	Por demostrar.
t.q.	Tal que.



# Notación

<b>Símbolo</b>	<b>Descripción</b>
$p$	Dimensión de un vector aleatorio.
$N$	Tamaño de muestra.
$\mathbb{R}$	Conjunto de los números reales, i.e. $\mathbb{R} = (-\infty, \infty)$ .
$\overline{\mathbb{R}}$	Conjunto de los números reales extendidos, i.e. $\overline{\mathbb{R}} = [-\infty, \infty]$ .
$\mathbb{N}$	Conjunto de los números naturales, i.e. $\mathbb{N} = \{1, 2, \dots\}$ .
$\mathbb{I}$	Intervalo unitario $[0, 1]$ .
$\text{card}(\cdot)$	Operador de cardinalidad de un conjunto, i.e. el número de elementos en éste.
$\mathbf{0}$	Vector nulo, i.e. $\mathbf{0} = (0, \dots, 0)^T \in \mathbb{R}^p$ .
$\mathbf{1}$	Vector unitario, i.e. $\mathbf{1} = (1, \dots, 1)^T \in \mathbb{R}^p$ .
$\mathcal{I}_p$	Matriz identidad de dimensiones $p \times p$ .
$\text{diag}(\cdot)$	Diagonal de una matriz cuadrada.
$\mathbf{P}$	Distribución de probabilidad.
$\mathbb{E}(\cdot)$	Esperanza de un v.a.
$\text{Cov}(\cdot)$	Matriz de covarianzas de un v.a.
$\text{Var}$	Varianza de una variable aleatoria.
$X, Y, X_j, Y_j$	Variabes aleatorias univariadas (vectores aleatorios de dimensión unitaria).
$\mathbf{X}$	Vector aleatorio (dimensión $p \geq 2$ ), v.g. $\mathbf{X} = (X_1, \dots, X_p)^T$ .
$\mathbf{X}_A$	Vector aleatorio cuyas variables perteneces al conjunto de índices en $A$ .
$x$	Observación de una variable aleatoria, i.e. $x \in \mathbb{R}$ (ó $\mathbf{x} \in \overline{\mathbb{R}}$ ).
$\mathbf{x}$	Observación de un vector aleatorio, i.e. $\mathbf{x} \in \mathbb{R}^p$ (ó $\mathbf{x} \in \overline{\mathbb{R}}^p$ ) con $p \geq 2$ .
$F$	FDA univariada genérica.
$F_j$	FDA univariada asociada a la variable $j$ de un v.a. $\mathbf{X} = (X_1, \dots, X_p)^T$ .
$H$	FDA multivariada genérica.
$F^{(-1)}$	Inversa generalizada de $F$ .
$f_j$	Función de densidad univariada de la variable aleatoria $X_j$ .
$f_A$	Función de densidad conjunta del v.a. $\mathbf{X}_A$ con $A$ un conjunto de índices.
$\text{Ran } F, \text{Ran } \mathbf{X}$	Rango de $F$ o Rango de $\mathbf{X}$ , en general el rango de un vector o variable aleatorios.
$\mathcal{G}$	Grafo no dirigido.
$\mathcal{A}$	Conjunto de aristas de un modelo gráfico no dirigido.
$\mathcal{V}$	Conjunto de vértices de un modelo gráfico no dirigido, v.g. $\mathcal{V} = \{1, \dots, p\}$ .
$\text{ady}(\cdot, \cdot)$	Conjunto de adyacencia (o de vecinos) de un vértice en un grafo no dirigido.
$\mathcal{N}$	Distribución gaussiana.
$\mathcal{T}$	Distribución T.
$\mathcal{GI}$	Distribución gama inversa.
$NPN$	Distribución <i>NonParaNormal</i> .
$\mu$	Vector de medias, i.e. $\mu = (\mu_1, \dots, \mu_p)^T$ .
$P$	Matriz de covarianzas.
$\Sigma$	Matriz de coeficientes de correlación lineal de Pearson.
$\Omega$	Matriz de precisión, $\Omega = P^{-1}$ u $\Omega = \Sigma^{-1}$ , este último para modelos estandarizados.
$\nu$	Grados de libertad, asociado a modelos de cópula $\mathcal{T}$ .

<b>Símbolo</b>	<b>Descripción</b>
$\rho_{\cdot}$	Coefficiente de correlación lineal de Pearson.
$\rho_{\cdot, \cdot}$	Coefficiente de correlación lineal parcial.
$\text{Corr}(\cdot, \cdot)$	Operador del coeficiente de correlación lineal de Pearson para un par de variables aleatorias.
$\text{Corr}(\cdot)$	Matriz de correlaciones lineales de Pearson de un v.a.
$\rho_S$	Coefficiente de concordancia Rho de Spearman.
$g(\cdot, \dots, \cdot)$	Densidad genérica de un modelo de distribución gaussiana.
$h(\cdot, \dots, \cdot)$	Densidad genérica de un modelo de distribución $\mathcal{T}$ .
$\phi$	Densidad gaussiana univariada estándar.
$\Phi$	FDA gaussiana univariada estándar.
$t_\nu$	FDA $\mathcal{T}$ univariada estándar, i.e. con media cero y varianza $\nu/(\nu - 2)$ .
$\Phi_\Sigma$	FDA gaussiana multivariada con matriz de correlaciones $\Sigma$ y vector de medias $\mathbf{0}$ .
$\mathcal{C}$	Función cópula.
$\Pi$	Cópula producto, i.e. $\Pi(u, v) = uv$ .
$\mathcal{C}^G$	Cópula gaussiana.
$\mathcal{C}^T$	Cópula $\mathcal{T}$ .
$\mathcal{C}^C$	Cópula Clayton.
$\beta$	Vector de coeficientes de regresión, i.e. $\beta = (\beta_1, \dots, \beta_p)^T$ .
$\beta_k^{(j)}$	Coefficiente de regresión asociado a la variable $X_k$ con variable de respuesta $X_j$ .
$\mathcal{B}$	Matriz de coeficientes de regresión, i.e. $\mathcal{B}_{jk} = \beta_k^{(j)}$ .
$\alpha, \lambda$	Parámetros de ajuste.
$L_p$	Distancia de una cópula $\mathcal{C}$ a $\Pi$ inducida por $p \in [1, \infty)$ .
$\ell_q$	Norma inducida por $q \in [1, \infty]$ .

# Introducción

La era de la información en la que vivimos desde hace un par de décadas, ha traído consigo *estructuras cada vez más complejas de datos* (c.f. [Donoho et al. \(2000\)](#)), las cuales se han traducido en nuevos *retos de inferencia estadística* y en general matemáticos. Actualmente la estructura de los datos está *dejando de ser compatible* con el diseño de las metodologías convencionales; lo anterior justifica que en años recientes hemos presenciado una gran explosión de nuevas propuestas que modifican o extienden, de algún modo, los paradigmas clásicos de inferencia con la finalidad de buscar herramientas afines a las estructuras modernas de datos. Datos que no pertenecen a espacios euclidianos o *cantidades masivas de variables*, éstas últimas sobre todo con *muy pocas observaciones en muestra*, son ejemplos de estructuras complejas. Hoy en día cada vez son más las áreas del conocimiento en donde se reconoce la presencia de datos complejos, v.g. en genética, medicina, finanzas, por sólo mencionar algunas. En particular la presente investigación gira en torno a los *modelos gráficos no dirigidos markovianos*, los cuales han atraído una enorme atención en años recientes como una *herramienta visualmente atractiva* para *resumir relaciones* existentes entre una *gran cantidad* de variables, problema de interés en diversos contextos modernos de inferencia estadística.

Un modelo gráfico no dirigido es una estructura que representa *relaciones de independencia condicional* entre variables aleatorias. Intuitivamente lo podemos concebir como la representación de una “*red social de variables*” (usualmente con un gran número de elementos): dos variables que no están “*conectadas*” directamente pueden ser dependientes a partir de un tercero que las conecte, entonces estas variables serán independientes dado el conocimiento de la tercera. Detectar al tercer elemento que induce dependencia entre variables que por su naturaleza no tienen ninguna relación entre sí, es la base para la determinación de relaciones espuria y una de las aplicaciones inmediatas de un grafo no dirigido. Además de la relevancia de un grafo no dirigido como *herramienta descriptiva*, éste tiene una gama de *aplicaciones* importantes para la inferencia en diversos modelos estadísticos, v.g. para la construcción de distribuciones conjuntas y estimación de parámetros. De igual forma, estos modelos se han convertido en una herramienta fundamental en distintas áreas del conocimiento, v.g. en conectividad cerebral y análisis de redes sociales, con gran potencial de aplicación en cualquier área donde el análisis involucre *datos multivariados*.

Los datos multivariados de antemano presentan *estructuras complejas de dependencia*, las cuales difícilmente podrían ser explicadas por los modelos clásicos de distribución conjunta (gausiana,  $\mathcal{T}$ , etc.). Esta problemática se presenta con mayor énfasis en situaciones de *inferencia a gran escala* (c.f. [Efron \(2010\)](#)), usualmente relacionadas con la presencia de *cantidades masivas de variables*, contexto de análisis que generalmente acarrea problemas *altamente dimensionales* (c.f. [Bühlmann and Van De Geer \(2011\)](#)), i.e. cuando el número de variables,  $p$ , supera (en ocasiones considerablemente) el tamaño de la muestra observada,  $N$  ( $N \ll p$ ); cabe destacar que actualmente esta estructura de datos surge de manera natural en diversos contextos de análisis. Pese a que los modelos gráficos no dirigidos tienen su *máximo potencial de aplicación* ante la estructura de datos antes planteada, el supuesto recurrente del cual se suele partir para su estimación es el de *distribución gaussiana multivariada*. Como modelo conjunto, una distribución gaussiana es demasiado restrictiva, no sólo por el hecho de que impone marginales en la misma familia, sino por la *estructura poco flexible de dependencia* que induce, la cual está inducida por el coeficiente de *correlación lineal de Pearson*, donde éste último es simplemente una medida de asociación lineal con un campo de acción bastante limitado (c.f. [Embrechts et al. \(1999\)](#)). El coeficiente de correlación de Pearson *no cumple con ser una medida de dependencia*, así que *en general* no es una herramienta pertinente para *detectar* relaciones de independencia o independencia condicional entre variables; sólo bajo el supuesto de distribución gaussiana la *correlación lineal nula* caracteriza la independencia. Sin embargo, y pese al conocimiento de la comunidad estadística de que el supuesto gaussiano multivariado es prácticamente inalcanzable con datos observados (c.f. [Patton \(2002\)](#)), debido a su *fácil manejo e interpretación*, los modelos gaussianos siguen vigente en distintos terrenos de la investigación contemporánea. Puntualmente el modelo gráfico no dirigido gaussiano tiene diversas *bondades prácticas*, dentro de las que destacan

las siguientes: las relaciones de independencia condicional modeladas en el grafo tienen una *codificación explícita* en los parámetros de la distribución (c.f. [Dempster \(1972\)](#)), lo anterior facilita la inferencia del grafo no dirigido; por otro lado, la estructura del grafo tiene una estrecha relación con el problema de *selección de variables* en el modelo de regresión lineal (c.f. [Lauritzen \(1996\)](#)), y así la estimación del modelo gráfico tiene una aplicación directa a otros modelos estadísticos. Pese al gran potencial que los modelos gráficos no dirigidos tienen en diversos análisis con datos multivariados, éstos pueden llegar a *perder terreno de acción, aplicabilidad e interpretación* cuando el modelo de probabilidad subyacente es bastante restrictivo como modelo conjunto.

Partiendo del creciente auge que los modelos gráficos no dirigidos siguen teniendo, tanto en la investigación así como en las aplicaciones, observamos que existe una gran necesidad de contar con metodologías de estimación del grafo cuya *inferencia no limite* la estructura de dependencia general del modelo de probabilidad asociado. Por otro lado, con base en la gran relevancia que desde hace ya algunas décadas han cobrado los *modelos de cópulas* por su capacidad para construir distribuciones multivariados bastante *flexibles* (c.f. [Durante and Sempi \(2010\)](#)), junto con el hecho de que resultan una vía natural para cuantificar el *grado de dependencia* en un vector aleatorio (c.f. [Mari and Kotz \(2001\)](#)), es que basar la inferencia de las relaciones de independencia condicional modeladas en un grafo no dirigido en la cópula subyacente de los datos resulta una vía bastante pertinente y atractiva. Motivados por la problemática antes planteada y el potencial de los modelos de cópulas, es que el *objetivo global* de la presente investigación es buscar la *estimación* del modelo gráfico no dirigido partiendo de la *cópula subyacente* del vector aleatorio de interés, a manera que el modelo de probabilidad subyacente sea lo *suficientemente flexible* como para representar *estructuras complejas de dependencia* que inevitablemente se encuentran en *datos multivariados*.

En una primera instancia, si asumimos *un solo modelo* de cópula paramétrica que explique la *estructura completa de dependencia* de un vector aleatorio, i.e. que la cópula subyacente al vector pertenezca a una familia paramétrica específica, para que la deducción del grafo no dirigido tenga sentido, tendremos que buscar familias de cópulas que cumplan con el *supuesto simplificador*. Diremos que una cópula o una familia de cópulas cumple el supuesto simplificador, si todos sus cópulas condicionales a pares dependen del condicionante sólo vía sus distribuciones condicionales (c.f. [Haff et al. \(2010\)](#)), en otras palabras, que sus cópulas condicionales no cambian con el valor del condicionante. Sin embargo, en [Stöber et al. \(2013\)](#) se asegura que las tres familias de cópulas que cumplen con el supuesto simplificador son la *gaussiana*, la  $\mathcal{T}$  y la *Clayton*. Así que las grandes interrogantes radican en descubrir si estas tres cópulas son lo suficientemente flexibles como modelos conjuntos y si además sus modelos gráficos no dirigidos cuentan con *utilidad práctica*, i.e. que los grafos involucrados no sean *casos triviales* (completos o vacíos). Estudiar a estos tres modelos de cópulas en términos de su estructura de dependencia, su comportamiento condicional y su grafo no dirigido asociado, para posteriormente, concluir sobre la flexibilidad del modelo conjunto y la utilidad del grafo, es una de las *contribuciones* que hace la investigación. Para el caso de las dos cópulas elípticas (gaussiana y  $\mathcal{T}$ ) haremos mención de algunos resultados en [Liu et al. \(2009\)](#), [Liu et al. \(2012a\)](#) y [Liu et al. \(2012b\)](#), cuestionando el campo de aplicación de este último. Por otro lado, con el creciente impacto que han ido ganando los modelos de *vine cópula* en años recientes (c.f. [Aas and Berg \(2009\)](#), [Czado \(2010\)](#) y [Min and Czado \(2011\)](#)), se *cuestiona rotundamente* el hecho de que una sola cópula multivariada perteneciente a una familia paramétrica sea capaz de modelar estructuras complejas de dependencia, sobre todo para un gran número de variables.

El diseño de cópulas sumamente flexibles hasta hace menos de una década se encontraba en su mayoría *limitado al caso bivariado*, ganando una diversidad de metodologías con gran potencial para esta dimensión. No obstante para la deducción de cópulas multivariadas ( $p \geq 3$ ), las extensiones del caso bivariado se han visto limitadas en cuanto a flexibilidad (c.f. [Aas et al. \(2009\)](#) y [Min and Czado \(2011\)](#) para discusiones al respecto). El gran paso para la inferencia de cópulas multivariadas con la capacidad de modelar diversas y complejas estructuras de dependencia fue con las *construcciones de cópulas a pares* en el enfoque de *vine cópula*, ya que en éste se logra la estimación de distribuciones conjuntas vía la inferencia de una serie de cópulas bivariadas (condicionales y no condicionales), logrado así *modelos multivariados notablemente flexibles*.

Dos de las *herramientas fundamentales* en *vine cópula* son la fórmula de [Joe \(1996\)](#) y el supuesto simplificador. La *fórmula de Joe* es una vía para la *extracción de observaciones* de cualquier *cópula condicional*, en particular aquellas involucradas en el grafo no dirigido. Haciendo uso de la fórmula de *Joe* y asumiendo el supuesto simplificador, es que se propone un *algoritmo iterativo para la estimación del grafo no dirigido basado en pruebas múltiples* de independencia, aplicadas a observaciones de las  $\binom{p}{2}$  cópulas condicionales asociadas al grafo, estas últimas extraídas de los datos originales. Como parte esencial de nuestro planteamiento, analizamos los *tipos de error* que induce nuestra



metodología de estimación y en particular para un *control en la tasa de descubrimientos falsos* (c.f. Benjamini and Hochberg (1995)) sugerimos el uso de algunos de los resultados en Guo and Rao (2008) para *pruebas dependientes*. La tasa de descubrimientos falsos en el fondo es una extensión multivariada del conocido error tipo I en pruebas de hipótesis, en el contexto de estimación del grafo, éste controla el *sobreajuste* de su estimación. La naturaleza de nuestra metodología induce la incorporación de un valor  $\alpha$  que domina la tasa de descubrimientos falsos y por ende el *sobreajuste*; a este valor lo conoceremos como *parámetro de ajuste* por ser aquel que modula el *grado de escasez* de la estimación, i.e. la cantidad de relaciones de independencia condicional representadas en el grafo. Además, aprovechando la gran flexibilidad y viabilidad computacional del *estimador Bernstein de la cópula* (c.f. Sancetta and Satchell (2004)) para el caso bivariado y el algoritmo formulado, proponemos una *estimación totalmente no paramétrica*: el estimador Bernstein del grafo no dirigido. Con la finalidad de evaluar nuestro planteamiento de estimación, en particular el desempeño del estimador Bernstein, llevamos a cabo un *estudio de simulación* basado en la inferencia del grafo no dirigido para muestras escasas de cópulas simplificadas donde el grafo no dirigido es útil como modelo. En el contexto del estudio de simulación, como *contribución adicional*, proponemos una vía para la *construcción de matrices de correlación positivas definidas* con patrones específicos de escasez, lo anterior sin imponer estructura de correlación alguna o limitar el grado máximo del grafo. El estimador Bernstein del grafo no dirigido conjuga los fundamentos de vine cópula junto con la flexibilidad que brinda el estimador Bernstein de cópulas bivariadas, logrando de igual manera flexibilidad en el modelo de probabilidad subyacente del grafo estimado; lo anterior constituye a este estimador como la *principal contribución* de la presente investigación.

Debido a la notable relevancia que han ganado los modelos gráficos no dirigidos para la inferencia en diversos contextos de aplicación donde surgen *datos altamente dimensionales* (c.f. Bühlmann and Van De Geer (2011)), como complemento a nuestra investigación, partiendo de la *propiedad de Markov del vector aleatorio*, *sentamos las bases conceptuales* para la estimación y aplicaciones del grafo no dirigido en altas dimensiones en un enfoque no paramétrico, con la finalidad de contar con modelos de probabilidad asociados que sean capaces de representar los patrones de dependencia provenientes de esta estructura moderna de datos. La estimación del modelo gráfico se propondrá buscando *penalizar la inferencia* por un parámetro de ajuste partiendo del *principio de escasez*; esta premisa ha resultado bastante útil para el desarrollo de nuevas metodologías y un supuesto realista ante este contexto de análisis. Por otro lado, más que hacia modelos penalizados de regresión lineal por la media, planteamos el uso del grafo no dirigido para la *construcción de distribuciones condicionales siguiendo el principio de parsimonia*; lo anterior en el sentido de explicar el comportamiento aleatorio de una *variable de respuesta* en función del *menor número de predictores* evitando redundancia de información. Naturalmente vía la distribución condicional se pueden conceptualizar regresiones por los cuantiles de ésta, logrando explicar de manera más global el comportamiento aleatorio de la variable de respuesta al cambio en los predictores. De igual manera, proponemos una extensión adicional a los principios de escasez y parsimonia, alineados con la problemática de *selección de modelo* en vine cópula. Pese a que estos últimos planteamientos se presentan simplemente a *nivel conceptual*, consideramos que pueden ser base para la extensión de algunos conceptos relacionados con la inferencia general en altas dimensiones y que en algunos casos *mantienen vigente* la aplicación del modelo gráfico no dirigido markoviano.

La presente investigación centra sus esfuerzos en el estudio y propuesta de alternativas de estimación del grafo no dirigido, cuyo modelo de probabilidad asociado sea lo suficientemente flexible como para ser capaz de responder a diversos patrones de dependencia que se encuentren en datos multivariados, buscando que los enfoques sean aplicables a contextos con cantidades masivas de variables y en particular a datos altamente dimensionales. En el Capítulo 1 se presentan a los modelos gráficos no dirigidos markovianos, enfatizando en lo limitado de la distribución gaussiana como modelo conjunto y la vía recurrente para la estimación del grafo. El Capítulo 2 sienta las bases de la teoría de cópulas multivariadas, puntualizando sobre su potencial para la construcción de distribuciones conjuntas y medidas de dependencia, se introduce el concepto de cópula condicional, el cual junto con el supuesto simplificador son el punto de partida para la conceptualización del grafo no dirigido vía cópulas. El Capítulo 3 recopila una serie de resultados con la finalidad de estudiar a las tres familias de cópulas simplificadas en tres aspectos esenciales: la estructura de dependencia que modelan con base en sus parámetros, el comportamiento de sus cópulas condicionales y el modelo gráfico no dirigido asociado a cada una de éstas; a grandes rasgos nuestro interés radica en evaluar la flexibilidad de los modelos de cópulas simplificadas y conocer la utilidad de su grafo no dirigido. Como parte central de este trabajo, en el Capítulo 4 se propone la estimación vía cópulas condicionales de donde surge una perspectiva totalmente no paramétrica: el estimador Bernstein del grafo no dirigido; además se complementan las propuestas con un estudio de simulación para evaluar su desempeño. Finalmente, el Capítulo 5 ofrece un panorama general sobre la problemática de la inferencia estadística en altas dimensiones, puntualizando en los enfoques existentes para la

estimación del grafo no dirigido, además se formulan líneas de investigación futura para la inferencia y aplicación del grafo no dirigido ante este contexto de análisis.

# Capítulo 1

## Modelos gráficos no dirigidos

**Resumen** Los modelos gráficos no dirigidos markovianos han cobrado una relevancia importante en los últimos años, sobre todo en contextos en donde se cuente con *cantidades masivas de variables* y el objetivo sea *inferir asociaciones* entre éstas. Un grafo no dirigido resume asociaciones entre variables vía *relaciones de independencia condicional* logrando un gran impacto visual. Además de ser una potencial herramienta descriptiva, dentro de las razones por las cuales estos modelos se han vuelto populares, destacan sus *aplicaciones* a la selección de variables y a la estimación de parámetros, principalmente en contextos de altas dimensiones ( $N \ll p$ ). Además se han convertido en la herramienta ad hoc para explicar algunas relaciones de interés en distintas áreas del conocimiento, v.g. conectividad cerebral e inteligencia artificial. Sin embargo, para lograr la estimación del grafo, el supuesto recurrente es el de *distribución gaussiana multivariada*, debido a su fácil manejo e interpretación, pese a que puede llegar a ser un modelo bastante restrictivo tanto en términos de la *estructura de dependencia* que representa, así como por el hecho que impone marginales en la misma familia. En este capítulo presentamos a los modelos gráficos no dirigidos, discutimos sobre sus aplicaciones y citamos ejemplos concretos en donde se han convertido en la herramienta fundamental de análisis, posteriormente, presentamos a los modelos gráficos gaussianos y argumentamos sobre sus ventajas prácticas y limitantes subyacentes del modelo conjunto.

### 1.1. Organización del capítulo

El presente capítulo tiene la intención de definir a los modelos gráficos no dirigidos markovianos, i.e. aquellos que cumplen las propiedades de Markov, para luego generar una discusión sobre su importancia y los supuestos que usualmente se hacen para su estimación. La Sección 1.2 da un panorama general sobre los modelos gráficos y su relevancia. La Sección 1.3 presenta a los modelos gráficos no dirigidos y puntualiza sobre las propiedades de Markov. La Sección 1.4 expone las aplicaciones más destacadas que tienen los modelos gráficos no dirigidos markovianos, para después, en la Sección 1.5 citar algunos artículos en donde éstos se han convertido en el modelo ad hoc para distintas áreas del conocimiento. En la Sección 1.6 se explica a lo que nos referiremos en lo subsecuente del documento con un grafo trivial y por que éste es importante para los objetivos de nuestra investigación. En la Sección 1.7 se presenta al modelo gráfico gaussiano y sus aplicaciones. Finalmente, en la Sección 1.8 se discute sobre las limitantes inmediatas que tiene un modelo de distribución gaussiana y se comenta brevemente sobre alternativas que se estudiarán en los capítulos subsecuentes para construir grafos no dirigidos markovianos con distribuciones conjuntas más flexibles.

### 1.2. Modelos gráficos

Los modelos gráficos son capaces de representar diversas estructuras de dependencia (o independencia) entre variables aleatorias, teniendo un amplio campo de aplicación, principalmente en problemas en donde se cuente con un gran número de variables que interactúen en sistemas complejos. Tomando en consideración que vivimos en la era de la información, que nos permite obtener y manipular grandes cantidades de datos de diversas fuentes, los modelos gráficos han cobrado relevancia en los últimos años debido a su atractivo visual para describir interrelaciones entre variables junto con las aplicaciones que tienen hacia otros modelos estadísticos, v.g. selección de variables. Por lo anterior es que los modelos gráficos son un tema actual de investigación en estadística, computación y matemáticas; además de que se han convertido en una herramienta importante en bastantes áreas del conocimiento,

v.g. neurociencias, inteligencia artificial, bioinformática y sociología, por sólo mencionar algunas. Algunos de los libros en donde se da un panorama general sobre el tema son: Lauritzen (1996), Edwards (2012), Koller and Friedman (2009) y Borgelt et al. (2009). Para los alcances de la presente investigación sólo consideraremos modelos de gráficas no dirigidas. Formalizaremos los conceptos fundamentales con las siguientes definiciones.

**Definición 1.2.1 Grafo.** Un *grafo*,  $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ , es una estructura que consiste de un conjunto finito de *vértices*, digamos  $\mathcal{V} := \{1, \dots, p\}$ , y un conjunto finito de *aristas*,  $\mathcal{A}$ . El conjunto de aristas es un subconjunto de  $(\mathcal{V} \times \mathcal{V}) \setminus \{(j, j) \mid j \in \mathcal{V}\}$ . Las aristas en  $\mathcal{A}$  representan enlaces entre los vértices de  $\mathcal{V}$ , es por eso que no se permiten los pares ordenados con elementos iguales como  $(j, j)$ .

**Definición 1.2.2 Modelo gráfico.** En un *modelo gráfico*,  $(\mathcal{G}, \mathbf{P})$ , los vértices en la gráfica  $\mathcal{G}$  representan a la colección de *variables aleatorias* en un *vector aleatorio* (v.a.) de dimensión  $p$ ,

$$\mathbf{X} = (X_1, \dots, X_p)^T \sim \mathbf{P},$$

donde  $\mathbf{P}$  es la distribución de probabilidad de  $\mathbf{X}$ .

El objeto de interés en esta investigación son los *modelos gráficos no dirigidos*, los cuales definimos a continuación.

**Definición 1.2.3 Modelo gráfico no dirigido.** Sea el modelo gráfico  $(\mathcal{G}, \mathbf{P})$  con  $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ . Diremos que una *arista*  $(j, k) \in \mathcal{A}$  es *no dirigida* si  $(k, j) \in \mathcal{A}$ . Por otro lado, una *gráfica*  $\mathcal{G}$  diremos que es no dirigida si *todas sus aristas son no dirigidas*. Finalmente el modelo  $(\mathcal{G}, \mathbf{P})$ , tal que  $\mathcal{G}$  sea no dirigida, diremos que es un *modelo gráfico no dirigido*.

#### Observaciones 1.2.1 :

**(Obs. 1)** En mucha de la literatura sobre modelos gráficos no dirigidos se ocupa la convención de que los pares ordenados que representan el mismo enlace entre las variables, v.g.  $(j, k) \in \mathcal{A}$  y su contraparte  $(k, j) \in \mathcal{A}$ , se consideren como el mismo elemento dentro del conjunto de aristas, i.e.  $(j, k) = (k, j)$  (c.f. (Edwards, 2012)). El anterior, aunque es un detalle de notación, es importante tenerlo en cuenta para no caer en definiciones ambiguas. Con la finalidad de lograr consistencia en nuestras definiciones y propuestas, el presente documento estará alineado a la Definición 1.2.3.

**(Obs. 2)** A partir de este momento, haremos uso de la abreviación v.a. para referirnos a un *vector aleatorio*.

Dado un v.a. de dimensión cinco,  $\mathbf{X} = (X_1, X_2, X_3, X_4, X_5)^T$ , la Figura 1.1 ejemplifica un posible modelo gráfico no dirigido asociado al modelo de probabilidad del v.a. en cuestión,  $\mathbf{P}$ . Entonces el modelo  $(\mathcal{G}, \mathbf{P})$  es t.q.  $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ , con  $\mathcal{V} = \{1, 2, 3, 4, 5\}$  y  $\mathcal{A} = \{(1, 4), (4, 1), (4, 3), (3, 4), (3, 5), (5, 3), (4, 5), (5, 4), (4, 2), (2, 4), (2, 5), (5, 2)\}$ .

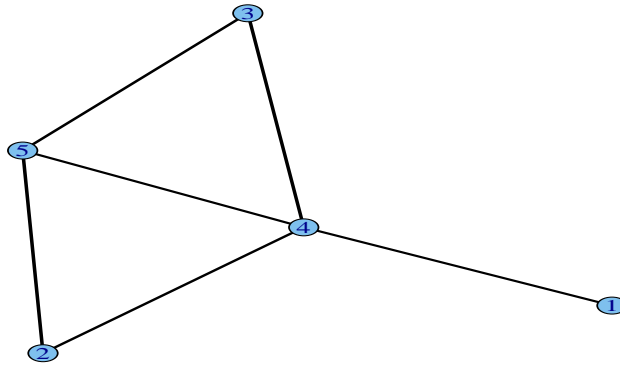


Figura 1.1: Ejemplo de grafo no dirigido en dimensión cinco.

### 1.3. Modelos gráficos no dirigidos

Una gráfica no dirigida es aquella donde todas sus aristas son no dirigidas. Asumiendo una propiedad de Markov en la distribución,  $\mathbf{P}$ , con respecto a la gráfica,  $\mathcal{G}$ , se pueden inferir algunas relaciones de independencia condicional entre las variables  $X_1, \dots, X_p$ .

#### Propiedades de Markov para gráficas no dirigidas

**Definición 1.3.1 Propiedad de Markov a pares.** Consideremos un modelo gráfico  $(\mathcal{G}, \mathbf{P})$ . Diremos que  $\mathbf{P}$  satisface la *propiedad de Markov a pares* con respecto a la gráfica no dirigida  $\mathcal{G}$  si para cualquier par de vértices no conectados,  $(j, k) \notin \mathcal{A}$  ( $j \neq k$ ),

$$X_j \perp\!\!\!\perp X_k | \mathbf{X}_{\mathcal{V} \setminus \{j, k\}}.$$

En otras palabras, una ausencia de arista entre los vértices  $j$  y  $k$  implica independencia entre  $X_j$  y  $X_k$  condicional al resto de las variables en el conjunto de vértices (c.f. [Studený \(2006\)](#)) para tener un amplio panorama de la relevancia que tienen las estructuras de independencia condicional, en particular se presenta una discusión en el campo de la inteligencia artificial); v.g. en la Figura 1.1  $X_1$  es independiente de  $X_3$  condicional en  $(X_2, X_4, X_5)^T$ . La propiedad de Markov a pares será el vehículo de estudio en la presente tesis.

#### Observaciones 1.3.1 :

**(Obs. 1)** Sea  $\mathcal{V} = \{1, \dots, p\}$ , el conjunto de vértices, y  $B \subset \mathcal{V}$  tal que  $|B| = m$ , ocuparemos la notación  $\mathbf{X}_B$  para referirnos al vector (o variable) aleatorio con las variables cuyos índices pertenecen a  $B$ , es decir

$$\mathbf{X}_B = (X_{j_1}, \dots, X_{j_m})^T,$$

donde  $j_1, \dots, j_m \in B$ ,  $j_r \neq j_s$  si  $r \neq s$ .

Entonces  $\mathbf{X}_{\mathcal{V} \setminus \{j, k\}}$  es el v.a. que incluye a todas las variables del vector original,  $\mathbf{X}$ , excluyendo a  $X_j$  y  $X_k$ .

A continuación presentamos dos sencillas definiciones que nos ayudan a extender la propiedad de Markov a pares hacia un atributo más general, clave para interpretar un grafo no dirigido.

**Definición 1.3.2 Camino.** Un camino es una secuencia de vértices  $\{k_1, \dots, k_\ell\} \subseteq \mathcal{V}$  tales que  $(k_i, k_{i+1}) \in \mathcal{A}$  para  $i = 1, \dots, \ell - 1$ .

Por ejemplo en la Figura 1.1,  $\{5, 2, 4\}$  es un camino.

**Definición 1.3.3 Separabilidad.** Consideremos tres subconjuntos disjuntos del conjunto de vértices,  $A, B, C \subseteq \mathcal{V}$ . Diremos que  $C$  separa a  $A$  y  $B$  si cualquier camino que va de  $j \in A$  a  $k \in B$  contiene al menos un vértice en  $C$ .

Por ejemplo en la Figura 1.1,  $\{4\}$  separa a  $\{1\}$  de  $\{2, 3, 5\}$ .

**Definición 1.3.4 Propiedad global de Markov.** Diremos que  $\mathbf{P}$  satisface la *propiedad global de Markov* con respecto a la gráfica no dirigida  $\mathcal{G}$  si para cualquiera tres subconjuntos disjuntos del conjunto de vértices  $A, B$  y  $C$  tales que  $C$  separa a  $A$  y  $B$ ,

$$\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B | \mathbf{X}_C.$$

En general la propiedad global de Markov implica al propiedad a pares, pero la implicación contraria aplica para una gran clase de modelos como se asegura en el siguiente Lema 1.3.1.

**Lema 1.3.1** Si la distribución  $\mathbf{P}$  tiene una densidad estrictamente positiva y continua con respecto a la medida de Lebesgue, entonces las propiedades a pares y global de Markov son equivalentes.

El Lema 1.3.1 se prueba en [Lauritzen \(1996\)](#) (pág. 36).

En la mayor parte de la literatura, un modelo gráfico no dirigido,  $(\mathcal{G}, \mathbf{P})$ , donde la propiedad de Markov a pares aplica se le conoce como *Gráfica de independencia condicional (GIC)* (c.f. [Edwards \(2012\)](#), Sección 1.3, para más detalles y discusión sobre las propiedades de Markov de gráficas no dirigidas).

## 1.4. Aplicaciones

La presente Sección busca dar un panorama general de algunos de los usos y aplicaciones que tiene un modelo gráfico no dirigido a partir de las propiedades de Markov que enunciamos en la Sección anterior.

### Análisis descriptivo

En una primera instancia, un grafo es una herramienta descriptiva que ayuda a tener una visión y comprensión detallada de algunas de las relaciones que existen entre distintas variables, v.g. en [Edwards \(2012\)](#) explican la paradoja de Simpson vía un grafo no dirigido. Pese a que la paradoja de Simpson puede llegar a ser un ejemplo sencillo, el uso de un grafo para explicarla es muy útil sobre todo cuando nos dirigimos a un público *no estadístico*. De esa manera los modelos gráficos se han convertido en un canal de comunicación entre un asesor estadístico y un usuario de la misma. Por lo anterior es que los modelos gráficos no dirigidos cobran bastante relevancia cuando se desea descubrir asociaciones *verdaderamente explicativas* y eliminar *relaciones espurias*, más aún en contextos donde se cuenta con cantidades masivas de variables y naturalmente es difícil explicar las relaciones existentes entre éstas, v.g. en microarreglos genéticos.

### Selección de variables

En muchos contextos el objeto de estudio es explicar el comportamiento de una variable en *función* de otras, i.e. la construcción de modelos de regresión, v.g. en modelos econométricos de elasticidades cruzadas de precios. Una correcta selección de variables en modelos de regresión es vital, ya que incorporar una gran cantidad de variables, que de alguna manera expliquen lo mismo, puede conllevar a un sobreajuste, y por otro lado, hacer un modelo simplista con muy pocas variables puede conducir a pérdida de información, i.e. se busca ajustar modelos que cumplan a medida de lo posible con el principio de parsimonia.

Para comprender la relevancia de un modelo gráfico como herramienta para la selección de variables definiremos el conjunto de adyacencia de una vértice en un grafo.

**Definición 1.4.1 Conjunto de adyacencia.** El conjunto de vecinos o de adyacencia del vértice  $j$  en el grafo  $\mathcal{G}$  se define como,

$$\text{ady}(\mathcal{G}, j) := \{k \in \mathcal{V} : (j, k) \in \mathcal{A} \text{ y } (k, j) \in \mathcal{A}\}. \quad (1.4.1)$$

Por ejemplo en la Figura 1.1,  $\text{ady}(\mathcal{G}, 1) = \{4\}$ , sin embargo,  $\text{ady}(\mathcal{G}, 4) = \{1, 2, 3, 5\}$ .

Una *premisa recurrente* en la construcción de modelos multivariados es asumir que el v.a. en cuestión cumple con la *propiedad de Markov*, la cual se enuncia a continuación en la Definición 1.4.2.

**Definición 1.4.2 Propiedad de Markov del v.a.** Dado un v.a.  $\mathbf{X} = (X_1, \dots, X_p)^T$ , se dice que el modelo de probabilidad subyacente,  $\mathbf{P}$ , sigue la propiedad de Markov si se cumple lo siguiente,

$$F_{X_j|\{X_k, k \neq j\}} = F_{X_j|\{X_k, k \in \text{ady}(\mathcal{G}, j)\}}, \quad \forall j \in \mathcal{V}, \quad (1.4.2)$$

donde  $\mathcal{V}$  es el conjunto de vértices de la GIC,  $\mathcal{G}$ , asociada a  $\mathbf{X}$  y  $\forall B \in \mathcal{V} \setminus \{j\}$ ,  $F_{X_j|\{X_k, k \in B\}}$  es la *función de distribución acumulada*<sup>1</sup> de  $X_j|\mathbf{X}_B$ .

Cuando una distribución cumple la propiedad de Markov, también se dice que es una *distribución (o medida) de Gibbs*.

Partiendo del hecho de que en la mayoría de los casos, en un modelo de regresión la relación funcional entre la variable dependiente y las explicatorias es una *característica de la distribución condicional*, v.g. una media o un

<sup>1</sup>La notación para la función de distribución acumulada empleada en la Definición 1.4.2 será modificada en capítulos subsecuentes, se decidió utilizar esta notación en este momento del documento para no entrar en detalles que se presentan más adelante.

cuantil condicionales, dada una GIC,  $\mathcal{G}$ , asociada a un v.a.  $\mathbf{X} = (X_1, \dots, X_p)^T$  y el problema de seleccionar variables para el modelo de regresión de  $X_j$  en función de  $X_{\mathcal{V} \setminus \{j\}}$ , asumiendo de la propiedad de Markov para  $\mathbf{P}$  y el hecho de que la GIC cumple la propiedad de Markov a pares del grafo, basta con considerar en el modelo a las variables representadas por los vértices en  $\text{ady}(\mathcal{G}, j)$ . El caso más clásico de selección de variables para un modelo de regresión lineal múltiple asociado a una GIC, es cuando  $\mathbf{X}$  sigue una distribución gaussiana multivariada; en la Sección 1.7 se puntualizará sobre un resultado que asocia al conjunto de adyacencia y los coeficientes de un modelo de regresión lineal en el caso gaussiano (c.f. [Lauritzen \(1996\)](#) Sección 5.1.3 y [Bühlmann and Van De Geer \(2011\)](#) Sección 13.4, para detalles sobre las relaciones entre el conjunto de adyacencia, la correlación parcial, los coeficientes de regresión y la matriz de precisión de un modelo gaussiano). En el Capítulo 5, Secciones 5.2 y 5.4, retomaremos el problema de selección de variables en el contexto de inferencia en *altas dimensiones* ( $N \ll p$ ).

## Estimación

Partiendo de un modelo de distribución gaussiana del v.a.  $\mathbf{X}$ , una de las aplicaciones más clásicas de los modelos gráficos no dirigidos es la de estimación restringida de parámetros al imponer relaciones de independencia condicional con base en la naturaleza del fenómeno de estudio (c.f. [Edwards \(2012\)](#), capítulo 3, presenta algunos enfoques clásicos de estimación de modelos gráficos gaussianos junto con un par de ejemplos).

Por otro lado, en el contexto de estadística en *altas dimensiones* muchas de las metodologías se basan en la estimación de la matriz de covarianza (o su inversa), así que esta aplicación sigue siendo vigente. [Bühlmann and Van De Geer \(2011\)](#), Sección 13.4.3, presenta alternativas de estimación de la matriz de covarianza y su inversa basadas en la gráfica de independencia condicional para el caso de altas dimensiones.

## Construcción de distribuciones conjuntas

En general asumir relaciones de independencia o independencia condicional entre variables aleatorias es una vía práctica para descomponer distribuciones conjuntas en formas más simples, v.g. en una cadena de Markov. En particular, las relaciones de independencia condicional contenidas en una GIC,  $\mathcal{G}$ , asociada a una distribución de probabilidad,  $\mathbf{P}$ , implican una factorización particular de la función de densidad conjunta del v.a. en cuestión,  $\mathbf{X}$ , tal como se enuncia en el Teorema de [Hammersley and Clifford \(1971\)](#). Antes de presentar el Teorema antes mencionado, introduciremos algunos conceptos fundamentales.

**Definición 1.4.3 Grafo completo.** Dado un modelo gráfico  $(\mathcal{G}, \mathbf{P})$ , diremos que  $\mathcal{G}$  es un *grafo completo* si,

$$\mathcal{A} = (\mathcal{V} \times \mathcal{V}) \setminus \{(j, j) \mid j \in \mathcal{V}\}. \quad (1.4.3)$$

El decir que un modelo gráfico tiene asociada una GIC completa implica que el conjunto de aristas,  $\mathcal{A}$ , contiene todas las posibles  $\binom{p}{2}$  combinaciones de enlaces entre las variables de  $\mathbf{X}$ .

**Definición 1.4.4 Clique.** Un subconjunto  $u \subseteq \mathcal{V}$  se le llama un *clique* si es completo maximal, i.e.  $u$  es completo y si  $u \subset w$ , entonces  $w$  no es completo.

En otras palabras, un *clique* es un conjunto maximal de variables en una gráfica que estén conectadas todas entre sí.

Por otro lado, diremos que una función potencial,  $\varphi(\cdot)$ , es toda aquella función no negativa en sus argumentos. Entonces el Teorema de [Hammersley and Clifford \(1971\)](#) asegura que la densidad conjunta de  $\mathbf{X}$  es proporcional al producto de funciones potenciales de los cliques en  $\mathcal{G}$ , como se enuncia a continuación.

**Teorema 1.4.1 Hammersley & Clifford.** Sea el v.a.  $\mathbf{X} = (X_1, \dots, X_p)^T$  con modelo gráfico no dirigido markoviano asociado  $(\mathcal{G}, \mathbf{P})$  y función de densidad conjunta  $f_{\mathbf{X}}(\cdot)$ . Si  $f_{\mathbf{X}}$  es no negativa en todo punto, i.e.  $f_{\mathbf{X}}(\mathbf{x}) > 0 \forall \mathbf{x} \in \mathbb{R}^p$ , entonces  $f_{\mathbf{X}}$  se factoriza sobre  $\mathcal{G}$ , i.e. existen funciones potenciales  $\varphi_c$  tales que

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \varphi_c(\mathbf{x}_c),$$

donde  $\mathcal{C}$  es el conjunto de cliques de  $\mathcal{G}$  y  $Z$  es una constante de normalización.

Para más detalles sobre el Teorema de [Hammersley and Clifford \(1971\)](#) y sus aplicaciones, véase [Besag \(1974\)](#), [Ripley \(1977\)](#) y [Murphy \(2006\)](#).

## 1.5. Casos de aplicación

En la presente sección enunciaremos algunas áreas de la investigación, junto con autores destacados, en donde los modelos gráficos no dirigidos markovianos se han convertido en una herramienta esencial para dar respuesta a preguntas de interés particular.

En el campo de la *conectividad cerebral*, a partir de datos funcionales provenientes de resonancias magnéticas, se ha logrado describir la arquitectura del sistema nervioso como un conjunto de *vértices*, que denotan regiones anatómicas o puntos en donde se encuentran electrodos, junto con *aristas* interconectadas, que representan conexiones funcionales o estructurales; algunos autores han llamado a estos modelos como *Brain Graphs*. Principalmente estos modelos han llamado la atención en este contexto, debido a que la neurociencia moderna trae consigo problemas de inferencia a gran escala, usualmente en el contexto de altas dimensiones. Para mayor referencia se pueden consultar los siguientes artículos: [Salvador et al. \(2005\)](#), [Valdés-Sosa et al. \(2005\)](#), [Varoquaux et al. \(2010\)](#) y [Bullmore and Bassett \(2011\)](#).

Junto con los famosos modelos gráficos conocidos como *gráficas acíclicas dirigidas*, DAG's <sup>2</sup> por sus siglas en inglés, los modelos gráficos no dirigidos juegan un rol importante en la construcción y representación de *sistemas expertos* en el terreno de la *inteligencia artificial*. Para un amplio panorama sobre las aplicaciones de modelos gráficos en sistemas expertos se puede consultar [Castillo et al. \(2012\)](#).

Una aplicación natural y que ha cobrado mayor relevancia en años recientes es el análisis de *redes sociales*, siendo de interés para las ciencias sociales y la antropología. Además este tipo de análisis están ganando terreno en empresas de consumo masivo, con estudios de mercado, e instituciones financieras, con análisis para detección de fraudes y prevención de lavado de dinero. Algunos modelos particulares se encuentran en [Wasserman and Pattison \(1996\)](#) y [Robins et al. \(2007\)](#), además se recomienda [Pircalabelu \(2015\)](#) capítulo 7.

Finalmente se recomienda ampliamente [Wainwright and Jordan \(2008\)](#), ya que profundiza sobre aplicaciones puntuales tanto de modelos gráficos no dirigidos markovianos como de DAG's en distintos contextos.

Cabe mencionar que en muchas de las aplicaciones que se encontraron se destaca la eficiencia computacional que se gana, una vez que se infieren las relaciones de independencia condicional modeladas en un grafo no dirigido, para la construcción de distribuciones conjuntas o algún otro objeto en donde interactúen múltiples variables, v.g. funciones de utilidad (c.f. [Bacchus and Grove \(1995\)](#) para una aplicación interesante relacionada con funciones de utilidad).

## 1.6. Modelos gráficos no dirigidos triviales

Hemos decidido incluir esta pequeña sección para darle más sentido y claridad a los objetivos que se pretenden alcanzar en los capítulos siguientes. Así que explicaremos a que nos referiremos con un *grafo trivial* en lo subsecuente del documento.

Para que un modelo gráfico no dirigido markoviano tenga la *utilidad* deseada y se puedan explotar las aplicaciones presentadas en la Sección 1.4, tendremos que asegurar que la distribución de probabilidad que impongamos,  $\mathbf{P}$ , en realidad considere relaciones de independencia condicional. Suponiendo que  $\mathbf{P}$  está caracterizada por alguna función de distribución acumulada multivariada,  $H$ , o una función de densidad,  $f$ , pertenecientes a una familia paramétrica, tendríamos que asegurar que la familia sea lo suficientemente flexible como para modelar independencias condicionales con alguna combinación de parámetros. En caso contrario, el grafo asociado a  $\mathbf{P}$  sería un *grafo completo* como en la Definición 1.4.3, lo cual carecería *utilidad* alguna. A un grafo completo lo vamos a considerar como

---

<sup>2</sup>Directed Acyclic Graphs.



uno de los dos casos de grafo trivial.

El caso totalmente opuesto a un grafo completo es un *grafo vacío*, el cual definimos a continuación.

**Definición 1.6.1 Grafo vacío.** Dado un modelo gráfico  $(\mathcal{G}, \mathbf{P})$ , diremos que  $\mathcal{G}$  es un *grafo vacío* si,

$$\mathcal{A} = \emptyset. \quad (1.6.1)$$

Si un v.a.  $\mathbf{X} = (X_1, \dots, X_p)^T$  tiene asociado un grafo vacío como en la Definición 1.6.1, implicaría que se tienen todas las relaciones de independencia condicional posibles modeladas en una GIC entre las variables de  $\mathbf{X}$ . Poder asegurar lo anterior, sería deseable para algunas de las aplicaciones expuestas en la Sección 1.4, v.g. simplificaría rotundamente la construcción de distribuciones conjuntas. Sin embargo, consideramos que tener un modelo conjunto en donde todas las variables sean independientes condicional en el resto de las mismas es poco realista y por ende el modelo de probabilidad per se carecería de *utilidad*. Así el segundo tipo de grafo trivial que consideraremos es el grafo vacío.

De manera muy recurrente en el documento haremos alusión a que la distribución de probabilidad de una GIC,  $\mathbf{P}$ , *no se asocie* con un grafo trivial con el hecho de que el modelo gráfico no dirigido tenga *utilidad*. Así que enfatizando en los objetivos de nuestra investigación, en los capítulos subsecuentes buscaremos estudiar y proponer modelos gráficos no dirigidos markovianos que tengan asociadas distribuciones conjuntas flexibles y que además sean *útiles*.

## 1.7. Modelo gráfico no dirigido gaussiano

Como en muchos de los modelos estadísticos existentes, un supuesto recurrente es el de distribución gaussiana, i.e.  $\mathbf{X} \sim \mathcal{N}_p(\mu, P)$ , con  $P$  una matriz de covarianzas. La presente sección tiene el objetivo de enunciar resultados clásicos sobre cómo es que se deducen las relaciones de independencia condicional contenidas en un grafo no dirigido, a partir de la estructura de dependencia que modela la distribución gaussiana. Se mencionarán brevemente algunos enfoques de estimación del grafo en este contexto, y finalmente, se comenta sobre algunas de las limitantes inmediatas del modelo gaussiano per se.

Esta Sección busca dar un panorama general sin entrar en detalles sobre la estructura de dependencia que modela una distribución gaussiana, más adelante en las Secciones 3.3 y 3.5 se profundiza sobre ese tema.

### Gráfica de independencia condicional gaussiana

En el caso de los modelos gráficos no dirigidos el supuesto de normalidad multivariada es ideal, ya que brinda una representación sencilla de la propiedad de Markov a pares codificada por la matriz de precisión,  $\Omega = P^{-1}$ , además de que las relaciones de independencia condicional incluidas en el grafo tienen una asociación natural con los coeficientes de regresión lineal entre las variables involucradas. Pero las grandes limitantes que tiene el modelo gaussiano al tomarlo como supuesto, en términos de la estructura de dependencia que representa, son factores que pueden *ocultar* asociaciones entre las variables o llevar al analista a conclusiones erróneas sobre la dependencia que guardan los datos. Pese a que es bien sabido que el supuesto de normalidad multivariada es muy difícil—sino es que imposible—de lograr, el fácil manejo de la distribución gaussiana junto con sus parámetros lo hace muy atractivo como modelo y lo mantiene vigente en el terreno de la investigación. Aún en el contexto de estadística en altas dimensiones, en donde los modelos gráficos no dirigidos han cobrado bastante relevancia en los últimos años, la distribución gaussiana multivariada sigue siendo un supuesto recurrente, pese a que a medida que el número de variables aumenta éste se vuelve cada vez menos realista. Hablar del modelo gráfico gaussiano y algunas otras propuestas de años recientes en el terreno de altas dimensiones es tema del Capítulo 5.

El resultado pionero que le dio vida a los modelos gráficos no dirigidos *probabilísticos* y en particular a los *gaussianos* es aquel en [Dempster \(1972\)](#):

$$(k, j), (j, k) \notin \mathcal{A} \Leftrightarrow X_k \perp\!\!\!\perp X_j | X_{\mathcal{V} \setminus \{k, j\}} \Leftrightarrow \Omega_{kj} = 0. \quad (1.7.1)$$

Entonces el problema de estimar el grafo bajo el supuesto de distribución gaussiana se reduce a estimar la matriz de precisión,  $\Omega$ , o al menos a detectar los elementos nulos de la matriz. En mucha de la literatura, sobre todo enfocada

al contexto de altas dimensiones, al problema de detectar las relaciones de independencia condicional del grafo se le conoce como detectar los *patrones de escasez*<sup>3</sup> entre las variables. Explicar con detalle a que se refiere el concepto de *escasez* es tarea puntual del Capítulo 5, ya que en el contexto de altas dimensiones ha sido más natural su aplicación.

Como se comentó en la Sección 1.4, la primera aplicación de estos modelos gráficos fue la de estimar la matriz de covarianzas o correlaciones a partir de enfoques de máxima verosimilitud restringida a relaciones de independencia condicional entre variables, i.e. restringida a imponer ciertos elementos iguales a cero en la matriz de precisión, para detalles sobre este enfoque de estimación se puede consultar [Edwards \(2012\)](#) capítulo 3. Más recientemente, en [Drton and Perlman \(2007\)](#) y [Drton and Perlman \(2008\)](#) se propone un procedimiento para detectar el patrón de escasez de la matriz de precisión basado en pruebas múltiples, lo anterior en el contexto de bajas dimensiones,  $p < n$ .

Por otro lado, los elementos de la matriz de precisión se relacionan de la siguiente manera con la correlación parcial entre las variables:

$$\rho_{jk|\mathcal{V}\setminus\{j,k\}} = -\frac{\Omega_{jk}}{\sqrt{\Omega_{jj}\Omega_{kk}}}. \quad (1.7.2)$$

El resultado anterior nos ayuda a comprender las equivalencias en (1.7.1), por el hecho de que las distribuciones condicionales de un modelo gaussiano multivariado son también gaussianas y además la independencia en este caso está representada por la correlación *lineal* nula.

Finalmente, consideremos el modelo de regresión,

$$X_j = \beta_k^{(j)} X_k + \sum_{s \in \mathcal{V} \setminus \{j,k\}} \beta_s^{(j)} X_s + \epsilon_j, \quad (1.7.3)$$

donde  $\mathbb{E}[\epsilon_j] = 0$  y a partir del supuesto gaussiano,  $\epsilon_j$  es independiente de  $\{X_s : s \in \mathcal{V} \setminus \{j\}\}$ . Es bien sabido que la correlación parcial está relacionada con los coeficientes de regresión de la siguiente forma,

$$\beta_k^{(j)} = -\frac{\Omega_{jk}}{\Omega_{jj}}, \quad \beta_j^{(k)} = -\frac{\Omega_{kj}}{\Omega_{kk}}. \quad (1.7.4)$$

En conclusión, para los modelos gráficos gaussianos tenemos las siguientes equivalencias que los hacen tan atractivos,

$$(j, k), (k, j) \in \mathcal{A} \Leftrightarrow \Omega_{jk} \neq 0 \Leftrightarrow \rho_{jk|\mathcal{V}\setminus\{j,k\}} \neq 0 \Leftrightarrow \beta_k^{(j)} \neq 0 \text{ y } \beta_j^{(k)} \neq 0. \quad (1.7.5)$$

Las equivalencias en (1.7.5) son las que asocian la estimación del grafo gaussiano con el problema de selección de variables como se mencionó en la Sección 1.4.

## 1.8. Discusión

En este capítulo exhibimos la relevancia de los modelos gráficos no dirigidos markovianos como una herramienta para comprender relaciones de independencia condicional entre variables aleatorias. Se expusieron las aplicaciones más importantes que estos modelos tienen y se ilustraron algunas de éstas citando casos explícitos de aplicación.

Por otro lado, como lo observamos en la sección anterior, un modelo gráfico gaussiano es ideal por su fácil manejo e interpretación, además de que es el modelo más estudiado en la literatura tanto de bajas como de altas dimensiones. Sin embargo, un modelo gaussiano es bastante restrictivo en el sentido de la estructura de dependencia que modela, junto con el hecho de que impone que todas sus marginales y condicionales sean gaussianas.

Partiendo de que las marginales de dimensión uno de un modelo gaussiano son gaussianas también, tendríamos que comenzar por hablar de las limitantes de una distribución gaussiana de dimensión uno, pero éste es un tema

<sup>3</sup>Sparsity patterns, en inglés.

ya bastante conocido por la comunidad estadística y está fuera de los alcances de la presente investigación; nuestro interés radica en exhibir las limitantes del modelo conjunto en términos de la estructura de dependencia que representa (c.f. Patton (2002) capítulo 1, para un panorama general sobre estudios asociados a variables económicas en donde el supuesto de normalidad tanto conjunta como marginal no es pertinente).

Una distribución gaussiana sólo es capaz de modelar dependencias a pares basadas en el coeficiente de correlación lineal de Pearson, donde éste es simplemente una medida de asociación lineal entre las variables con un campo de acción bastante limitado (c.f. Embrechts et al. (1999)), en la Sección 3.5 generaremos una discusión al respecto. Además el coeficiente de correlación lineal *no cumple con ser una medida de dependencia* así que, fuera del modelo de distribución gaussiana, no es una herramienta pertinente si lo que se busca es detectar relaciones de independencia o independencia condicional, en la Sección 2.6 presentamos las condiciones que debe cumplir una medida de dependencia. En la Sección 3.5 se profundiza sobre las limitantes del modelo gaussiano en términos más generales, i.e. visto como modelo de cópula gaussiana.

Cuando el motor de estudio es la inferencia de estructuras de dependencia o independencia entre variables aleatorias, en los últimos años los modelos de cópulas han cobrado mucha relevancia por su capacidad para construir distribuciones conjuntas bastante flexibles. Así que basar la búsqueda de relaciones de independencia condicional en la cópula subyacente a los datos es una vía bastante atractiva, y en particular será la base de la investigación en la presente tesis.

Como se mencionó en la introducción del documento, el objetivo principal de esta investigación es estudiar, y en caso de ser necesario proponer, modelos gráficos no dirigidos markovianos que sean de *utilidad* cuya distribución de probabilidad asociada,  $\mathbf{P}$ , sea *más flexible* como modelo conjunto al menos con respecto a una distribución gaussiana. Entonces el Capítulo 2 tiene la intención de estudiar a las funciones cópula multivariadas y sus propiedades, buscando comprender cómo es que se pueden deducir relaciones de independencia condicional basadas en la cópula subyacente de vectores aleatorios, esperando encontrar modelos interesantes de cópulas para los cuales se puedan deducir GIC's de modelos gráficos markovianos.



## Capítulo 2

# Cóputas: multivariadas y condicionales

**Resumen** Hoy en día es grande la lista de artículos en donde se exhibe el potencial de las funciones cópula para modelar patrones complejos de dependencia que *inevitablemente* aparecen en datos multivariados, además de que la investigación para lograr modelos de cópulas cada vez más flexibles es un tema vigente con bastante relevancia para las aplicaciones (c.f. [Min and Czado \(2011\)](#) presenta una introducción donde se da un panorama general sobre la relevancia de los modelos de cópulas multivariadas y cómo ha ido evolucionado su estudio a lo largo de más de una década). Con el auge que han alcanzado los modelos de cópulas, cada vez son más las aplicaciones en donde se han desplazado a modelos clásicos, usualmente basados en el supuesto de distribución gaussiana donde la medida de asociación entre las variables es el coeficiente de correlación lineal, por modelos conjuntos basados en cópulas que logren explicar de manera efectiva la estructura de dependencia entre las variables, más allá de simples relaciones lineales. Matemáticamente, una cópula simplemente es una función de distribución multivariada en el hipercubo unitario,  $\mathbb{I}^n$ , con marginales uniformes, pero es a partir del Teorema [Sklar \(1959\)](#) que el concepto cobra una relevancia bastante importante para la construcción de modelos conjuntos, ya que permite modelar por separado la estructura de dependencia entre variables y los comportamientos marginales de las mismas. Además de que naturalmente, la cópula subyacente de un v.a. se ha convertido en una herramienta potencial para medir el grado de dependencia entre las variables involucradas; en particular para la detección de relaciones de independencia e independencia condicional (c.f. [Schweizer and Wolff \(1981\)](#) para una discusión profunda sobre medidas de dependencia basadas en cópulas). Entonces, si buscamos modelos gráficos no dirigidos markovianos para distribuciones conjuntas flexibles, la cópula subyacente del v.a. es una herramienta ad hoc para la detección de las relaciones de independencia condicional representadas en el grafo. Por otro lado, si asumimos un solo modelo de cópula que explique la estructura completa de dependencia, i.e. que dado un v.a.  $\mathbf{X}$ , la cópula subyacente de  $\mathbf{X}$ ,  $C_{\mathbf{X}}$ , pertenezca a una familia paramétrica específica, tendremos que buscar familias de cópulas que cumplan con el *supuesto simplificador*; supuesto recurrente para la construcción de distribuciones conjuntas en la teoría de *vine cópula*, la cual cada vez llama más la atención en las aplicaciones por su capacidad para construir cópulas bastante flexibles, que no necesariamente pertenezcan a una familia paramétrica específica (c.f. [Aas et al. \(2009\)](#) es el primer artículo en donde se propone un algoritmo general para la construcción de cópulas multivariadas partiendo de construcciones *vine*). Sin embargo, en [Stöber et al. \(2013\)](#) se asegura que las tres familias de cópulas que cumplen con el supuesto simplificador son la gaussiana, la T y la Clayton. Así que la gran interrogante con la que cierra este capítulo, es si estas familias son lo suficientemente flexibles como modelos conjuntos y si además las GIC's asociadas tienen la *utilidad* deseada como modelos gráficos, como para dar por alcanzados los objetivos de nuestra investigación.

### 2.1. Organización del capítulo

El presente capítulo es una introducción general a las funciones cópula multivariadas, enfatizando sobre su relación con la construcción de distribuciones de probabilidad conjuntas y medidas de dependencia, además se expone sobre el concepto de cópula condicional, puntualizando sobre el supuesto simplificador necesario para la deducción de una GIC. La Sección 2.2 sienta las bases conceptuales sobre funciones de distribución multivariadas. En la Sección 2.3 se exponen la definición y propiedades elementales de una cópula. Como parte central del capítulo, en la Sección 2.4 se presenta el Teorema de [Sklar \(1959\)](#) e implicaciones importantes del mismo. En la Sección 2.5 se presentan ejemplos puntuales de cópulas subyacentes de v.a. particulares. La Sección 2.6 enuncia las condiciones que debe

cumplir una medida de dependencia según [Schweizer and Wolff \(1981\)](#), presenta medidas de dependencia basadas en cópulas y expone sobre la dependencia en colas. Además en la Sección 2.7 se da un panorama general sobre el concepto de cópula condicional y se comenta sobre su relevancia actual, se presenta la extensión del Teorema de Sklar hacia cópulas condicionales, se define el supuesto simplificador y se comenta sobre familias de cópulas simplificadas. Finalmente, en la Sección 2.8 se genera una discusión sobre los temas presentados en el capítulo.

## 2.2. Preliminares: vectores aleatorios y funciones de distribución multivariadas

A continuación recordamos la definición y propiedades esenciales de una *función de distribución acumulada* (FDA) de un *vector aleatorio* (v.a.). La intención de esta sección es presentar lo mínimo necesario para entender el concepto de cópula y su utilidad.

Comenzaremos estableciendo notación básica. Sea  $p \in \mathbb{N}$ , como se implementó en el capítulo anterior, denota la dimensión del v.a. correspondiente. De aquí en adelante, denotaremos con  $\mathbf{x}$  a un vector  $(x_1, \dots, x_n)$  que pertenece a  $\mathbb{R}^p$  (o a  $\overline{\mathbb{R}}^p$ ); por lo general,  $\mathbf{x}$  representará la observación particular de un vector aleatorio  $\mathbf{X}$  de dimensión  $p$ . En lo subsecuente,  $\mathbb{I}$  denotará al intervalo unitario  $[0, 1]$ .

Una de las maneras de caracterizar la ley de un v.a.  $\mathbf{X}$ ,  $\mathbf{P}_{\mathbf{X}}$ , es vía su función de distribución acumulada (o simplemente función de distribución), la cual se define a continuación.

**Definición 2.2.1 FDA.** Dado un vector aleatorio  $\mathbf{X} = (X_1, \dots, X_p)^T$ , su *función de distribución*  $H_{\mathbf{X}} : \overline{\mathbb{R}}^p \rightarrow \mathbb{I}$  está definida por

$$H_{\mathbf{X}}(x_1, \dots, x_n) = \mathbb{P}(X_1 \leq x_1, \dots, X_p \leq x_p) = \mathbb{P}\left(\bigcap_{i=1}^p \{X_i \leq x_i\}\right), \quad (2.2.1)$$

si todos los elementos  $x_i \in \mathbb{R}$ , mientras que,

**(FDA 1)**  $H_{\mathbf{X}}(x_1, \dots, x_n) = 0$ , si al menos uno de sus argumentos es igual a  $-\infty$ ,

**(FDA 2)**  $H_{\mathbf{X}}(\infty, \dots, \infty) = 1$ .

Con el propósito de enunciar las propiedades de una FDA, necesitaremos el concepto de  $G$ -volumen para una función acotada,  $G$ , el cual enunciamos a continuación. Cabe destacar que las expresiones tales como  $\min\{\mathbf{x}, \mathbf{y}\}$  ó  $\mathbf{x} \leq \mathbf{y}$  harán referencia a operadores entrada a entrada de los vectores  $\mathbf{x}$ ,  $\mathbf{y}$ .

**Definición 2.2.2** Dados dos puntos,  $\mathbf{a}, \mathbf{b} \in \overline{\mathbb{R}}^p$ , tales que  $\mathbf{a} = (a_1, \dots, a_p)$  y  $\mathbf{b} = (b_1, \dots, b_p)$ , con  $\mathbf{a} \leq \mathbf{b}$ , un  $p$ -cubo  $[\mathbf{a}, \mathbf{b}]$  es el producto cartesiano

$$[\mathbf{a}, \mathbf{b}] = \times_{i=1}^p [a_i, b_i]. \quad (2.2.2)$$

Por otro lado, sean  $A$  un conjunto convexo en  $\overline{\mathbb{R}}^p$  y  $G$  una función acotada,  $G : A \rightarrow \overline{\mathbb{R}}$ . Definimos el  $G$ -volumen,  $V_G$ , del  $p$ -cubo  $[\mathbf{a}, \mathbf{b}]$  como

$$V_G([\mathbf{a}, \mathbf{b}]) := \sum_{\mathbf{v} \in \times_{i=1}^p \{a_i, b_i\}} (-1)^{N(\mathbf{v})} G(\mathbf{v}), \quad (2.2.3)$$

donde  $N(\mathbf{v}) = \text{card}\{j | v_j = a_j, j = 1, \dots, p\}$ , siempre que  $a_i < b_i$  para todo  $i \in \{1, \dots, p\}$ . En otro caso,  $V_G([\mathbf{a}, \mathbf{b}]) = 0$ .

**Ejemplo 2.2.1** Sean  $p=2$ ,  $\mathbf{a} = (a_1, a_2)$ ,  $\mathbf{b} = (b_1, b_2)$ , con  $\mathbf{a} < \mathbf{b}$  y  $G$  una función acotada de  $\mathbb{R}^2$  a  $\overline{\mathbb{R}}$ , entonces

$$V_G([\mathbf{a}, \mathbf{b}]) = G(a_1, a_2) - G(a_1, b_2) - G(b_1, a_2) + G(b_1, b_2). \quad (2.2.4)$$

**Teorema 2.2.1** La FDA  $H_{\mathbf{X}}$  de un v.a.  $\mathbf{X} = (X_1, \dots, X_p)^T$  tiene las siguientes propiedades:

**(FDA 3)**  $H_{\mathbf{X}}$  es isotónica, i.e.  $H_{\mathbf{X}}(\mathbf{x}) \leq H_{\mathbf{X}}(\mathbf{y}) \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^p, \mathbf{x} \leq \mathbf{y}$ ;

**(FDA 4)**  $\forall i \in \{1, \dots, p\}$  y  $\forall (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_p) \in \mathbb{R}^{p-1}$ , la función

$$t \mapsto H_{\mathbf{X}}(x_1, \dots, x_{i-1}, t, x_{i+1}, \dots, x_p)$$

para  $t \in \mathbb{R}$ , es continua por la derecha;

**(FDA 5)** el  $H_{\mathbf{X}}$ -volumen,  $V_{H_{\mathbf{X}}}$ , de cada  $p$ -cubo  $[\mathbf{a}, \mathbf{b}]$  es positivo, i.e.  $V_{H_{\mathbf{X}}}([\mathbf{a}, \mathbf{b}]) \geq 0$ .

La condición (FDA 5) en el Teorema 2.2.1 es conocida como la propiedad  $p$ -creciente de una FDA  $p$ -dimensional; esta propiedad asegura una probabilidad positiva de que el v.a.  $\mathbf{X}$  tome valores en el cubo  $[\mathbf{a}, \mathbf{b}]$ .

En el Ejemplo 2.2.1 notemos que los puntos  $(a_1, a_2)$ ,  $(a_1, b_2)$ ,  $(b_1, a_2)$  y  $(b_1, b_2)$  son los vértices del rectángulo  $[a_1, b_1] \times [a_2, b_2]$  (un rectángulo es un 2-cubo), en cuyo caso si (2.2.4) es positivo, la condición 2-creciente de FDA's bivariadas asegura una probabilidad positiva en rectángulos.

A continuación enunciamos la noción y notación de distribuciones marginales, la cual será esencial en lo subsecuente del documento.

**Definición 2.2.3** Sea  $H$  una FDA  $p$ -dimensional, con  $p \geq 2$ . Tomemos  $\sigma = \{j_1, \dots, j_m\} \in \{1, \dots, p\}$ , con  $1 \leq m \leq p-1$ . Definamos a la función  $H_\sigma : \mathbb{R}^m \rightarrow \mathbb{I}$ , como

$$H_\sigma(x_1, \dots, x_m) = H(y_1, \dots, y_p),$$

donde

$$y_j = \begin{cases} x_j, & j \in \{j_1, \dots, j_m\}, \\ \infty & j \notin \{j_1, \dots, j_m\}. \end{cases}$$

A las FDA's  $H_\sigma$  las llamaremos  $\sigma$ -marginales.

De aquí en adelante, ocuparemos la notación  $i : j := \{i, i+1, \dots, j\}$ , con  $i < j$ , para referirnos a conjuntos de índices consecutivos. Esta notación será de mucha utilidad tanto para los casos condicionales como para los no condicionales.

### Observaciones 2.2.1 :

**(Obs. 1)** En particular, cuando  $\sigma = \{j\}$  cambiaremos la notación para las FDA's, ya que ocuparemos la notación  $F_j$  para referirnos a las marginal *uno dimensionales*. Si  $\text{card}(\sigma) \geq 2$ , entonces se ocupará la notación  $H_\sigma$  para referirnos a la FDA correspondiente. Así fácilmente logramos una distinción entre las distribuciones marginales uno dimensionales,  $F$ , y las conjuntas,  $H$ .

**(Obs. 2)** Notemos que la definición de marginales no abarca el caso cuando  $\sigma = 1 : p$ , pero de igual forma, en lo subsecuente ocuparemos la notación  $H_{1:p}$  para referirnos a la FDA de un vector  $p$ -dimensional,  $\mathbf{X} = (X_1, \dots, X_p)^T$ , de tal forma que  $H_{\mathbf{X}} = H_{1:p}$ ; en algunas ocasiones se ocuparán de manera indistinta ambas notaciones.

**(Obs. 3)** Regresando a los conceptos de modelos gráficos, recordemos que el conjunto de vértices  $\mathcal{V} = \{1, \dots, p\} = 1 : p$ , así que la notación  $H_\sigma$  aplicará para todo subconjunto  $\sigma \subseteq \mathcal{V}$ .

A partir del siguiente Teorema es que se caracteriza la independencia entre variables aleatorias a partir de su FDA conjunta, este Teorema es vital para los fines de la presente Tesis.

**Teorema 2.2.2** Sean  $X_1, \dots, X_p$  variables aleatorias independientes y denotaremos con  $F_j$  a la FDA de cada  $X_j$  ( $j = 1, \dots, p$ ), entonces la FDA del v.a.  $\mathbf{X} = (X_1, \dots, X_p)^T$ ,  $H_{1:p}$ , es el producto de las marginales uno dimensionales, i.e.  $\forall x_1, \dots, x_p \in \mathbb{R}$ ,

$$H_{1:p}(x_1, \dots, x_p) = \prod_{j=1}^p F_j(x_j). \quad (2.2.5)$$

### 2.3. Cópulas: definición y propiedades elementales

En esta sección definimos a las funciones cópula, enunciamos sus propiedades elementales y presentamos algunos ejemplos.

**Definición 2.3.1** Sea  $p \geq 2$ , una *cópula*  $p$ -dimensional,  $\mathcal{C}_p$ , es una FDA sobre  $\mathbb{I}^p$ , tal que sus marginales uno dimensionales son uniformemente distribuidas sobre  $\mathbb{I}$ .

Entonces cada cópula  $p$ -dimensional,  $\mathcal{C}_p$ , puede estar asociada con un v.a.  $\mathbf{U} = (U_1, \dots, U_p)^T$  tal que  $U_j \sim \mathcal{U}(0, 1)$   $\forall j \in \{1, \dots, p\}$  y  $\mathbf{U} \sim \mathcal{C}_p$ .

Partiendo del hecho de que una cópula es una FDA, y con base en los resultados mostrados en la Sección 2.2, de manera equivalente podemos caracterizar a las funciones cópula con el siguiente Teorema.

**Teorema 2.3.1** Una función  $\mathcal{C}_p : \mathbb{I}^p \rightarrow \mathbb{I}$  es una cópula si y sólo si cuenta con las siguientes propiedades,

(C 1) sea  $\mathbf{u} = (u_1, \dots, u_p) \in \mathbb{I}^p$ , entonces  $\forall j \in \{1, 2, \dots, p\}$ ,  $\mathcal{C}_p(\mathbf{u}) = u_j$  si  $u_k = 1 \forall k \neq j$ ;

(C 2)  $\mathcal{C}_p$  es isotónica, i.e.  $\mathcal{C}_p(\mathbf{u}) \leq \mathcal{C}_p(\mathbf{v})$ ,  $\forall \mathbf{u}, \mathbf{v} \in \mathbb{I}^p$ , tales que  $\mathbf{u} \leq \mathbf{v}$ ;

(C 3)  $\mathcal{C}_p$  es  $p$ -creciente.

#### Observaciones 2.3.1 :

(Obs. 1) Una consecuencia del Teorema 2.3.1 es la siguiente propiedad,

(C 4) Sea  $\mathbf{u} = (u_1, \dots, u_p) \in \mathbb{I}^p$ , tenemos que  $\mathcal{C}_p(\mathbf{u}) = 0$  si  $u_j = 0$  al menos para algún  $j \in \{1, 2, \dots, p\}$ ; en ocasiones se dice que la cópula está fijada por cumplir con esta condición.

(Obs. 2) Partiendo de un v.a.  $\mathbf{U} = (U_1, \dots, U_p)^T$  distribuido sobre  $\mathbb{I}^p$  y con marginales uniformes en  $\mathbb{I}$ , tal que  $\mathbf{U} \sim \mathcal{C}_p$ , para alguna cópula  $p$ -dimensional. La condición (C 1) proporciona la vía para obtener las distribuciones marginales uno dimensionales de  $\mathcal{C}_p$ . Pero de igual forma se pueden obtener las cópulas  $\sigma$ -marginales  $m$ -dimensionales,  $1 < m < p$ , para  $\sigma = \{j_1, \dots, j_m\}$  como en la Definición 2.2.3

$$\mathcal{C}_\sigma(u_1, \dots, u_m) = \mathcal{C}_p(v_1, \dots, v_p), \quad (2.3.1)$$

donde

$$v_j = \begin{cases} u_j, & j \in \{j_1, \dots, j_m\}, \\ 1 & j \notin \{j_1, \dots, j_m\}. \end{cases}$$

Hasta antes de la Ecuación (2.3.1) el subíndice de la cópula hacía alusión a la dimensión, pero en este caso hace referencia a un conjunto de variables aleatorias, lo cual es congruente por el hecho de que una cópula es una FDA en algún hipercono unitario. Más adelante enfatizaremos en la notación subsecuente para funciones cópula.



El resultado siguientes proporciona cotas, superior e inferior, para las funciones cópula.

**Teorema 2.3.2 (Cotas de Fréchet-Hoeffding).** Sea  $C_p$  una cópula  $p$ -dimensional, entonces  $\forall \mathbf{u} = (u_1, \dots, u_p) \in \mathbb{I}^p$

$$\mathcal{W}_p(\mathbf{u}) = \max \left\{ \sum_{j=1}^p u_j - p + 1, 0 \right\} \leq C_p(\mathbf{u}) \leq \mathcal{M}_p(\mathbf{u}) = \min \{u_1, \dots, u_p\}.$$

A las funciones  $\mathcal{W}_p(\cdot)$  y  $\mathcal{M}_p(\cdot)$  se les conoce como las cotas de *Fréchet-Hoeffding*.

**Ejemplo 2.3.1** Sea  $\mathbf{u} \in \mathbb{I}^p$ , algunos ejemplos fundamentales de cópulas son los siguientes:

- $\Pi_p(\mathbf{u}) = \prod_{j=1}^p u_j$ ,  $\forall p \geq 2$ . Esta cópula es conocida como la *cópula independencia* o *cópula producto*, más adelante profundizaremos en la relevancia de esta cópula.
- $\mathcal{M}_p(\mathbf{u})$ ,  $\forall p \geq 2$ .
- $\mathcal{W}_2(\mathbf{u})$ . Cabe mencionar que  $\mathcal{W}_p(\mathbf{u})$  no es cópula  $\forall p \geq 3$ .

Partiendo del Teorema 2.3.1 se puede probar que combinaciones convexas de cópulas también son cópulas, como se enuncia en siguiente Ejemplo 2.3.2.

**Ejemplo 2.3.2 (Combinación convexa de cópulas).** Sean  $C_p^1, C_p^2$  dos cópulas  $p$ -dimensionales y sea  $\alpha \in \mathbb{I}$ , entonces

$$C_p^* = \alpha C_p^1 + (1 - \alpha) C_p^2,$$

también es una cópula  $p$ -dimensional (c.f. [Durante and Sempi \(2010\)](#), ejemplo 1.3.1, para detalles sobre combinaciones convexas de cópulas.).

Para más detalles sobre las propiedades de las funciones cópula y más ejemplos de éstas, véase [Nelsen \(2007\)](#) y [Durante and Sempi \(2010\)](#).

## 2.4. Teorema de Sklar

El Teorema de [Sklar \(1959\)](#) es el pilar de la teoría de cópulas y el que le da relevancia al concepto. Del Teorema de Sklar, 2.4.1, podemos comprender a la cópula como la relación funcional entre la función de distribución conjunta y las correspondientes funciones de distribución marginal uno dimensionales.

**Teorema 2.4.1 Teorema de Sklar.** Sea  $H$  una FDA  $p$ -dimensional con marginales uno dimensionales  $F_1, \dots, F_p$ . Entonces existe una cópula  $\mathcal{C}$  tal que  $\forall x_1, \dots, x_p \in \mathbb{R}$ ,

$$H(x_1, \dots, x_p) = \mathcal{C}(F_1(x_1), \dots, F_p(x_p)). \quad (2.4.1)$$

Si las funciones  $F_1, \dots, F_p$  son todas continuas, entonces  $\mathcal{C}$  es única; en otro caso,  $\mathcal{C}$  queda únicamente determinada sobre  $\text{Ran } F_1 \times \dots \times \text{Ran } F_p$ .

El siguiente Teorema es importante para la construcción de modelos estadísticos en donde se considere, por separado, el comportamiento marginal univariado y la estructura de dependencia de un v.a. por medio de una cópula.

**Teorema 2.4.2** Si  $\mathcal{C}$  es una cópula  $p$ -dimensional y  $F_1, \dots, F_p$  son FDA's uno dimensionales, entonces la función  $H$  definida por (2.4.1) es una FDA conjunta con marginales  $F_1, \dots, F_p$ .

Si contamos con un v.a.  $\mathbf{X} = (X_1, \dots, X_p)^T$  dos problemas son de interés:

- el estudio del comportamiento probabilístico de cada uno de sus componentes;
- investigar las relaciones que existen entre las variables involucradas.

La teoría de cópulas es una herramienta potencial para brindar respuestas al segundo problema. El Teorema 2.4.2 brinda una vía natural para extender los catálogos de distribuciones conjuntas, ya que cualquier cópula puede ser combinada con diferentes FDA's univariadas vía (2.4.1), logrando así modelos multivariados flexibles capaces de capturar estructuras de dependencia complejas. En contraste con los modelos clásicos como el de distribución T, gaussiana, gamma, etc. en donde las marginales—y las condicionales en algunos casos—pertenecen a la misma familia de distribución, una de las grandes ventajas de los modelos de cópulas es que son útiles en situaciones en donde se necesitan diferentes distribuciones para cada marginal en términos de comportamientos en las colas, asimetrías, etc. En particular, más adelante explicaremos como vía cópulas podremos inferir las relaciones de independencia condicional para la construcción de un grafo no dirigido.

Si contamos con una FDA conjunta,  $H$ , con marginales uno dimensionales  $F_1, \dots, F_p$ , partiendo de (2.4.1) es inmediato extraer la cópula correspondiente, evaluando a la función de distribución conjunta en cada argumento con las correspondientes inversas de las FDA's univariadas,  $F_j^{-1}$   $j \in \{1, \dots, p\}$ , siempre y cuando éstas existan. El siguiente corolario del Teorema de Sklar, 2.4.1, muestra esta relación.

**Corolario 2.4.1** Sean  $H$ ,  $F_1, \dots, F_p$  y  $\mathcal{C}$  como en el Teorema 2.4.1, además si cada una de las FDA's univariadas,  $F_j$ 's, es continua, entonces  $\forall (u_1, \dots, u_p) \in \mathbb{I}^p$ ,

$$\mathcal{C}(u_1, \dots, u_p) = H(F_1^{-1}(u_1), \dots, F_p^{-1}(u_p)). \quad (2.4.2)$$

Pero si una de las marginales uno dimensionales no es estrictamente creciente, entonces ésta no posee una inversa en los términos usuales. Por lo tanto necesitaremos recordar el concepto de *inversa generalizada* para FDA's univariadas.

**Definición 2.4.1** Sea  $F$  una función de distribución uno dimensional. Entonces una *inversa generalizada* de  $F$  es una función  $F^{(-1)}$  con dominio  $\mathbb{I}$  tal que,

1. si  $t \in \text{Ran } F$ , entonces  $F^{(-1)}(t)$  es cualquier número  $x \in \overline{\mathbb{R}}$  tal que  $F(x) = t$ , i.e.  $\forall t \in \text{Ran } F$ ,

$$F(F^{(-1)}(t)) = t;$$

2. si  $t \notin \text{Ran } F$ , entonces

$$F^{(-1)}(t) = \inf \{x | F(x) \geq t\} = \sup \{x | F(x) \leq t\}.$$

Cabe mencionar que si  $F$  es estrictamente creciente ésta tiene una única inversa generalizada, la inversa usual, en tal caso se usará la notación acostumbrada  $F^{-1}$ .

Para resultados más generales partiendo de inversas generalizadas y subcópulas, concepto que generaliza a las funciones cópula, véase [Nelsen \(2007\)](#).

**Observaciones 2.4.1 :**

**(Obs. 1)** Dado un v.a.  $\mathbf{X} = (X_1, \dots, X_p)^T$  constituido por variables continuas, las funciones cópula son una vía para transformar  $\mathbf{X}$  en otro v.a.  $\mathbf{U}$  cuyas marginales son uniformes sobre  $\mathbb{I}$ , t.q.

$$\mathbf{U} = (U_1, \dots, U_p)^T = (F_1(X_1), \dots, F_p(X_p))^T$$

y donde este último preserva la estructura de dependencia entre los componentes del primer vector. El Lema 2.5.1 de la siguiente sección justifica esta última aseveración.

## 2.5. Cópulas y vectores aleatorios

En esta sección se presentan algunos ejemplos elementales sobre la cópula asociada a estructuras de dependencia particulares, junto con algunas propiedades probabilísticas de las funciones cópula.

El primero, es un resultado importante que utilizaremos recurrentemente en lo subsecuente del presente capítulo, éste asegura la invarianza de la cópula ante transformaciones monótonas crecientes de las variables (c.f. [Nelsen \(2007\)](#), Teoremas 2.4.3 y 2.4.4, para la demostración bivariada del siguiente Lema, además cuando al menos se tiene una transformación estrictamente decreciente se presentan resultados para  $\mathcal{C}_T$  como simples transformaciones de  $\mathcal{C}_X$  nuevamente en el caso bivariado).

**Lema 2.5.1** Sean  $\mathbf{X} = (X_1, \dots, X_p)^T$  un v.a. con FDA conjunta continua,  $H$ , y cópula  $\mathcal{C}_X$ . Por otro lado, sean  $T_1, \dots, T_p$  funciones estrictamente crecientes de  $\mathbb{R}$  a  $\mathbb{R}$ . Si  $\mathcal{C}_T$  es la cópula del v.a.  $\mathbf{T} = (T(X_1), \dots, T(X_p))^T$ , entonces  $\mathcal{C}_X = \mathcal{C}_T$ .

Los siguientes resultados sirven para caracterizar estructuras particulares de dependencia de v.a. en términos de las cópulas  $\Pi_p$ ,  $\mathcal{M}_p$  y  $\mathcal{W}_2$ .

**Lema 2.5.2** Sea  $\mathbf{X} = (X_1, \dots, X_p)^T$  un v.a. con FDA continua,  $H$ . Entonces  $\mathcal{C}_X = \Pi_p$ , si y sólo si,  $X_1, \dots, X_p$  son independientes.

**Lema 2.5.3** Sea  $\mathbf{X} = (X_1, \dots, X_p)^T$  un v.a. con FDA continua,  $H$ . Entonces  $\mathcal{C}_X = \mathcal{M}_p$ , si y sólo si, existe una variable aleatoria  $Z$  y funciones estrictamente crecientes  $T_1, \dots, T_p$  tales que  $\mathbf{X} = (T_1(Z), \dots, T_p(Z))^T$  casi seguramente.

**Lema 2.5.4** Sea  $\mathbf{X} = (X_1, X_2)$  un v.a. con FDA continua,  $H$ . Entonces  $\mathcal{C}_X = \mathcal{W}_2$ , si y sólo si, para alguna función estrictamente decreciente,  $T$ ,  $X_1 = T(X_2)$  casi seguramente.

**Observaciones 2.5.1 :**

(Obs. 1) El Lema 2.5.2 surge a partir del Teorema de Sklar, 2.4.1, y el Teorema 2.2.2.

(Obs. 2) El Lema 2.5.4 no puede ser extendido para  $p \geq 3$ , a partir del hecho de que no es posible que cada una de las  $p$  variables aleatorias sea función decreciente de las restantes.

(Obs. 3) Notemos que en la presente sección ocupamos la notación  $\mathcal{C}_X$  para referirnos a la cópula subyacente del v.a.  $\mathbf{X}$ , esta notación es usual en la literatura y nosotros la ocuparemos esporádicamente. Por facilidad de notación será más común que ocupemos  $\mathcal{C}_A$ , con  $A \subseteq \mathcal{V}$ , para referirnos a la cópula de  $\mathbf{X}_A$ ,  $\mathcal{C}_{\mathbf{X}_A}$ , a menos que por el contexto exista ambigüedad sobre el vector asociado a la cópula.

## 2.6. Dependencia vía cópulas

El Lema 2.5.2 brinda una caracterización particular para la independencia de variables aleatorias vía la cópula  $\Pi$  y partiendo del hecho de que la cópula contiene la información conjunta de un v.a. particular, i.e. la información sobre la estructura de dependencia entre las variables, es natural pensar que a partir de la cópula subyacente de un par de variables aleatorias se puede llegar a cuantificar el grado de dependencia entre éstas como lo estudiaremos en esta sección. Comenzaremos por enunciar las propiedades que debe cumplir una medida de dependencia.

### Propiedades de una medida de dependencia

En [Schweizer and Wolff \(1981\)](#) proponen la siguiente definición de medida de dependencia.

**Definición 2.6.1** Sean  $X_1, X_2$  cualesquiera variables aleatorias continuas con cópula subyacente  $\mathcal{C}$ . Una medida numérica de asociación para este par de variables aleatorias es *medida de dependencia*, denotada por  $\mu_{X_1, X_2}$  ó  $\mu_{\mathcal{C}}$ ; si satisface las siguiente propiedades:

1.  $\mu_{X_1, X_2}$  está definida para cualquier par de variables aleatorias continuas;
2.  $0 \leq \mu_{X_1, X_2} \leq 1$ ;

3.  $\mu_{X_1, X_2} = \mu_{X_2, X_1}$ ;
4.  $X_1, X_2$  son independientes si y sólo si  $\mu_{X_1, X_2} = \mu_{\Pi} = 0$ ;
5.  $\mu_{X_1, X_2} = 1$  si y sólo si  $X_1, X_2$  son cada una casi seguramente función estrictamente monótona de la otra; es decir, si  $\mathcal{C}$  es igual a  $\mathcal{M}$  o  $\mathcal{W}$ ;
6. Si  $\alpha$  y  $\beta$  son casi seguramente funciones estrictamente monótonas en  $\text{Ran } X_1$  y  $\text{Ran } X_2$ , respectivamente, entonces  $\mu_{\alpha(X_1), \beta(X_2)} = \mu_{X_1, X_2}$ ;
7. Si  $\{(X_{1,n}, X_{2,n})\}$  es una sucesión de variables aleatorias continuas con cópula  $\mathcal{C}_n$ , respectivamente, y  $\{\mathcal{C}_n\}$  converge puntualmente a  $\mathcal{C}$ , entonces  $\lim_{n \rightarrow \infty} \mu_{\mathcal{C}_n} = \mu_{\mathcal{C}}$ .

Cabe mencionar que en general ni  $\rho$ , el coeficiente de correlación lineal de Pearson, ni  $\rho_S$ , el coeficiente de concordancia de Spearman<sup>1</sup>, cumplen con ser medidas de dependencia. Una de las propiedades más importantes que no cumplen es la *cuatro*, ya que no son capaces de representar la independencia; sin embargo cabe destacar que bajo *dependencia monótona*<sup>2</sup>, estructura de las cual suelen partir diversos modelos de probabilidad,  $\rho_S = 0$  si implica independencia, el detalle de esta aseveración se comenta en la sección de Observaciones 2.6.1. Entonces tanto  $\rho$  como  $\rho_S$  en términos generales *no son herramientas ideales* para detectar relaciones de independencia o independencia condicional, siendo éste un objetivo primordial de la presente investigación. En particular el coeficiente de correlación lineal siempre se ve afectado por efectos marginales, lo anterior no es deseable para una medida de dependencia debido a que la estructura de dependencia entre las variables sólo está contenida en la cópula. Por otro lado, pese a que el coeficiente de concordancia de Spearman sólo depende de la cópula subyacente, desviaciones positivas de la cópula  $\Pi$  pueden anular desviaciones negativa, o viceversa, haciendo que el valor de cero no se reserve para la independencia. Cabe destacar que ambas medidas son capaces de explicar correctamente el grado de asociación lineal y el grado de concordancia/discordancia entre las variables, respectivamente, el problema es que para efectos prácticos las estructuras de dependencia que aparecen en datos multivariados en general son mucho más complejas y recurrentemente se abusa del uso de estas medidas de asociación.

Profundizar sobre las propiedades, ventajas y desventajas de las medidas de asociación  $\rho$  y  $\rho_S$  está fuera de los alcances de la presente investigación. Pero se recomienda Nelsen (2007), sección 5.1, para una discusión sobre la concordancia/discordancia como un tipo de dependencia, además se presentan algunas medidas de concordancia como rho de Spearman, también se recomienda Kruskal (1958) como una referencia clásica con resultados interesantes sobre medidas de concordancia. Por otro lado, en la Sección 3.5 se puntualiza sobre algunas limitantes del coeficiente de correlación lineal de Pearson como parámetro de una cópula gaussiana.

## Medidas de dependencia basadas en cópulas

En Schweizer and Wolff (1981) comentan:

Ya que  $X_1$  y  $X_2$  son independientes si y sólo si  $\mathcal{C}(u, v) = uv$ , esta observación sugiere que cualquier medida de distancia, adecuadamente estandarizada, entre las superficies  $z = \mathcal{C}(u, v)$  y  $z = uv$ , por ejemplo, cualquier distancia  $L_p$ , debiera proporcionar una medida simétrica y no paramétrica de dependencia.

Partiendo del comentario en Schweizer and Wolff (1981) es que la siguiente propuesta de distancia hacia la cópula  $\Pi$  define una *medida de dependencia*.

**Definición 2.6.2** Para cualquier  $p \in [1, \infty)$  la distancia  $L_p$  entre una cópula  $\mathcal{C}$  y  $\Pi$  esta dada por:

$$\mathcal{V}_p(\mathcal{C}) = \left( k_p \int \int_{\mathbb{I}^2} |\mathcal{C}(u, v) - uv|^p du dv \right)^{\frac{1}{p}}, \quad (2.6.1)$$

donde  $k_p$  es una constante elegida de tal modo que (2.6.1) sea igual a 1 cuando  $\mathcal{C} = \mathcal{M}$  ó  $\mathcal{W}$ .

<sup>1</sup>Dado un par de variables aleatorias  $X_1$  y  $X_2$  con FDA's  $F_1$  y  $F_2$ , respectivamente, el coeficiente de concordancia rho de Spearman,  $\rho_S$ , entre  $X_1$  y  $X_2$ , se define como:  $\rho_S(X_1, X_2) = \text{Corr}(F_1(X_1), F_2(X_2))$ . Notemos de su definición, que  $\rho_S$  depende de observaciones de la cópula más que de las variables originales, por lo tanto se elimina el efecto marginal y su valor sólo depende de la cópula subyacente.

<sup>2</sup>Si dada una cópula  $\mathcal{C}$ ,  $\mathcal{C}(u, v) \geq \Pi(u, v) \forall u, v \in \mathbb{I}$  (cuadrante positiva) ó  $\mathcal{C}(u, v) \leq \Pi(u, v) \forall u, v \in \mathbb{I}$  (cuadrante negativa), entonces se dice que  $\mathcal{C}$  modela una dependencia monótona.

De [Nelsen \(2007\)](#), tenemos la siguiente expresión para  $k_p$ .

**Teorema 2.6.1** Para cada  $p \in [1, \infty)$  la distancia  $L_p$  calculada mediante  $\mathcal{V}_p(C)$  es una medida de dependencia, en donde:

$$k_p = \frac{\Gamma(2p+3)}{2[\Gamma(p+1)]^2}. \quad (2.6.2)$$

Algunos casos particulares estudiados son:

1.  $\mathcal{V}_1(C) = \sigma_{SW}(\mathbf{X}, \mathbf{Y})$ , la medida de dependencia propuesta en [Schweizer and Wolff \(1981\)](#), donde

$$\mathcal{V}_1(C) = \sigma_{SW}(X_1, X_2) = \sigma_{SW}(C) = 12 \int \int_{\mathbb{I}^2} |\mathcal{C}(u, v) - uv| dudv. \quad (2.6.3)$$

2.  $(\mathcal{V}_2(C))^2 = \phi^2(X_1, X_2)$ , mejor conocido como el índice de dependencia de [Hoeffding \(1940\)](#), donde

$$\mathcal{V}_2(C) = \phi(X_1, X_2) = \phi(C) = \left( 90 \int \int_{\mathbb{I}^2} |\mathcal{C}(u, v) - uv|^2 dudv \right)^{\frac{1}{2}}. \quad (2.6.4)$$

**Observaciones 2.6.1 :**

**(Obs. 1)** Al asumir dependencia monótona se sigue que  $|\rho_S| = \sigma_{SW}$ , lo anterior justifica que sólo bajo esta estructura de dependencia  $\rho_S = 0$  implica independencia.

## Dependencia en colas

A continuación presentamos un tipo particular de dependencia conocido como *dependencia en colas*<sup>3</sup>. El estudio de la dependencia en colas ha cobrado bastante relevancia en las últimas dos décadas, principalmente en contextos en donde la dependencia entre variables aleatorias cambia ante eventos extremos de alguna(s) variable(s), v.g. en finanzas (c.f. [Nikoloulopoulos et al. \(2012\)](#) sección 7, se presenta un caso de aplicación a datos de retornos financieros en donde se exhibe lo complejo que puede llegar a ser modelar la estructura de dependencia, en particular la dependencia en colas). Hemos decidido puntualizar en este tipo de dependencia con la finalidad de profundizar más adelante sobre la incapacidad para modelar dependencia en colas que tienen los modelos de cópula gaussiana, y en particular los de distribución gaussiana, los cuales hoy en día siguen siendo bastante utilizados para modelar fenómenos en donde la dependencia en colas es inherente. Además en el Capítulo 3 los modelos de cópula gaussiana cobran una relevancia bastante importante en la construcción de la GIC de un modelo gráfico no dirigidos, así que comprender las limitantes que tiene una cópula gaussiana como modelo conjunto es un factor crucial en el siguiente capítulo.

**Definición 2.6.3** Sean  $X_1$  y  $X_2$  variables aleatorias continuas con funciones de distribución  $F_1$  y  $F_2$ , respectivamente. Denotaremos con  $\lambda_U$  al parámetro de *dependencia en la cola superior*,  $\lambda_U$  es el límite (si éste existe) de la probabilidad condicional de que  $X_2$  sea mayor que el percentil al 100% de  $F_2$  dado que  $X_1$  es mayor que el percentil al 100% de  $F_1$ , cuando  $t$  se aproxima a 1 por la izquierda, i.e.

$$\lambda_U = \lim_{t \rightarrow 1^-} \mathbb{P} \left[ X_2 > F_2^{(-1)}(t) | X_1 > F_1^{(-1)}(t) \right]. \quad (2.6.5)$$

Similarmente, denotemos con  $\lambda_L$  al parámetro de *dependencia en la cola inferior*,  $\lambda_L$  es el límite (si éste existe) de la probabilidad condicional de que  $X_2$  sea menor o igual que el percentil al 0% de  $F_2$  dado que  $X_1$  es menor o igual que el percentil al 0% de  $F_1$ , cuando  $t$  se aproxima a 0 por la derecha, i.e.

$$\lambda_L = \lim_{t \rightarrow 0^+} \mathbb{P} \left[ X_2 \leq F_2^{(-1)}(t) | X_1 \leq F_1^{(-1)}(t) \right]. \quad (2.6.6)$$

---

<sup>3</sup>Tail dependence.

Estos parámetros dependen sólo de la cópula entre  $X_1$  y  $X_2$ , como se enuncia en el siguiente Teorema.

**Teorema 2.6.2** Sean  $X_1, X_2, F_1, F_2, \lambda_U$  y  $\lambda_L$  como en la Definición 2.6.3, además sea  $\mathcal{C}$  la cópula entre  $X_1$  y  $X_2$ . Si los límites en (2.6.5) y (2.6.6) existen, entonces

$$\lambda_U = 2 - \lim_{t \rightarrow 1^-} \frac{1 - \mathcal{C}(t, t)}{1 - t} = 2 - \delta'_\mathcal{C}(1^-),$$

además,

$$\lambda_L = \lim_{t \rightarrow 0^+} \frac{\mathcal{C}(t, t)}{t} = \delta'_\mathcal{C}(0^+),$$

donde  $\delta_\mathcal{C}(t) = \mathcal{C}(t, t)$  y se le conoce como la sección diagonal de la cópula  $\mathcal{C}$ .

Para más detalles sobre la dependencia en colas y el cálculo explícito de los parámetros  $\lambda_U$  y  $\lambda_L$  para distintas familias de cópulas, se puede consultar [Nelsen \(2007\)](#), sección 5.4. Si se desea profundizar en los distintos tipos de dependencia que pueden ser de interés en un modelo conjunto, además de la dependencia en colas, se recomienda [Mari and Kotz \(2001\)](#) para una investigación detallada sobre diversas estructuras de dependencia que pueden ser modeladas.

## 2.7. Cópulas condicionales

Regresando al objetivo primordial de nuestra investigación, la presente sección tiene el objetivo de dar un panorama general sobre el contexto de las funciones cópula condicionales, además de mostrar una extensión al Teorema de Sklar, 2.4.1, hacia distribuciones condicionales. Finalmente se explica la relevancia del supuesto simplificador para inferir la estructura de una GIC.

### Panorama general

En la última década el concepto de cópula condicional ha logrado captar la atención en muchos terrenos de la investigación, debido a que muchos modelos estadísticos se basan en distribuciones condicionales entre las variables de estudio, v.g. modelos de regresión múltiple, modelos de series de tiempo, modelos gráficos, etc., y es muy común que se recurra a supuestos de distribuciones conjuntas en donde es fácil la obtención de las distribuciones condicionales, v.g. una distribución gaussiana o una distribución T, modelos que en algunos contextos no son lo suficientemente flexibles para captar estructuras complejas de dependencia.

Las finanzas son el campo en donde más se ha explotado el potencial de las cópulas (no condicionadas) debido a lo complejo que puede llegar a ser modelar la dependencia entre activos financieros, y también es el terreno donde se han generado más publicaciones de aplicaciones con cópulas condicionales, v.g. [Palaro and Hotta \(2006\)](#), [Rodríguez \(2007\)](#), [Huang et al. \(2009\)](#) y [Hu \(2010\)](#). Asimismo, cabe destacar el trabajo pionero en [Patton \(2001a\)](#), [Patton \(2001b\)](#) y [Patton \(2002\)](#) sobre el estudio de cópulas que varían en el tiempo. Él introduce el concepto de cópula condicional, y en [Patton \(2006\)](#) lo aplica al estudio de asimetrías en la estructura de dependencia para variables de tipo de cambio; en ese mismo artículo se presenta una discusión general sobre cópulas condicionales.

Por otro lado, para la teoría de *vine cópula* el concepto de cópula condicional es vital para la construcción de distribuciones conjuntas vía cópulas a pares. En [Czado \(2010\)](#) se presenta un panorama general sobre esta teoría. En el Capítulo 4 ocuparemos algunas ideas de la teoría de *vine cópula* para los fines de estimar una GIC.

En el contexto de estadística *no paramétrica* en [Gijbels et al. \(2011\)](#) y [Veraverbeke et al. \(2011\)](#) se discuten estimadores para cópulas condicionales bivariadas con condicionantes de dimensión uno, pero no es muy clara la extensión a condicionantes de mayor dimensión—como los necesitamos en la presente investigación—.

## Extensión del Teorema de Sklar

A continuación se enuncia la extensión del Teorema de Sklar hacia distribuciones condicionales. El Teorema se enunciará de manera general y se ocupará la notación de la Definición de marginales, 2.2.3, asociada a los conjuntos de vértices y aristas.

**Teorema 2.7.1** Sea  $\mathbf{X} = (X_1, \dots, X_p)^T$  un v.a. de dimensión  $p$ . Por otro lado, sean  $A \subset \mathcal{V} = \{1, \dots, p\}$  tal que  $\text{card}(A) = m \geq 2$ , y  $B \subset \mathcal{V}$  tales que  $A \cap B = \emptyset$ . Tomemos  $x_{j_i} \in \text{Ran } X_{j_i}$  para  $j_i \in A$ ,  $i = 1, \dots, m$ , además  $\mathbf{v}_B \in \text{Ran } \mathbf{X}_B$ .

Si  $H_{A|B}(\cdot, \dots, \cdot | \mathbf{v}_B)$  es la FDA condicional de  $X_A$  dado  $X_B = \mathbf{v}_B$ , y además  $F_{j_1|B}(\cdot | \mathbf{v}_B), \dots, F_{j_m|B}(\cdot | \mathbf{v}_B)$  son las FDA's condicional de  $X_{j_1}, \dots, X_{j_m}$ , respectivamente, dado  $X_B = \mathbf{v}_B$ , entonces existe una cópula condicional  $\mathcal{C}_{A|B}(\cdot, \dots, \cdot; \mathbf{v}_B)$  tal que  $\forall x_{j_1}, \dots, x_{j_m} \in \overline{\mathbb{R}}$ ,

$$H_{A|B}(x_{j_1}, \dots, x_{j_m} | \mathbf{v}_B) = \mathcal{C}_{A|B}(F_{j_1|B}(x_{j_1} | \mathbf{v}_B), \dots, F_{j_m|B}(x_{j_m} | \mathbf{v}_B); \mathbf{v}_B). \quad (2.7.1)$$

De manera análoga al Teorema de Sklar, 2.4.1, si las funciones  $F_{j_1|B}(\cdot | \mathbf{v}_B), \dots, F_{j_m|B}(\cdot | \mathbf{v}_B)$  son todas continuas, la cópula  $\mathcal{C}_{A|B}(\cdot, \dots, \cdot; \mathbf{v}_B)$  es única.

### Observaciones 2.7.1 :

**(Obs. 1)** La notación  $\mathcal{C}_{A|B}(\cdot, \dots, \cdot; \mathbf{v}_B)$  hace referencia a la cópula de  $X_A$  condicional al valor  $\mathbf{v}_B \in \text{Ran } \mathbf{X}_B$ . En lo subsecuente del documento se implementará toda la notación contenida en el Teorema 2.7.1, lo anterior para facilitar la interpretación hacia el modelo gráfico.

Como se comentó en la Sección 2.6 las cópulas son una herramienta potencial para modelar la estructura de dependencia entre variables aleatorias, vía la extensión del Teorema de Sklar, 2.7.1, las cópulas se convierten en la opción ideal para modelar también la dependencia condicional entre variables. De esta forma el Teorema de extensión se vuelve fundamental para los fines de la presente investigación. Para determinar las relaciones de independencia condicional modeladas en una GIC a partir de datos, bastará considerar cópulas condicionales a pares y de alguna manera inferir la presencia de independencia condicional como lo veremos en el Capítulo 4.

Es importante mencionar que el Corolario 2.4.1 del Teorema de Sklar, 2.4.1 también aplica para extraer cópulas condicionales de distribuciones conjuntas y marginales condicionales. En algunos desarrollos subsecuentes haremos uso de la extensión de este corolario.

Notemos que en general una cópula condicional depende del valor del condicionante,  $\mathbf{v}_B$  en el caso del Teorema de extensión, 2.7.1. De cierto modo podemos entender al valor del condicionante como un parámetro de la cópula condicional. Cuando la familia de cópulas a la que pertenezca la cópula no condicional cumpla con el *supuesto simplificador* eliminaremos de la cópula condicional el valor del condicionante, i.e.

$$\mathcal{C}_{A|B}(\cdot, \dots, \cdot; \mathbf{v}_B) = \mathcal{C}_{A|B}(\cdot, \dots, \cdot).$$

## Supuesto simplificador

Diremos que una cópula o una familia de cópulas cumple el *supuesto simplificador*<sup>4</sup> si *todas* sus cópulas condicionales a pares<sup>5</sup> dependen del condicionante sólo vía sus distribuciones condicionales (c.f. [Haff et al. \(2010\)](#) para una discusión sobre los alcances del supuesto simplificador). En otras palabras las cópulas condicionales no cambian con el valor del condicionante.

El supuesto simplificador es de mucha relevancia para la presente investigación. El hecho es que si partimos de un v.a.  $\mathbf{X} = (X_1, \dots, X_p)^T$ , para que se logre la independencia entre  $X_j$  y  $X_k$  condicional en  $\mathbf{X}_{\mathcal{V} \setminus \{j,k\}}$ , tiene que ocurrir que

<sup>4</sup>Simplifying assumption.

<sup>5</sup>Pair copulas.

$$X_j \perp\!\!\!\perp X_k | \mathbf{X}_{\mathcal{V} \setminus \{j,k\}} = \mathbf{v}, \quad \forall \mathbf{v} \in \text{Ran } \mathbf{X}_{\mathcal{V} \setminus \{j,k\}},$$

en otras palabras,  $\forall u_1, u_2 \in \mathbb{I}$ ,

$$\mathcal{C}_{j,k|\mathcal{V} \setminus \{j,k\}}(u_1, u_2; \mathbf{v}) = \mathcal{C}_{j,k|\mathcal{V} \setminus \{j,k\}}(u_1, u_2) = \Pi(u_1, u_2), \quad \forall \mathbf{v} \in \text{Ran } \mathbf{X}_{\mathcal{V} \setminus \{j,k\}}. \quad (2.7.2)$$

Entonces necesitamos que *al menos algunas* de las cópulas condicionales a pares cumplan con ser invariantes al cambio en el valor del condicionante. Por decirlo de un modo informal, necesitamos que *parcialmente* se cumpla el supuesto simplificador. Si partimos de un modelo de cópula o de una familia de cópulas en donde no se cumpla el supuesto simplificador (al menos parcialmente) simplemente será *imposible* deducir las relaciones de independencia condicional de una GIC.

Si tomamos un enfoque de estimación del grafo en el cual partimos de suponer que una *sola cópula paramétrica* explica la estructura completa de dependencia entre las variables de un v.a.  $\mathbf{X}$ , entonces necesitaremos recurrir a familias paramétricas de cópulas en donde se cumpla el supuesto simplificador: cópulas simplificadas.

Hay tres familias paramétricas de cópulas absolutamente continuas<sup>6</sup> que cumplen con el supuesto simplificador (c.f. Stöber et al. (2013) para la deducción de este resultado). Las tres familias de cópulas simplificadas se enuncian a continuación:

- gaussiana.
- $\mathcal{T}$ .
- Clayton.

Considerando que el caso de cópulas que no son absolutamente continuas rara vez es relevante en la práctica, nos acotaremos a estudiar los modelos absolutamente continuos.

## 2.8. Discusión

A lo largo de este capítulo presentamos a las funciones cópula multivariadas y su relación con la construcción de modelos conjuntos de variables aleatorias, discutimos sobre sus propiedades y se expusieron ejemplos concretos con la finalidad de comprender su capacidad para generar distribuciones multivariadas flexibles, donde se modela por separado la estructura de dependencia entre las variables y los comportamientos marginales. Se puntualizó sobre la existencia de medidas de dependencia basadas en cópulas, las cuales capturan eficientemente el grado de dependencia entre pares de variables aleatorias, en contraste con las medidas  $\rho$  y  $\rho_S$  que miden sólo asociaciones particulares entre variables.

Finalmente se introdujo el concepto de cópula condicional, vital para nuestra investigación, y se comentó sobre el supuesto simplificador necesario para poder deducir relaciones de independencia condicional en una GIC. Adicionalmente se enunciaron las tres familias de cópulas multivariadas que cumplen con este supuesto. Así que, como lo adelantamos en la introducción de este capítulo, si buscamos modelos gráficos no dirigidos markovianos que tengan asociadas distribuciones conjuntas flexibles, la pregunta con la que cierra este capítulo es: ¿los modelos de cópulas gaussiana,  $\mathcal{T}$  y Clayton son lo suficientemente flexibles como para modelar estructuras complejas de dependencia? En caso afirmativo, nos interesa saber si las GIC's asociadas no son casos triviales. Dar respuesta a estas interrogantes es la tarea fundamental del siguiente Capítulo 3.

<sup>6</sup>Un v.a.  $\mathbf{X}$ , y equivalentemente su FDA  $H_{\mathbf{X}}$ , es absolutamente continuo si éste tiene densidad. De igual forma, la cópula subyacente de un v.a. absolutamente continuo es una cópula absolutamente continua.



## Capítulo 3

# Modelos gráficos no dirigidos de cópulas simplificadas

**Resumen** En una primera instancia, analizaremos el enfoque de deducción de una GIC asumiendo un solo modelo paramétrico de cópula que represente la estructura completa de dependencia del v.a. en cuestión. El presente capítulo conjuga la definición de una GIC en términos de la propiedad de Markov a pares con la estructura de dependencia que modela cada una de las tres cópulas simplificadas. Estudiaremos estas familias de cópulas en tres términos generales: la estructura de dependencia que modelan con bases en sus parámetros, el comportamiento de sus cópulas condicionales y el modelo gráfico asociado a cada una de éstas. Al estudiar el modelo gráfico asociado nos enfocaremos en comprender en qué medida un grafo no dirigido, estimado a partir de una muestra proveniente de cada una de estas cópula, cumple con la propiedad de Markov a pares. Una de las grandes conclusiones del capítulo, es que si deseamos imponer una sola cópula paramétrica al vector  $\mathbf{X}$ , el único modelo para el cual técnicamente se deducen *grafos no triviales* es el de cópula gaussiana y por ende el de distribución gaussiana. Sin embargo, dada la relación que existe entre la cópula  $\mathcal{T}$  y la gaussiana, se logra dar una interpretación y *utilidad* a un grafo de cópula  $\mathcal{T}$ . Por otro lado, el grafo Clayton se descarta como una buena opción de modelo por siempre ser un grafo trivial. Independientemente de la estructura de dependencia que sean capaces de modelar las dos cópulas que resultan tener modelos gráficos no dirigidos *útiles*, al final del capítulo se cuestiona el hecho de imponer un sólo modelo de cópula paramétrica que explique la estructura completa de dependencia de un v.a.

### 3.1. Organización del capítulo

En términos generales, el presente capítulo busca comprender la *utilidad* de un modelo gráfico no dirigido de cópula simplificada para cada una de las tres familias correspondientes, junto con la *flexibilidad* que tiene cada una de éstas como modelo conjunto. La Sección 3.2 da un panorama general sobre alternativas existentes en la literatura de modelos gráficos más flexibles que el de distribución gaussiana y comenta brevemente sobre sus alcances. La Sección 3.3 estudia a las cópulas elípticas simplificadas, i.e. a las familias de cópulas gaussiana y  $\mathcal{T}$ , en términos de la estructura de dependencia que modelan, el comportamiento de sus cópulas condicionales y la *utilidad* que tiene su modelo gráfico no dirigido asociado; por otro lado, la Sección 3.4 hace lo propio con la cópula Clayton como el miembro simplificado de la clase de cópulas Arquimedianas. En la Sección 3.5 se enfatiza sobre las limitantes que tiene una cópula gaussiana con respecto a otros modelos de copulas. Se concluye este capítulo, con la Sección 3.6, discutiendo sobre los alcances que tienen los modelos gráficos de cópula paramétrica simplificada que resultaron *útiles*, i.e. los modelos gráficos de cópula gaussiana y  $\mathcal{T}$ , además sobre las limitantes per se que tiene imponer un solo modelo de cópula paramétrica como modelo conjunto.

### 3.2. Modelos gráficos no dirigidos de cópulas simplificadas

Como lo comentamos en el Capítulo 1 el modelo gráfico por excelencia es el gaussiano, modelo ya bastante estudiado en la literatura tanto de altas como de bajas dimensiones; en el Capítulo 5 se profundizará sobre este modelo en altas dimensiones. Dadas las limitantes intrínsecas de un modelos gaussiano, en años recientes se han presentado

propuestas que prometen ser más flexibles en términos de la estructura de dependencia que modelan, en particular estos artículos han estado enfocados en la estimación del grafo en altas dimensiones. Nos referimos a la extensión del modelo gráfico gaussiano hacia el modelo de cópula gaussiana, Liu et al. (2009), y al planteamiento de los mismos autores para modelos gráficos de cópula elíptica, Liu et al. (2012b). Estos modelos asumen que la distribución de probabilidad del modelo gráfico,  $\mathbf{P}$ , tiene asociada una cópula gaussiana o una cópula elíptica, respectivamente; en otras palabras,  $\mathcal{C}_{\mathbf{X}}$  es gaussiana o elíptica.

Una familia de cópulas pertenece a la clase de cópulas elípticas si sus miembros provienen del Corolario 2.4.1 cuando  $H$  es una distribución elíptica particular (c.f. Durante and Sempi (2010), sección 1.6.1, para la definición formal y detalles sobre la clase de cópulas elípticas). Las familias de cópulas gaussiana y  $\mathcal{T}$  (las cópulas subyacentes de una distribución gaussiana y  $\mathcal{T}$ , respectivamente) son las únicas en la clase de las elípticas que cumplen con el supuesto simplificador como se prueba en Stöber et al. (2013). Dicho en otros términos, son las únicas dos familias de cópulas elípticas en donde tentativamente se podría deducir una GIC que tuviera *utilidad*. Así que quedan en duda los alcances y aplicación de la propuesta en Liu et al. (2012b) para cópulas elípticas en general. Más adelante en este capítulo buscaremos darle una interpretación a la estimación del grafo en el caso de cópula  $\mathcal{T}$  y retomaremos la discusión sobre la propuesta en Liu et al. (2012b).

Otra gran clase de familias de cópulas es la Arquimediana, misma que cuenta con un catálogo extenso de cópulas para el caso bivariado, las cuales modelan diversas estructuras de dependencia, además de ser una clase de cópulas que atrae particular atención debido a las numerosas propiedades que las hacen fáciles de analizar (c.f. Nelsen (2007) y Joe (1997), para revisiones generales sobre cópulas Arquimedianas y extensiones multivariadas). Sin embargo, las extensiones multivariadas que se han propuesto no necesariamente son demasiado flexibles como lo comentaremos brevemente en la Sección 3.4. Nuevamente, con base en Stöber et al. (2013) la única cópula Arquimediana simplificada es la Clayton, pero resulta que el grafo asociado *es el trivial* para cualquier valor de  $\theta$ , el parámetro de la cópula, como lo veremos en la Sección 3.4.

Recordemos de la Sección 1.6 que cuando decimos que la GIC de una cópula tenga *utilidad*, nos referimos a que en el modelo de cópula en realidad existan relaciones de independencia condicional que puedan ser representadas por una gráfica no dirigida y no caer en un grafo completo, o en su defecto tener una cópula en donde se presenten todas las relaciones de independencia condicional posibles y caer en un grafo vacío; nos referiremos a los casos triviales cuando el grafo sea completo o vacío. Más adelante cuando hablemos de los modelos gráficos de cópula  $\mathcal{T}$  y Clayton quedará más claro a lo que nos referimos.

Como lo hemos venido observando a lo largo del documento, el enfoque usual en la literatura de modelos gráficos no dirigidos, tanto en altas como en bajas dimensiones, es asumir un *solo modelo de cópula o de distribución* para los datos y con base en éste inferir la estructura del grafo. Asumir un *solo modelo de cópula* puede llegar a ser poco realista para algunos contextos en donde se requieren modelos más flexibles, más aún cuando se cuenta con un número grande de variables (c.f. Czado (2010) para una discusión sobre la pertinencia y necesidad de construir modelos más flexibles), pero entrar en esa discusión es tema del Capítulo 4.

Si quisiéramos tomar este enfoque de estimación del grafo en donde una sola cópula represente la estructura completa de dependencia del v.a. de análisis, a lo largo de este capítulo estudiaremos brevemente a las familias de cópulas simplificadas, con la finalidad de comprender cómo se comportan sus cópulas condicionales a pares y poder concluir si un modelo gráfico no dirigido es (o no) una herramienta ad hoc para explicar la estructura de dependencia de cada modelo de cópula.

Comenzaremos por estudiar a las dos cópulas elípticas simplificadas para luego pasar a la cópula Arquimediana que cumple el supuesto también.

### 3.3. Cópulas elípticas simplificadas

En esta sección evaluaremos la *utilidad* de un modelo gráfico no dirigido de cópula simplificada elíptica para cada una de las dos familias para las cuales aplica el supuesto simplificador.

A manera de referencia, queremos comentar que a los modelos de cópula elíptica en la literatura comúnmente se les conoce como modelos de distribución *metaelíptica* (c.f. Fang et al. (2002) para una revisión general sobre distribuciones metaelípticas), además en años recientes Liu et al. (2012b) introduce un modelo equivalente al cual le da el nombre de distribución *transelíptica*.

Comenzamos con un resultado que será de mucha relevancia en lo subsecuente del presente capítulo.

### Independencia en cópulas elípticas

La siguiente proposición asegura que la *única familia de cópulas* bivariadas elípticas que incluye a la cópula  $\Pi$  es la familia gaussiana, i.e. la única opción para caracterizar a la independencia en el contexto de cópula elíptica es vía una cópula gaussiana.

**Proposición 3.3.1** Sea el v.a.  $\mathbf{X} = (X_1, X_2)^T$  con cópula subyacente  $\mathcal{C}_{\mathbf{X}}$ , tal que  $\mathcal{C}_{\mathbf{X}}$  pertenece a la clase de cópulas elípticas, entonces si  $X_1$  y  $X_2$  son independientes se sigue que  $\mathcal{C}_{\mathbf{X}}$  es gaussiana.

Para detalles de la demostración de la Proposición 3.3.1 se puede consultar la Proposición 1.1 en Abdous et al. (2005), junto con el Lema 5 en Kelker (1970).

Pese a que la Proposición 3.3.1 está dirigida a relaciones de independencia no condicional, será vital cuando estudiemos el comportamiento de las cópulas condicionales a pares de los modelos de cópula gaussiana y T. De hecho este resultado es fundamental para las conclusiones del presente capítulo.

### Cópula gaussiana

Denotemos con  $g_{1:p}(\cdot, \dots, \cdot; \Sigma)$  a la densidad conjunta de un modelo  $\mathcal{N}_p(\mathbf{0}, \Sigma)$  con  $\Sigma \in \mathbb{R}^{p \times p}$  una matriz de correlaciones<sup>1</sup> positiva definida tal que  $\text{diag}(\Sigma) = \mathbf{1}$ . Entonces  $\forall \mathbf{x} = (x_1, \dots, x_p) \in \mathbb{R}^p$ ,

$$g_{1:p}(x_1, \dots, x_p; \Sigma) = g_{1:p}(\mathbf{x}; \Sigma) = \frac{1}{\sqrt{(2\pi)^p |\Sigma|}} \exp\left(-\frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x}\right). \quad (3.3.1)$$

La cópula gaussiana de dimensión  $p$  está dada por,

$$\begin{aligned} \mathcal{C}_{1:p}^G(u_1, \dots, u_p; \Sigma) &= \int_{-\infty}^{\Phi^{-1}(u_1)} \dots \int_{-\infty}^{\Phi^{-1}(u_p)} g_{1:p}(x_1, \dots, x_p; \Sigma) dx_1 \dots dx_p \\ &= \Phi_{\Sigma}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_p)), \end{aligned} \quad (3.3.2)$$

donde  $\Phi$  es la función de distribución (univariada) normal estándar y  $\Phi_{\Sigma}$  es la función de distribución conjunta de un modelo  $\mathcal{N}_p(\mathbf{0}, \Sigma)$ .

Notemos que si  $\mathbf{X} \sim \mathcal{N}_p(\mathbf{0}, \Sigma)$ , entonces a partir del Corolario 2.4.1 se sigue que la cópula de  $\mathbf{X}$ ,  $\mathcal{C}_{\mathbf{X}}$ , es aquella en (3.3.2).

#### Observaciones 3.3.1 :

**(Obs. 1)** Partiendo del hecho de que la cópula es invariante ante transformaciones monótonas crecientes de las variables, con base en el Lema 2.5.1, se sigue que si  $\mathbf{Y} \sim \mathcal{N}_p(\mu, P)$ , con  $P = \text{Cov}(\mathbf{Y})$  y  $\Sigma = \text{Corr}(\mathbf{Y})$ , y por otro lado  $\mathbf{X} \sim \mathcal{N}_p(\mathbf{0}, \Sigma)$ , con  $\Sigma = \text{Corr}(\mathbf{X})$ , entonces  $\mathcal{C}_{\mathbf{Y}} = \mathcal{C}_{\mathbf{X}}$ .

**(Obs. 2)** Notemos de (3.3.2) que una matriz de correlaciones lineales,  $\Sigma$ , es el parámetro que modela la dependencia en una cópula gaussiana, sin embargo cabe hacer distinción entre el parámetro de la cópula y la correlación entre las variables originales. Si un v.a.  $\mathbf{X} = (X_1, \dots, X_p)^T$  tiene asociada una cópula gaussiana  $\mathcal{C}^G$  como en (3.3.2), entonces la matriz  $\Sigma$  representa el nivel de asociación lineal entre una transformación particular de las variables  $F_1(X_1), \dots, F_p(X_p)$ , así que basta con la existencia de las FDA's  $F_1, \dots, F_p$  para que el

<sup>1</sup>Correlación lineal de Pearson.

parámetro  $\Sigma$  exista y el modelo de cópula gaussiana tenga sentido, aún cuando no exista la correlación entre las variables originales; v.g. si cada una de las variables  $X_1, \dots, X_p$  tuviese una distribución Cauchy univariada la correlación lineal entre cada una de éstas no existe, sin embargo las FDA's de las  $X_i$ 's si existen y por ende el modelo de cópula gaussiana aplicaría para explicar la dependencia entre las variables de interés. Más adelante se obtienen explícitamente las transformaciones que se aplican a las FDA's para la obtención de la matriz de correlaciones  $\Sigma$ .

### Estructura de dependencia: cópula gaussiana

Para comprender la estructura de dependencia que representa un modelo de cópula gaussiana, comenzaremos por definir al modelo de distribución *NonParaNormal* (Normal No Paramétrica) propuesto en Liu et al. (2009), el cual resulta ser un modelo equivalente al de cópula gaussiana.

**Definición 3.3.1 Distribución NonParaNormal**<sup>2</sup>. Diremos que  $\mathbf{X}$  sigue una distribución NonParaNormal (NPN),  $\mathbf{X} \sim NPN_p(\Sigma^0, f)$ , si existen un conjunto de funciones *monótonas crecientes* univariadas,  $f = \{f_1, \dots, f_p\}$ , y  $\Sigma^0 \in \mathbb{R}^{p \times p}$  una matriz de correlaciones (latente) positiva definida con  $\text{diag}(\Sigma^0) = \mathbf{1}$  de tal forma que  $f(\mathbf{X}) \sim \mathcal{N}_p(\mathbf{0}, \Sigma^0)$ .

Además el Lema 3.3.1 determina la forma de las funciones en  $f$  de la Definición 3.3.1.

**Lema 3.3.1** Consideremos un v.a.  $\mathbf{X} = (X_1, \dots, X_p)^T$  con FDA's marginal estrictamente crecientes  $F_1, \dots, F_p$ . Si existe un conjunto de funciones estrictamente crecientes  $f = \{f_j\}_{j=1}^p$  tales que  $f(\mathbf{X}) \sim \mathcal{N}_p(\mathbf{0}, \Sigma^0)$ , donde  $\text{diag}(\Sigma^0) = \mathbf{1}$ , tendremos que  $f_j = \Phi^{-1}(F_j)$ .

Finalmente, el Lema 3.3.2 asegura la equivalencia entre un modelo NPN y uno de cópula gaussiana.

**Lema 3.3.2** Sea  $\mathbf{X} = (X_1, \dots, X_p)^T$  un v.a. con FDA's marginal estrictamente crecientes  $F_1, \dots, F_p$ , entonces  $\mathbf{X} \sim NPN_p(\Sigma^0, f^0)$  si y sólo si

$$C_{\mathbf{X}}(u_1, \dots, u_p) = \Phi_{\Sigma^0}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_p)).$$

Podemos comprender la equivalencia de los modelos de cópula gaussiana y NPN vía el Lema 2.5.1 tomando a  $f = \{f_j(\cdot) = \Phi^{-1}(F_j(\cdot))\}_{j=1}^p$  como el conjunto de funciones monótonas crecientes, véase Apéndice A para demostraciones de los Lemas 3.3.1 y 3.3.2.

La relevancia del modelo NPN, es que toda vez que tengamos un v.a.  $\mathbf{X}$  con cópula  $\mathcal{C}_{1;p}$  como en (3.3.2), se asegura la existencia de un v.a. *latente*  $f(\mathbf{X})$  con distribución gaussiana y la misma estructura de dependencia que  $\mathbf{X}$ . En otras palabras, un modelo de cópula gaussiana representa la *misma* estructura de dependencia que un modelo de distribución gaussiana, pero permitiendo que las distribuciones marginales no necesariamente pertenezcan a esta familia. Entonces la estructura de dependencia entre las variables de  $\mathbf{X}$  está caracterizada por una matriz de correlaciones (lineales),  $\Sigma$ , que explica el grado de asociación lineal entre las variables aleatorias gaussianas  $\Phi^{-1}(F_1(X_1)), \dots, \Phi^{-1}(F_p(X_p))$ .

Entonces los modelos de cópula gaussiana sólo son capaces de explicar dependencias a pares vía el coeficiente de correlación lineal entre transformaciones particulares de las variables aleatorias. Como caso particular, la independencia entre variables está caracterizada por la correlación lineal nula ( $\rho = 0$ ) como se enuncia en el Lema 3.3.3.

<sup>2</sup>Observamos que la definición de la distribución NPN fue sufriendo modificaciones por sus autores a lo largo de una serie de artículos: Liu et al. (2009), Liu et al. (2012a) y Han and Liu (2014). De hecho notamos cierta *inconsistencia* en los artículos referentes a la estimación del modelo gráfico no dirigido en altas dimensiones vía NPN, Liu et al. (2009) y Liu et al. (2012a), en el sentido de que no limitaban a los elementos del conjunto de funciones,  $f$ , a ser monótonas estrictamente crecientes, imposibilitando así la aplicación directa del Lema 2.5.1, y en tal caso se perdía claridad sobre la equivalencia entre NPN y un modelo de cópula gaussiana. Posteriormente en Han and Liu (2014) consolidan las dos definiciones que aparecen en los artículos anteriores y corrigen el tema con el sentido de las funciones; pese a que este último artículo no está relacionado con grafos no dirigidos sino con *análisis de componentes principales*, las metodologías aplicadas están estrechamente relacionadas con las propuestas de estimación del grafo en los artículos al respecto.

**Lema 3.3.3** Sea  $\mathbf{X} = (X_1, X_2)^T$  con cópula  $C_{\mathbf{X}}$  como en (3.3.2) con  $p = 2$ . Entonces  $X_1$  y  $X_2$  son independientes si y sólo si  $\Sigma_{12} = \rho_{12} = 0$ .

La prueba de la proposición anterior es inmediata a partir de la definición de cópula gaussiana y el hecho de que independencia implica correlación lineal nula. La correlación nula no implica independencia en *términos generales*, dentro del *contexto de cópulas elípticas* esto sólo ocurre en el caso de cópula gaussiana. Más adelante, en la sección de Observaciones 3.3.2 proponemos una clase amplia de distribuciones bivariadas, fuera de la estructura de dependencia elíptica, en donde la correlación nula también caracteriza a la independencia; sin embargo sólo presentamos esta clase de distribuciones con un objetivo ilustrativo ya que como modelo conjunto es bastante restrictivo.

Por otro lado, una cópula gaussiana es incapaz de modelar dependencia en colas como en la Definición 2.6.3, ya que los parámetros de dependencia en colas inferior,  $\lambda_L^G$ , y superior,  $\lambda_U^G$ , de una cópula gaussiana son nulos, i.e.  $\lambda_L^G = \lambda_U^G = 0$ . Cabe mencionar que los parámetros de dependencia en colas para cópulas subyacente de distribuciones elípticas y simétricas siempre coinciden, i.e.  $\lambda_L = \lambda_U$ .

Finalmente, de (3.3.2) podemos observar que las cópulas marginales de una cópula gaussiana son también gaussianas; a partir de la observación anterior, cabe hacer énfasis en que el supuesto de cópula gaussiana implica aseverar que ningún par de variables tendrá dependencia en colas, supuesto que puede llegar a ser poco realista en diversos contextos. A continuación estudiaremos el comportamiento de las cópulas condicionales de cópula gaussiana.

#### Observaciones 3.3.2 :

**(Obs. 1)** Sabemos que en el contexto de cópulas elípticas la familia gaussiana de cópulas es la única que contiene a la independencia y que ésta es caracterizada por la correlación nula (Lemas 3.3.1 y 3.3.3). Sin embargo fuera de la estructura de dependencia elíptica, un problema de interés es determinar si existen distribuciones bivariadas que no provengan de una cópula elíptica tal que la correlación nula de Pearson implique independencia; lo anterior para comprender bajo qué otros modelos la correlación de Pearson puede llegar a ser una buena opción para la detección de relaciones de independencia. Como posible alternativa proponemos la siguiente clase de distribuciones bivariadas absolutamente continuas.

Sean  $U_1, U_2 \sim \mathcal{U}(0, 1)$  con cópula  $\mathcal{C}$ , t.q.  $\mathcal{C}$  modela dependencia monótona (de cuadrante positiva o negativa). Si  $F_1$  y  $F_2$  son las FDA's de  $U_1$  y  $U_2$  respectivamente, tenemos que  $F_1(U_1) = U_1$  y  $F_2(U_2) = U_2$ , entonces

$$\begin{aligned} \rho_S(U_1, U_2) &= \text{Corr}(F_1(U_1), F_2(U_2)) \\ &= \text{Corr}(U_1, U_2) \\ &= \rho(U_1, U_2), \end{aligned}$$

i.e. bajo este contexto marginal y de estructura de dependencia el coeficiente de correlación lineal de Pearson,  $\rho$ , y el coeficiente de concordancia de Spearman,  $\rho_S$ , son iguales. Por lo tanto, bajo dependencia monótona y marginales uniformes en el intervalo  $(0, 1)$  la correlación lineal de Pearson nula si implica independencia.

#### Cópulas condicionales de cópula gaussiana

Partiendo del hecho de que nos interesa estudiar las relaciones de independencia condicional contenidas en un modelo de cópula gaussiana, comenzaremos por comprender el comportamiento de las cópulas condicionales del modelo, para luego analizar cómo es que se pueden deducir las relaciones de independencia condicional de interés.

**Proposición 3.3.2** Sea el v.a.  $\mathbf{X} = (X_1, \dots, X_p)^T$  con cópula gaussiana,  $C_{1-p}$ , como en (3.3.2). Por otro lado, sean  $\mathcal{V} = \{1, \dots, p\}$ ,  $A \subset \mathcal{V}$  tal que  $\text{card}(A) = m \geq 2$ ,  $B \subset \mathcal{V}$  donde  $A \cap B = \emptyset$  y  $\mathbf{v}_B \in \text{Ran } \mathbf{X}_B$ , entonces

$$C_{A|B}(u_1, \dots, u_m; \mathbf{v}_B) = C_{A|B}(u_1, \dots, u_m) = \Phi_{\Sigma^{A|B}}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_m)),$$

donde  $(\Sigma^{A|B})_{ij} = \rho_{ij|B}$ , con  $i, j \in A$ , i.e.  $\Sigma^{A|B}$  es la matriz de correlaciones parciales de las variables en  $A$  con respecto a  $B$ .

La Proposición 3.3.2 reitera sobre el supuesto simplificador que cumple la cópula gaussiana, y además asegura que todas las cópulas condicionales de cópula gaussiana son cópulas gaussianas con parámetro de dependencia la matriz de correlaciones parciales correspondiente. En el Apéndice B proponemos una prueba a esta Proposición; de igual forma, la Proposición con la notación de un grafo no dirigido per se es una aportación de la presente investigación.

### El modelo gráfico de cópula gaussiana

Al igual que en un modelo de distribución gaussiana las condicionales son distribuciones gaussianas nuevamente, en un modelo de cópula gaussiana las cópulas condicionales también son gaussianas. Además en ambos modelos la dependencia condicional es caracterizada por correlaciones parciales y como vimos en (1.7.2), la correlación parcial entre dos variables aleatorias con respecto a las demás en el modelo es proporcional a los elementos de la matriz de precisión,  $\Omega = \Sigma^{-1}$ , con  $\Sigma$  una matriz de correlaciones.

Sea el v.a.  $\mathbf{X} = (X_1, \dots, X_p)^T$  con cópula  $\mathcal{C}_{1:p}(\cdot, \dots, \cdot; \Sigma)$  como en (3.3.2) con matriz de correlaciones  $\Sigma$ , a partir de la Proposición 3.3.2 y el Lema 3.3.3 tenemos las siguientes equivalencias hacia un modelo gráfico de cópula gaussiana,

$$(j, k), (k, j) \in \mathcal{A} \Leftrightarrow \Sigma_{j,k}^{-1} \neq 0 \Leftrightarrow \rho_{jk|\mathcal{V} \setminus \{j,k\}} \neq 0. \quad (3.3.3)$$

En otras palabras,  $X_k \perp\!\!\!\perp X_j | X_{\mathcal{V} \setminus \{k,j\}}$  si y sólo si la correlación parcial correspondiente entre las variables es nula, lo anterior surge por el hecho de que las cópulas condicionales de cópula gaussianan son nuevamente gaussianas.

Notemos que las equivalencias en (3.3.3) son análogas al caso de distribución gaussiana (1.7.1), entonces la inferencia sobre la estructura del grafo en el caso de cópula gaussiana también está basada en la estimación de una matriz de precisión,  $\Omega$ , o al menos en la detección de elementos nulos en esta matriz, al igual que en el caso de distribución gaussiana. La diferencia es que la matriz de correlaciones asociada a la de precisión es la correspondiente a transformaciones particulares de las variables originales, como lo comentamos previamente.

Retomando la definición de cópula gaussiana como una distribución NPN, a partir del hecho de que toda vez que tengamos un v.a.  $\mathbf{X} = (X_1, \dots, X_p)^T$  con cópula  $\mathcal{C}_{1:p}$  como en (3.3.2), existirá un v.a. *latente*  $f(\mathbf{X})$  con distribución gaussiana y la misma estructura de dependencia que  $\mathbf{X}$ , se asegura que si  $\mathcal{G}$  y  $\mathcal{G}'$  son los grafos no dirigidos asociados a  $\mathbf{X}$  y  $f(\mathbf{X})$ , respectivamente, se seguirá que  $\mathcal{G} = \mathcal{G}'$ . Lo anterior es una vía alternativa para comprender la igualdad en las equivalencias (1.7.1) y (3.3.3).

A manera de conclusión con respecto a este modelo, podemos decir que un modelo gráfico de cópula gaussiana deducido a partir de correlaciones parciales nulas entre las variables cumple perfectamente con la propiedad de Markov a pares y además no necesariamente el grafo asociado es un modelo trivial (vacío o completo). Entonces la GIC de cópula gaussiana es de total *utilidad* y pertinencia para explicar las relaciones de independencia condicional del modelo.

## Cópula T

La construcción de la cópula  $\mathcal{T}$  es análoga a la de la cópula gaussiana con respecto a una distribución multivariada  $\mathcal{T}$  de Student (c.f. [Kotz and Nadarajah \(2004\)](#) como una referencia útil sobre la distribución  $\mathcal{T}$  multivariada).

Denotemos con  $h_{1:p}(\cdot, \dots, \cdot; \nu, \Sigma)$  a la densidad conjunta de un modelo  $\mathcal{T}_p(\nu, \mathbf{0}, \Sigma)$ , con  $\nu > 2$  y  $\Sigma \in \mathbb{R}^{p \times p}$  una matriz de correlaciones positiva definida con  $\text{diag}(\Sigma) = \mathbf{1}$ . Entonces  $\forall \mathbf{x} = (x_1, \dots, x_p) \in \mathbb{R}^p$ ,

$$h_{1:p}(x_1, \dots, x_p; \nu, \Sigma) = h_{1:p}(\mathbf{x}; \nu, \Sigma) = \frac{\Gamma\left(\frac{\nu+p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \sqrt{(\pi\nu)^p |\Sigma|}} \left(1 + \frac{\mathbf{x}^T \Sigma^{-1} \mathbf{x}}{\nu}\right)^{-\frac{\nu+p}{2}}. \quad (3.3.4)$$

La cópula  $\mathcal{T}$  de dimensión  $p$  está dada por,

$$\begin{aligned} \mathcal{C}_{1:p}^{\mathcal{T}}(u_1, \dots, u_p; \nu, \Sigma) &= \int_{-\infty}^{t_{\nu}^{-1}(u_1)} \dots \int_{-\infty}^{t_{\nu}^{-1}(u_p)} h_{1:p}(z_1, \dots, z_p; \nu, \Sigma) dz_1 \dots dz_p \\ &= t_{\nu, \Sigma}(t_{\nu}^{-1}(u_1), \dots, t_{\nu}^{-1}(u_p)), \end{aligned} \quad (3.3.5)$$

donde  $t_\nu$  es la FDA univariada estándar  $\mathcal{T}$  con  $\nu$  grados de libertad, esperanza 0 y varianza  $\nu/(\nu - 2)$ , por otro lado,  $t_{\nu, \Sigma}$  es la FDA conjunta  $\mathcal{T}$  de dimensión  $p$ , con  $\nu$  grados de libertad y matriz de correlaciones  $\Sigma$  (c.f. Demarta and McNeil (2005) para la construcción formal de la cópula  $\mathcal{T}$  multivariada y resultados generales sobre la misma).

Notemos que si  $\mathbf{X} \sim \mathcal{T}_p(\nu, \mathbf{0}, \Sigma)$ , entonces a partir del Corolario (2.4.1) se sigue que la cópula de  $\mathbf{X}$ ,  $\mathcal{C}_{\mathbf{X}}$ , es aquella en (3.3.5). Por otro lado, aplica el comentario análogo al de la Observación 3.3.1 para cópulas  $\mathcal{T}$ , en el sentido de que si  $\mathbf{Y} \sim \mathcal{T}_p(\nu, \mu, P)$  y  $\mathbf{X} \sim \mathcal{T}_p(\nu, \mathbf{0}, \Sigma)$ , entonces  $\mathcal{C}_{\mathbf{Y}} = \mathcal{C}_{\mathbf{X}}$  cuando  $\Sigma$  es la matriz de correlaciones que subyace de la matriz de dispersión  $P$ .

### Estructura de dependencia: cópula $\mathcal{T}$

Antes de hablar de la estructura de dependencia que modela una cópula  $\mathcal{T}$  en general, comenzaremos por entender la estructura que modela el miembro más esencial de esta familia: el modelo de distribución  $\mathcal{T}$ .

Sea  $\mathbf{X} \sim \mathcal{T}_p(\nu, \mathbf{0}, \Sigma)$  con  $\nu > 2$  y  $\Sigma$  una matriz de correlaciones positiva definida, entonces

$$\mathbf{X} \stackrel{d}{=} \sqrt{W} \mathbf{Z}, \quad (3.3.6)$$

donde  $\mathbf{Z} \sim \mathcal{N}_p(\mathbf{0}, \Sigma)$  y  $W$  satisface que  $\nu/W \sim \chi_\nu^2$ , además  $W$  y  $\mathbf{Z}$  son independientes; equivalentemente  $W$  tiene una distribución gama inversa,  $W \sim \mathcal{GI}(\nu/2, \nu/2)$ .

Consideremos un v.a.

$$\mathbf{X} = (X_1, \dots, X_p)^T = \sqrt{W} \mathbf{Z} = \sqrt{W} (Z_1, \dots, Z_p)^T, \quad (3.3.7)$$

donde  $W \sim \mathcal{GI}(\nu/2, \nu/2)$  y  $\mathbf{Z} = (Z_1, \dots, Z_p)^T \sim \mathcal{N}_p(\mathbf{0}, \Sigma)$  con  $\Sigma$  una matriz de correlaciones, además  $W$  y  $\mathbf{Z}$  independientes. Entonces  $\forall j, k \in \{1, \dots, p\}$ ,

$$\begin{aligned} \text{Cov}(X_j, X_k) &= \mathbb{E}[X_j X_k] \\ &= \mathbb{E}[W Z_j Z_k] \\ &= \mathbb{E}[W] \mathbb{E}[Z_j Z_k] \\ &= \frac{\nu}{\nu - 2} \Sigma_{jk}, \end{aligned}$$

se sigue que

$$\text{Corr}(X_j, X_k) = \Sigma_{jk}. \quad (3.3.8)$$

Notemos de (3.3.7) que  $\mathbf{X} \sim \mathcal{T}_p(\nu, \mathbf{0}, \Sigma)$ . Por otro lado, la igualdad en (3.3.8) sirve para comprender que toda vez que tengamos un modelo multivariado de distribución  $\mathcal{T}$  tendremos un v.a. *latente* gaussiano con la misma estructura de correlación, cuyos componentes son afectados por un factor común el cual es proporcional a una distribución  $\chi^2$ . Forzosamente la estructura de dependencia entre las variables en el modelo se verá afectada por este factor, que de algún modo relaciona a cada una de éstas.

Intuitivamente, a partir de que la dependencia entre las variables está influenciada por el factor entre las mismas, se justifica el hecho de que no correlación no implique independencia. En otras palabras, pese a que las variables *latentes* gaussianas estén no correlacionadas y por ende sean independientes, las variables en el modelo de distribución  $\mathcal{T}$  siempre tendrán un grado no nulo de dependencia caracterizado por  $\sqrt{W}$  vía  $\nu$ . De hecho resulta de la Proposición 3.3.1 que la independencia *no es un miembro* de la familia de distribución  $\mathcal{T}$ .

Por otro lado, de (3.3.4) tenemos que las distribuciones marginales de una distribución  $\mathcal{T}$  multivariada son  $\mathcal{T}$  también. Más aún, el parámetro de grados de libertad  $\nu$  es el mismo para todas las marginales independientemente de su dimensión. Sin embargo, en breve observaremos que la estructura y grado de dependencia no sólo dependen de  $\nu$ , sino también de la correlación lineal de cada par de variables.

Regresando a los modelos de cópula  $\mathcal{T}$ , de manera análoga a los modelos de cópula gaussiana podemos comprender a los de cópula  $\mathcal{T}$  como aquellos que representan la *misma* estructura de dependencia que una distribución  $\mathcal{T}$ , pero permitiendo que las distribuciones marginales no necesariamente pertenezcan a la misma familia. De hecho en Liu et al. (2012b), definición 3.1, extienden a la familia NPN hacia la familia de distribución *transelíptica*, la cual es equivalente a un modelo de cópula elíptica. Partiendo de un v.a.  $\mathbf{X} = (X_1, \dots, X_p)^T$  con un modelo subyacente de cópula  $\mathcal{T}$  como en (3.3.5), visto como un modelo de distribución transelíptica, se asegura la existencia de un v.a. *latente*,  $f(\mathbf{X})$ , con distribución  $\mathcal{T}$  y la misma estructura de dependencia que  $\mathbf{X}$  (c.f. Liu et al. (2012b) para detalles de la distribución transelíptica).

En contraste con los modelos de cópula gaussiana, y factor por el cual estos modelos han cobrado relevancia en la última década sobre todo para aplicaciones en finanzas, es por el hecho de que son capaces de modelar dependencia en ambas colas. Análogo al caso de cópula gaussiana, el hecho de que la distribución  $\mathcal{T}$  es elíptica y simétrica implica que sus parámetros de dependencia en colas sean iguales, i.e.  $\lambda_L^{\mathcal{T}} = \lambda_U^{\mathcal{T}}$ . El Lema 3.3.4 presenta los parámetros de dependencia en colas para cópula  $\mathcal{T}$ .

**Lema 3.3.4** Para variables aleatorias continuas con cópula  $\mathcal{C}_{1,2}(\cdot, \cdot; \nu, \Sigma)$  como en (3.3.5), los coeficientes de dependencia en colas están dados por,

$$\lambda_L^{\mathcal{T}} = \lambda_U^{\mathcal{T}} = 2t_{\nu+1} \left( -\sqrt{\nu+1} \sqrt{1-\rho} / \sqrt{1+\rho} \right), \quad (3.3.9)$$

donde  $\rho$  es el elemento fuera de la diagonal de  $\Sigma$ .

Para una prueba del Lema 3.3.4 véase Demarta and McNeil (2005), proposición 1.

Notemos de (3.3.9) que  $\lambda_L^{\mathcal{T}}, \lambda_U^{\mathcal{T}} > 0 \forall \nu \in (2, \infty)$ , entonces técnicamente el modelo es capaz de representar dependencia en ambas colas para cualquier valor admisible de  $\nu$ , sin embargo,

$$\lim_{\nu \rightarrow \infty} \lambda_L^{\mathcal{T}} = \lim_{\nu \rightarrow \infty} \lambda_U^{\mathcal{T}} = 0. \quad (3.3.10)$$

El Cuadro 3.1 muestra el comportamiento de los parámetros de dependencia en colas de cópula  $\mathcal{T}$  al variar  $\nu$  y  $\rho$ . Pese a que el cambio en los parámetros depende del valor de  $\rho$ , podemos observar que a medida que  $\nu$  aumenta éstos tienden a cero; la justificación de lo anterior se dará más adelante con el Corolario 3.3.1.

$\nu/\rho$	-0.9	-0.7	-0.5	-0.3	0	0.3	0.5	0.7	0.9
2	0.004	0.025	0.057	0.099	0.181	0.293	0.391	0.519	0.717
4	$\approx 0$	0.003	0.011	0.028	0.075	0.161	0.253	0.390	0.629
6	$\approx 0$	$\approx 0$	0.002	0.008	0.033	0.093	0.170	0.303	0.563
8	$\approx 0$	$\approx 0$	$\approx 0$	0.002	0.014	0.055	0.117	0.239	0.508
10	$\approx 0$	$\approx 0$	$\approx 0$	$\approx 0$	0.006	0.033	0.081	0.191	0.462
$\infty$	0	0	0	0	0	0	0	0	0

Cuadro 3.1: parámetros de dependencia en colas de cópula  $\mathcal{T}$ .

Para finalizar, de (3.3.5) se sigue que las cópulas marginales de cópula  $\mathcal{T}$  son nuevamente cópulas  $\mathcal{T}$ . Como ocurre para una distribución  $\mathcal{T}$ , todas las cópulas marginales de cópula  $\mathcal{T}$  comparten el mismo parámetro de grados de libertad,  $\nu$ ; la observación anterior junto con la simetría en la dependencia en colas son las dos grandes restricciones que tienen los modelos de cópula  $\mathcal{T}$  (c.f. Aas and Berg (2009) y Fischer et al. (2009) para discusiones al respecto y atractivas vías de extensión al modelo que cópula  $\mathcal{T}$ ). A continuación estudiaremos el comportamiento de las cópulas condicionales de cópula  $\mathcal{T}$ .

### Cópulas condicionales de cópula $\mathcal{T}$

A partir de la siguiente proposición comprenderemos el comportamiento de las cópulas condicionales de cópula  $\mathcal{T}$ .



**Proposición 3.3.3** Sea el v.a.  $\mathbf{X} = (X_1, \dots, X_p)^T$  con cópula  $\mathcal{T}$ ,  $\mathcal{C}_{1:p}$ , como en (3.3.5). Por otro lado, sean  $\mathcal{V} = \{1, \dots, p\}$ ,  $A \subset \mathcal{V}$  tal que  $\text{card}(A) = m \geq 2$ ,  $B \subset \mathcal{V}$  donde  $A \cap B = \emptyset$  y  $\mathbf{v}_B \in \text{Ran } \mathbf{X}_B$ , entonces

$$C_{A|B}(u_1, \dots, u_m; \mathbf{v}_B) = C_{A|B}(u_1, \dots, u_m) = t_{\nu^*, \Sigma^{A|B}}(t_{\nu^*}^{-1}(u_1), \dots, t_{\nu^*}^{-1}(u_m)),$$

donde  $(\Sigma^{A|B})_{ij} = \rho_{ij|B}$ , con  $i, j \in A$ , i.e.  $\Sigma^{A|B}$  es la matriz de correlaciones parciales de las variables en  $A$  con respecto a aquellas en  $B$  y  $\nu^* = \nu + \text{card}(B)$ , con  $\nu$  los grados de libertad de la cópula no condicional.

La Proposición 3.3.3 reitera sobre el supuesto simplificador que cumple la cópula  $\mathcal{T}$  y asegura que todas las cópulas condicionales de cópula  $\mathcal{T}$  son cópulas  $\mathcal{T}$ , con parámetro de dependencia la matriz de correlaciones parciales correspondiente, además los grados de libertad aumentan con respecto al número de variables en el condicionante. Partiendo del hecho de que las condicionales de distribución  $\mathcal{T}$  son distribuciones  $\mathcal{T}$  con una matriz de dispersión asociada y grados de libertad con respecto al número de variables en el condicionante, la prueba de la Proposición 3.3.3 es análoga a la de la Proposición 3.3.2 referente a las condicionales de cópula gaussiana (c.f. [Kotz and Nadarajah \(2004\)](#), sección 1.11, y [Roth \(2012\)](#), sección 5, para detalles sobre condicionales de distribución  $\mathcal{T}$  multivariada).

### El modelo gráfico de cópula $\mathcal{T}$

A continuación se presenta uno de los resultados más relevantes de la presente sección. La prueba a la Proposición 3.3.4 es inmediata a partir de las Proposiciones 3.3.3 y 3.3.1.

**Proposición 3.3.4** Sea  $\mathbf{X} = (X_1, \dots, X_p)^T$  con cópula  $\mathcal{C}_{1:p}$  como en (3.3.5), entonces la GIC asociada a  $\mathbf{X}$  es el *grafo completo* como en la Definición 1.4.3, i.e.  $\mathcal{A} = (\mathcal{V} \times \mathcal{V}) \setminus \{(j, j) \mid j \in \mathcal{V}\}$ .

Recordemos que si un modelo gráfico no dirigido tiene asociada una GIC completa implica que el conjunto de aristas,  $\mathcal{A}$ , contiene todas las posibles  $\binom{p}{2}$  combinaciones de enlaces entre las variables de  $\mathbf{X}$ . En otras palabras, que en la estructura de dependencia asociada a la distribución de probabilidad del modelo,  $\mathbf{P}$ , no existe una sola relación de independencia condicional.

A partir de la Proposición 3.3.4 diríamos que un modelo gráfico de cópula  $\mathcal{T}$  no es en realidad *útil*, debido a que *en teoría*, sin importar la combinación de parámetros de la cópula, el modelo siempre será el mismo: el grafo completo. Pese a que la Proposición 3.3.1 asegura que la única familia de cópulas elípticas que contiene a la cópula  $\Pi$  es la gaussiana, la pregunta de fondo es: ¿qué tan lejos/cerca se encuentra una cópula  $\mathcal{T}$  bivariada, parametrizada de algún modo, de la cópula  $\Pi$ ? Más aún, si contamos con una muestra de cópula  $\mathcal{T}$  multivariada y de algún modo obtenemos  $\widehat{\mathcal{G}}$ : ¿será posible que  $\widehat{\mathcal{G}}$  no sea completo? En caso afirmativo: ¿ $\widehat{\mathcal{G}}$  podría llegar a ser una buena aproximación para explicar las relaciones de independencia condicional en la muestra?

Recordemos que en el fondo un modelo de cópula  $\mathcal{T}$ , en términos de estructura de dependencia, es análogo a una distribución  $\mathcal{T}$ , el cual a su vez tiene la estructura de un modelo de distribución gaussiana donde sus elementos son afectados por un factor común, entonces surge el siguiente planteamiento:

Si contamos con un v.a.  $\mathbf{X} = (X_1, X_2)^T \sim \mathcal{T}_2(\nu, \mathbf{0}, \mathcal{I}_2)$ , i.e.  $X_1$  y  $X_2$  variables aleatorias no correlacionadas de distribución conjunta  $\mathcal{T}$ . Sabemos que existe un v.a. latente  $\mathbf{Z} = (Z_1, Z_2)^T \sim \mathcal{N}_2(\mathbf{0}, \mathcal{I}_2)$ , i.e.  $Z_1$  y  $Z_2$  variables aleatorias no correlacionadas de distribución conjunta gaussiana, entonces  $Z_1$  y  $Z_2$  son independientes. Partiendo de (3.3.6), surge la siguiente pregunta: ¿en qué medida modifica la dependencia entre  $Z_1$  y  $Z_2$  el factor común,  $\sqrt{W}$ ? En otras palabras, ¿qué tan lejos/cerca se encuentra de la independencia un modelo de distribución  $\mathcal{T}$  bivariada con elementos no correlacionados? Partiendo del hecho de que  $\mathcal{C}_{\mathbf{X}}$  es cópula  $\mathcal{T}$  como en (3.3.5) con  $p = 2$  y  $\Sigma = \mathcal{I}_2$ , buscaremos dar respuesta a los cuestionamientos previos estudiando el comportamiento que tiene la distancia de la cópula  $\mathcal{T}$  bivariada con  $\rho = 0$  a la cópula  $\Pi$ , con respecto al cambio en  $\nu$ .

Recordemos la medida de dependencia  $\sigma_{SW}$  de Schweizer and Wolff como en (2.6.3) que cuantifica la distancia de una cópula particular a  $\Pi$ , la Figura 3.1 muestra la distancia  $\sigma_{SW}(\mathcal{C}_{\nu}^{\mathcal{T}})$  al cambio en  $\nu$ , con  $\mathcal{C}_{\nu}^{\mathcal{T}}$  como en (3.3.5) con  $p = 2$ ,  $\Sigma = \mathcal{I}_2$  y  $\nu$  grados de libertad.

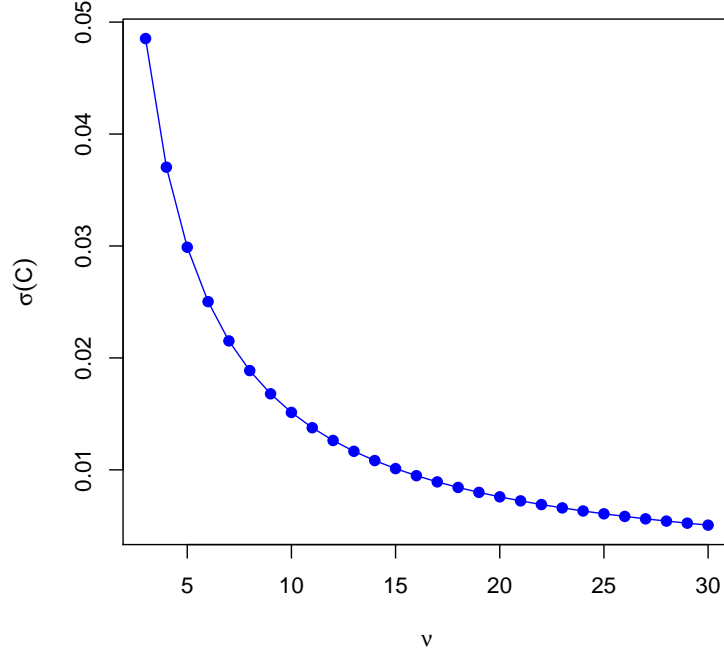


Figura 3.1: Distancia entre cópula  $\mathcal{T}$  con  $\rho = 0$  y  $\Pi$ .

De la Figura 3.1 observamos que a medida que  $\nu$  aumenta,  $\mathcal{C}_\nu^{\mathcal{T}}$  se encuentra cada vez más cerca de  $\Pi$ . Además, aún para valores cercanos a 2,  $\mathcal{C}_\nu^{\mathcal{T}}$  no se encuentra demasiado lejos de  $\Pi$ . En otras palabras, a medida que  $\nu$  aumenta, el impacto en el grado de dependencia que provoca  $\sqrt{W}$  es cada vez menor y aún con valores cercanos a 2 de  $\nu$  el grado de dependencia no se ve sustancialmente afectado.

Para comprender más a profundidad y reforzar lo observado en la Figura 3.1 enunciaremos la siguiente Proposición 3.3.5, que asegura que cuando  $\nu \rightarrow \infty$  un modelo de distribución  $\mathcal{T}$  converge a un gaussiano.

**Proposición 3.3.5** Sean  $g_{1:p}(\cdot, \dots, \cdot; \Sigma)$  como en (3.3.1), la densidad de una distribución gaussiana multivariada, y  $h_{1:p}(\cdot, \dots, \cdot; \nu, \Sigma)$  como en (3.3.4), la densidad conjunta de una distribución  $\mathcal{T}$ , con  $\Sigma$  una matriz de correlaciones positiva definida y  $\nu$  grados de libertad. Se sigue que  $\forall (x_1, \dots, x_p) \in \mathbb{R}^p$ ,

$$\lim_{\nu \rightarrow \infty} h_{1:p}(x_1, \dots, x_p; \nu, \Sigma) = g_{1:p}(x_1, \dots, x_p; \Sigma).$$

El hecho de que toda vez que tengamos modelos de cópula gaussiana y  $\mathcal{T}$ , tendremos v.a. latentes de distribución gaussiana y  $\mathcal{T}$ , respectivamente, con los mismos parámetros, asegura que el resultado en la Proposición 3.3.5 se extiende como se enuncia en el siguiente Corolario 3.3.1.

**Corolario 3.3.1** Sean las cópulas  $\mathcal{C}_{1:p}^G(u_1, \dots, u_p; \Sigma)$  como en (3.3.2) y  $\mathcal{C}_{1:p}^{\mathcal{T}}(u_1, \dots, u_p; \nu, \Sigma)$  como en (3.3.5), con  $\Sigma$  una matriz de correlaciones definida positiva y  $\nu$  grados de libertad. Se sigue que  $\forall (u_1, \dots, u_p) \in \mathbb{I}^p$ ,

$$\lim_{\nu \rightarrow \infty} \mathcal{C}_{1:p}^{\mathcal{T}}(u_1, \dots, u_p; \nu, \Sigma) = \mathcal{C}_{1:p}^G(u_1, \dots, u_p; \Sigma).$$

Si bien es un hecho que la correlación nula no implica independencia para modelos de cópula  $\mathcal{T}$ , en realidad el grado de dependencia en este caso es tan bajo, como se observa en la Figura 3.1, que si contamos con una muestra de cópula  $\mathcal{T}$  multivariada con elementos no correlacionados con un valor de  $\nu$ , grados de libertad, relativamente

alto ( $\approx 10$ ) y de algún modo hacemos pruebas de independencia, muy probablemente se obtenga evidencia para no rechazar la hipótesis.

Más aún, en la Proposición 3.3.3 se observa que los grados de libertad de cópulas condicionales de cópula  $\mathcal{T}$  aumentan con el número de variables en el condicionante, entonces a partir del Corolario 3.3.1 podemos deducir que a medida que el número de variables en el condicionante aumente las cópulas condicionales de cópula  $\mathcal{T}$  estarán cada vez más cerca de una cópula gaussiana. En otras palabras, pese a que las condicionales de cópula  $\mathcal{T}$  son  $\mathcal{T}$  nuevamente y la Proposición 3.3.1 asegure que la familia de cópulas  $\mathcal{T}$  no incluye a  $\Pi$ , la realidad es que si de alguna manera logramos detectar correlaciones parciales nulas y el número de variables en el condicionante crece estaremos muy cerca de tener independencia condicional. Cabe destacar que, al igual que en el modelo de cópula gaussiana, cuando nos referimos a la correlación parcial no es de las variables originales sino de las variables transformadas  $t_\nu^{-1}(F_1(X_1)), \dots, t_\nu^{-1}(F_p(X_p))$ ; análogamente a la distribución NPN, es a partir del modelo transeléptico que se puede comprender lo antes mencionado para el caso de cópula  $\mathcal{T}$  (c.f. Liu et al. (2012b) para detalles del modelo transeléptico).

Dicho lo anterior, es que podemos concluir que si contamos una muestra de cópula  $\mathcal{T}$  multivariada con presencia de correlaciones parciales nulas en las variables transformadas y si de algún modo realizamos pruebas de independencia condicional, entonces muy probablemente se obtenga evidencia para no rechazar las hipótesis. En tal caso, un grafo estimado a partir de correlaciones parciales nulas es un modelo *no completo* que aproxima bastante bien las relaciones de independencia condicional en un modelo de cópula  $\mathcal{T}$ . Aunque en el fondo no podríamos decir que sería una GIC propiamente porque técnicamente no se estaría cumpliendo la propiedad de Markov a pares pero, con base en lo discutido en esta sección, si podemos decir que sería una aproximación *útil* de la estructura de dependencia condicional. Entonces la propuesta de estimación del grafo para la familia de distribución transeléptica en Liu et al. (2012b), la cual se basa en la detección de patrones de correlación parcial nula, tiene sentido al menos para los casos de cópula gaussiana y  $\mathcal{T}$  en donde de antemano se cumple el supuesto simplificador.

### 3.4. Cópulas Arquimedianas simplificadas

La extensión más común y simple de cópulas Arquimedianas para dimensiones mayores a dos es la conocida como *cópulas Arquimedianas intercambiables*, la cual es extremadamente restrictiva ya que un solo parámetro domina la estructura de dependencia sin importar la dimensión de la cópula, además de que no es posible extender todas las cópulas Arquimedianas bivariadas a su correspondiente versión intercambiable (c.f. Frey and McNeil (2003) para detalles). Para otras extensiones de cópulas Arquimedianas a dimensiones mayores a dos se recomienda Embrechts et al. (2001), Whelan et al. (2004), McNeil (2008), Savu and Trede (2010).

Como lo mencionamos en la Sección 3.2 la cópula Clayton multivariada, la única cópula simplificada Arquimediana, tiene asociado un *grafo no dirigido trivial* para cualquier valor de  $\theta$ , su parámetro de dependencia, por ende no es un modelo gráfico *útil*. Por lo anterior es que no profundizaremos demasiado en los modelo de cópula Arquimediana y el modelo Clayton lo analizaremos superficialmente, pero consideramos que las referencias bibliográficas que sustentan este capítulo alrededor del tema dan un panorama bastante amplio sobre estos modelos.

#### Cópula Clayton

La cópula Clayton<sup>3</sup> de dimensión  $p$  está dada por,

$$C_{1:p}^C(u_1, \dots, u_p) = \{\max[u_1^{-\theta} + \dots + u_p^{-\theta} - 1, 0]\}^{1/\theta}, \quad (3.4.1)$$

para  $\theta \in \left[-\frac{1}{p-1}, \infty\right) \setminus \{0\}$ .

Notemos de (3.4.1) que la estructura de dependencia de una cópula Clayton multivariada está caracterizada por un solo parámetro,  $\theta$ , sin importar la dimensión de la cópula,  $p$ . Lo anterior hace que este modelo sea sumamente

<sup>3</sup>La cópula Clayton es la subyacente de las distribuciones Pareto multivariada (c.f. Mardia (1962)) y Burr multivariada (c.f. Takahasi (1965)).

restrictivo, en particular veremos más adelante que las cópulas a pares condicionales comparten el mismo parámetro, lo que hace que el grafo asociado sea el trivial sin importar el valor de  $\theta$

### Estructura de dependencia: cópula Clayton

Con base en [Nelsen \(2007\)](#), capítulo 4, la cópula Clayton bivariada al cambio en sus parámetros atraviesa por distintos grados de dependencia como se muestra en (3.4.2).

$$C_{1:2}^C(\cdot, \cdot; \theta = -1) = \mathcal{W}, \quad C_{1:2}^C(\cdot, \cdot; \theta = 0) = \Pi, \quad C_{1:2}^C(\cdot, \cdot; \theta = \infty) = \mathcal{M}. \quad (3.4.2)$$

Como lo vimos en (3.4.1) el valor  $\theta = 0$  no es admisible como parámetro de la cópula Clayton, lo que implica que la cópula independencia no sea un miembro explícito de esta familia. Sin embargo, el valor de  $\theta$  se puede aproximar a cero a manera que  $\Pi$  prácticamente se considere dentro de la familia.

Con respecto a la dependencia en colas, la cópula Clayton se caracteriza por modelar dependencia en la cola inferior (c.f. [Nelsen \(2007\)](#) ejemplo 5.22).

Finalmente, las marginales de cópula Clayton son nuevamente cópulas Clayton (c.f. [Takahasi \(1965\)](#)).

### Cópulas condicionales de cópula Clayton

Como se asegura en [Mesfioui and Quesy \(2008\)](#) y se reafirma en [Stöber et al. \(2013\)](#), las cópulas condicionales de una cópula Clayton multivariada,  $C_{1:p}^C$  como en (3.4.1), son nuevamente cópulas Clayton donde el parámetro de dependencia de éstas,  $\theta(m)$ , sólo depende de  $\theta$  y del número de variables en el condicionante,  $m$ , como en (3.4.3).

$$\theta(m) = \frac{\theta}{m\theta + 1}. \quad (3.4.3)$$

Notemos que para  $\theta \geq -1/(d-1)$ , se sigue que  $\theta/(m\theta + 1) \geq -1/(d-m-1)$ , entonces el espacio parametral es consistente.

En [Stöber et al. \(2013\)](#) sección 3, se prueba que para  $p \geq 4$  la cópula Clayton es la *única* en la clase de cópulas Arquimedianas que cumple con el supuesto simplificador.

### El modelo gráfico de cópula Clayton

Con base en (3.4.2) y el hecho de que las cópulas a pares condicionales de cópula Clayton son nuevamente Clayton, técnicamente la GIC asociada a esta cópula es un modelo *completo* para cualquier valor admisible de  $\theta$ .

Por otro lado, partiendo de una cópula Clayton,  $C_{1:p}^C$  como en (3.4.1), con  $p \geq 3$ , se sigue de (3.4.3) que el parámetro de *todas las cópulas condicionales a pares* es,

$$\theta(p-2) = \frac{\theta}{(p-2)\theta + 1}. \quad (3.4.4)$$

Entonces partiendo de una muestra de cópula Clayton, el grafo no dirigido estimado,  $\widehat{\mathcal{G}}^C$ , será *vacío* a medida que  $\theta(p-2)$  se aproxime a cero, y por otro lado, será *completo* a medida que  $\theta(p-2)$  se aleje de cero. Pese a que se pueda lograr una GIC que cumpla perfectamente la propiedad de Markov a pares, ésta no tiene *utilidad* alguna por ser un modelo *trivial* para cualquier valor del parámetro de dependencia de la cópula no condicional.

## 3.5. Limitantes de los modelos de cópula gaussiana

Como se concluyó en la Sección 3.3, el modelo paramétrico *ideal* para un grafo no dirigido es el de cópula gaussiana. Pese a que es un modelo más flexible que el de distribución gaussiana, en el sentido que permite que las marginales no pertenezcan necesariamente a la misma familia, éste sigue siendo restrictivo en términos de la estructura de

dependencia que modela. Se decidió incorporar esta pequeña sección debido a que en nuestra investigación nos encontramos diversos artículos que prometían proponer diversas metodologías partiendo de modelos más flexibles a los de distribución gaussiana, sin embargo, acotaban su campo de acción a modelos de cópula gaussiana; principalmente se encontraron artículos en el contexto de altas dimensiones en donde se vuelve aún menos realista el supuesto, v.g. Liu et al. (2012a), Han et al. (2013) y Mai and Zou (2015), por sólo mencionar algunos. Así en esta sección comentaremos brevemente sobre las limitantes del modelo, enfatizando en su incapacidad para modelar dependencia en colas.

En aplicaciones reales un modelo de cópula gaussiana puede resultar bastante restrictivo. A continuación enunciamos algunas desventajas evidentes que estos modelos tienen (c.f. Embrechts et al. (1999) para una discusión sobre modelos basados en el coeficiente de correlación lineal).

### Desventajas

- Sólo modelan dependencia a pares vía  $\rho$ , el coeficiente de correlación lineal de Pearson.
- $\rho$  no es medida de dependencia.
- $\rho$  no es invariante ante transformaciones monótonas crecientes.
- $\rho$  se ve influenciada por efectos marginales de las variables.
- Ante evidencia de dependencia monótona perfecta, i.e.  $\mathcal{C}_{\mathbf{X}} = \mathcal{M}$  ó  $\mathcal{C}_{\mathbf{X}} = \mathcal{W}$ , la correlación lineal no necesariamente toma el valor de 1 ó -1.
- Incapaces de modelar dependencia en colas (tail-dependence).

Las desventajas anteriores son evidencia de la gran necesidad de validar el supuesto de cópula gaussiana, lo cual en la práctica en muy pocas ocasiones se lleva a cabo. En caso de que no se cumpla el supuesto, las conclusiones de modelos subyacentes pueden no ser correctas.

La Figura 3.2 muestra contornos de una densidad bivariada basada en una cópula gaussiana con ciertas marginales. La Figura 3.2 es un ejemplo de como un modelo de cópula gaussiana sigue teniendo limitaciones análogas a un modelo con distribución gaussiana, en cuanto a la estructura de dependencia que representa, en particular su incapacidad para modelar dependencia en colas.

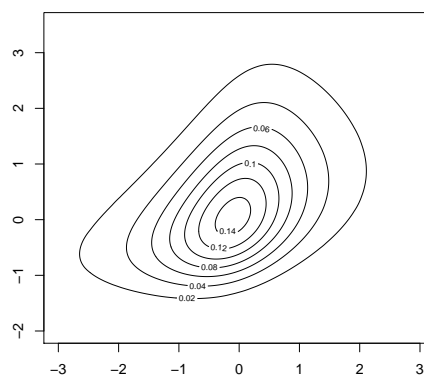


Figura 3.2: Cópula gaussiana.

En contraste, hay modelos de cópulas que son más flexibles. La Figura 3.3 muestra contornos de densidades bivariadas, con las mismas marginales que en la Figura 3.2, pero con cópulas T (izquierda) y Clayton (derecha). Como se comentó en este capítulo, la cópula T es capaz de modelar dependencias en ambas colas, mientras que la cópula Clayton puede modelar dependencias en la cola inferior.

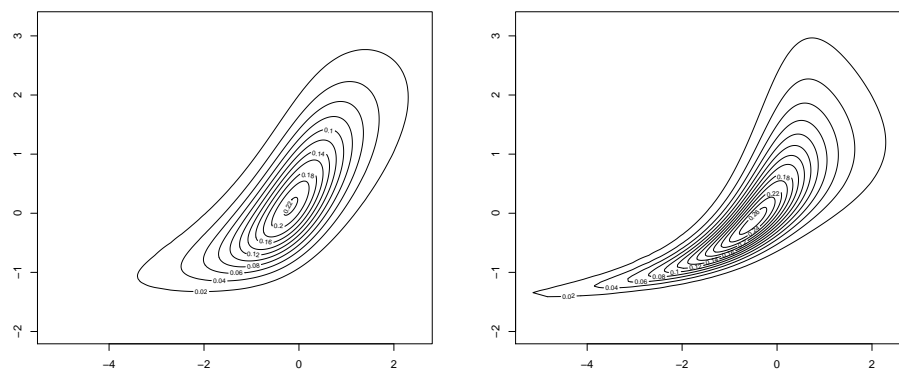


Figura 3.3: Cópula T (izquierda). Cópula Clayton (derecha).

### 3.6. Discusión

A lo largo de este capítulo estudiamos a los tres modelos de cópulas multivariadas simplificadas, con tres propósitos muy puntuales: comprender la flexibilidad de la estructura de dependencia que modelan, estudiar el comportamiento de sus cópulas condicionales, y finalmente, determinar la *utilidad* de su modelo gráfico no dirigido asociado. Las conclusiones generales del capítulo son: los modelos de cópula gaussiana son para los cuales un grafo no dirigido tiene *total pertinencia* para representar las relaciones de independencia condicional, sin embargo, una cópula gaussiana no es ideal como modelo conjunto por su falta de flexibilidad; para los modelos de cópula  $\mathcal{T}$  un modelo gráfico no dirigido explica de forma *aproximada* relaciones de independencia condicional, teniendo una *utilidad* práctica para las aplicaciones de un grafo, además de ser modelos capaces de representar dependencia en colas; finalmente, para el caso de cópula Clayton un modelo gráfico no dirigido *carece de utilidad*. Entonces los modelos gráficos no dirigidos paramétricos están limitados a los miembros gaussiano y  $\mathcal{T}$  de la clase de cópulas elípticas. Cabe destacar que este capítulo se sustenta ampliamente en los resultados obtenidos en [Stöber et al. \(2013\)](#) sobre familias multivariadas de cópulas simplificadas, se recomienda la revisión de este artículo para detalles técnicos alrededor de estos resultados.

Sin embargo, con el creciente auge que han ido ganando los modelos de *vine cópula* en años recientes (c.f. [Aas et al. \(2009\)](#), [Czado \(2010\)](#) y [Min and Czado \(2011\)](#)), se cuestiona rotundamente el hecho de que una sola cópula multivariada perteneciente a una familia paramétrica sea capaz de modelar estructuras de dependencias complejas, que de antemano están presentes en datos multivariados, sobre todo cuando contamos con un número muy grande de variables.

Así que partiendo de nuestro interés en proponer modelos gráficos no dirigidos *más flexibles*, en términos de la estructura de dependencia asociada a la distribución de probabilidad,  $\mathbf{P}$ , en el Capítulo 4 *se propone* una vía de estimación basada en cópulas condicionales junto con un enfoque *no paramétrico* de estimación de la GIC, basados en la extracción de observaciones de cópulas condicionales a pares, tal como se hace en el paradigma de construcción de cópulas multivariadas *vine cópula*.

## Capítulo 4

# Modelos gráficos no dirigidos vía cópulas condicionales: estimador Bernstein

**Resumen** La construcción de cópulas *flexibles* capaces de modelar estructuras complejas de dependencia es un área que, hasta hace poco menos de una década, estaba en su mayoría limitada al caso bivariado, logrando así una gama amplia de metodologías con bastante potencial para las aplicaciones en esta dimensión, v.g. pese a que el pegado de cópulas en Siburg and Stoimenov (2008) se extiende hacia cualquier dimensión, su campo más tangible de aplicación es cuando  $p=2$ . Sin embargo, el paso hacia generalizaciones para el caso multivariado ( $p \geq 3$ ) no ha sido necesariamente sencillo en la mayoría de los casos (c.f. Aas and Berg (2009), Czado (2010) y Min and Czado (2011) para discusiones al respecto). De hecho hace no muchos años atrás, fuera de las cópulas gaussianas y  $\mathcal{T}$ , o en general los modelos de cópulas elípticas, los modelos propuestos en la literatura eran demasiado limitados (c.f. Frahm et al. (2003)) o su investigación aún era reciente (c.f. Joe (1996), Bedford and Cooke (2001), Bedford and Cooke (2002) y Kurowicka and Cooke (2006) como algunos de los pioneros en la teoría de *vine cópula*). El gran paso hacia la construcción de cópulas multivariadas capaces de modelar diversas estructuras de dependencia, se ha dado con las *construcciones de cópulas a pares* (CCP's)<sup>1</sup> en el paradigma de *vine cópula*, donde se explota el potencial existente para la construcción de cópulas bivariadas logrando modelos multivariados sumamente flexibles (c.f. Aas et al. (2009) como el primer artículo donde se propone un marco general para la construcción de cópulas multivariadas vía *vines*). En el presente capítulo hacemos uso de una de las herramientas fundamentales en *vine cópula* para la obtención de observaciones de cópulas condicionales, nos referimos a la fórmula de Joe (1996). La fórmula de Joe (1996) es una vía para extraer observaciones de cualquier cópula condicional, en particular aquellas cópulas condicionales bivariadas asociadas con las relaciones de independencia (condicional) modeladas en un grafo no dirigido. Como una de las partes más importantes del capítulo, partiendo del supuesto simplificador proponemos un algoritmo iterativo para la estimación del grafo no dirigido, el  $\alpha$ -estimador del grafo, el cual depende de un parámetro de ajuste,  $\alpha$ , que controla el *grado de escasez* de la estimación. Nuestra propuesta está basada en *pruebas múltiples de independencia para observaciones de cópulas a pares condicionales* extraídas vía la fórmula de Joe (1996), así que como parte esencial de nuestro planteamiento, a lo largo del capítulo discutimos sobre los *tipos de error* involucrados en la estimación y en particular para un control estadístico en la tasa de descubrimientos falsos haremos uso de algunos de los resultados en Guo and Rao (2008) para pruebas dependientes. Por otro lado, los modelos *no paramétricos de cópula Bernstein* han tenido un creciente auge para las aplicaciones sobre todo en el caso bivariado (c.f. Sancetta and Satchell (2004) y Erdely and Diaz-Viera (2010)) debido a su gran flexibilidad como estimadores suaves de la cópula, potencial por el cual siguen vigentes en la investigación (c.f. Janssen et al. (2015)). Aprovechando la flexibilidad del estimador Bernstein de la cópula en el caso bivariado y el algoritmo planteado, se propone una *estimación no paramétrica: el estimador Bernstein del grafo no dirigido*. Al final del capítulo, presentamos un estudio de simulación para evaluar y comprender el desempeño de nuestra propuesta de estimación del grafo vía el algoritmo planteado, enfatizando en el comportamiento del estimador Bernstein, basados en la inferencia del grafo para muestras de cópulas gaussianas y  $\mathcal{T}$  con presencia de relaciones de independencia condicional (no correlación parcial en el caso de cópula  $\mathcal{T}$ ). En el contexto del estudio de simulación, como aportación adicional, proponemos una vía para la construcción de matrices de correlación positivas definidas con patrones específicos en las relaciones de independencia condicional (o no correlación parcial), lo anterior sin la necesidad de imponer estructura de correlación alguna o restringir el

---

<sup>1</sup>Pair copula constructions (PPC's).

grado máximo del grafo, la propuesta aplica a los casos de cópula gaussiana y  $\mathcal{T}$ . Cabe destacar que el estimador Bernstein del grafo no dirigido es *la principal contribución* que hace la presente investigación, en el sentido de que logra aumentar la flexibilidad del modelo de probabilidad subyacente de la estimación con respecto a las propuestas paramétricas.

## 4.1. Organización del capítulo

Partiendo del objetivo fundamental de investigación de buscar estimaciones del grafo no dirigido cuyo modelo de probabilidad subyacente sea capaz de representar estructuras flexibles de dependencia, como capítulo central del presente documento, en éste se propone la *estimación no paramétrica* del grafo no dirigido haciendo uso de la fórmula de (Joe, 1996), el supuesto simplificador y el estimador Bernstein de la cópula para el caso bivariado, además el desempeño del estimador propuesto es evaluado mediante un estudio de simulación. En la Sección 4.2 se genera una discusión sobre la relevancia que cobra el supuesto simplificador para la inferencia de relaciones de independencia condicional partiendo de una muestra observada, además se presenta la fórmula de Joe (1996). En la Sección 4.3 se expone el algoritmo que proponemos para la estimación del grafo vía cópulas condicionales. En la Sección 4.4 se genera una discusión sobre la necesidad de implementar un control estadístico del error al encontrarnos ante el problema de contrastar pruebas múltiples, se comenta sobre los tipos de error en los que se puede incurrir al estimar el grafo, se presentan algunos de los resultados obtenidos en Guo and Rao (2008) para hipótesis dependientes sobre las cotas para la *tasa de descubrimientos falsos* definida por Benjamini and Hochberg (1995), y finalmente, alineados a los resultados en Guo and Rao (2008) se propone el control para un tipo de error en la estimación del grafo. La Sección 4.5 presenta las herramientas necesarias para implementar el algoritmo planteado en la Sección 4.3, tomando un enfoque paramétrico de cópula gaussiana y  $\mathcal{T}$  multivariadas (cópulas simplificadas). En la Sección 4.6 se deduce la propuesta del *estimador Bernstein del grafo no dirigido*. En la Sección 4.7 se describe el estudio de simulación junto con los aspectos que tomamos en consideración para su desarrollo. La Sección 4.8 resume los resultados del estudio, en términos generales se analizan tres resultados subyacentes de las estimaciones propuestas: el comportamiento de las *curvas ROC* asociadas; el cambio en la *tasa de error global* (TEG) ante la variación del parámetro de ajuste; y finalmente, se genera una discusión sobre los aspectos que pueden *afectar el control en la tasa de descubrimientos falsos*. La Sección 4.9 resalta sobre las implicaciones que tiene el supuesto simplificador en nuestra propuesta de estimación. Finalmente, en la Sección 4.10 se desarrolla una discusión general sobre los resultados presentados y las conclusiones obtenidas a lo largo del capítulo, puntualiza sobre tópicos para investigación futura y se concluye con los alcances que tiene nuestra propuesta en el contexto de altas dimensiones ( $N \ll p$ ).

## 4.2. Preliminares

La fórmula de Joe (1996) junto con el supuesto simplificador son la base de nuestra propuesta de estimación, en esta sección exhibimos su relevancia práctica para lograr inferir las relaciones de independencia condicional modeladas en un grafo no dirigido.

### Independencia condicional vía cópulas

Para los fines de estimar el modelo gráfico de interés, es necesario contar con una representación explícita para la independencia condicional como la presentamos en (2.7.2). Sin embargo, partiendo de una muestra de datos observados, validar (digamos con una prueba de hipótesis) que  $\forall u_1, u_2 \in \mathbb{I}, \mathcal{C}_{j,k|\mathcal{V}\setminus\{j,k\}}(u_1, u_2; \mathbf{v}) = \Pi(u_1, u_2) \forall \mathbf{v} \in \text{Ran } \mathbf{X}_{\mathcal{V}\setminus\{j,k\}}$  es prácticamente imposible. En primer lugar, en una muestra no observamos todos los valores  $\mathbf{v} \in \text{Ran } \mathbf{X}_{\mathcal{V}\setminus\{j,k\}}$ , y por otro lado, no es inmediato saber cómo podríamos obtener una muestra representativa de observaciones de las cópulas condicionales  $\mathcal{C}_{j,k|\mathcal{V}\setminus\{j,k\}}(\cdot, \cdot; \mathbf{v})$ . Para aclarar lo antes comentado, comencemos con un ejemplo con tres variables.

Supongamos que tenemos  $N$  observaciones independientes de un vector trivariado  $(X_1, X_2, X_3)^T$ ,

$$\{(x_{1i}, x_{2i}, x_{3i})\}_{i=1}^N, \quad (4.2.1)$$

y nos hacemos la siguiente pregunta:  $\zeta(1, 3) \in \mathcal{A}$ ?



Para el elemento  $i$  de la muestra comenzamos por preguntarnos cómo es la función de distribución de  $X_1$  y  $X_3$  condicional en  $X_2 = x_{2i}$  o más aún cómo es su cópula condicional. A partir del Teorema de extensión, 2.7.1, del Teorema de Sklar (1959), 2.4.1, tenemos lo siguiente al evaluar la FDA condicional en los elementos  $x_{1i}$  y  $x_{3i}$  de la muestra:

$$\begin{aligned} H(x_{1i}, x_{3i}|x_{2i}) &= \mathcal{C}_{13|2}(F_{1|2}(x_{1i}|x_{2i}), F_{3|2}(x_{3i}|x_{2i}); x_{2i}) \\ &= \mathcal{C}_{13|2}(u_{1|2,i}, u_{3|2,i}; x_{2i}). \end{aligned}$$

De aquí en adelante, cuando sea necesario, denotaremos con  $u_{k|s,i}$  a las funciones de distribución univariada de la variable  $k$  condicional al valor que toman las variables pertenecientes a  $s \subset \mathcal{V} \setminus \{k\}$  en el elemento  $i$  de la muestra, en otras palabras, observaciones de las cópulas condicionales correspondientes. En particular para nuestro ejemplo  $s = \{2\}$  y por cada elemento  $i$  en la muestra tenemos un v.a. bivariado  $(U_{3|2,i}, U_{1|2,i})$  con marginales uniformemente distribuidas en el intervalo  $(0, 1)$ , tales que,

$$(U_{3|2,i}, U_{1|2,i}) \sim \mathcal{C}_{13|2}(\cdot, \cdot; x_{2,i}), \quad \forall i \in \{1, \dots, N\}. \quad (4.2.2)$$

Finalmente, notamos que por cada cópula condicional,  $\mathcal{C}_{13|2}(\cdot, \cdot; x_{2,i})$ , tenemos una sola observación  $(u_{3|2,i}, u_{1|2,i})$  del v.a.  $(U_{3|2,i}, U_{1|2,i})$ . A partir de la observación anterior surge la siguiente pregunta: ¿cómo validar la independencia condicional cuando sólo tenemos una observación de cada cópula condicional?

De hecho el problema de estudiar la independencia condicional no es sencillo y usualmente se recurre a supuestos restrictivos para abordarlo, v.g. asumir que las variables presentan relaciones lineales con errores aditivos gaussianos (i.e. que las variables siguen un modelo de factores) (c.f. Zhang et al. (2012) para una introducción con un panorama general sobre los enfoques existentes para probar la independencia condicional). Otro de los supuestos usuales a los que recurre para lograr hacer pruebas de independencia condicional es al *supuesto simplificador*, aquel que introducimos en la Sección 2.7. Asumiendo el supuesto simplificador en nuestro ejemplo, todas las cópulas en (4.2.2) serían la misma, i.e.  $\mathcal{C}_{13|2}(\cdot, \cdot; x_{2,i}) = \mathcal{C}_{13|2}(\cdot, \cdot) \forall i \in \{1, \dots, N\}$ , y así lograríamos tener una muestra de tamaño  $N$  de observaciones de la cópula condicional  $\mathcal{C}_{13|2}$ :

$$\{(u_{1|2,i}, u_{3|2,i})\}_{i=1}^N. \quad (4.2.3)$$

Finalmente, bastaría aplicar una prueba de independencia (no condicional) a la muestra en (4.2.3) y con base en ésta rechazar o no la pertenencia de la arista  $(1, 3)$  al conjunto de aristas,  $\mathcal{A}$ .

Pero la siguiente pregunta natural es: ¿cómo obtenemos las observaciones  $(u_{1|2,i}, u_{3|2,i}) \ i \in \{1, \dots, N\}$ ?

Para dar respuesta a esa pregunta necesitamos introducir la fórmula de Joe (1996), la cual presentamos a continuación.

## Fórmula de Joe

Por practicidad para los fines de nuestra investigación, enunciaremos la fórmula de Joe (1996) con la notación de un modelo gráfico no dirigido.

**Proposición 4.2.1 Fórmula de Joe.** Sea  $\mathbf{X} = (X_1, \dots, X_p)^T$  un v.a. con conjunto de vértices asociado  $\mathcal{V}$ . Además sean  $A, \{k\} \subset \mathcal{V}$ , tales que  $\text{card}(A) = m \geq 1$  y  $k \notin A$ . Digamos que los vértices en  $A$  son  $w_1, \dots, w_m$  y tomemos  $\mathbf{v} = (v_{w_1}, \dots, v_{w_m}) \in \text{Ran } \mathbf{X}_A$ . Entonces bajo condiciones de regularidad, para algún  $j \in \{1, \dots, m\}$ ,

$$F_{k|A}(x|\mathbf{v}) = \frac{\partial \mathcal{C}_{w_j k|A-j}(F_{w_j|A-j}(\mathbf{v}_j|\mathbf{v}_{-j}), F_{k|A-j}(x|\mathbf{v}_{-j}); \mathbf{v}_{-j})}{\partial F_{w_j|A-j}(\mathbf{v}_j|\mathbf{v}_{-j})}, \quad (4.2.4)$$

donde  $A_{-j} = A \setminus \{w_j\}$ ,  $\mathbf{v}_j = v_{w_j}$ . Finalmente, si  $j \notin \{1, m\}$  entonces  $\mathbf{v}_{-j} = (v_{w_1}, \dots, v_{w_{j-1}}, v_{w_{j+1}}, \dots, v_{w_m})$ ; si  $j = 1$  entonces  $\mathbf{v}_{-j} = (v_{w_2}, \dots, v_{w_m})$ ; y si  $j = m$  entonces  $\mathbf{v}_{-j} = (v_{w_1}, \dots, v_{w_{m-1}})$ .

La fórmula de Joe asegura que las distribuciones condicionales que se necesitan como argumentos de cópulas al nivel  $j$  se obtienen como derivadas parciales de cópulas al nivel  $j - 1$ . Una propuesta alternativa de prueba a la fórmula de Joe (1996) se encuentra en el Apéndice C; lo anterior es una aportación de esta investigación.

Regresando a nuestro ejemplo, a partir de (4.2.4) se sigue que con base en la parcial de  $\mathcal{C}_{12}$  con respecto a  $F_2$  se pueden extraer las observaciones  $u_{1|2,i}$ , y de igual forma, se extraen las observaciones  $u_{3|2,i}$  con base en la parcial de  $\mathcal{C}_{32}$  con respecto a  $F_2$ , dando así respuesta a la última pregunta planteada.

En particular la fórmula de Joe es una vía para obtener observaciones de cada cópula condicional al conjunto de vértices en  $A \in \mathcal{V}$ , a partir de la parcial de la cópula condicional (o no condicional en el caso de que  $\text{card}(A) = 1$ ) al conjunto de vértices en  $A \setminus \{w_j\}$  para algún  $w_j \in A$ . Puntualmente, dado un par de vértices  $k, j \in \mathcal{V}$  podemos tomar  $A = \mathcal{V} \setminus \{k, j\}$  y extraer las observaciones de las cópulas condicionales de interés para el grafo no dirigido, lo anterior es la base para el algoritmo que se propone en la siguiente sección para estimar el grafo.

### 4.3. $\alpha$ -estimador del grafo no dirigido vía cópulas condicionales

En esta sección presentamos el algoritmo propuesto para la estimación del grafo y explicamos su aplicación.

Dada una muestra observada de tamaño  $N$  de un v.a.  $p$ -variado  $\mathbf{X} = (X_1, \dots, X_p)^T$ ,

$$\{(x_{1i}, \dots, x_{pi})\}_{i=1}^N, \quad (4.3.1)$$

el problema de estimar la GIC,  $\widehat{\mathcal{G}}$ , es el fondo buscar un estimador del conjunto de aristas,  $\widehat{\mathcal{A}}$ , i.e. inferir sobre la presencia/ausencia de cada una de las relaciones de independencia condicional modeladas en el grafo. A la cantidad de aristas en un grafo no dirigido la entenderemos como el grado de *escasez*<sup>2</sup> del modelo.

Mediante la obtención de *observaciones de cópulas condicionales* y partiendo de un parámetro  $\alpha$  que controla el *grado de escasez* del grafo, nuestra propuesta de estimación se presenta en el Algoritmo 4.3.1.

**Algoritmo 4.3.1**  $\alpha$ -estimador del grafo no dirigido vía cópulas condicionales:  $\widehat{\mathcal{G}}_\alpha$ .

1. Asumiendo el supuesto simplificador, *extraer iterativamente* observaciones de las  $\binom{p}{2}$  cópulas condicionales  $\mathcal{C}_{jk|\mathcal{V}\setminus\{jk\}}$  para  $j, k \in \mathcal{V}$ , t.q.  $j < k$ .
2. Fijar  $\alpha \in [0, 1]$  como valor que controla el *grado de escasez* de la estimación: a  $\alpha$  le llamaremos *parámetro de ajuste*<sup>3</sup>.
3. A partir de  $\alpha$  *contrastar* cada uno de los siguientes pares de hipótesis,

$$\mathcal{H}_{0,jk} := \mathcal{C}_{jk|\mathcal{V}\setminus\{jk\}} = \Pi, \quad \mathcal{H}_{1,jk} := \mathcal{C}_{jk|\mathcal{V}\setminus\{jk\}} \neq \Pi, \quad j, k \in \mathcal{V}, \text{ t.q. } j < k, \quad (4.3.2)$$

con base en las muestras de observaciones de cópulas condicionales obtenidas en el primer paso.

4. Hacer  $\widehat{\mathcal{G}}_\alpha = (\mathcal{V}, \widehat{\mathcal{A}}_\alpha)$ , donde

$$\widehat{\mathcal{A}}_\alpha := \{(j, k), (k, j) \in \mathcal{V} \times \mathcal{V} \mid \text{Se rechaza } \mathcal{H}_{0,jk}, j, k \in \mathcal{V}, \text{ t.q. } j < k\}. \quad (4.3.3)$$

<sup>2</sup>Sparsity, en inglés.

<sup>3</sup>En mucha de la literatura estadística de habla inglesa a este tipo de parámetros se les conoce como: *tuning parameter*.

**Observaciones 4.3.1 :**

**(Obs. 1)** En la literatura estadística moderna, el concepto de *escasez* es utilizado en diversos contextos sobre todo cuando la naturaleza de la metodología involucra una gran cantidad de variables. Aunque este concepto puede tener distintas interpretaciones puntuales según el contexto de análisis, en general el principio de escasez asume que sólo un pequeño número de predictores contribuyen a una variable de respuesta. Como ya lo mencionamos, en el contexto de modelos gráficos no dirigidos, la escasez se asocia con la cantidad de aristas en el modelo, i.e. con la cantidad de relaciones de independencia condicional subyacentes del v.a. Así el modelo gráfico no dirigido *completo implica escasez nula*, y de igual forma, el modelo gráfico no dirigido *vacío induce escasez completa*. Retomaremos de manera más profunda el concepto en el Capítulo 5.

**(Obs. 2)** En el Paso 3 del Algoritmo 4.3.1, el *contraste* de hipótesis vía un valor  $\alpha$  es en un sentido amplio en donde  $\alpha$  domina el contraste conjunto entre las hipótesis y no sólo el individual.

**(Obs. 3)** Nuevamente en el Paso 3 del Algoritmo 4.3.1, una vez obtenidas las observaciones de las cópulas condicionales a pares involucradas en el grafo, se puede hacer uso de una prueba de independencia no condicional con las observaciones de las cópulas condicionales y así contrastar la independencia condicional. Para los fines de la presente investigación, la prueba que se ocupará es aquella en [Genest and Rémillard \(2004\)](#) la cual se basa en el proceso de la cópula empírica.

Dentro del Algoritmo 4.3.1 un paso crucial es el primero, así que a continuación detallaremos sobre éste, puntualizando en la necesidad del supuesto simplificador.

**Extracción iterativa de observaciones de cópulas condicionales**

Partiendo de una muestra de tamaño  $N$  de un v.a. de dimensión  $p$  como en (4.3.1), digamos que para un par arbitrario de vértices  $j, k \in \mathcal{V}$  ( $j < k$ ) necesitamos extraer las observaciones,

$$\left\{ \left( u_{j|\mathcal{V} \setminus \{j,k\}, i}, u_{k|\mathcal{V} \setminus \{j,k\}, i} \right) \right\}_{i=1}^N, \quad (4.3.4)$$

de la cópula condicional  $\mathcal{C}_{jk|\mathcal{V} \setminus \{j,k\}}$ . Notemos que de antemano estamos asumiendo el supuesto simplificador al no poner a la cópula condicional dependiente del valor del condicionante.

Por practicidad de notación pongamos un orden (en realidad arbitrario) a los vértices en el conjunto del condicionante, digamos  $\mathcal{V} \setminus \{j, k\} = \{w_1, \dots, w_{p-2}\}$ .

Primero para la variable  $X_k$ , podemos comenzar por obtener una muestra de observaciones de la *cópula no condicional*  $\mathcal{C}_{kw_1}$  con base en las FDA's marginales y a partir de éstas inferir sobre la cópula, como lo representamos en (4.3.5). La cópula estimada la denotaremos con  $\widehat{\mathcal{C}}_{kw_1}$ . La inferencia puede ser paramétrica o no paramétrica.

$$\left\{ \left( u_{k,i}, u_{w_1,i} \right) \right\}_{i=1}^N = \left\{ \left( F_k(x_{k,i}), F_{w_1}(x_{w_1,i}) \right) \right\}_{i=1}^N \longrightarrow \widehat{\mathcal{C}}_{kw_1}. \quad (4.3.5)$$

Posteriormente, podemos obtener observaciones de las distribuciones de  $X_k$  condicional en las observaciones de  $X_{w_1}$  vía (4.2.4), como en (4.3.6).

$$F_{k|w_1}(x_{k,i}|x_{w_1,i}) = \frac{\partial \widehat{\mathcal{C}}_{kw_1}(F_k(x_{k,i}), F_{w_1}(x_{w_1,i}))}{\partial F_{w_1}(x_{w_1,i})}, \quad i \in \{1, \dots, N\}. \quad (4.3.6)$$

Realizando los procedimientos en (4.3.5) y (4.3.6) respectivos al vértice  $j$  se obtienen las observaciones  $F_{j|w_1}(x_{j,i}|x_{w_1,i})$ ,  $i \in \{1, \dots, N\}$ . A la obtención de las observaciones en (4.3.7) lo llamaremos la *primera iteración* para el par de vértices  $j, k \in \mathcal{V}$  ( $j < k$ ).

$$\left\{ \left( F_{j|w_1}(x_{j,i}|x_{w_1,i}), F_{k|w_1}(x_{k,i}|x_{w_1,i}) \right) \right\}_{i=1}^N. \quad (4.3.7)$$

Análogamente, como en (4.3.5) podemos inferir sobre la cópula  $C_{w_2 w_1}$  y posteriormente obtener observaciones de las distribuciones de  $X_{w_2}$  condicional en las observaciones de  $X_{w_1}$  vía (4.2.4),  $F_{w_2|w_1}(x_{w_2,i}|x_{w_1,i})$ ,  $i \in \{1, \dots, N\}$ .

Asumiendo el supuesto simplificador, los valores estimados de las FDA's de  $X_k$  y  $X_{w_2}$  condicionales en los valores de  $X_{w_1}$  son observaciones de la cópula condicional  $C_{kw_2|w_1}$ , por lo tanto las podemos utilizar para obtener  $\widehat{C}_{kw_2|w_1}$ , como lo representamos en (4.3.8).

$$\left\{ \left( F_{k|w_1}(x_{k,i}|x_{w_1,i}), F_{w_2|w_1}(x_{w_2,i}|x_{w_1,i}) \right) \right\}_{i=1}^N \longrightarrow \widehat{C}_{kw_2|w_1}. \quad (4.3.8)$$

A partir de  $\widehat{C}_{kw_2|w_1}$  podemos obtener observaciones de las FDA's de  $X_k$  condicional en los valores de  $\mathbf{X}_{\{w_1, w_2\}}$  vía (4.2.4), como lo mostramos en (4.3.9),

$$F_{k|w_1, w_2}(x_{k,i}|x_{w_1,i}, x_{w_2,i}) = \frac{\partial \widehat{C}_{kw_2|w_1}(F_{k|w_1}(x_{k,i}|x_{w_1,i}), F_{w_2}(x_{w_2,i}|x_{w_1,i}))}{\partial F_{w_2}(x_{w_2,i}|x_{w_1,i})}, \quad i \in \{1, \dots, N\}. \quad (4.3.9)$$

Realizando los procedimientos en (4.3.8) y (4.3.9) respectivos al vértice  $j$  se obtienen las observaciones  $F_{j|w_1, w_2}(x_{j,i}|x_{w_1,i}, x_{w_2,i})$ ,  $i \in \{1, \dots, N\}$ . A la obtención de las observaciones en (4.3.10) lo llamaremos la *segunda iteración* para el par de vértices  $j, k \in \mathcal{V}$  ( $j < k$ ).

$$\left\{ \left( F_{j|w_1, w_2}(x_{j,i}|x_{w_1,i}, x_{w_2,i}), F_{k|w_1, w_2}(x_{k,i}|x_{w_1,i}, x_{w_2,i}) \right) \right\}_{i=1}^N. \quad (4.3.10)$$

En general, a la obtención de las observaciones en (4.3.11) lo llamaremos la  $r$ -ésima iteración,

$$\left\{ \left( F_{j|w_1, w_2, \dots, w_r}(x_{j,i}|x_{w_1,i}, x_{w_2,i}, \dots, x_{w_r,i}), F_{k|w_1, w_2, \dots, w_r}(x_{k,i}|x_{w_1,i}, x_{w_2,i}, \dots, x_{w_r,i}) \right) \right\}_{i=1}^N. \quad (4.3.11)$$

Entonces se requieren  $p - 2$  iteraciones para obtener la muestra en (4.3.4), por lo tanto, realizando las  $p-4$  iteraciones subsecuentes se obtienen las observaciones deseadas para realizar la prueba de independencia (condicional) para el vértice  $(j, k)$  asumiendo el supuesto simplificador.

Cabe recordar que el orden que le dimos a los vértices en  $\mathcal{V} \setminus \{j, k\}$  fue sólo por practicidad de notación, notemos del proceso iterativo antes descrito que el orden en el cual se van agregando las variables al condicionante es totalmente arbitrario. Notemos además que la extracción iterativa de observaciones de cópulas condicionales *no impone* un modelo multivariado de cópula y en realidad sólo depende de inferencias sobre cópulas a pares en cada iteración, justo como hemos comentado que se procede en el paradigma de construcción de cópulas multivariadas *vine cópula*.

#### 4.4. Control en la tasa de descubrimientos falsos (CTDF)

Como observamos en la Sección 4.3, la estimación del grafo trae consigo el problema de contrastar simultáneamente un número finito de hipótesis nulas:  $\mathcal{H}_{0,jk}$ ,  $j, k \in \mathcal{V}$ , t.q.  $j < k$ . Partiendo de que la  $\text{card}(\mathcal{V}) = p$ , el número de hipótesis a contratar es  $\binom{p}{2}$ , el cual puede ser demasiado grande al crecer el valor de  $p$ . En este contexto de *inferencia a gran escala* (c.f. Efron (2010)), el contraste individual vía un valor  $\alpha \in [0, 1]$  que defina el *tamaño* o *nivel* de cada prueba por separado es insuficiente para asegurar un control del *error conjunto* entre las pruebas de interés (c.f. Casella and Berger (2002), Definición 8.3.5 y 8.3.6, sobre el tamaño y nivel, respectivamente, de una prueba de hipótesis). Para una hipótesis particular, digamos  $\mathcal{H}_{0,jk}$ , los errores tipo I y II asociados a la estimación del grafo se presentan en el Cuadro 4.1.

	$(j, k) \notin \widehat{\mathcal{A}}$	$(j, k) \in \widehat{\mathcal{A}}$
$(j, k) \notin \mathcal{A}$	(Verdadero Negativo) $(j, k) \notin \mathcal{A}$ y $(j, k) \notin \widehat{\mathcal{A}}$	Error tipo I (Falso Positivo) $(j, k) \notin \mathcal{A}$ pero $(j, k) \in \widehat{\mathcal{A}}$
$(j, k) \in \mathcal{A}$	Error tipo II (Falso Negativo) $(j, k) \in \mathcal{A}$ pero $(j, k) \notin \widehat{\mathcal{A}}$	(Verdadero Positivo) $(j, k) \in \mathcal{A}$ y $(j, k) \in \widehat{\mathcal{A}}$

Cuadro 4.1: Errores asociados a la estimación del grafo para una prueba particular.

Así que habrá que buscar extensiones de los errores tipo I y II, que consideren de manera conjunta entre las hipótesis la cantidad de *falsos positivos* (FP) y *falsos negativos* (FN) que subyacen de la estimación del grafo y definir controles para los mismos. En el contexto de comparaciones múltiples, la mayor parte de la literatura está concentrada en definir extensiones y controles del error tipo I. Dentro de los enfoques que existen para controlar los efectos de la multiplicidad en pruebas de hipótesis (c.f. [Efron \(2010\)](#)), en [Benjamini and Hochberg \(1995\)](#) sugieren que la *tasa de descubrimientos falsos* (TDF)<sup>4</sup> es una vía apropiada para extender el error tipo I en muchos de los problemas de hipótesis múltiples. Consideremos lo siguiente,

$$R := \# \text{ de hipótesis rechazadas} = \frac{\text{card}(\widehat{\mathcal{A}})}{2}, \quad (4.4.1)$$

$$V := \# \text{ de hipótesis nulas rechazadas que son verdaderas} = \frac{\# \text{ de aristas en } \widehat{\mathcal{A}} \text{ que no están en } \mathcal{A}}{2}, \quad (4.4.2)$$

donde  $\mathcal{A}$  y  $\widehat{\mathcal{A}}$  son los conjuntos de aristas real y estimado, respectivamente. La *proporción de descubrimientos falsos* se define como,

$$Q := \begin{cases} V/R, & \text{si } R > 0, \\ 0, & \text{si } R = 0. \end{cases} \quad (4.4.3)$$

Entonces la TDF se define como el *valor esperado de Q*, i.e.

$$\text{TDF} = \mathbb{E}[Q]. \quad (4.4.4)$$

En el contexto de la estimación del grafo no dirigido, un *descubrimiento* será la detección de la *presencia de una arista* particular en el conjunto de aristas, i.e. descubrir que un par de variables *no son independientes condicional* en el resto de éstas. Un detalle a destacar sobre la metodología propuesta en [Benjamini and Hochberg \(1995\)](#), es que parte del supuesto de que las estadísticas de prueba, y por ende los p-valores, de las hipótesis involucradas son independientes. Por otro lado, la naturaleza de nuestra propuesta de estimación del grafo (Sección 4.3), vía observaciones estimadas de cópulas condicionales, induce dependencia entre las estadísticas de prueba. Entonces si deseamos tener un control en la tasa de descubrimientos falsos pertinente para el Algoritmo 4.3.1, necesitamos una metodología que considere la dependencia que puede existir entre los p-valores subyacentes de las pruebas de independencia condicional.

En [Guo and Rao \(2008\)](#) proponen alternativas para controlar la tasa de descubrimientos falsos sin suponer ninguna estructura de dependencia entre las estadísticas de prueba, así que esta metodología está más alineada con nuestro contexto de estudio. A continuación presentamos una breve reseña sobre algunas de las cotas obtenidas en la literatura para la TDF para pruebas dependientes, puntualizando en uno de los Teoremas principales en [Guo and Rao \(2008\)](#).

<sup>4</sup>False Discovery Rate (FDR), por sus siglas en inglés.

## Cotas para la TDF ante pruebas dependientes

Comencemos por describir los procesos de pasos *hacia-abajo*<sup>5</sup> y *hacia-arriba*<sup>6</sup> con los cuales se busca controlar la TDF.

Supongamos que deseamos contrastar  $\mathcal{H}_{0,1}, \dots, \mathcal{H}_{0,m}$ <sup>7</sup> y contamos con los p-valores correspondientes  $P_1, \dots, P_m$ . Denotemos con  $I := \{1, \dots, m\}$ , al conjunto de índices asociados a las hipótesis nulas a contrastar, y a  $I_0 := \{i \mid \mathcal{H}_{0,i} \text{ es verdadera, } i \in I\}$ , como el conjunto de índices asociados a las hipótesis nulas verdaderas. Claramente  $\text{card}(I) = m$  y denotemos con  $m_0 = \text{card}(I_0)$ , a la cardinalidad de  $I_0$ , entonces  $m_1 = m - m_0$  es el número de hipótesis nulas que son falsas.

Asumiremos que cada p-valor correspondiente a las hipótesis nulas verdaderas satisface,

$$\mathbb{P}(P_i \leq x) \leq x \quad \forall x \in (0, 1) \text{ y } i \in I_0,$$

donde la distribución conjunta de  $(P_1, \dots, P_m)^T$  es arbitraria.

Sean los p-valores ordenados denotados por  $P_{(1)} \leq \dots \leq P_{(m)}$  y las hipótesis asociadas por  $\mathcal{H}_{(1)}, \dots, \mathcal{H}_{(m)}$ , además supongamos que  $\alpha_1 \leq \dots \leq \alpha_m$  es una secuencia no decreciente de constantes críticas.

El proceso *hacia-abajo* basado en las constantes  $\alpha_1 \leq \dots \leq \alpha_m$  funciona de la siguiente manera: si  $P_{(1)} > \alpha_1$ , entonces ninguna de las hipótesis nulas se rechaza; en otro caso, se rechazan las hipótesis  $\mathcal{H}_{(1)}, \dots, \mathcal{H}_{(r)}$  t.q.  $r$  es el índice más grande que satisface que  $P_{(1)} \leq \alpha_1, \dots, P_{(r)} \leq \alpha_r$ .

El proceso *hacia-arriba* basado en las constantes  $\alpha_1 \leq \dots \leq \alpha_m$  funciona de la siguiente manera: si  $P_{(m)} \leq \alpha_{(m)}$ , entonces se rechazan todas las hipótesis nulas; en otro caso, se rechazan las hipótesis  $\mathcal{H}_{(1)}, \dots, \mathcal{H}_{(r)}$  t.q.  $r$  es el menor índice que satisface  $P_{(m)} > \alpha_m, \dots, P_{(r+1)} > \alpha_{r+1}$ ; si  $P_{(r)} > \alpha_r \forall r \in I$  entonces no se rechaza ninguna de las hipótesis nulas.

Entonces un proceso *hacia-abajo* comienza con la hipótesis más significativa y continua rechazando hipótesis nulas mientras sus p-valores correspondientes sean menores o iguales a las constantes críticas, en contraste, un proceso *hacia-arriba* comienza con la hipótesis menos significativa y continua no rechazando hipótesis nulas mientras sus p-valores correspondientes sean mayores a las constantes críticas.

En [Benjamini and Hochberg \(1995\)](#) se propone un proceso *hacia-arriba* con constantes críticas  $\alpha_i = (i/m)\alpha, i \in I$ , y se prueban que  $\text{TDF} \leq (m_0/m)\alpha$ , asumiendo *independencia entre los p-valores*. Posteriormente, en [Benjamini and Yekutieli \(2001\)](#) se prueba que la  $\text{TDF} \leq (m_0/m)\alpha \sum_{j=1}^m (1/j)$ , para el mismo procedimiento que en [Benjamini and Hochberg \(1995\)](#) pero *sin la imposición de una estructura de dependencia entre los p-valores*. En [Guo and Rao \(2008\)](#) se prueba que la cota propuesta por [Benjamini and Yekutieli \(2001\)](#) no es mejorable para el procedimiento *hacia-arriba*, en el sentido de que existe una distribución de los p-valores en donde la cota es alcanzada, además se presenta el siguiente Teorema 4.4.1 (Teorema 4.1 en [Guo and Rao \(2008\)](#))

**Teorema 4.4.1** La TDF para el procedimiento *hacia-abajo* con constantes críticas  $\alpha_i = (i/m)\alpha, i \in I$ , satisface la siguiente desigualdad:

$$\text{TDF} \leq \frac{m_0}{m} \alpha \left\{ \sum_{j=1}^{m_1+1} \frac{1}{j} + \frac{m_1}{m_1+1} - \frac{m_1}{m} \right\}. \quad (4.4.5)$$

Notemos que la cota en (4.4.5) para el procedimiento *hacia-abajo* es menor que aquella en [Benjamini and Yekutieli \(2001\)](#) para el procedimiento *hacia-arriba* y las mismas constantes críticas. Por otro lado, del Teorema 4.4.1 se

<sup>5</sup>Step-down, en inglés.

<sup>6</sup>Step-up, en inglés.

<sup>7</sup>En el contexto de estimación del grafo  $m = \binom{p}{2}$ .

desprende el siguiente Corolario 4.4.1.

**Corolario 4.4.1** Sea,

$$D(m) = \max_{1 \leq m_0 \leq m} \frac{m_0}{m} \left\{ \sum_{j=1}^{m_1+1} \frac{1}{j} + \frac{m_1}{m_1+1} - \frac{m_1}{m} \right\}. \quad (4.4.6)$$

Entonces la TDF para el procedimiento *hacia-abajo* con las constantes críticas  $\alpha'_i = (i/m) \alpha / D(m)$ ,  $i \in I$ , satisface la siguiente desigualdad:

$$\text{TDF} \leq \alpha. \quad (4.4.7)$$

El hecho de que  $m_0$  sea en realidad un valor desconocido en la práctica implica que para el procedimiento *hacia-abajo*, con constantes críticas  $\alpha_i = (i/m) \alpha$  como en el Teorema 4.4.1, podemos asegurar es que  $\text{TDF} \leq \alpha D(m)$ ; donde ésta siempre es una cota mayor a la de (4.4.7) por el hecho de que  $D(m) \geq 1$ ,  $\forall m \geq 2$ .

En la Sección 4.8 comparamos el desempeño de los procedimientos en el Teorema 4.4.1 y el Corolario 4.4.1, para el control en la tasa de descubrimientos falsos en el contexto de estimación del grafo no dirigido a partir de hipótesis múltiples como en el Algoritmo 4.3.1, basados en el estudio de simulación que proponemos en la Sección 4.7. También en la Sección 4.8 analizaremos los factores que pueden afectar o favorecer el control en nuestro contexto de estimación.

Cabe remarcar que estamos implementando el CTDF, el cual extiende el control convencional que se le da al error tipo I, simplemente como una propuesta para controlar el error al estimar el grafo no porque consideremos que controlar este tipo de error es más importante que controlar alguna extensión del error tipo II. Más adelante se continuará con esta discusión sobre el control para los distintos tipos de error.

## 4.5. Enfoque paramétrico de cópula multivariada

Asumiendo un sólo modelo de copula multivarida para el v.a.  $\mathbf{X} = (X_1, \dots, X_p)^T$ , como lo comentamos en el Capítulo 3, los casos de cópula simplificada en los cuales un modelo gráfico es *útil* son los de cópulas gaussiana y  $\mathcal{T}$ . Por otro lado, como lo presentamos en la Sección 4.3, necesitamos las parciales de cópulas a pares condicionales (y no condicionales) para la estimación del grafo. Partiendo del hecho de que las cópulas condicionales de cópula gaussiana y  $\mathcal{T}$  pertenecen a la misma familia (Proposición 3.3.2 y 3.3.3, respectivamente), bastará con estudiar cómo se comportan las parciales de cópulas bivariadas de estas familias para poder implementar el Algoritmo 4.3.1 en un *enfoque paramétrico de cópula multivariada*. En esta sección, en las Proposiciones 4.5.1 y 4.5.2, presentamos las expresiones para extraer observaciones de cópulas condicionales vía (4.2.4) para los casos de cópula gaussiana y  $\mathcal{T}$  multivariadas.

**Proposición 4.5.1 Parciales de cópula gaussiana bivariada.** Sea el v.a.  $\mathbf{X} = (X_1, X_2)^T$  con cópula subyacente  $\mathcal{C}_{12}^G$  como en (3.3.2) con  $p=2$  y  $\Sigma$  matriz de correlaciones, tal que  $\Sigma_{21} = \Sigma_{12} = \rho_{12}$ . Entonces  $\forall x_1 \in \text{Ran } X_1$  y  $\forall x_2 \in \text{Ran } X_2$  la extracción de las observaciones  $F_{1|2}(x_1|x_2)$  vía la fórmula (4.2.4) es la siguiente:

$$\begin{aligned} F_{1|2}(x_1|x_2) &= \frac{\partial \mathcal{C}_{12}^G(F_1(x_1), F_2(x_2))}{\partial F_2(x_2)} \\ &= \Phi \left( \frac{\Phi^{-1}(F_1(x_1)) - \rho_{12} \Phi^{-1}(F_2(x_2))}{\sqrt{1 - \rho_{12}^2}} \right), \end{aligned} \quad (4.5.1)$$

donde  $\Phi$  es la FDA gaussiana estándar univariada.

**Proposición 4.5.2 Parciales de cópula  $\mathcal{T}$  bivariada.** Sea el v.a.  $\mathbf{X} = (X_1, X_2)^T$  con cópula subyacente  $\mathcal{C}_{12}^{\mathcal{T}}$  como en (3.3.5) con  $p=2$  y  $\Sigma$  matriz de correlaciones, tal que  $\Sigma_{21} = \Sigma_{12} = \rho_{12}$ , y  $\nu_{12} > 2$  grados de libertad. Entonces  $\forall x_1 \in \text{Ran } X_1$  y  $\forall x_2 \in \text{Ran } X_2$  la extracción de las observaciones  $F_{1|2}(x_1|x_2)$  vía la fórmula (4.2.4) es la siguiente:

$$\begin{aligned} F_{1|2}(x_1|x_2) &= \frac{\partial \mathcal{C}_{12}^{\mathcal{T}}(F_1(x_1), F_2(x_2))}{\partial F_2(x_2)} \\ &= t_{\nu_{12}+1} \left( \frac{t_{\nu_{12}}^{-1}(F_1(x_1)) - \rho_{12} t_{\nu_{12}}^{-1}(F_2(x_2))}{\sqrt{\left( \nu_{12} + (t_{\nu_{12}}^{-1}(F_2(x_2)))^2 \right) (1 - \rho_{12}^2)}} \right), \end{aligned} \quad (4.5.2)$$

donde  $t_{\nu_{12}}$  es la FDA  $\mathcal{T}$  estándar univariada con  $\nu_{12}$  grados de libertad, esperanza 0 y varianza  $\frac{\nu_{12}}{\nu_{12}-2}$ .

En el enfoque *paramétrico de cópula multivariada* notamos que la *extracción iterativa de observaciones de cópulas condicionales*, presentada en la Sección 4.3, depende simplemente de los parámetros de cópulas a pares (bivariadas) condicionales o no condicionales según sea el caso. En las Proposiciones 4.5.1 y 4.5.2, estos parámetros los hemos denotado como  $\rho_{12}$  y  $(\rho_{12}, \nu_{12})$  para cópula gaussiana y  $\mathcal{T}$ , respectivamente.

En términos de estimación de parámetros, en este enfoque en donde se asume un modelo completo de cópula simplificada, la estimación simplemente es a nivel no condicional y *los parámetros condicionales se deducen de los no condicionales*. De manera más puntual, las matrices de correlación parcial en las Proposiciones 3.3.2 y 3.3.3, se calculan de manera iterativa como observamos en el Apéndice B para el caso de cópula gaussiana (el caso de cópula  $\mathcal{T}$  es análogo), por otro lado, es claro que el parámetro de grados de libertad para cópulas condicionales en la Proposición 3.3.3 sólo dependen de la estimación no condicional. En la Sección 4.8 discutiremos sobre las implicaciones, en términos de los errores asociados a la estimación del grafo, de este enfoque.

## 4.6. Estimador Bernstein del grafo no dirigido

Esta sección es la parte central del presente capítulo, en ésta proponemos la inferencia del modelo gráfico no dirigido en un enfoque *totalmente no paramétrico*, basada en la estimación de cópulas a pares vía *polinomios de Bernstein*. Bajo este enfoque de estimación del grafo, se espera que las cópulas a pares estimadas representen de manera más precisa la *estructura de dependencia observada* en los datos condicionales y no condicionales.

Como se comenta en [Janssen et al. \(2015\)](#), los estimadores Bernstein han atraído considerable atención como estimadores no paramétricos *suaves* de funciones de distribución, densidad, cópulas y densidades cópula. La gran flexibilidad con la cuentan los mantiene vigentes en la investigación, v.g. en el mismo [Janssen et al. \(2015\)](#) estudian sus propiedades al estimar distribuciones condicionales y regresiones por cuantiles.

En el caso particular de la estimación de la cópula, ésta se basa en un suavizamiento de la cópula empírica. Pese a que la definición del estimador Bernstein de la cópula se extiende a cualquier dimensión (c.f. [Sancetta and Satchell \(2004\)](#) para una investigación general sobre el estimador Bernstein de la cópula extendido a cualquier dimensión), el hecho de que su definición dependa de la cópula empírica, hace que su implementación se vuelva inviable en el sentido computacional para dimensiones mayores a tres, sobre todo cuando el tamaño de la muestra es grande. En la presente propuesta de estimación no paramétrica del grafo no dirigido, explotamos la gran flexibilidad y viabilidad computacional del estimador Bernstein de la cópula para el caso bivariado. Comenzaremos por la definición de cópula empírica, para posteriormente introducir al estimador Bernstein de la cópula, ambos en el caso bivariado.

**Definición 4.6.1 Cópula empírica.** Dada una muestra de observaciones  $\{(x_{1,i}, x_{2,i})\}_{i=1}^N$  del v.a.  $(X_1, X_2)^T$ , la *cópula empírica* asociada a la muestra está dada por,

$$\mathcal{C}_N \left( \frac{i}{N}, \frac{j}{N} \right) = \frac{\# \text{ de pares } (x, y) \text{ en la muestra con } x \leq x_{(i)}, y \leq y_{(j)}}{N}, \quad (4.6.1)$$

donde  $x_{(i)}, y_{(j)}$ , con  $1 \leq i, j \leq N$ , denotan los estadísticos de orden de la muestra.



**Definición 4.6.2 Estimador Bernstein de la cópula.** Dada una muestra de observaciones  $\{(x_{1,i}, x_{2,i})\}_{i=1}^N$  del v.a.  $(X_1, X_2)^T$ , con cópula empírica  $\mathcal{C}_N$  como en (4.6.1), el suavizamiento vía polinomios de Bernstein de  $\mathcal{C}_N$  es el siguiente,

$$\tilde{\mathcal{C}}(u, v) = \sum_{i=0}^N \sum_{j=0}^N \mathcal{C}_N \left( \frac{i}{N}, \frac{j}{N} \right) \binom{N}{i} u^i (1-u)^{N-i} \binom{N}{j} v^j (1-v)^{N-j}, \quad (4.6.2)$$

$\forall (u, v) \in \mathbb{I}^2$ .

Por otro lado, notamos en la Sección 4.3 que, la extracción de observaciones de cópulas condicionales sólo depende de la inferencia de cópulas bivariadas. Así nuestra *propuesta de estimación no paramétrica* del grafo no dirigido, es hacer una estimación Bernstein bivariada en cada iteración del proceso iterativo para extraer observaciones de las  $\binom{p}{2}$  cópulas condicionales involucradas en el grafo. Por ejemplo, en (4.3.5), (4.3.8) y las iteraciones subsecuentes las cópulas serían estimadores Bernstein. El hecho de que la cópula Bernstein sea un estimador suave de la cópula permite el cálculo de parciales como se requiere para la extracción de observaciones de cópulas condicionales vía (4.2.4). Entonces en el *paso uno* del Algoritmo 4.3.1 lograríamos tener, para cada una de las cópulas condicionales de interés, observaciones estimadas vía polinomios de Bernstein sin la imposición de algún modelo paramétrico y con éstas proceder a efectuar las pruebas de hipótesis. Al estimador que resulte de aplicar el Algoritmo 4.3.1 con el paso uno como lo acabamos de describir lo llamaremos: *el estimador Bernstein del grafo no dirigido*. A este estimador del grafo lo denotaremos como:  $\tilde{\mathcal{G}} = (\mathcal{V}, \tilde{\mathcal{A}})$ .

Para la implementación del estimador Bernstein del grafo no dirigido simplemente necesitamos conocer la expresión de la parcial de la cópula Bernstein bivariada, como la presentamos en (4.6.3), y con ésta ejecutar la fórmula (4.2.4) toda vez que sea necesaria.

$$\frac{\partial \tilde{\mathcal{C}}(u, v)}{\partial v} = \sum_{i=0}^N \sum_{j=0}^N \mathcal{C}_N \left( \frac{i}{N}, \frac{j}{N} \right) \binom{N}{i} u^i (1-u)^{N-i} \binom{N}{j} v^{j-1} (1-v)^{N-j} \left( \frac{j-vN}{v(1-v)} \right). \quad (4.6.3)$$

Seguendo la fórmula (4.2.4), en la Proposición 4.6.1 se presenta el *estimador Bernstein de la FDA condicional* con condicionante de dimensión uno, i.e. el *estimador Bernstein de las observaciones de cópulas condicionales* como las que están involucradas en la extracción iterativa de observaciones de cópulas condicionales.

**Proposición 4.6.1 Estimador Bernstein de la FDA condicional.** Dada una muestra de observaciones  $\{(x_{1,i}, x_{2,i})\}_{i=1}^N$  del v.a.  $(X_1, X_2)^T$  con cópula Bernstein  $\tilde{\mathcal{C}}_{12}$ , el estimador Bernstein de las observaciones de las FDA's de  $X_1$  condicional en  $X_2$  es el siguiente,

$$\tilde{F}_{1|2}(x_{1,i}|x_{2,i}) = \frac{\partial \tilde{\mathcal{C}}_{12}(F_1(x_{1,i}), F_2(x_{2,i}))}{\partial F_2(x_{2,i})}.$$

En contraste con el enfoque paramétrico de cópula multivariada simplificada, el estimador Bernstein bivariado se calcula en cada iteración de la extracción de observaciones de cópulas condicionales (Sección 4.3), i.e. cada iteración induce la estimación de la cópula correspondiente a la misma. Lo anterior trae consigo ciertas implicaciones en cuanto a la acumulación de error en la estimación final del grafo no dirigido, las cuales serán comentadas en la Sección 4.7. El desempeño del estimador Bernstein será evaluado mediante el estudio de simulación que proponemos a continuación.

## 4.7. Estudio de simulación

Esta sección tiene la finalidad de explicar aspectos generales que tuvimos que considerar para estudiar el *desempeño empírico* con datos sintéticos de la estimación subyacente del Algoritmo 4.3.1, en particular se puntualiza en el comportamiento del estimador Bernstein. También al final de esta sección se expone la estructura general que tuvo el estudio. Los datos sintéticos son muestras simuladas de cópulas gaussianas y  $\mathcal{T}$  con *presencia de escasez*, i.e. con presencia de relaciones de independencia condicional (o no correlación parcial en el caso de cópula  $\mathcal{T}$ ) como aquellas que se modelan en un grafo no dirigido. Cabe mencionar que las muestras simuladas son observaciones directas de

las copulas correspondientes, i.e. muestras de v.a.'s con marginales uniformes en el intervalo  $(0, 1)$ . Antes de describir la estructura del estudio, hay dos aspectos que debemos tomar en consideración: el primero, es sobre cómo controlar el grado de escasez y permitir aleatoriedad en los patrones de escasez de las muestras simuladas; el segundo, tiene que ver con cómo asegurar la construcción de matrices de correlación definidas positivas con patrones de escasez previamente especificados. El segundo punto es también una aportación importante que hacemos en nuestra investigación, la cual busca no imponer estructura de correlación parcial alguna o restringir el grado<sup>8</sup> máximo del grafo, para lograr la construcción de matrices de correlación definidas positivas con un patrón de escasez particular<sup>9</sup>.

### Grado de escasez y aleatoriedad en el patrón de escasez

El procedimiento que implementaremos para simular grafos no dirigidos  $p$ -dimensionales con diversos patrones de escasez es aquel propuesto en Meinhshausen and Bühlmann (2006). Esta propuesta también ha sido implementada por otros autores como Liu et al. (2009) y Liu et al. (2012a). El proceso lo presentamos en el Algoritmo 4.7.1.

#### Algoritmo 4.7.1 Simulación de grafos no dirigidos escasos.

1. Asociaremos a cada vértice  $j \in \mathcal{V}$  con un v.a. bivariado  $(Y_j^{(1)}, Y_j^{(2)})$ , tales que  $Y_1^{(k)}, \dots, Y_p^{(k)} \sim \mathcal{U}(0, 1)$  para  $k = 1, 2$ .
2. Simulamos una observación  $y_j = (y_j^{(1)}, y_j^{(2)})$  de cada v.a.  $(Y_j^{(1)}, Y_j^{(2)})$ ,  $j \in \mathcal{V}$ .
3. A cada par de vértices  $(j, k) \in \mathcal{V} \times \mathcal{V}$  le asignamos la siguiente probabilidad de pertenencia al conjunto de aristas,

$$p_{jk} = \mathbb{P}((j, k) \in \mathcal{A}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\|y_j - y_k\|^2}{2s}\right),$$

donde  $s$  modula el grado de escasez de la simulación y  $\|\cdot\|$  denota la distancia euclidiana.

4. Hacer  $s = 0.125$  para controlar el grado de escasez de la gráfica simulado.
5. Por cada par de vértices  $(j, k) \in \mathcal{V} \times \mathcal{V}$ , t.q.  $j < k$ , realizar un ensayo Bernoulli con probabilidad de éxito  $p_{jk}$ , éste denotarlo con  $b_{jk}$ .
6. El grafo no dirigido escaso simulado,  $\check{\mathcal{G}} = (\mathcal{V}, \check{\mathcal{A}})$ , es tal que,

$$\check{\mathcal{A}} = \{(j, k), (k, j) \in \mathcal{V} \times \mathcal{V} \mid b_{jk} = 1, j < k\}.$$

En el *paso cuatro* del Algoritmo 4.7.1, el valor que toma el parámetro que controla el grado de escasez del grafo simulado,  $s = 0.125$ , se fijó simplemente por temas de tiempo de proceso en nuestro estudio de simulación y por ser un valor recurrente que toman los autores que hemos citado previamente. Sin embargo, para estudios posteriores se podrían evaluar diferentes grados de escasez para las simulaciones. La Figura 4.1 muestra grafos simulados para distintos grados de escasez con  $p = 15$ , lo anterior con la finalidad de brindarle al lector sensibilidad sobre las *diferentes estructuras que induce* el grado de escasez en un modelos gráfico no dirigido.

<sup>8</sup>Sobre la definición de grado, recordemos la Definición 1.4.1 sobre el conjunto de adyacencia o los vecinos de un vértice en el grafo, digamos  $j$ , el grado del vértice  $j$  es igual a la cardinalidad de su conjunto de adyacencia, i.e.  $\text{grado}(j) = \text{card}(\text{ady}(\mathcal{G}, j))$ . Entonces el grado máximo del grafo, denotado con  $m_{\mathcal{G}}$ , se define como  $m_{\mathcal{G}} := \max_{j \in \mathcal{V}} \{\text{grado}(j)\}$ .

<sup>9</sup>En la literatura que consultamos en donde se presentaron estudios de simulación, observamos que la vía recurrente era imponer una estructura de correlación parcial o restringir el grado máximo del grafo para lograr matrices definidas positivas, v.g. Liu et al. (2009) y Liu et al. (2012a).

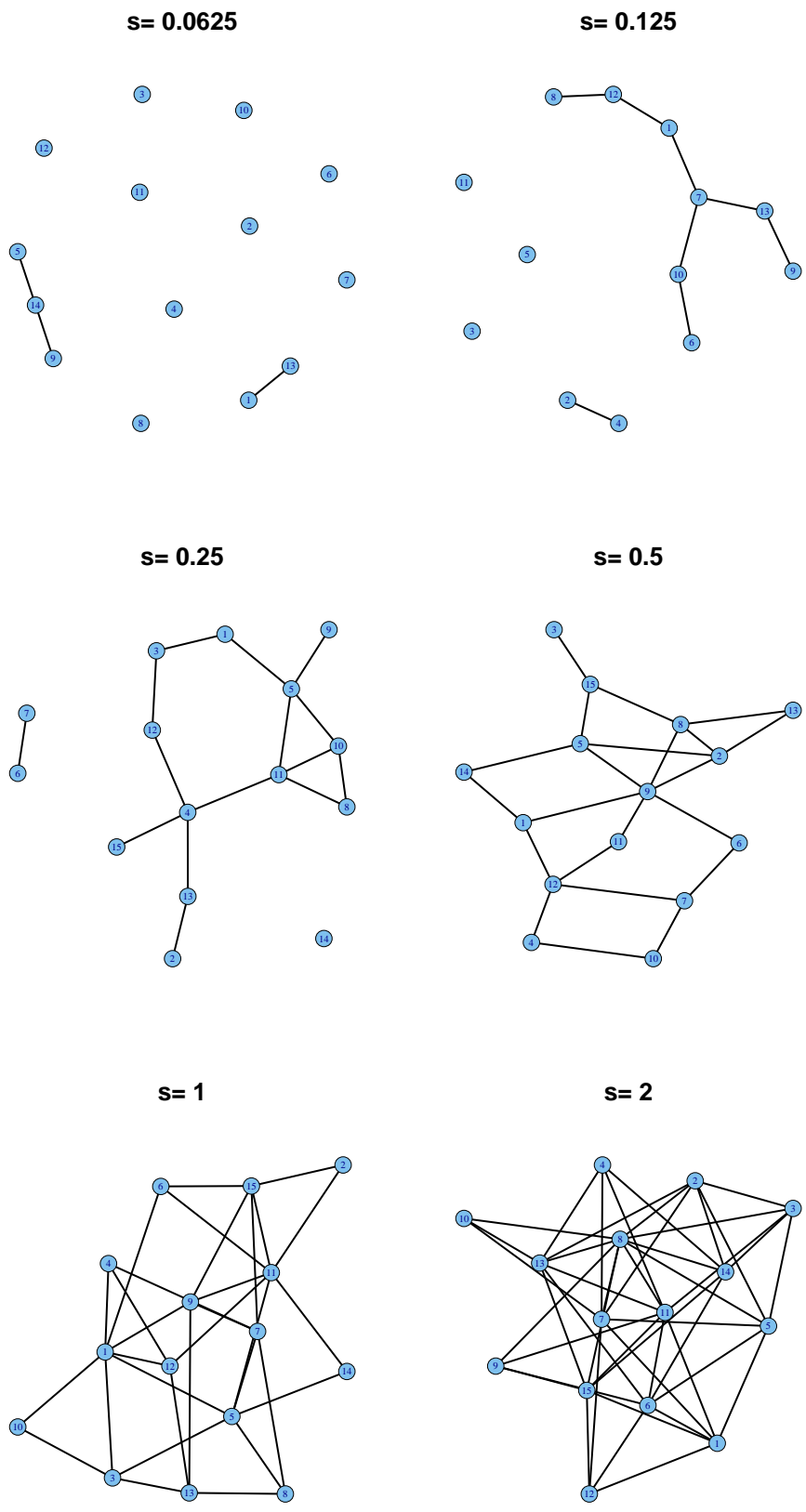


Figura 4.1: Grados de escasez en el grafo simulado para  $p = 15$ .

## Matrices de correlación positivas definidas: cópulas gaussiana y $\mathcal{T}$

A continuación nuestra propuesta para la *construcción de matrices de correlación positivas definidas con patrones de escasez previamente definidos* para los casos de cópulas gaussianas y  $\mathcal{T}$ . Nuestra propuesta está alineada con resultados subyacentes de la teoría de *modelos de ecuaciones estructurales*. Comenzaremos el planteamiento para el caso de *distribución gaussiana*.

Sea  $\mathbf{X} = (X_1, \dots, X_p)^T \sim \mathcal{N}_p(\mathbf{0}, P)$ , con  $P$  una *matriz de covarianzas positiva definida*. Consideremos el modelo de regresión en (1.7.3). Si  $\mathcal{B}$  es una matriz de dimensiones  $p \times p$ , t.q.  $\mathcal{B}_{jk} = \beta_k^{(j)}$  para  $j \neq k$  y  $\text{diag}(\mathcal{B}) = \mathbf{0}$ , y además  $\epsilon = (\epsilon_1, \dots, \epsilon_p)^T$ , se sigue que,

$$\mathbf{X} = \mathcal{B}\mathbf{X} + \epsilon. \quad (4.7.1)$$

Entonces,

$$\begin{aligned} \mathbf{X} - \mathcal{B}\mathbf{X} &= \epsilon \\ (\mathcal{I}_p - \mathcal{B})\mathbf{X} &= \epsilon \end{aligned} \quad (4.7.2)$$

$$\mathbf{X} = (\mathcal{I}_p - \mathcal{B})^{-1} \epsilon. \quad (4.7.3)$$

Se sigue que,

$$P = \text{Var}(\mathbf{X}) = [(\mathcal{I}_p - \mathcal{B})^{-1}] \mathcal{D} [(\mathcal{I}_p - \mathcal{B})^{-1}]^T, \quad (4.7.4)$$

donde  $\mathcal{D}$  es una matriz diagonal t.q.  $\mathcal{D} = \text{Var}(\epsilon)$ . Entonces  $\mathbf{X}$  definido como en (4.7.1) asegura que  $P$  sea invertible y por ende positiva definida, como se observa en (4.7.4).

### Observaciones 4.7.1 :

**(Obs. 1)** La matriz  $(\mathcal{I}_p - \mathcal{B})$  en (4.7.2) siempre es invertible sin importar la estructura de  $\mathcal{B}$ . Lo anterior justifica la igualdad en (4.7.3).

Ahora, recordemos las relaciones en (1.7.5). Entonces toda vez que queramos construir una matriz de correlaciones positiva definida a partir de la expresión en (4.7.1), pero que considere las relaciones de independencia condicional de un modelo gráfico no dirigido particular, digamos  $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ , bastará con hacer,

$$\mathcal{B}_{jk} = \beta_k^{(j)} \neq 0, \forall (j, k) \in \mathcal{A} \quad y \quad \mathcal{B}_{jk} = 0, \forall (j, k) \in (\mathcal{V} \times \mathcal{V}) \setminus \mathcal{A}. \quad (4.7.5)$$

Por otro lado, siguiendo las igualdades en (1.7.2) y (1.7.4) que asocian a los coeficientes de correlación parcial con los coeficientes del modelo de regresión, se concluye que vía (4.7.5) se puede construir cualquier estructura posible de correlación parcial que considere los patrones de escasez inducidos por  $\mathcal{A}$ .

Naturalmente la metodología aplica para la construcción de matrices de correlación positivas definidas y patrones de escasez específicos como parámetro de cópulas gaussianas, por el hecho de que toda vez que tengamos una cópula gaussiana tenemos una distribución gaussiana (latente) con la misma estructura de dependencia (definida vía una matriz de correlaciones), véase Lema 3.3.2.

De igual forma, toda vez que se cuenta con un modelo de cópula  $\mathcal{T}$  se tiene un v.a. (latente) con distribución  $\mathcal{T}$  y la misma estructura de dependencia que induce el modelo de cópula, el cual a su vez tiene la misma estructura de correlación de un v.a. (latente) gaussiano, véase (3.3.8). Entonces la metodología es aplicable también para la construcción de matrices de correlación positivas definidas con patrones de escasez específicos como parámetro de cópulas  $\mathcal{T}$ .

## Simulación de muestras escasas de cópula gaussiana y $\mathcal{T}$

El Algoritmo 4.7.2 describe la forma en la que simulamos una muestra multivariada de observaciones de cópula gaussiana o  $\mathcal{T}$  con presencia de patrones de escasez. El algoritmo planteado conjuga nuestra propuesta para la construcción de matrices de correlación positivas definidas con el Algoritmo 4.7.1.

### Algoritmo 4.7.2 Simulación de muestras escasas de cópulas gaussianas y $\mathcal{T}$ .

1. Ejecutar el Algoritmo 4.7.1 para obtener un grafo no dirigido simulado:  $\check{\mathcal{G}} = (\mathcal{V}, \check{\mathcal{A}})$ , donde  $\mathcal{V} = \{1, \dots, p\}$ .
2. Por cada elemento  $(j, k) \in \check{\mathcal{A}}$  simular una observación de una distribución  $\mathcal{N}(0, 1)$ :  $w_{jk}$ .
3. Construir la matriz  $\check{\mathcal{B}}$  de dimensiones  $p \times p$  como,

$$\check{\mathcal{B}}_{jk} = w_{jk}, \forall (j, k) \in \check{\mathcal{A}} \quad y \quad \check{\mathcal{B}}_{jk} = 0, \forall (j, k) \in (\mathcal{V} \times \mathcal{V}) \setminus \check{\mathcal{A}}.$$

4. Calcular,

$$\check{P} = [(\mathcal{I}_p - \check{\mathcal{B}})^{-1}] [(\mathcal{I}_p - \check{\mathcal{B}})^{-1}]^T.$$

5. Estandarizar  $\check{P}$  para obtener una matriz de correlaciones positiva definida con los mismos patrones de escasez que  $\check{\mathcal{A}}$ ,  $\check{\Sigma}$ , i.e.

$$\check{\Sigma} = \frac{\check{P}_{jk}}{\sqrt{\check{P}_{jj}\check{P}_{kk}}}.$$

6. Simular una muestra de observaciones de cópula gaussiana (o  $\mathcal{T}$ ) con parámetro de correlaciones  $\check{\Sigma}$ .

## El desempeño del estimador

Para evaluar el desempeño del estimador Bernstein y compararlo con la estimación paramétrica vía el Algoritmo 4.3.1, proponemos las siguientes métricas.

$$\text{TFP}(\alpha) = \frac{\# \text{ de aristas en } \widehat{\mathcal{A}}_\alpha \text{ que no están en } \mathcal{A}}{\text{card}(\mathcal{A})}. \quad (4.7.6)$$

$$\text{TFN}(\alpha) = \frac{\# \text{ de aristas en } \mathcal{A} \text{ que no están en } \widehat{\mathcal{A}}_\alpha}{[2 * \binom{p}{2} - \text{card}(\mathcal{A})]}. \quad (4.7.7)$$

Entonces  $\text{TFP}(\alpha)$  y  $\text{TFN}(\alpha)$  son las tasas de *falsos positivos* y *negativos*, respectivamente, subyacentes del  $\alpha$ -estimador en el Algoritmo 4.3.1,  $\widehat{\mathcal{G}}_\alpha$ .

Por otro lado, definimos

$$\text{TEG}(\alpha) = \text{TFP}(\alpha) + \text{TFN}(\alpha).$$

Entonces  $\text{TEG}(\alpha)$  es la *tasa de error global* subyacentes del  $\alpha$ -estimador en el Algoritmo 4.3.1,  $\widehat{\mathcal{G}}_\alpha$ .

Mientras que la  $\text{TFP}(\alpha)$  y la  $\text{TFN}(\alpha)$  cuantifican los niveles de *sobreajuste* y *subajuste*, respectivamente, la  $\text{TEG}(\alpha)$  cuantifica el *error total* de la estimación  $\widehat{\mathcal{G}}_\alpha$ .

Como comentario adicional, notemos que la  $\text{TFP}(\alpha)$  está estrechamente relacionada con la  $\text{TDF}(\alpha)$  (aunque no son lo mismo), así que en el fondo el control en la tasa de descubrimientos falsos que proponemos en la Sección 4.4 busca controlar el sobreajuste del grafo estimado. En la Sección 4.8 retomamos esta discusión.

## Estructura del estudio de simulación

Se simularon muestras con presencia de escasez para las siguientes tres cópulas:

1. Gaussiana;
2.  $\mathcal{T}$  con  $\nu = 4$ ;
3.  $\mathcal{T}$  con  $\nu = 8$ .

Para cada una de estas tres cópulas se combinaron distintos valores de tamaño de muestra y número de variables. Las combinaciones posibles se dieron entre los siguientes valores:

- $N = 50, 100, 200$ .
- $p = 10, 15$ .

Para cada combinación de cópula, tamaño de muestra y número de variables, (cópula,  $N, p$ ), se realizaron 100 repeticiones.

El estudio de simulación tiene como objetivo general evaluar y comprender el desempeño de nuestra propuesta de estimación del grafo vía el Algoritmo 4.3.1, enfatizando en el comportamiento del estimador Bernstein. Para lograr dicho objetivo general, nos planteamos los siguientes tres objetivos particulares:

1. Comparar las estimaciones del grafo entre las tres diferentes cópulas: gaussiana,  $\mathcal{T}$  con  $\nu = 4$  y  $\mathcal{T}$  con  $\nu = 8$ . A grandes rasgos queremos observar el impacto que tiene el parámetro de grados de libertad en los modelos de cópula  $\mathcal{T}$  ante la presencia de relaciones de no correlación parcial, el cual de antemano induce dependencia condicional entre las observaciones, y cómo éste pierde impacto al aumentar el número de variables en el condicionante.
2. Comparar los resultados del enfoque paramétrico de estimación como se describió en la Sección 4.5, con los resultados del estimador Bernstein de la Sección 4.6. Por cada muestra simulada se estimarán grafos:
  - a) en el enfoque paramétrico (con la cópula paramétrica correspondiente a la simulación):  $\widehat{\mathcal{G}}_P$ ;
  - b) vía el enfoque Bernstein:  $\widetilde{\mathcal{G}}$ .
3. Evaluar el desempeño y utilidad de los controles del error que proponemos en la Sección 4.4. Por cada enfoque, paramétrico y Bernstein, se estimarán tres grafos cambiando el procedimiento de control del error:
  - a) con el contraste individual convencional, en donde cada p-valor se compara con un valor  $\alpha \in [0, 1]$  fijo para todas las hipótesis;
  - b) con el proceso *hacia-abajo* del Teorema 4.4.1, en algunas ocasiones nos referiremos a este procedimiento como “*B & H hacia-abajo*” o simplemente como “*B & H*”;
  - c) con el proceso *hacia-abajo* del Corolario 4.4.1, en algunas ocasiones nos referiremos a este procedimiento como “*G & R hacia-abajo*” o simplemente como “*G & R*”.

Entonces para cada muestra simulada mediante el Algoritmo 4.7.2 y un valor de  $\alpha$  que domine el contraste de las hipótesis, se estiman seis grafos diferentes con base en el Algoritmo 4.3.1:

1. estimador paramétrico con contraste individual vía  $\alpha$ ;
2. estimador paramétrico vía “*B & H hacia-abajo*”;
3. estimador paramétrico vía “*G & R hacia-abajo*”;
4. estimador Bernstein con contraste individual vía  $\alpha$ ;
5. estimador Bernstein vía “*B & H hacia-abajo*”;
6. estimador Bernstein vía “*G & R hacia-abajo*”.

Para cada uno de los tres objetivos, es de interés conocer la sensibilidad de los estimadores ante efectos de cambios en el tamaño de muestra y número de variables en el modelo.

## 4.8. Resultados: curvas ROC, TEG y CTDF

Buscando lograr los objetivos que nos hemos planteado, en esta sección analizaremos las *curvas ROC*<sup>10</sup> y las *tasas de error global* subyacentes de las estimaciones propuestas del grafo no dirigido. Además analizamos el comportamiento que tuvieron las TDF's *estimadas* con respecto a las cotas superiores teóricas que se espera que cumplan, lo anterior es de relevancia debido a la relación que existe entre la TFP y la TDF.

### Curvas ROC

Evaluaremos el comportamiento de las *curvas ROC* generadas al estimar el grafo para distintos valores<sup>11</sup> de  $\alpha \in [0, 1]$ . La curva ROC se define como,

$$(TFP(\alpha), TVP(\alpha)) = (TFP(\alpha), 1 - TFN(\alpha)),$$

donde  $TVP(\alpha)$  es la *tasa de verdaderos positivos* de la estimación inducida por  $\alpha$ . Dada la definición de la curva ROC, ésta nos brinda un *panorama global del desempeño* de cada estimador sin favorecer algún tipo de error: sobreajuste o subajuste. Para ilustrar el desempeño general de los distintos estimadores sobre los posibles valores de  $\alpha$ , las *curvas ROC promedio* sobre las 100 repeticiones realizadas se muestran en las:

- Figuras 4.2, 4.3 y 4.4 para cópula gaussiana;
- Figuras 4.5, 4.6 y 4.7 para cópula  $\mathcal{T}$  con  $\nu = 4$ ;
- Figuras 4.8, 4.9 y 4.10 para cópula  $\mathcal{T}$  con  $\nu = 8$ .

Además con la finalidad de lograr mayor sensibilidad sobre el comportamiento de los estimadores para las distintas cópulas, en el Cuadro 4.2 se muestran las *áreas bajo las curvas ROC* provenientes de las figuras antes mencionadas. El área bajo las curvas ROC está restringida al rango de la TFP observada, por lo tanto las áreas entre enfoques de estimación, junto con sus controles de error, no son comparables por la falta de completez de algunas curvas en donde se implementó el CTDF. Sin embargo, para el mismo enfoque y control del error la comparación de áreas entre cópulas nos ayudará a cumplir con el primer objetivo planteado: comparar las estimaciones del grafo entre las tres diferentes cópulas. Cabe mencionar que el área calculada es la región que comprende entre la curva ROC y la función identidad, así que técnicamente el área máxima posible es de 0.5 ( $TVP = 1$  para cualquier valor en la TFP, cuando la TFP es de rango completo).

A continuación presentamos una serie de observaciones generales sobre las curvas ROC provenientes de la estimación del grafo no dirigido:

- (ROC 1)** La primer observación a destacar, que ocurre para toda configuración (cópula,  $N, p$ ) y ante cualquier procedimiento de contraste de hipótesis, es el hecho rotundo de que las *estimaciones presentan un mejor desempeño en el enfoque paramétrico con respecto al estimador Bernstein*.
- (ROC 2)** Naturalmente *el desempeño de los estimados se ve favorecido por el aumento en el tamaño de muestra*, ante cualquier configuración de estimación. De hecho los estimadores para  $N=50$ , aunque por encima, en general se encuentran muy cerca de la línea que representa la clasificación aleatoria. La clasificación aleatoria en nuestro contexto significa incluir, de manera independiente, cada arista en el conjunto de aristas estimado con una probabilidad de 0.5. En términos globales lo que se observa, es que la curva ROC crece más rápido hacia valores cercanos a uno en la TVP al aumento en el tamaño de muestra.
- (ROC 3)** El *potencial de los estimadores*, tanto en el enfoque paramétrico así como en el Bernstein, se ve *afectado considerablemente al aumentar el número de variables* en el modelo. Sin embargo, como lo comentamos en el punto anterior, al aumentar el tamaño de muestra esta aficción se reduce, en el sentido de que el área bajo las curvas para  $p=15$  crece al aumentar el tamaño de muestra y éstas están cada vez más cerca de aquellas con  $p=10$ . Cabe destacar que las curvas para  $p=15$  y  $p=10$  están más próximas en el enfoque paramétrico con respecto al Bernstein.

<sup>10</sup>Receiver Operating Characteristic.

<sup>11</sup>Se tomó una rejilla equiespaciada por 0.001 de  $\alpha \in [0, 1]$ .

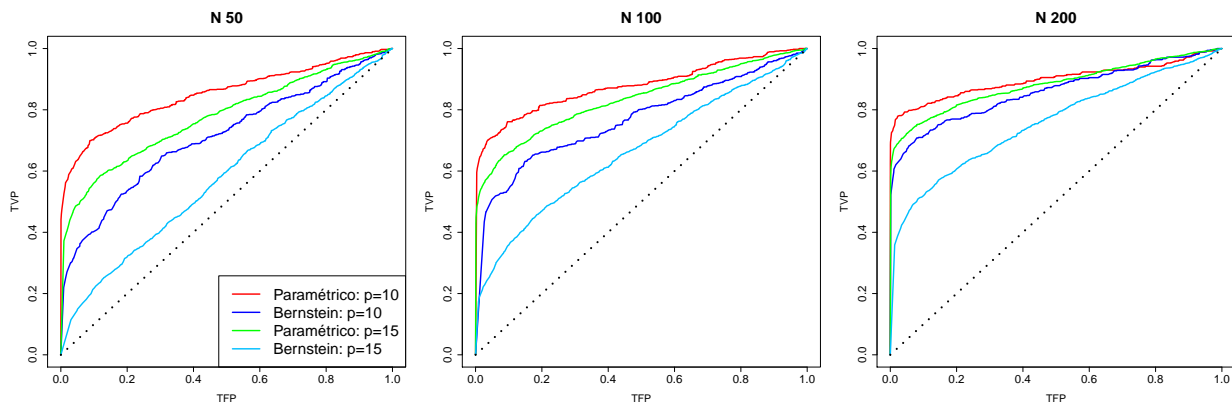


Figura 4.2: Cópula gaussiana: contraste individual vía  $\alpha$ .

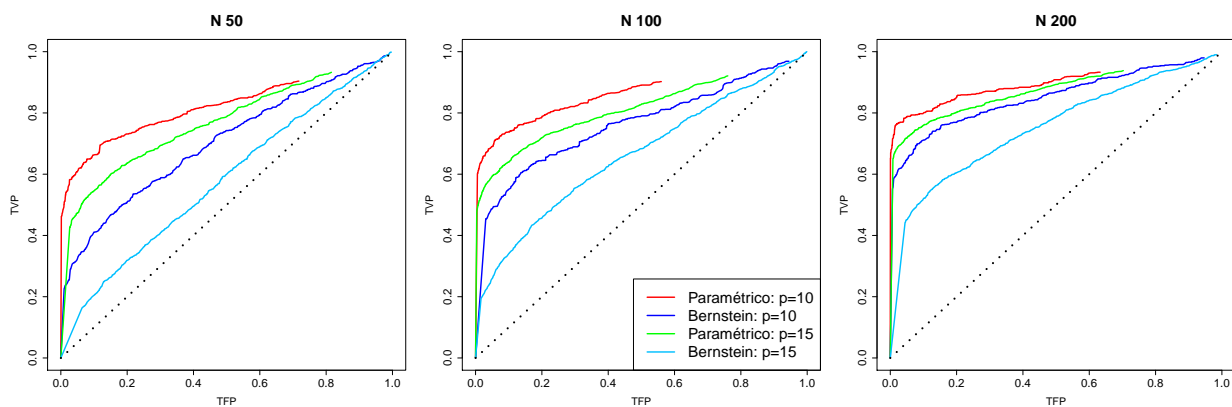


Figura 4.3: Cópula gaussiana: Benjamini & Hochberg *hacia-abajo*.

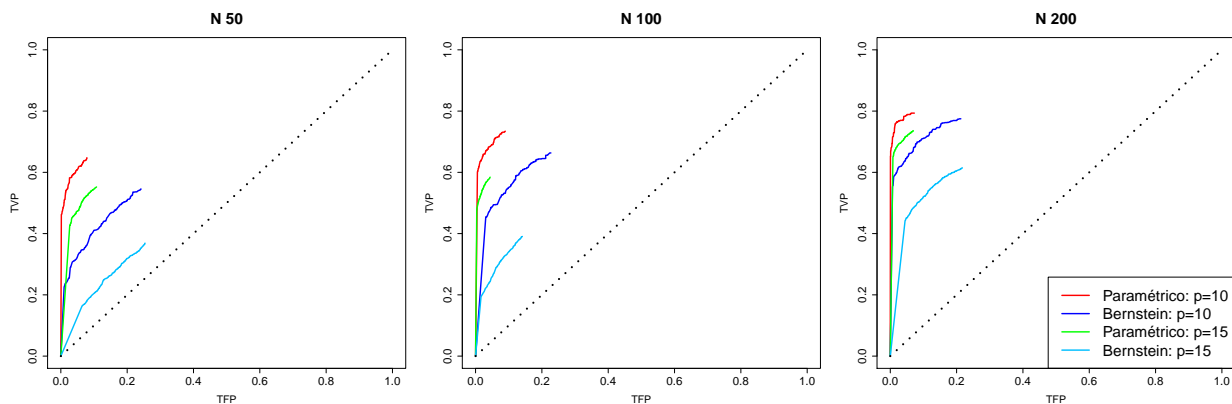


Figura 4.4: Cópula gaussiana: Guo & Rao *hacia-abajo*.

**(ROC 4)** Pese a la aficción que los estimadores tienen al aumento en variables, en términos generales, *los estimadores paramétricos con  $p=15$  presentan un desempeño superior con respecto a los estimadores Bernstein con  $p=10$ .*

**(ROC 5)** En el enfoque Bernstein, observamos que las curvas de *contraste individual* y *B & H* tiene un *desempeño comparable*, en el sentido de que presentan un comportamiento similar en las distintas configuraciones del estudio. Lo anterior se observa de manera muy clara en el Cuadro 4.2.



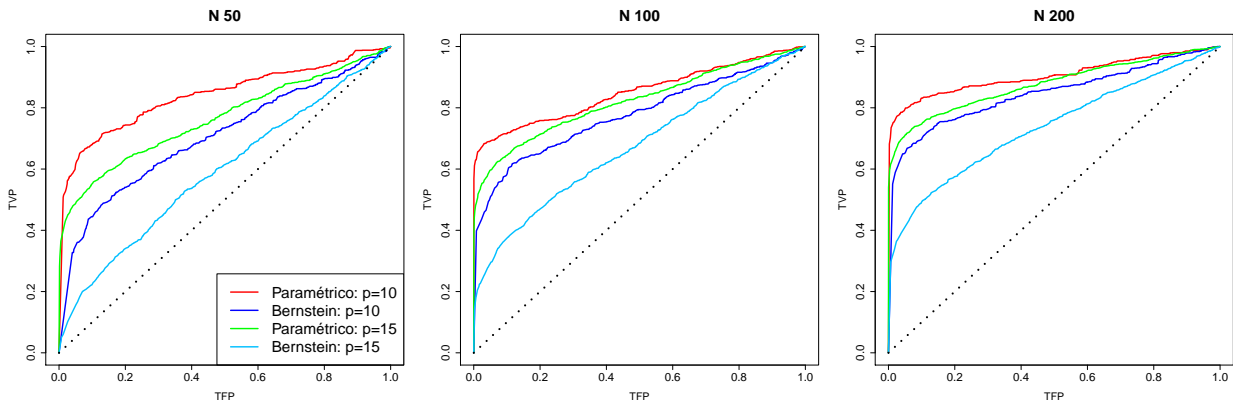


Figura 4.5: Cópula  $\mathcal{T}$  con  $\nu = 4$ : contraste individual vía  $\alpha$ .

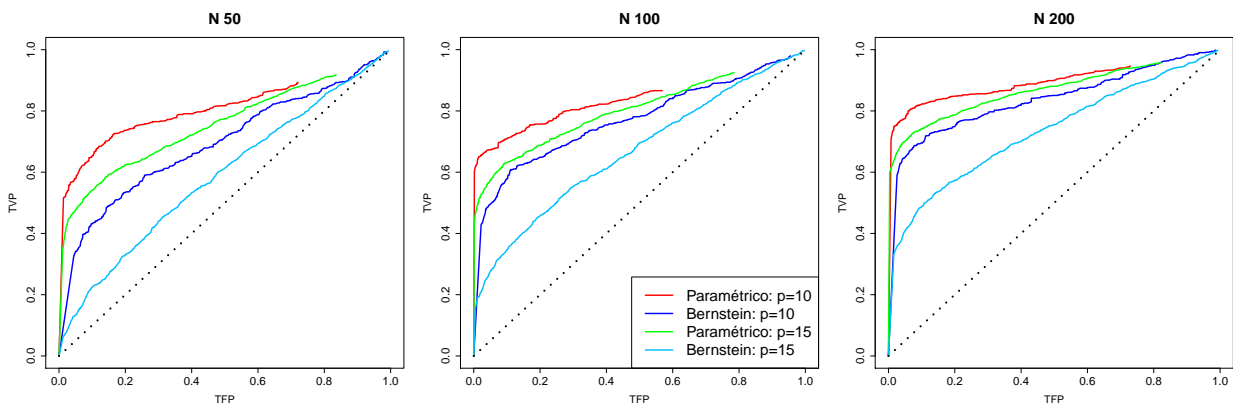


Figura 4.6: Cópula  $\mathcal{T}$  con  $\nu = 4$ : Benjamini & Hochberg *hacia-abajo*.

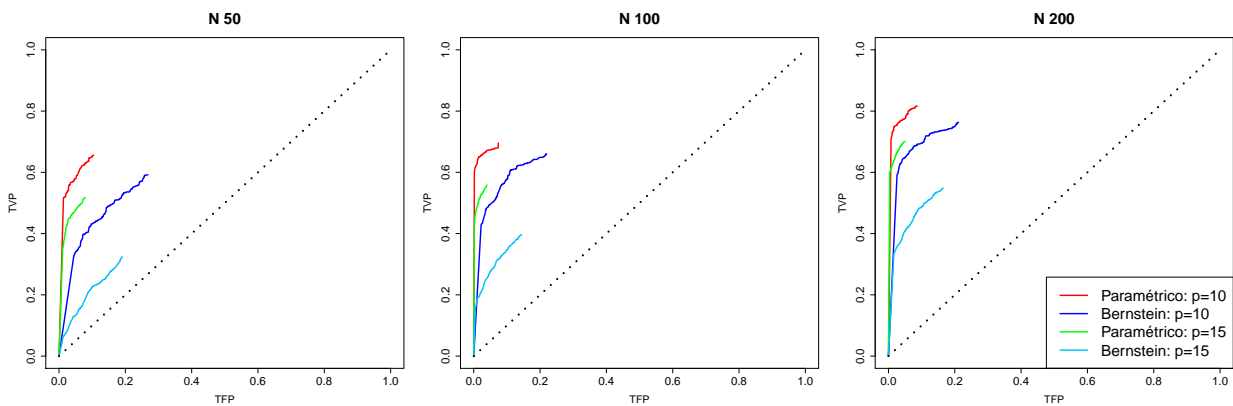


Figura 4.7: Cópula  $\mathcal{T}$  con  $\nu = 4$ : Guo & Rao *hacia-abajo*.

**(ROC 6)** En el enfoque paramétrico, observamos que las curvas  $B$  &  $H$  son *incompletas*. De igual forma, las curvas  $G$  &  $R$  son *incompletas* en ambos enfoques. La justificación de lo anterior es el hecho de la implementación del control en la TDF que estos estimadores incluyen, debido a la estrecha relación que existe entre la TDF y la TFP. Cabe mencionar que las curvas  $G$  &  $R$  son mucho más cortas que las curvas  $B$  &  $H$ .

**(ROC 7)** En el enfoque paramétrico, los rangos de la TFP crecen al aumentar el número de variables en el modelo para las curvas  $B$  &  $H$ , caso contrario, en las curvas  $G$  &  $R$  el rango en la TFP disminuye al aumentar el número de variables en el modelo. En el enfoque Bernstein, el mismo comentarios aplica para las curvas  $G$  &  $R$  pero sólo para los casos de cópula  $\mathcal{T}$ .

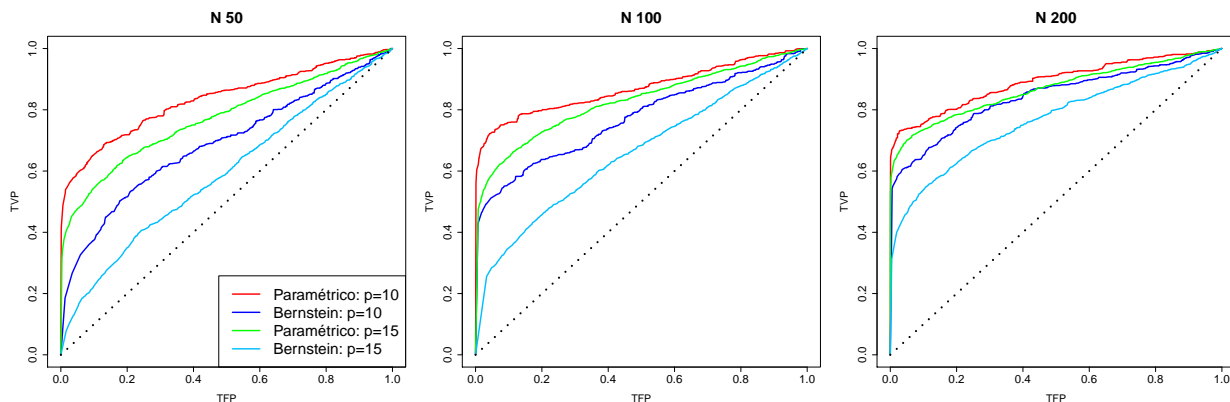


Figura 4.8: Cópula  $\mathcal{T}$  con  $\nu = 8$ : contraste individual vía  $\alpha$ .

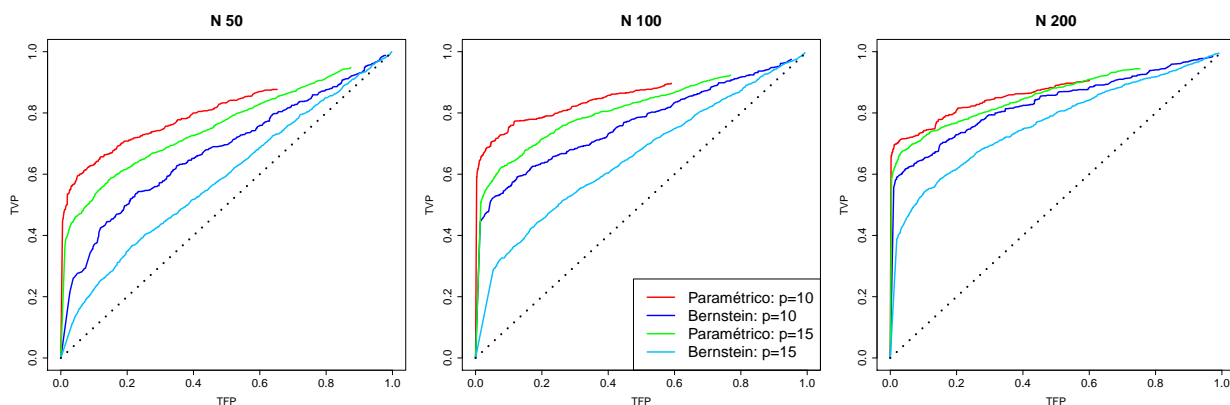


Figura 4.9: Cópula  $\mathcal{T}$  con  $\nu = 8$ : Benjamini & Hochberg *hacia-abajo*.

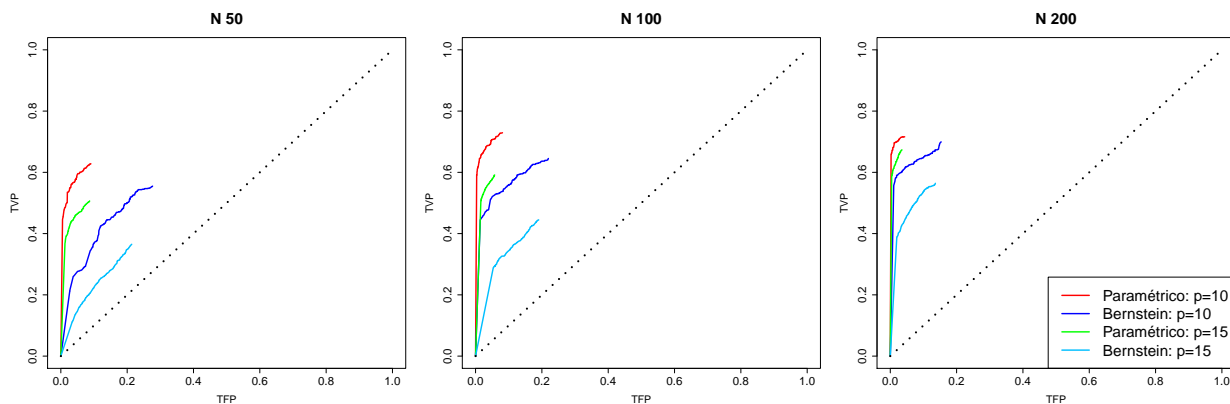


Figura 4.10: Cópula  $\mathcal{T}$  con  $\nu = 8$ : Guo & Rao *hacia-abajo*.

**(ROC 8)** En términos generales, consideramos que las pequeñas diferencias en las áreas bajo las curvas ROC presentadas en el Cuadro 4.2, se asocian a la variabilidad intrínseca entre las diferentes muestras simuladas, *más que con cambios significativos en el desempeño de los estimadores entre los tres modelos de cópula*. Entonces para las configuraciones que propusimos para la cópula  $\mathcal{T}$  de número de variables y grados de libertad, este último parámetro no tuvo mayor impacto en la dependencia condicional de aquellas variables con presencia de correlación parcial nula.

Cópula	N	Estimador	CTDF					
			C. individual		B & H		G & R	
			p=10	p=15	p=10	p=15	p=10	p=15
Gaussiana	50	Paramétrico	0.351	0.280	0.297	0.259	0.043	0.040
		Bernstein	0.212	0.080	0.203	0.111	0.070	0.058
	100	Paramétrico	0.375	0.333	0.295	0.296	0.056	0.022
		Bernstein	0.267	0.165	0.264	0.160	0.098	0.029
	200	Paramétrico	0.399	0.383	0.347	0.342	0.053	0.044
		Bernstein	0.358	0.257	0.350	0.260	0.125	0.076
$\mathcal{T}$ con $\nu = 4$	50	Paramétrico	0.343	0.264	0.290	0.249	0.053	0.029
		Bernstein	0.221	0.091	0.212	0.087	0.098	0.019
	100	Paramétrico	0.349	0.322	0.280	0.292	0.046	0.019
		Bernstein	0.278	0.173	0.269	0.166	0.094	0.033
	200	Paramétrico	0.403	0.376	0.368	0.357	0.061	0.031
		Bernstein	0.346	0.236	0.339	0.233	0.118	0.056
$\mathcal{T}$ con $\nu = 8$	50	Paramétrico	0.335	0.273	0.273	0.255	0.046	0.036
		Bernstein	0.190	0.091	0.196	0.112	0.089	0.049
	100	Paramétrico	0.369	0.328	0.301	0.295	0.051	0.023
		Bernstein	0.273	0.160	0.266	0.151	0.096	0.034
	200	Paramétrico	0.391	0.367	0.313	0.339	0.028	0.021
		Bernstein	0.344	0.268	0.331	0.268	0.084	0.054

Cuadro 4.2: Área bajo las curvas ROC.

**(ROC 9)** Tomando una postura en donde nos interesa controlar *únicamente el error de sobreajuste*. Podemos concluir que *los estimadores G & R son preferibles* si lo que se desea es controlar el sobreajuste vía la TFP, ya que observamos que las curvas ROC asociadas cuentan con un rango en la TFP bastante limitado para cualquier valor  $\alpha \in [0, 1]$ . Esta conclusión era de esperarse por el hecho del control con el que este estimador cuenta.

**(ROC 10)** Tomando una *postura más neutra* en donde nos interese controlar con la misma ponderación el sobreajuste así como el subajuste. Por el hecho de la aparición de curvas incompletas, a partir de las curvas ROC no es totalmente claro qué enfoque de estimación sería deseable. Más adelante analizamos la TEG *para los distintos enfoques de estimación* y buscaremos darle respuesta a esta cuestión.

Para complementar el análisis de las curvas ROC, buscaremos darle interpretación a algunos de los hallazgos.

Como lo comentamos en la Sección 4.5, la estimación paramétrica del grafo sólo depende de la inferencia de los parámetros no condicionales, mientras que en la estimación Bernstein del grafo se involucran  $p - 2$  estimaciones no paramétricas de las cópulas correspondientes (condicionales y no condicionales), como se mencionó en la Sección 4.6. La implicación inmediata de lo anterior *para el enfoque Bernstein es la acumulación de error en cada iteración*, que impacta en las estimaciones de las observaciones de las cópulas condicionales involucradas en el grafo y por ende a la estimación final del mismo. La acumulación del error a la que nos referimos en el enfoque Bernstein es más bien en un sentido numérico, ya que partiendo de (4.6.2) observamos que el estimador Bernstein de la cópula está sujeto al cálculo de potencias para valores entre  $(0, 1)$  los cuales pueden ser muy pequeños y ocasionar imprecisión numérica en las estimaciones, misma que se acumula iteración a iteración. En el caso paramétrico el error de estimación sólo proviene de la primer iteración. Con base en la Observación **(ROC 1)** esta situación se hizo evidente en el estudio de simulación desarrollado.

Con respecto a la Observación **(ROC 2)**, en el enfoque paramétrico los parámetros estimados no condicionales *convergen a los reales* cuando  $N \rightarrow \infty$ . De igual forma, con base en [Ruschendorf \(1976\)](#), *la cópula empírica converge a la cópula verdadera*, entonces el estimador Bernstein de la cópula naturalmente también mejora al incremento en el tamaño de muestra. Por otro lado con base en la prueba de independencia que se implementó (c.f. [Genest and Rémillard \(2004\)](#)), en ambos enfoques de estimación, las estadísticas de prueba asociadas se basan en el proceso de la cópula empírica, por lo tanto el aumento en el tamaño de muestra también favorece la precisión de la prueba.

Sobre la Observación **(ROC 3)**. Como ya lo comentamos para el enfoque Bernstein, el aumento en el número de variables naturalmente induce aumento en el error para las estimaciones de observaciones de cópulas condicionales y por ende aumento en el error de estimación del grafo. Pero más aún para ambos enfoques al incrementar el número de variables, en el fondo estamos aumentando el número de hipótesis a contrastar, así los errores subyacentes de la prueba de independencia se potencializan. Por otro lado, en el enfoque paramétrico, pese a que éste sólo depende de parámetros no condicionales, la inferencia de éstos se puede complicar en el sentido de que la estimación resulte no definida positiva. De hecho en el enfoque paramétrico para muchas de las muestras simuladas en el estudio, sobre todo cuando  $p=15$  o  $N=50$ , la estimación de las matrices de correlación resultaban no positivas definidas<sup>12</sup>.

Finalmente, en los términos de los puntos que se abordan en las Observaciones **(ROC 9)** y **(ROC 10)**, reconocemos que el tipo de error que se deseé controlar depende totalmente del análisis que se lleve a cabo, en otras palabras, depende de la finalidad que tenga el modelo gráfico no dirigido en cuestión. Cabe recordar que en esta investigación propusimos el CTDF como un control que sólo domina un tipo de error. Así que tomando una postura más neutra, como se comenta en la Observación **(ROC 10)**, a continuación analizaremos la TEG de las estimaciones del grafo no dirigido en el estudio de simulación.

### Tasa del error global (TEG)

Técnicamente, la *tasa de error global* (TEG) que analizaremos tiene la interpretación de ser una *tasa promedio*, en el sentido de que se calcula con las observaciones con las cuales se generaron las curvas ROC (promedio) antes presentadas. Las Figuras 4.11, 4.12, 4.13 y 4.14 muestran el comportamiento de la TEG al cambio en  $\alpha$ , el parámetro de ajuste, para los enfoques paramétrico y Bernstein para  $p=10$  y  $p=15$ , sólo para el caso de cópula gaussiana. Decidimos no analizar los casos de cópula  $\mathcal{T}$  debido al comportamiento tan similar de los estimadores del grafo entre las tres cópulas. En cada una de las figuras se muestra el comportamiento de la TEG para cada uno de los tres controles del error (contrate individual, B & H y G & R) y se compara su comportamiento al cambio en el tamaño de muestra. Adicionalmente se resaltan los valores mínimos por cada control: verde, contraste individual; rosa, B & H; turquesa, G & R.

Para los tres controles, la  $TEG = 1$  si  $\alpha = 0$ , la justificación es que si  $\alpha = 0$  ninguna hipótesis es rechazada, entonces  $TFP = 0$  y la  $TFN = 1$ . Por otro lado, para el contraste individual si  $\alpha = 1$  también la  $TEG = 1$ , pero en este caso la justificación es el caso inverso, ya que si  $\alpha = 1$  entonces todas las hipótesis se rechazan, lo que implica que la  $TFP = 1$  y la  $TFN = 0$ , lo anterior no aplica en general para los otros dos controles del error. A continuación presentamos una serie de observaciones con respecto a la TEG observada ante las diferentes estimaciones del grafo:

- (TEG 1)** Como era de esperarse el estimador *Bernstein* presenta *tasas de error global más altas* con respecto al enfoque paramétrico, además sus valores mínimos son considerablemente más grandes en la mayoría de los casos.
- (TEG 2)** Alineados a las Observaciones **(ROC 2)** y **(ROC 3)**, el *aumento en el tamaño de la muestra favorece* notablemente la disminución en la TEG, caso contrario, el *aumento en el número de variables en el modelo aumenta* la TEG.
- (TEG 3)** La TEG para el caso del contraste individual toma su valor mínimo para *valores pequeños de  $\alpha$*  (desde 0.015 y hasta 0.12) y posteriormente *crece de manera lineal* hacia el uno.
- (TEG 4)** Para el contraste B & H, la TEG tiene un comportamiento similar a aquel en el contraste individual con la salvedad de que *cae más lentamente* hacia su valor mínimo (alcanzándolo en valores de  $\alpha$  desde 0.12 y hasta 0.52), y de igual forma, *crece de manera más lenta* después de su menor valor, además para el caso paramétrico después de alcanzar su valor mínimo no logra una  $TEG = 1$  aún para  $\alpha = 1$ .
- (TEG 5)** Observamos que el comportamiento de la TEG al cambio en  $\alpha$  para el control B & H *es cercano al comportamiento* en el contraste individual para el estimador Bernstein, lo anterior está alineado con la Observación **(ROC 5)**.

<sup>12</sup>Por practicidad, cuando una estimación resultaba no positiva definida, la muestra correspondiente se reemplazaba con una nueva simulación. Aunque ya no fueron implementadas, ante la presencia de matrices de correlación estimadas no positivas definidas se exploraron opciones para el cálculo de matrices G-inversas (inversas generalizadas), tal que las correlaciones parciales subyacentes de las mismas mantuvieran las propiedades deseadas, en particular analizamos los resultados en Radhakrishna Rao (1981).

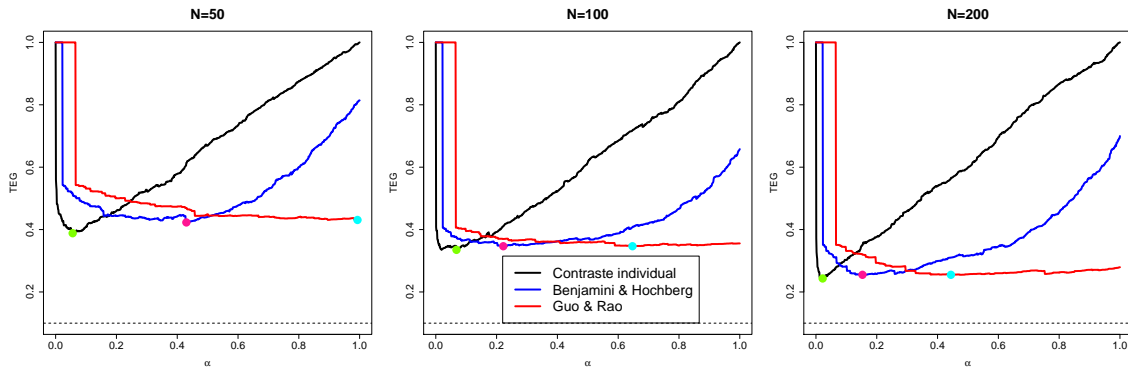


Figura 4.11: Tasa de error global: estimador paramétrico con  $p=10$ .

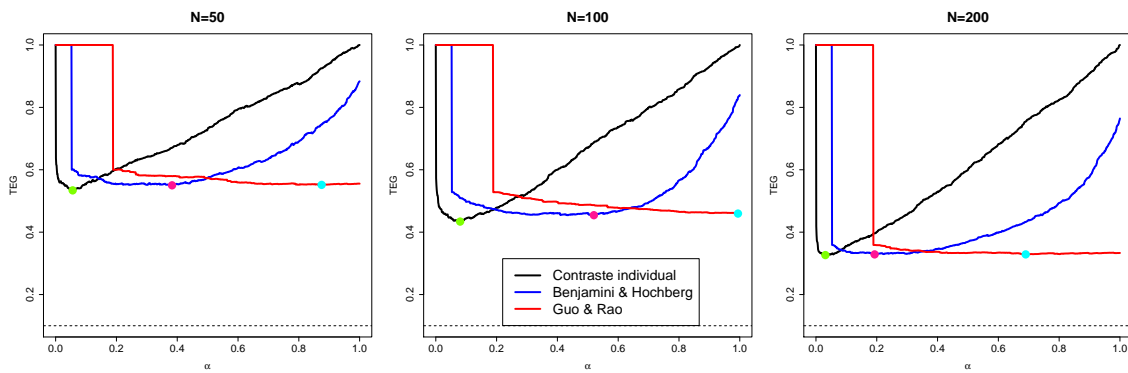


Figura 4.12: Tasa de error global: estimador paramétrico con  $p=15$ .

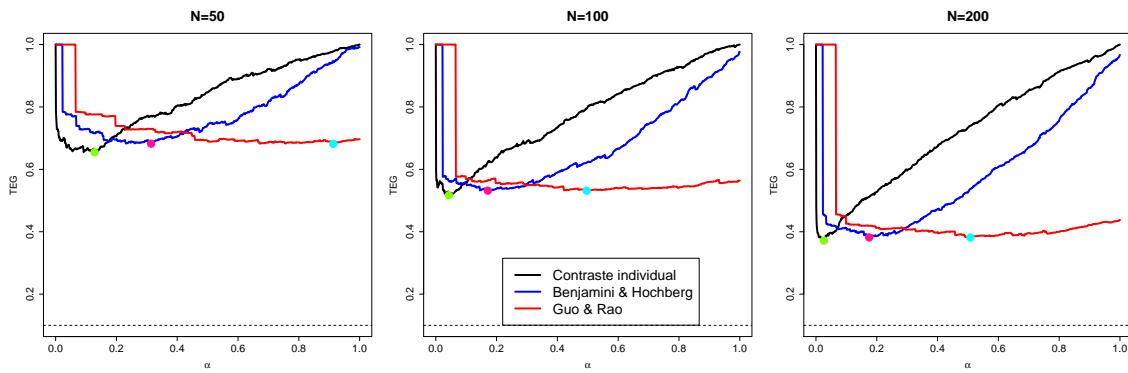


Figura 4.13: Tasa de error global: estimador Bernstein con  $p=10$ .

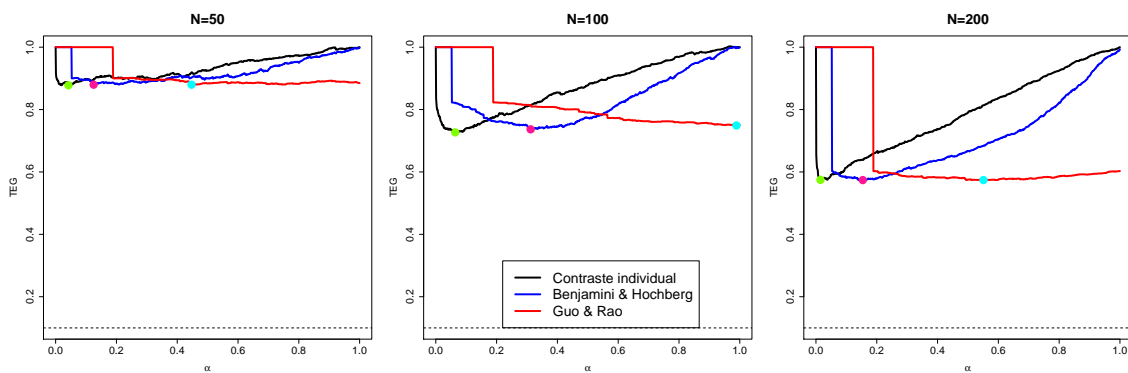


Figura 4.14: Tasa de error global: estimador Bernstein con  $p=15$ .

**(TEG 6)** Para el caso del contraste G & R la caída hacia su valor mínimo es *considerablemente más lenta* con respecto a los otros dos controles (alcanzándolo en valores de  $\alpha$  desde 0.44), de hecho en algunos casos lo alcanza en valores muy cercanos a uno de  $\alpha$  (hasta 0.99), pero posterior a que la TEG logra su menor valor su crecimiento es poco significativo o simplemente permanece en ese valor al incremento en  $\alpha$ . Si bien para valores pequeños de  $\alpha$  la TEG no logra su mínimo valor, en términos generales, con valores de  $\alpha \approx 0.1$  para  $p = 10$  y  $\alpha \approx 0.2$  para  $p = 15$  la TEG ya es cercana a éste. Sin ser tan estrictos en la detección de manera puntual del valor mínimo de la tasa, consideramos que una descripción más realista de su comportamiento es que una vez que la TEG se encuentra cercana a su valor mínimo ( $\alpha \approx 0.1$  para  $p = 10$  y  $\alpha \approx 0.2$  para  $p = 15$ ) el comportamiento de la tasa con respecto al cambio en  $\alpha$  se asemeja a un *decaimiento lineal relativamente lento* al crecer este último, en tal caso su valor mínimo sería logrado en  $\alpha = 1$  o antes y posteriormente permanecería en el mismo valor (en la Figura 4.12 se observa claramente este comportamiento), el hecho de que no se haya observado exactamente el comportamiento antes descrito lo atribuimos a la variabilidad intrínseca de las muestras simuladas en el estudio.

**(TEG 7)** Para la misma configuración de enfoque de estimación, número de variables y tamaño de muestra, los valores mínimos alcanzados por la TEG son muy cercanos para los tres tipos de control del error.

En algunos contextos de análisis, es de interés examinar el grafo resultante,  $\mathcal{G}_\alpha$ , recorriendo distintos valores posibles del parámetro de ajuste,  $\alpha \in [0, 1]$ , lo anterior permite tener un panorama a distintos niveles del estudio en cuestión. En el contexto de la estimación del grafo no dirigido, recorrer el parámetro de ajuste sobre sus posibles valores, resulta en una herramienta útil para comparar las relaciones de independencia condicional a distintos niveles. En tales casos, contar con un enfoque de estimación que asegure una TEG baja para la mayoría de los distintos valores de  $\alpha$  es deseable, entonces la estimación con el control G & R es preferible en estos casos. Por otro lado, tomando una postura donde simplemente se busquen enfoques de estimación donde para un valor fijo de  $\alpha$  se logre la menor TEG posible, los tres tipos de control del error tendrían un alcance indistinto.

## CTDF en el estudio de simulación

Para cerrar la sección de resultados, a continuación comentamos sobre la relación que existe entre la TFP y la TDF, además analizamos los factores que pueden afectar o favorecer el control de ésta última con base en el estudio de simulación que se llevó a cabo.

Partiendo de la definición de la proporción de descubrimientos falsos en (4.4.3),  $Q$ , observamos que el numerador de ésta coincide con el mismo en la TFP, (4.7.6), sin embargo, el denominador en la primera es el número de hipótesis rechazadas, i.e. el número *estimado de relaciones de no independencia condicional*, y por otro lado, el denominador en la TFP es el número *verdadero de relaciones de no independencia condicional*. Debido a la estrecha relación entre la TFP y  $Q$  en la definición de la TDF, (4.4.4), es que el CTDF gobierna de igual forma el sobreajuste del grafo vía la TFP, por lo tanto es de interés comprender en qué medida se satisfacen las cotas teóricas para la TDF, Teorema 4.4.1 y Corolario 4.4.1, ante las distintas configuraciones en el estudio de simulación.

Para cada valor de  $\alpha$ , un enfoque particular (paramétrico o Bernstein) y un CTDF específico, un estimador para la TDF es el promedio de las proporciones de falsos positivos,  $\overline{\text{TDF}} = \overline{Q}$ , sobre las 100 repeticiones que se realizaron en el estudio. Las Figuras 4.15, 4.16, 4.17 y 4.18 muestran el comportamiento de la  $\overline{\text{TDF}}$  al cambio en  $\alpha$  para los estimadores paramétrico y Bernstein con  $p = 10$  y  $p = 15$ , además se compara el comportamiento al cambio en el tamaño de muestra. La intención es evaluar en qué medida la  $\overline{\text{TDF}}$  cumple las cotas teóricas, mismas que se presentan como referencia en las figuras antes enunciadas. Nuevamente, sólo se presentarán los resultados para cópula gaussiana.

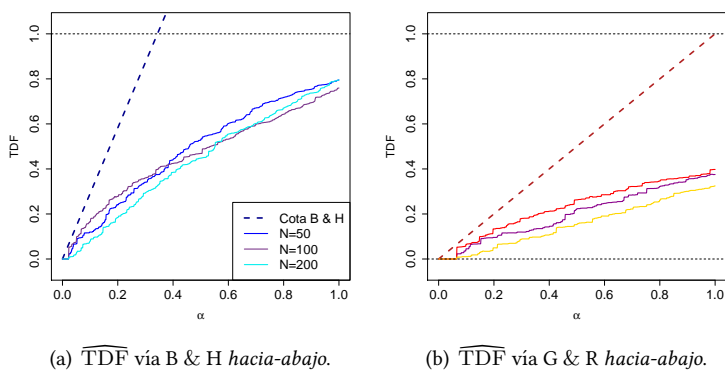


Figura 4.15: Estimador paramétrico con p=10.

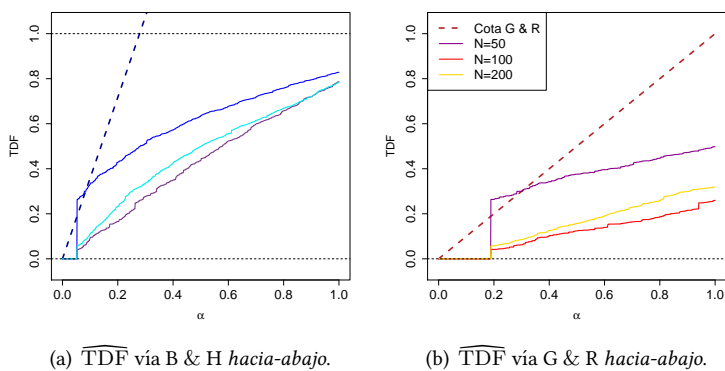
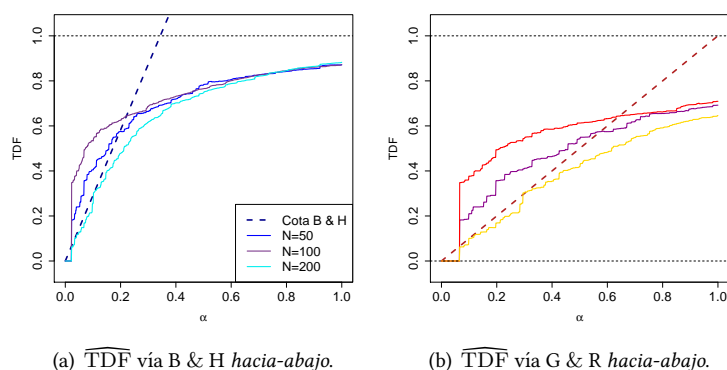
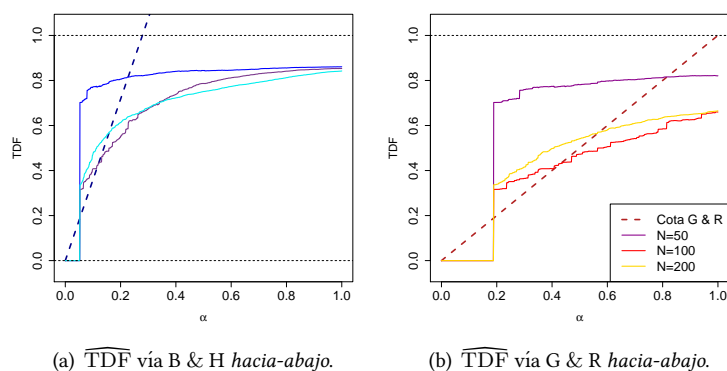


Figura 4.16: Estimador paramétrico con p=15.

Las cotas para cada uno de los dos controles son:

- para el control B & H *hacia-abajo*, con base en (4.4.5) la  $TDF \leq \alpha D(m)$  con  $D(\cdot)$  como en (4.4.6). Si  $p = 10$ , entonces  $m = 45$  y  $D(45) = 2.91$ . Si  $p = 15$ , entonces  $m = 105$  y  $D(105) = 3.58$ . El hecho de que  $D(m) > 1$  implica que para valores de  $\alpha \geq 1/D(m)$  el control se vuelve vano y naturalmente la cota siempre se logra;
- para el control G & R *hacia-abajo*, con base en (4.4.7) la  $TDF \leq \alpha$ . Entonces el control tiene sentido y utilidad  $\forall \alpha \in (0, 1)$ .

En el enfoque paramétrico las cotas se cumplen casi perfectamente, sólo para  $p = 15$  y  $N = 50$  en unos cuantos valores de  $\alpha$  la  $\overline{TDF}$  supera ligeramente la cota para ambos controles. Caso contrario para el estimador Bernstein, la cota no se logra tan fácilmente sobre todo para  $p = 15$ , se observa lo siguiente: para  $N=200$ , con  $p = 10$  las cotas se cumplen perfectamente en ambos enfoques y para  $p = 15$  la cota no se logra para intervalos considerablemente amplios de valores de  $\alpha$ ; para  $N = 100$ , en ambos enfoques con  $p = 10$  se presentaron las tasas estimadas más lejanas de las cotas correspondientes a lo largo de la mayoría de los valores  $\alpha$  donde la cota tiene utilidad, sin embargo para  $p = 15$ , pese a que la cota es superada por intervalos de  $\alpha$  éstos son ligeramente más estrechos que aquellos para  $N = 200$ ; para  $N = 50$ , cuando  $p = 15$  la cota es superada para ambos enfoques en la gran mayoría de los valores de  $\alpha$  donde el control tiene utilidad, sin embargo cuando  $p = 10$ , pese a que la cota es superada por ciertos intervalos de  $\alpha$  éstos son ligeramente más estrechos que aquellos para  $N = 100$ . Observamos un comportamiento inestable para la  $\overline{TDF}$  en la mayoría de los casos con el estimador Bernstein, una posible causa puede ser que necesitemos un mayor número de repeticiones para el cálculo de  $\overline{Q}$  con la finalidad de ganar precisión en el estimador de la TDF.

Figura 4.17: Estimador Bernstein con  $p=10$ .Figura 4.18: Estimador Bernstein con  $p=15$ .

## 4.9. El supuesto simplificador en la estimación del grafo no dirigido

En esta breve sección comentamos sobre las implicaciones que tiene asumir el supuesto simplificador en la estimación del grafo y concluimos sobre el tipo de error de estimación que el supuesto favorece.

La premisa general de la cual parte cualquier prueba de independencia (en particular la prueba en [Genest and Rémillard \(2004\)](#)), es que las observaciones a las cuales se les aplica la prueba provienen de la misma población, en otras palabras para el caso bivariado, si contamos con una muestra,  $\{(x_{1,i}, x_{2,i})\}_{i=1}^N$ , la premisa es que cada una de las observaciones  $(x_{1,i}, x_{2,i})$  proviene del mismo v.a., digamos  $\mathbf{X} = (X_1, X_2)^T$ . Para los efectos del estudio de simulación planteado, podíamos estar seguros que las muestras a la cuales se le aplicó la prueba de independencia eran observaciones de la misma población (de a misma cópula), por el hecho de que eran muestras de cópulas simplificadas, sin embargo en términos más generales, éste es el supuesto más fuerte que subyace de nuestra propuesta de estimación. De hecho no sólo estamos asumiendo el supuesto simplificador al hacer la prueba de independencia, sino en cada una de las iteraciones de la extracción de observaciones de cópulas condicionales vía la fórmula de [Joe \(1996\)](#) presentada en la Sección 4.3. Un proceder ideal sería poder validar el supuesto simplificador, i.e. poder asegurar que en cada una de las iteraciones a partir de la segunda la muestra obtenida proviene de la misma población, sin embargo actualmente hay muy poca literatura sobre cómo probar el supuesto simplificador (c.f. [Killiches et al. \(2015\)](#) como una propuesta bastante reciente para probar el supuesto simplificador). Por otro lado, en [Haff et al. \(2010\)](#) se exhibe el potencial del supuesto simplificador, en el sentido de probar que las *construcciones de cópulas a pares simplificadas son una buena aproximación* aún cuando el modelo verdadero se encuentre lejos del supuesto. En nuestro contexto



de estimación del grafo buscaremos comprender cuál es el tipo de error al que el supuesto simplificador favorece, por el hecho de no verificar que las observaciones a las cuales se le aplica la prueba provengan de la misma población.

Para un par de vértices, digamos  $j, k \in \mathcal{V}$ , partiendo de que contamos con las observaciones (4.3.4) a las cuales se les aplicará la prueba de independencia, éstas pueden o no proceder de la misma población, sin embargo en nuestra propuesta de estimación simplemente asumimos que las observaciones pertenecen a la misma población y procedemos a aplicar la prueba en [Genest and Rémillard \(2004\)](#). Entonces podemos incurrir en las siguientes cuatro situaciones:

	No Rechazamos $\mathcal{H}_{0,jk}$	Rechazamos $\mathcal{H}_{0,jk}$
$\left\{ \left( u_{j \mathcal{V} \setminus \{j,k\}, i}, u_{k \mathcal{V} \setminus \{j,k\}, i} \right) \right\}_{i=1}^N$ es una población única	Escenario (A)	Escenario (B)
$\left\{ \left( u_{j \mathcal{V} \setminus \{j,k\}, i}, u_{k \mathcal{V} \setminus \{j,k\}, i} \right) \right\}_{i=1}^N$ son distintas poblaciones	Escenario (C)	Escenario (D)

Cuadro 4.3: Escenarios en los que se puede incurrir ante el supuesto simplificador (para una prueba particular).

Para los escenarios (A) y (B), los errores tipo I y II serán simplemente los correspondientes a la prueba particular, ya que estamos asegurando la premisa de la que parte la prueba de independencia. Sin embargo, ante la presencia de multiplicidad de poblaciones la prueba de independencia debería ser rechazada automáticamente, en otras palabras, si existiera certeza de dicha multiplicidad simplemente se estaría perdiendo la premisa fundamental de la prueba y ésta perdería sentido, así el escenario (D) sería la acción correcta en este caso, por lo tanto el escenario (C) estaría potenciando el error tipo II para una prueba individual (no rechazar  $\mathcal{H}_{0,jk}$  cuando ésta es falsa). Considerando las múltiples pruebas involucradas en la estimación del grafo, podemos concluir que *el supuesto simplificador potencializa el subajuste* de la misma. Analizar la sensibilidad de incurrir o no en el escenario (C) de la prueba de independencia en [Genest and Rémillard \(2004\)](#) (u otra que se proponga) está fuera de los alcances de la presente investigación, sin embargo éste puede ser un problema a estudiar en el futuro. Por el momento nuestra intención era reconocer el tipo de error que se está favoreciendo al asumir el supuesto simplificador y dejar el cuestionamiento abierto sobre qué tipo de control podríamos implementar para evitar incurrir en el escenario (C) en cada una de las pruebas involucradas.

## 4.10. Discusión

A lo largo del capítulo se estudiaron *diversas modificaciones de nuestra propuesta de estimación basada en cópulas condicionales asumiendo el supuesto simplificador*: Algoritmo 4.3.1. En particular se presenta el *enfoque paramétrico* de cópula multivariada simplificada y el *estimador Bernstein* del grafo no dirigido; ésta última como una propuesta totalmente no paramétrica y la principal contribución de esta investigación. La naturaleza de nuestra propuesta basada en *pruebas múltiples* de independencia condicional nos llevó a estudiar los tipos de error en los que se puede incurrir al estimar el grafo, para posteriormente proponer un *control en la tasa de descubrimientos falsos* (CTDF) el cual gobierna el sobreajuste del grafo; en particular partimos de los resultados en [Guo and Rao \(2008\)](#) para pruebas dependientes. También discutimos sobre el hecho de que el tipo de error (sobreajuste o subajuste) que se busque controlar para la estimación del grafo no dirigido depende totalmente del objetivo del análisis, así el CTDF es sólo una propuesta para dominar un tipo de error.

Para evaluar el desempeño de nuestra propuesta de estimación, sobre todo el comportamiento del estimador Bernstein, propusimos un estudio de simulación en el cual se buscó inferir la estructura del grafo no dirigido para datos sintéticos provenientes de cópulas gaussiana o  $\mathcal{T}$  con presencia de escasez, i.e. con presencia de relaciones de independencia condicional o no correlación parcial. El gran resultado del estudio es que *el desempeño del estimador paramétrico supera considerablemente al estimador Bernstein*. Otras de las conclusiones más relevantes se resumen a continuación: el aumento en el número de variables afecta considerablemente la estimación del grafo, sobre todo en el enfoque Bernstein; por otro lado, el aumento en el tamaño de muestra favorece la estimación en ambos enfoques; el potencial de los estimadores es indistinto entre los modelos de cópulas gaussiana y  $\mathcal{T}$ , i.e. no se observa que el parámetro de grados de libertad implique un impacto considerable sobre la dependencia condicional para variables

con presencia de correlación parcial nula; en general observamos que los CTDF propuestos disminuyen la TFP, i.e. previenen el sobreajuste; las tasas de error global (TEG) se presentaron más altas para el estimador Bernstein con respecto al paramétrico; pese a que los valores mínimos para la TEG entre los tres tipos de control del error (contraste individual, B & H y G & R) eran similares, el contraste vía G & R aseguraba TEG's bajas para la mayoría de los distintos valores de  $\alpha$ , el parámetro de ajuste, lo cual puede ser deseable en diversos contextos de análisis; finalmente, observamos que el cumplimiento de las cotas teóricas para la TDF se vio fuertemente afectado en el enfoque Bernstein. Como complemento a nuestra propuesta de estimación y a los resultados observados en el estudio de simulación, analizamos el impacto que tiene el supuesto simplificador en la acumulación de error de estimación, de donde concluimos que éste *favorece el error tipo II* para cada prueba individual.

Pese a que consideramos que el estimador Bernstein del grafo no dirigido, propuesto en la Sección 4.6, es una vía pertinente para lograr mayor flexibilidad en el modelo de probabilidad subyacente,  $\mathbf{P}$ , dada la naturaleza de su construcción, reconocemos que las observaciones estimadas de las cópulas condicionales a las cuales se les aplica la prueba de independencia traen consigo un *error de estimación acumulado iteración a iteración*. Aún cuando con base en los resultados del estudio de simulación pudimos concluir sobre algunas de las características del estimador Bernstein del grafo, consideramos que como trabajo futuro se pueden analizar sus propiedades teóricas; donde estas últimas estarán íntimamente ligadas con las propiedades de las estimaciones Bernstein (iterativas) de las observaciones de aquellas cópulas condicionales involucradas en el grafo no dirigido. Conociendo teóricamente el comportamiento del estimador Bernstein del grafo, se podrían *proponer ajustes al mismo a manera de controlar el error que se acumula iterativamente* en la observaciones de las cópulas condicionales.

Adicionalmente, es importante destacar que los estimadores Bernstein de la cópula son *incapaces de modelar dependencia en colas* (c.f. [Sancetta and Satchell \(2004\)](#)). Así que aún cuando el estimador Bernstein de la cópula puede lograr modelos sumamente flexibles cuando  $p = 2$ , siguen estando limitados para algunas aplicaciones. Sin embargo, aunque no es una lista muy extensa, existen en la literatura diversas *propuestas de modelos no paramétricos* de cópulas aplicables con cantidades moderadas de variables (c.f. [Joe \(2014\)](#), sección 5.10.3). Así que de igual manera como líneas de investigación futura, se puede explorar nuestra propuesta de inferencia del grafo no dirigido (Algoritmo 4.3.1) con otros modelos no paramétricos de estimación de cópulas bivariadas, estudiar el comportamiento del estimador del grafo asociado y la flexibilidad que puede tener el modelo de probabilidad subyacente, en particular la dependencia en colas; v.g. las estimaciones de la densidad cópula vía B-splines en [Shen et al. \(2008\)](#) y [Kauermann et al. \(2013\)](#) han demostrado ser mejores opciones que los estimadores Bernstein y los enfoques kernel, así que una alternativa atractiva de investigación posterior es ocupar nuestra propuesta de inferencia del grafo pero cambiando las estimaciones Bernstein de las cópulas bivariadas por estimadores B-splines.

Por otro lado, pese a que no se mencionó explícitamente en el cuerpo de la Sección 4.5, el estimador paramétrico así como se presentó *está limitado al caso de bajas dimensiones*, cuando el tamaño de la muestra es mayor a número de variables en el modelo ( $N > p$ ), debido a que su deducción parte de la inferencia de una matriz de correlaciones, la cual *no necesariamente cumple con propiedades deseables y necesarias en el contexto de altas dimensiones* ( $N \ll p$ ), i.e. cuando el tamaño de la muestra es mucho menor al número de variables. Sin embargo, como objetivo adicional en nuestra investigación nos planteamos estudiar las alternativas existentes para la inferencia del grafo no dirigido en altas dimensiones. Entonces el Capítulo 5 es el responsable de dar un *panorama general* sobre la inferencia del modelo gráfico no dirigido en altas dimensiones, además en este contexto de estimación, se comentará sobre posibles vías de investigación futura para la inferencia de grafos flexibles en términos de la estructura de dependencia subyacente.

## Capítulo 5

# Modelos gráficos no dirigidos en altas dimensiones: perspectivas de análisis

**Resumen** Técnicamente el concepto de *altas dimensiones* se refiere a la inferencia estadística cuando el número de parámetros desconocidos,  $R$ , es de *uno o varios ordenes de magnitud* mayor al tamaño de muestra,  $N$ , i.e.  $N \ll R$ . Actualmente los *datos altamente dimensionales* forman parte del conjunto de los denominados por muchos autores *datos complejos*<sup>1</sup>, los cuales en su mayoría son consecuencia de la era de la información propia de la modernidad. Hemos decidido motivar el estudio de la estadística en altas dimensiones citando algunos autores relevantes en el área:

The coming century is surely the century of data. A combination of blind faith and serious purpose makes our society invest massively in the collection and processing of data of all kinds, on scales unimaginable until recently. [. . .] We are seeing examples where the observations gathered on individual instances are curves, or spectra, or images, or even movies, so that a single observation has dimensions in the thousands or billions, while there are only tens or hundreds of instances available for study. Classical methods are simply not designed to cope with this kind of explosive growth of dimensionality[. . .]

[Donoho et al. \(2000\)](#)

High-dimensional data are nowadays rule rather than exception in areas like information technology, bioinformatics or astronomy, to name just a few.

[Bühlmann and Van De Geer \(2011\)](#)

Como lo comentamos a lo largo del Capítulo 1, los modelos gráficos no dirigidos son una herramienta que ha cobrado bastante atención en años recientes ante contextos donde se cuenta con *cantidades masivas de variables* y en particular en el marco de *altas dimensiones*, debido a su atractivo visual para resumir interrelaciones entre grandes cantidades de variables, además de las aplicaciones que tienen hacia otros modelos estadísticos como lo presentamos en la Sección 1.4. Algunas áreas a destacar donde se ha vuelto imprescindible el manejo de datos en altas dimensiones son: tecnología de la información, con portales de Internet; ciencias genómicas, con microarreglos de ADN; medicina, con imágenes de resonancias magnéticas; finanzas, con los llamados datos *tick by tick* (en tiempo real). Sin embargo, como se comenta en [Donoho et al. \(2000\)](#), las metodologías clásicas simplemente no contemplan esta estructura de datos. Actualmente, la premisa más recurrente de la cual se suele partir en las metodologías modernas aplicables a datos altamente dimensionales es asumir el *principio de escasez*, mismo que ha resultado bastante útil para el desarrollo de nuevas propuestas metodológicas, además de ser un supuesto bastante realista ante este contexto de análisis. Por otro lado, pese a la gran complejidad que acarrea trabajar con datos en altas dimensiones, en términos de los diversos patrones de dependencia que se puedan encontrar en los datos, los modelos multivariados en su mayoría siguen siendo los clásicos pero ahora con la incorporación de un *parámetro de penalización que regula el grado de*

---

<sup>1</sup>Complex data, en inglés.

*escasez de la inferencia*, usualmente denotado por  $\lambda$  y también conocido como *parámetro de ajuste* (tuning parameter). De algún modo, el parámetro de penalización *restringe* los modelos hacia cierta clase de éstos en donde la inferencia tenga sentido. Así es que las metodologías más reconocidas para la estimación del grafo en altas dimensiones se fundamentan en el supuesto de *distribución gaussiana* y la inferencia penalizada de la matriz de precisión,  $\Omega$ , con base en el conocido resultado en Dempster (1972), (1.7.1). Reconociendo las grandes limitantes de un modelo de distribución gaussiana, es que en años recientes se han propuesto extensiones al modelo de *cópula gaussiana*, Liu et al. (2009) y Liu et al. (2012a), o más general al modelo de *cópula elíptica*, Liu et al. (2012b). Pero ya ha sido tarea del Capítulo 3 estudiar a fondo lo que implica asumir este tipo de modelos conjuntos y comentar sobre los alcances de estas propuestas, de donde se destaca el hecho rotundo de que son modelos que se vuelven cada vez menos realistas al incremento en el número de variables junto con la *falta de utilidad* que tiene el grafo no dirigido fuera de los modelos de *cópula gaussiana* y  $\mathcal{T}$ . De igual forma, ya fue labor del Capítulo 4 buscar la inferencia del grafo no dirigido cuyo modelo de probabilidad subyacente *no indujese* a priori una estructura de dependencia particular, sin embargo, *no resultó una tarea sencilla* hacer inferencia sobre *todas y cada una de las relaciones de independencia condicional* modeladas en un grafo no dirigido en un enfoque totalmente no paramétrico. Así este capítulo cuenta con dos grandes objetivos puntuales: el primero, es discutir sobre las ventajas, alcances y limitaciones que un modelo gráfico no dirigido tiene actualmente ante esta estructura moderna de datos multivariados; para posteriormente, proponer líneas de investigación futura para la estimación y sobre todo la aplicación del grafo no dirigido en altas dimensiones cuyo modelo de probabilidad subyacente sea flexible, lo anterior en pro de buscar modelos que cumplan con el principio de parsimonia, el cual relacionaremos con el principio de escasez, vía relaciones de independencia condicional.

## 5.1. Organización del capítulo

Con este último capítulo se concluye nuestra investigación. Además de plantear la problemática general que subyace de la inferencia del grafo no dirigido en altas dimensiones y comentar sobre los enfoques más destacados de estimación, alineados al objetivo primordial que ha tenido hasta ahora nuestra investigación de buscar la inferencia del grafo no dirigido para modelos de probabilidad flexibles, se proponen líneas de investigación futura para su inferencia y aplicación. La Sección 5.2 expone la problemática general de la inferencia estadística en altas dimensiones. La Sección 5.3 comenta sobre algunos de los enfoques más destacados en la literatura para la inferencia del grafo no dirigido en altas dimensiones. En la Sección 5.4 se plantean extensiones del principio de escasez para la inferencia de modelos que cumplan con el principio de parsimonia y con base en éstas se comenta sobre posibles líneas de investigación futura para la inferencia y aplicación del grafo no dirigido en altas dimensiones. Finalmente, la Sección 5.5

## 5.2. La problemática en altas dimensiones

Dentro de las situaciones que engloba la estadística en altas dimensiones encontramos problemas tanto en el ámbito de modelos supervisados, v.g. modelos de regresión o clasificación donde el número de covariables sea mayor que el tamaño de muestra, así como para aquellos no supervisados, v.g. análisis de segmentación<sup>2</sup> o *modelos gráficos* cuando el número de variables,  $p$ , supere el número de observaciones,  $N$ , i.e.  $N \ll p$ . Notemos que para el problema de estimar el grafo no dirigido estamos asumiendo de antemano que por cada variable en el modelo al menos habrá un parámetro desconocido a estimar, así por practicidad a lo largo del documento nos hemos referido al contexto de altas dimensiones como  $N \ll p$ , en contraste con  $N \ll R$  (donde  $R$  es el número de parámetros desconocidos a ser estimados en los modelos). Ante el contexto de inferencia en altas dimensiones una vía con bastante potencial, tanto conceptual como de aplicación, es *asumir el principio de escasez*.

Como ya lo hemos comentado antes, el concepto de escasez es bastante general y puede cambiar de interpretación puntual según el contexto de análisis. Sin embargo, a continuación proponemos una definición del principio de escasez en términos bastante abstractos y genéricos, pero que serán la base para explicar la problemática de la inferencia en altas dimensiones.

---

<sup>2</sup>Cluster analysis.

**Definición 5.2.1 Principio de escasez<sup>3</sup>.** Ante la presencia de cientos o miles de variables, sobre todo en el contexto de altas dimensiones, el principio de escasez asume que *sólo un pequeño número de predictores contribuyen a una variable de respuesta* o que el *impacto de cada uno de éstos se encuentra contraído (o disminuido)*, lo anterior con la finalidad de *evitar (a medida de lo posible) el sobreajuste* del modelo.

Ante el contexto de inferencia que plantea la Definición 5.2.1, asumir el principio de escasez hace más sentido que postular que todas y cada una de las variables tendrán un impacto directo o significativo sobre otra en particular.

Comencemos la motivación al cambio de paradigma en altas dimensiones con el clásico modelo de regresión lineal. Partiendo de un v.a.  $p$ -dimensional,  $\mathbf{X} = (X_1, \dots, X_p)^T$ , y sin pérdida de generalidad asumiremos a  $X_1$  como variable de respuesta, tenemos el típico modelo de regresión lineal como en (5.2.1),

$$X_1 = \sum_{s \in \mathcal{V} \setminus \{1\}} \beta_s X_s + \epsilon, \quad (5.2.1)$$

donde  $\epsilon$  es una variable aleatoria que representa el error, partiendo del hecho de que el conocimiento de  $\mathbf{X}_{-1}$  no implicará que podamos conocer exactamente a  $X_1$  bajo el modelo lineal. Además  $\epsilon$  se asume independiente de  $\mathbf{X}_{-1}$  y t.q.  $\mathbb{E}[\epsilon] = 0$ . Usualmente, por simplicidad y sin pérdida de generalidad se asume que el intercepto es cero y que todas las covariables están centradas y medidas en la misma escala, estos supuestos se alcanzan aproximadamente estandarizando los datos previo al análisis. Por otro lado, denotaremos con  $\beta = (\beta_2, \dots, \beta_p)^T$ , entonces existen  $(p-1)$  parámetros desconocidos en el modelo. Aquí la notación  $\mathbf{X}_{-1}$  está alineada con aquella en la Proposición 4.2.1, i.e.  $\mathbf{X}_{-1}$  es el v.a. con las variables en  $\mathbf{X}$  pero eliminando  $X_1$ .

Supongamos que contamos con una muestra observada,  $\{(x_{1i}, \dots, x_{pi})\}_{i=1}^N$ , tal que  $N < (p-1)$ . Ante este contexto de altas dimensiones, la estimación del vector  $\beta$  vía *mínimos cuadrados ordinarios* es un problema *mal condicionado* en el sentido de que el *estimador no es único y éste sobreajusta fuertemente* los datos. Partiendo de la estructura de la muestra observada y con base en la Definición 5.2.1 es en cierto modo claro, al menos intuitivamente, que el vector  $\beta$  puede ser estimado (o aproximado) *razonablemente bien* asumiendo el principio de escasez, i.e. que el vector  $\beta$  es escaso bajo alguna definición.

En el modelo de regresión lineal, bajo la idea de cuantificar el grado de escasez de la estimación, la vía usual es hacer uso de *normas*  $\ell_q$ , para  $1 \leq q \leq \infty$ . Una norma  $\ell_q$  para  $\beta$  se define como sigue,

$$\|\beta\|_q := \left( \sum_{i=2}^p |\beta_i|^q \right)^{\frac{1}{q}}, \quad 1 \leq q < \infty,$$

además,

$$\|\beta\|_\infty := \lim_{q \rightarrow \infty} \|\beta\|_q = \max_{i=2, \dots, p} |\beta_i|.$$

En lo subsecuente, de igual forma, ocuparemos la notación,

$$\|\beta\|_q^q := \sum_{i=2}^p |\beta_i|^q, \quad 1 \leq q < \infty.$$

Entonces el modelo de regresión en (5.2.1) *penalizado* por la norma  $\ell_q$  es aquel que restringe la estimación,  $\widehat{\beta}_q(\lambda)$ , a valores tales que  $\|\beta\|_q \leq S_\lambda$ , donde existe una relación uno a uno entre  $\lambda$  y  $S_\lambda$ . Así  $\lambda$  es el *parámetro de ajuste que modula el grado de escasez de la estimación* vía  $S_\lambda$ , en el sentido de que *contrae (o disminuye)* los elementos del estimador por mínimos cuadrados ordinarios. La definición de la estimación penalizada es la siguiente,

<sup>3</sup>Sparsity principle, en inglés.

$$\widehat{\beta}_q(\lambda) = \underset{\beta}{\operatorname{argmín}} \left( \frac{\|\mathbf{x}_1 - \mathbf{x}_{-1}\beta\|_2^2}{n} + \lambda \|\beta\|_q^q \right), \quad (5.2.2)$$

donde a partir de la muestra observada,  $\mathbf{x}_1 = (x_{11}, \dots, x_{1N})^T$ , un vector  $N \times 1$ , y  $\mathbf{x}_{-1}$  es la matriz de diseño de dimensiones  $N \times (p-1)$ , i.e.  $(\mathbf{x}_{-1})_{ij} = x_{ji}$ ; equivalentemente

$$\widehat{\beta}_q(\lambda) = \underset{\beta: \|\beta\|_q \leq S_\lambda}{\operatorname{argmín}} \left\{ \frac{\|\mathbf{x}_1 - \mathbf{x}_{-1}\beta\|_2^2}{n} \right\}. \quad (5.2.3)$$

Notemos que la definición de  $\widehat{\beta}_q(\lambda)$ , (5.2.3), no impone modelo de distribución alguno para los errores o las variables de análisis en (5.2.1). En ese sentido el cálculo del estimador  $\widehat{\beta}_q(\lambda)$  es un problema de optimización de la verosimilitud penalizada, el cual puede ser formulado para una gran gama de modelos, v.g. un supuesto recurrente para la construcción de modelos de regresión como en (5.2.1) es que los elementos en la muestra son independientes entre sí sin embargo la extensión penalizada por escasez hacia procesos estacionarios es posible.

Los dos estimadores penalizados más reconocidos y estudiados son: RIDGE,  $\widehat{\beta}_2(\lambda)$ , y LASSO<sup>4</sup>,  $\widehat{\beta}_1(\lambda)$ . Notemos que RIDGE, [Hoerl and Kennard \(1970\)](#), es el estimador subyacente de la penalización  $\ell_2$ , mientras LASSO, [Tibshirani \(1996\)](#), es aquel de la norma  $\ell_1$ . La gráfica clásica donde se representa cómo es que los coeficientes de regresión son contraídos en los estimadores RIDGE y LASSO es aquella en [Tibshirani \(1996\)](#), Figura 2, donde se ilustra el comportamiento de los contornos de la verosimilitud involucrada en (5.2.3) junto con los espacios de restricción que induce cada penalización. Una observación importante a destacar es que las penalizaciones  $\ell_1$  y  $\ell_2$  gozan de *factibilidad computacional* al inducir problemas de optimización convexa.

Ante la presencia de *multicolinealidad*, la cual surge de manera casi natural dada la presencia de una gran cantidad de predictores, el estimador por mínimos cuadrado ordinarios *pierde precisión de predicción* por el hecho de que éste presenta una *varianza considerablemente alta* pese a que goza de *insesgadez*. Una vía recurrente que usualmente mejora la precisión de predicción es *contraer o eliminar algunos coeficientes*. Motivados por la problemática antes planteada es que en [Hoerl and Kennard \(1970\)](#) proponen el estimador RIDGE, el cual *sacrifica el sesgo* del estimador con la finalidad de disminuir la varianza y *ganar precisión de predicción*, sin embargo el estimador *RIDGE en la mayoría de los casos carece de interpretación* en el sentido de que todos los coeficientes estimados en  $\widehat{\beta}_2(\lambda)$  son distintos de cero. Por otro lado, motivado por la falta de interpretación en la que puede incurrir el estimador RIDGE es que [Tibshirani \(1996\)](#) propone el estimador LASSO, el cual induce una *selección de variables* en el sentido de que contrae algunos coeficientes y para otros la estimación es *exactamente cero* ( $\widehat{\beta}_{2,j}(\lambda) = 0$ , para algunas  $j$ 's), dependiendo del grado de escasez que induzca  $\lambda$ <sup>5</sup> se espera que LASSO incluya en el modelo aquellos predictores *suficientemente relevantes* para explicar de la manera más precisa a la variable de respuesta. Además recordemos las equivalencias en (1.7.5), donde se asocia a los coeficientes nulos de regresión con las relaciones de independencia condicional modeladas en el grafo no dirigido para el caso de errores gaussianos, o de igual forma, distribución gaussiana del v.a.  $\mathbf{X}$ . Es así que la peculiaridad de LASSO de estimar coeficientes nulos es la razón por la que este estimador, y en general la penalización  $\ell_1$ , han sido un parteaguas para la estimación del grafo no dirigido en altas dimensiones (c.f. [Meinshausen and Bühlmann \(2006\)](#) y [Friedman et al. \(2008\)](#) como dos de las referencias más relevantes sobre la implementación de LASSO y la normal  $\ell_1$  para estimar el modelo gráfico no dirigido en altas dimensiones).

### 5.3. Algunos enfoques destacados de estimación

Actualmente existen dos grandes referencia que compilan los distintos enfoques de estimación y aplicación del modelo gráfico no dirigido en altas dimensiones: [Bühlmann and Van De Geer \(2011\)](#) (Capítulo 13) y [Pourahmadi \(2013\)](#) (Capítulo 5). En particular [Pourahmadi \(2013\)](#) está enfocado en una de las grandes aplicaciones de los modelos gráficos no dirigidos en el caso de distribución gaussiana: la estimación de la matriz de covarianzas; además con base en [Liu et al. \(2009\)](#) y [Liu et al. \(2012a\)](#) todas las metodologías son aplicables a cópula gaussiana, véase Definición

<sup>4</sup>Least Absolute Shrinkage and Selection Operator.

<sup>5</sup>En general el parámetro de ajuste,  $\lambda$ , se elige mediante validación cruzada buscando modelos con buena precisión de predicción (o alguna otra medida de riesgo que se defina). Lo anterior tanto para regresión LASSO así como para RIDGE.

3.3.1 (NonParaNormal). Debido a la existencia de estas dos referencias que brindan un panorama bastante amplio sobre el contexto actual de los modelos gráficos no dirigidos en altas dimensiones, es que la presente sección tiene la intención de simplemente citar algunos autores destacados y comentar sobre propuestas de estimación que se relacionan con tópicos estudiados en esta investigación.

En contraste con las discusiones que hemos desarrollado en los capítulos anteriores, en donde nuestro motor ha sido la inferencia del grafo no dirigido sin imponer un modelo que no se capaz de representar estructuras complejas de dependencia que de antemano están presentes en datos multivariados, a grandes rasgos la mayoría de los enfoques de estimación del grafo no dirigido en altas dimensiones se basan en la *detección de elementos nulos de la matriz de precisión*,  $\Omega$ , partiendo de (1.7.1) o en un caso más general de (3.3.3), por tanto son enfoques paramétricos limitados a aquellos modelos cuya estructura de dependencia condicional esté representada por una matriz de correlaciones lineales y que además cumplan el supuesto simplificador, i.e. cópula gaussiana y  $\mathcal{T}$ .

### Modelos gráficos escasos de cópula elíptica: gaussiana y $\mathcal{T}$

Detectamos que en la literatura existen tres grandes vías para la estimación del grafo no dirigido en altas dimensiones alineadas básicamente al supuesto de distribución gaussiana: la estimación penalizada de  $\Omega$ , la matriz de precisión; vía regresión LASSO directamente; bajo el supuesto de fidelidad<sup>6</sup>. A continuación citamos algunas de las referencias más destacadas en cada enfoque, ya sea por ser pioneros en la propuesta, por ser planteamientos con eficiencia computacional o por ser compendios generales.

**Estimación penalizada de  $\Omega$ :** [Friedman et al. \(2008\)](#), [Yuan \(2010\)](#) y [Cai et al. \(2011\)](#).

**Regresión LASSO:** [Meinshausen and Bühlmann \(2006\)](#).

**Supuesto de fidelidad:** [Bühlmann and Van De Geer \(2011\)](#), sección 13.9.

Pese a que la mayoría de las metodologías originales parten del supuesto de distribución gaussiana, al menos implícitamente al buscar estimar la matriz de precisión, las metodologías son extendibles hacia cópula gaussiana vía [Liu et al. \(2009\)](#) y [Liu et al. \(2012a\)](#), y hacia cópula  $\mathcal{T}$  vía [Liu et al. \(2012b\)](#).

Como lo comentamos en la Sección 3.3, la propuesta en [Liu et al. \(2012b\)](#) busca extender las metodologías hacia cópula elíptica en general, sin embargo sabemos que éstas sólo tendrían sentido para cópulas gaussiana y  $\mathcal{T}$ . Para el resto de las cópulas elípticas la extensión no cumpliría con la promesa inicial de un grafo no dirigido de modelar relaciones de independencia condicional, éste sólo representaría relaciones de no correlación parcial nula sin tener una interpretación clara hacia la dependencia modelada por estas cópulas ya que no cumplen el supuesto simplificador.

Fuera de los modelos de cópula elíptica, encontramos una propuesta interesante basada en el modelo de factores: [Fan et al. \(2015\)](#). El área de oportunidad de esta metodología es la asunción de relaciones lineales entre las variables.

## 5.4. Escasez vía relaciones de independencia condicional subyacentes del grafo no dirigido

Con esta sección damos por concluida nuestra investigación, en ésta explicamos nuestra perspectiva del principio de escasez la cual busca tener aplicaciones, más que para la inferencia penalizada de modelos lineales de regresión, para la construcción de *distribuciones condicionales siguiendo el principio de parsimonia*, en el sentido de explicar el comportamiento aleatorio de la variable de respuesta en función del *menor número de predictores evitando redundancia de información*. La búsqueda de parsimonia la propondremos vía las relaciones de independencia condicional modeladas en el grafo no dirigido *partiendo de la propiedad de Markov de la distribución conjunta*, Definición 1.4.2.

Esta es una sección de *propuestas de investigación futura* donde pretendemos simplemente dejar planteamientos sobre la mesa con respecto a la *inferencia*, y más aún, la *aplicación* del grafo no dirigido ante la presencia de cantidades

<sup>6</sup>Faithfulness assumption, en inglés. También se le conoce como *faithful distribution*: distribución fiel. Para detalles consultar sección 13.9 en [Bühlmann and Van De Geer \(2011\)](#).

masivas de variables y sobre todo en altas dimensiones. Las propuestas están alineadas a la postura que hemos tomado en el resto de la investigación: la búsqueda de modelos flexibles capaces de representar estructuras complejas de dependencia. Aunque lo dejaremos a nivel de concepto, consideramos que las propuestas pueden llegar a tener una aplicación tangible para la inferencia en altas dimensiones.

Comenzamos asociando uno de los aspectos en la definición del principio de escasez, Definición 5.2.1, con las relaciones de independencia condicional modeladas en el grafo no dirigido, lo cual nos da una noción de parsimonia partiendo de la propiedad de Markov de la distribución conjunta. Posteriormente, proponemos una estimación penalizada de las relaciones de independencia condicional del grafo. Finalmente, damos una observación adicional sobre una posible extensión de la escasez y el principio de parsimonia en general, ésto último alineado a las construcciones de cópulas a pares en vine cópula.

### Parsimonia vía el principio de escasez: la propiedad de Markov de la distribución

Partiendo del problema con el cual se motivó la inferencia en altas dimensiones en la Sección 5.2, supongamos un v.a.  $p$ -dimensional  $\mathbf{X} = (X_1, \dots, X_p)^T$  donde  $X_j$  es una variable de respuesta la cual buscamos explicar en función de las variables en  $\mathbf{X}_{-j}$  (o un subconjunto de éstas). Por otro lado, supongamos que contamos con el modelo gráfico no dirigido asociado a  $\mathbf{X}$ ,  $(\mathcal{G}, \mathbf{P})$ . Asumiendo que  $\mathbf{X}$  cumple con la propiedad de Markov en la Definición 1.4.2, el problema de seleccionar el *menor número de variables* en  $\mathbf{X}_{-j}$  que sean *suficientes* para explicar el comportamiento aleatorio de  $X_j$ , se reduce a observar el conjunto de adyacencia de  $X_j$  en  $\mathcal{G}$  y seguir la igualdad en (1.4.2), justo como lo comentamos en la Sección 1.4. En ese sentido estamos entendiendo a la distribución condicional como la vía idónea para explicar el comportamiento aleatorio de una variable de respuesta en función de predictores observados. Entonces la escasez la conceptualizamos como la *reducción (o contracción)* de la distribución condicional de  $X_j$  vía las relaciones de independencia condicional subyacentes del grafo no dirigido,  $\mathcal{G}$ , partiendo de la propiedad de Markov del v.a., donde naturalmente la distribución condicional escasas nos dan una *noción de parsimonia* sobre el modelo.

Más que simples regresiones lineales por la media, podemos conceptualizar *regresiones por cuantiles* subyacentes de la distribución condicional que induzca la escasez del grafo  $\mathcal{G}$  toda vez que tengamos una variable de respuesta  $X_j$ , como lo presentamos en la Definición 5.4.1 de regresiones  $\delta$ -cuantil.

**Definición 5.4.1** Sea el v.a.  $\mathbf{X} = (X_1, \dots, X_p)^T$ , con variable de respuesta  $X_j$ . Para  $\mathbf{v}_{-j} \in \text{Ran } \mathbf{X}_{-j}$  y  $\delta \in (0, 1)$ , sea  $x_j^\delta = Q_\delta(\mathbf{v}_{-j})$  que denota una solución a la ecuación,

$$\mathbb{P}[X_j \leq x_j^\delta \mid X_k, k \neq j] = \delta.$$

Asumiendo que  $\mathbf{X}$  cumple la propiedad de Markov, Definición 1.4.2, y que  $\mathcal{G}$  es la GIC subyacente de  $\mathbf{X}$ , se sigue que  $x_j^\delta$  denota una solución a la ecuación,

$$\mathbb{P}[X_j \leq x_j^\delta \mid X_k, k \in \text{ady}(\mathcal{G}, j)] = \delta.$$

Entonces  $Q_\delta$ , es la curva  $\delta$ -cuantil de  $X_j$  condicional en su conjunto de adyacencia.

Las regresiones por cuantiles nos proporcionan de manera natural *bandas de probabilidad* del comportamiento aleatorio de la variable de respuesta,  $X_j$ , en función de sus vecinos,  $\text{ady}(\mathcal{G}, j)$ .

Siguiendo la filosofía de inferencia en altas dimensiones, la cual está en su gran mayoría basada en modelos penalizados con la incorporación de un parámetro de ajuste,  $\lambda$ , generalmente un grafo no dirigido estimado tiene un grado de escasez asociado e inducido por el parámetro de ajuste,  $\mathcal{G}_\lambda$ , entonces en el fondo la parsimonia de los modelos estará fuertemente influenciada por  $\lambda$ . En nuestra propuesta para la inferencia del grafo en el Algoritmo 4.3.1,  $\alpha$  es el parámetro de ajuste<sup>7</sup> que controla el grado de escasez de la estimación.

<sup>7</sup>En el Algoritmo 4.3.1 ocupamos  $\alpha$  para alinearnos con la notación usual en pruebas de hipótesis, sin embargo, en términos generales  $\alpha$  tiene la interpretación de un parámetro de ajuste,  $\lambda$ , en el paradigma de modelos penalizados.



Habiendo exhibido la relevancia de un modelo gráfico no dirigido para la construcción de distribuciones condicionales que cumplan con el principio de parsimonia vía nuestra conceptualización de escasez, a continuación comentamos en términos abstractos cómo podría ser la inferencia penalizada del grafo no dirigido en altas dimensiones sin imponer un modelo paramétrico de cópula o distribución, partiendo del hecho de que los enfoques actuales de inferencia parten de supuestos que, como hemos comentado en los capítulos anteriores, son restrictivos en términos de la estructura de dependencia que inducen, véase Sección 5.3.

### Estimación no paramétrica vía de la matriz de dependencia condicional penalizada

Nuestra propuesta de estimación penalizada del grafo no dirigido comienza con los pasos uno y dos del Algoritmo 4.3.1, en un enfoque no paramétrico: la extracción iterativa de las observaciones de las  $\binom{p}{2}$  cópulas condicionales involucradas en el grafo y la elección de un parámetro de ajuste  $\lambda$ . Así que antes de describir el planteamiento de estimación, comentaremos sobre alternativas a considerar como estimadores no paramétricos de cópulas bivariadas, dado el limitado desempeño que tuvo el estimador Bernstein del grafo no dirigido en el Capítulo 4.

Como se comenta en [Bühlmann and Van De Geer \(2011\)](#), previo a considerar modelos con parámetros de regularización por escasez, una vía recurrente para la inferencia en altas dimensiones era la incorporación de restricciones por suavidad para modelos no paramétricos, y que actualmente, muchas de las propuestas de penalización llegan a incorporar ambas restricciones: por escasez y por suavidad. Así que nuestra propuesta es agregarle al estimador Bernstein de la cópula bivariada un parámetro de ajuste,  $\gamma$ , que domine el *grado de suavidad* de la estimación, consideramos que esta modificación puede controlar los errores numéricos que subyacen del estimador Bernstein de la cópula y que están presentes en las estimaciones de las observaciones de las cópulas condicionales involucradas en el grafo no dirigido. De igual forma, se pueden explorar otras opciones no paramétricas para estimar las cópulas bivariadas como aquellas que se presentan en [Joe \(2014\)](#), sección 5.10.3; en particular las propuestas vía B-splines en [Shen et al. \(2008\)](#) y [Kauermann et al. \(2013\)](#). Estas modificaciones pueden ser incorporadas también a nuestro planteamiento en el Algoritmo 4.3.1.

En contraste con la propuesta basada en pruebas múltiples del Capítulo 4, la siguientes está alineada a una perspectiva de estimación de parámetros. En la Definición 5.4.2 presentamos a la matriz de *dependencia condicional*,  $\Omega^{SW}$ , basada en la medida de dependencia  $\sigma_{SW}$  de [Schweizer and Wolff \(1981\)](#) presentada en (2.6.3).

**Definición 5.4.2 Matriz de dependencia condicional.** Dado un v.a.  $\mathbf{X} = (X_1, \dots, X_p)^T$ , la matriz de *dependencia condicional*,  $\Omega^{SW}$ , basada en la medida de dependencia  $\sigma_{SW}$  en (2.6.3), se define como,

$$\Omega_{jk}^{SW} = \begin{cases} \sigma_{SW}(X_j, X_k | X_{\mathcal{V} \setminus \{j,k\}}), & \text{si } j \neq k, \\ 1, & \text{si } j = k, \end{cases}$$

$\forall j, k \in \mathcal{V}$ , con  $\mathcal{V}$  el conjunto de vértices de  $\mathbf{X}$ .

Una versión muestral de  $\Omega^{SW}$  se presenta en la Definición 5.4.3.

**Definición 5.4.3 Estimador de la matriz de dependencia condicional.** Dada una muestra observada de un v.a.  $\mathbf{X} = (X_1, \dots, X_p)^T$ ,

$$\underline{\tilde{x}} := \{(x_{1i}, \dots, x_{pi})\}_{i=1}^N,$$

supongamos que contamos con las extracciones de observaciones estimadas de cópulas condicionales como en la Sección 4.3,

$$\underline{\tilde{u}}_{jk} := \{(u_{j|\mathcal{V} \setminus \{j,k\}, i}, u_{k|\mathcal{V} \setminus \{j,k\}, i})\}_{i=1}^N, \quad j, k \in \mathcal{V}, \text{ t.q. } j < k,$$

entonces un *estimador para la matriz de dependencias* condicionales es,

$$\widehat{\Omega}_{jk}^{SW} = \begin{cases} \frac{12}{N^2-1} \sum_{i=1}^N \sum_{j=1}^N \left| \mathcal{C}_N^{jk} \left( \frac{i}{N}, \frac{j}{N} \right) - \frac{i}{N} \cdot \frac{j}{N} \right|, & \text{si } j < k, \\ \widehat{\Omega}_{kj}^{SW} & \text{si } j > k, \\ 1, & \text{si } j = k, \end{cases}$$

$\forall j, k \in \mathcal{V}$ , con  $\mathcal{V}$  el conjunto de vértices de  $\mathbf{X}$ , donde  $C_N^{jk}$  es la cópula empírica (Definición 4.6.1) asociada a cada muestra estimada  $\underline{u}_{jk}$  t.q.  $j < k$ .

Sin embargo en el contexto de altas dimensiones, el estimador  $\widehat{\Omega}^{SW}$  en la Definición 5.4.3 puede presentar *sobreajuste* en los valores puntuales de la dependencia condicional vía  $\sigma_{SW}$ . Así que *contraer algunos de los elementos* de la matriz estimada y *eliminar el resto* es una opción para evitar el sobreajuste de  $\Omega^{SW}$  en pro de la parsimonia. Por otro lado, notemos de (5.2.3) que el estimador penalizado en el modelo de regresión lineal dependen básicamente de dos cosas: la primera, una función de pérdida (o utilidad) que involucra a los elementos en muestra y al parámetro a estimar en el enfoque penalizado; y segunda, la restricción por escasez. Haciendo uso de la norma  $\ell_1$ , que de antemano sabemos que por la geometría de su restricción impone elementos nulos, una estimación penalizada para  $\Omega^{SW}$ , sería la siguiente,

$$\widehat{\Omega}^{SW}(\lambda) := \underset{\Omega^{SW} : \|\Omega^{SW}\|_1 \leq S_\lambda}{\text{argmín}} \left\{ J(\Omega^{SW}, \underline{x}) \right\},$$

donde

$$\|\Omega^{SW}\|_1 := \sum_{j < k} |\Omega_{jk}^{SW}|,$$

$\underline{x}$  es la muestra observada de un v.a.  $\mathbf{X}$ ,  $S_\lambda$  es la restricción numérica asociada con  $\lambda$  y  $J(\Omega^{SW}, \underline{x})$  es una función de pérdida relacionada con el ajuste del modelo. De igual forma se puede cambiar la función de pérdida por una de utilidad y modificar la optimización por un proceso de maximización.

Finalmente, el modelo gráfico no dirigido estimado mediante la penalización  $\ell_1$  aplicada a la matriz  $\Omega^{SW}$  sería,

$$\widehat{\mathcal{A}}_\lambda := \{(j, k), (k, j) \in \mathcal{V} \times \mathcal{V} \mid \widehat{\Omega}_{jk}^{SW}(\lambda) = 0, j, k \in \mathcal{V}, \text{ t.q. } j < k\}.$$

## Una extensión adicional del principio de escasez y de parsimonia

Como lo hemos venido comentado, para una muestra observada de un v.a.  $p$ -dimensional el paradigma de estimación de modelos multivariados vía vine cópula se basa en la construcción de cópulas a pares condicionales, en el sentido de que se *factoriza la densidad conjunta* en el producto de las  $p$  densidades marginales del v.a. y densidades cópula bivariadas condicionales a los *distintos niveles posibles*. En donde por niveles nos referimos al número de variables en el condicionante, las cuales van desde cero variables, i.e. cópulas no condicionales, siendo éstas siempre  $p - 1$  cópulas bivariadas, y hasta  $p - 2$  variables en el condicionante, i.e. las cópulas relacionadas con el grafo no dirigido, siendo ésta una sola cópula en cada factorización (c.f. [Aas and Berg \(2009\)](#)). En este esquema de factorización las relaciones de independencia condicional a cualquier nivel son deseables, en el sentido de que le *restan complejidad al modelo* sin perder información vital para el mismo. Así que consideramos que la *mayor representación* de un modelo que cumpla el principio de parsimonia sería aquel que incluyese el *mayor número posible* de relaciones de independencia condicional a cualquier nivel, no sólo al nivel  $p - 2$  como en el grafo no dirigido, i.e. con base en los datos observados elegir aquella factorización de la densidad conjunta en la cual se estimasen el mayor número de *densidades cópula igual a uno* (la densidad de la cópula producto). Lo anterior sin tener que recurrir a ninguna premisa para la construcción del modelo, tal como la propiedad de Markov del v.a. De igual forma, el grado de escasez de un modelo multivariado se relacionaría con las *densidades cópula igual a la unidad*, en analogía con los elementos nulos de una matriz de precisión o de dependencia condicional, esta última como en nuestra propuesta de estimación penalizada.

Pese a que actualmente se reconoce a vine cópula como una herramienta con un gran potencial para la construcción de modelos *bastante flexibles* capaces de representar estructuras de dependencia complejas, los problemas que resuelve hoy en día no involucran cantidades masivas de variables o datos altamente dimensionales. Una de las grandes problemáticas en vine cópula es la *selección de modelo*, debido a que el número de posibles factorizaciones de la densidad conjunta *crece rápidamente* al aumentar la cantidad de variables. Así que el estudio de la selección de modelo en vine cópula es actualmente uno de los temas más estudiados en el área, v.g. [Min and Czado \(2011\)](#). Por lo tanto en el esquema de construcción de modelos multivariados vía vine cópula, la búsqueda del modelo que cumpla

con el principio de parsimonia con base en nuestro planteamiento, cómo inducir escasez vía cópulas densidad a pares unitarias o cualquier otra propuesta que pudiera ser aplicada al contexto de altas dimensiones y/o ante problemas con grandes cantidades de variables involucradas, siguen siendo *problemas abiertos* en la literatura.

## 5.5. Discusión

En el presente capítulo explicamos la problemática de la inferencia estadística en el contexto altamente dimensional, enfatizando en la necesidad de un cambio de paradigma con respecto a las metodologías clásicas de análisis. Se propone una definición que pretende ser general sobre el principio de escasez y se reconoce como un supuesto bastante realista para este contexto de inferencia, con potencial para diversas aplicaciones. Se estudia el panorama actual de estimación del grafo no dirigido en altas dimensiones, de donde se concluye que la gran mayoría de los enfoques propuestos basan la inferencia en supuestos poco realistas, de distribución conjunta o de cópula, ante la gama de diversos patrones de dependencia que se pueden encontrar en datos multivariados. Pese a que las metodologías propuestas pretenden aplicar a cantidades masivas de variables en el contexto altamente dimensional, los modelos conjuntos se reducen a cópulas gaussianas y  $\mathcal{T}$ , o las propuestas se limitan a la inferencia de un grafo que modela correlaciones parciales nulas más que independencias condicionales, perdiendo así toda interpretación y aplicabilidad.

Finalmente, se propone una vía para conceptualizar el principio de escasez partiendo de la propiedad de Markov del v.a. y las relaciones de independencia condicional modeladas en un grafo no dirigido, lo anterior tiene implicaciones sobre la construcción de modelos para explicar el comportamiento aleatorio de una variable de respuesta en función de predictores observados siguiendo el principio de parsimonia. Así la inferencia del modelo gráfico no dirigido tiene aplicaciones para la construcción de modelos escasos de distribución condicional, de donde se desprenden modelos de regresión por cuantiles mucho más informativos que regresiones lineales por la media. Buscando que la propuesta tenga impacto y aplicación para la inferencia de modelos en altas dimensiones se plantea, a nivel de concepto, cómo podría ser la estimación penalizada del modelo gráfico no dirigido en un enfoque totalmente no paramétrico. De igual forma, se comenta sobre una posible extensión adicional de los principios de escasez y parsimonia, lo anterior alineado a las construcciones de cópulas a pares en vine cópula. Todo lo previamente mencionado se deja como propuestas de investigación futura.

Se concluye la investigación destacando lo complejo de proponer en general modelos conjuntos multivariados flexibles, sobre todo cuando la inferencia depende de grandes cantidades de variables en contextos altamente dimensionales. Sin embargo, dado el gran potencial de las herramientas y conceptos existentes tanto para la inferencia en altas dimensiones como en vine cópula, consideramos que conjugar ambos enfoques de análisis es un área de investigación bastante fértil y poco explorado aún, pero que podría traer consigo modelos bastante flexibles y aplicables a las estructuras modernas de datos.



# Conclusiones

A lo largo de esta investigación exploramos diversas vías para la estimación del grafo dirigido cuyo modelo de probabilidad asociado fuese lo suficientemente flexible, en el sentido de representar patrones complejos de dependencia, las propuestas analizadas estuvieron ligadas a la inferencia basada en la cópula del vector aleatorio de interés, reconociendo a la cópula subyacente como aquella que contiene toda la información sobre las relaciones de independencia condicional modeladas en un grafo no dirigido. En particular se puntualizó en metodologías aplicables al contexto de altas dimensiones. La motivación general de nuestra investigación es el hecho de que hoy en día los modelos gráficos no dirigidos son una herramienta con mucho potencial para el análisis estadístico multivariado, sobre todo con datos altamente dimensionales, sin embargo, su inferencia por lo general está restringida a modelos de probabilidad sumamente limitados. A grandes rasgos se reconocieron dos grandes enfoques de estimación: asumiendo un modelo de cópula simplificada multivariada y desde una perspectiva totalmente no paramétrica partiendo del supuesto simplificador.

Del enfoque de cópula multivariada simplificada, destacamos que estamos acotados simplemente a dos modelos de cópulas elípticas: gaussina y  $\mathcal{T}$ . Pese a que para estas dos cópulas las relaciones de independencia condicional representadas en el grafo no dirigido cuentan con una codificación explícita vía sus parámetros, éstos son modelos multivariados que de antemano *están muy acotados con respecto a la estructura de dependencia* que representan. Por otro lado, en términos generales es cuestionable el hecho de que *un sólo modelo de cópula paramétrica* sea capaz de explicar los diversos patrones de dependencia que se encuentran indudablemente en datos multivariados, sobre todo ante un gran número de variables.

Con respecto al enfoque totalmente no paramétrico, en contraste con el paramétrico, *no contamos con una representación explícita* de las relaciones de independencia condicional de interés; lo anterior se constituye como el mayor reto para este enfoque. Es por lo antes mencionado, que consideramos a la *extracción iterativa de las observaciones de las  $\binom{p}{2}$  cópulas condicionales* involucradas en el grafo no dirigido, vía la fórmula de Joe y el supuesto simplificador, como un *procedimiento con gran potencial para la estimación precisa y flexible* del mismo. Por un lado, la precisión se puede lograr ya que contamos con una *muestra explícita* de las observaciones asociadas directamente con las relaciones de independencia condicional que necesitamos inferir; por otro lado y en términos de flexibilidad de modelo, existen diversas posibilidades metodológicas para construir *cópulas a pares que expliquen efectivamente* la dependencia observada. Sin embargo, la extracción de observaciones involucra una *acumulación iterativa de error de estimación* proveniente de la inferencia de las cópulas a pares, tal como se observó en el desempeño del estimador Bernstein del grafo no dirigido en el Capítulo 4, razón que impacta en el *error global de estimación del grafo*. Partiendo de esta *nueva problemática de estimación* del modelo gráfico no dirigido, entre los Capítulos 4 y 5 comentamos sobre posibles vías de solución alineadas a nuestra propuesta no paramétrica. Se mencionó que se puede comenzar por estudiar las propiedades teóricas del estimador Bernstein de las observaciones de cópulas condicionales, para conocer su sensibilidad ante aumentos en tamaño de muestra o número de variables y con base en su comportamiento proponer modificaciones al estimador de las cópulas, o más aún, al procedimiento de estimación del grafo, v.g. agregarle al estimador Bernstein de la cópula bivariada un parámetro que controle su suavidad. Otra opción es, en nuestro planteamiento de inferencia del grafo se puede cambiar el estimador Bernstein de las cópulas a pares por otra alternativa no paramétrica, v.g. un estimador por B-splines. Así *dejamos abierta una brecha de investigación* para proponer modelos de cópulas a pares o metodologías en general, que de antemano sean capaces de reducir el error acumulado de estimación del modelo gráfico no dirigido.



# Apéndices





## Apéndice A

# Distribución NonParaNormal y el modelo de cópula gaussiana

En este apéndice se plantean demostraciones a los Lemas 3.3.1 y 3.3.2.

### A.1. Transformaciones en NPN

**Demostración.** Primero, notemos que  $f_j(X_j) \sim \mathcal{N}(0, 1)$ . Ahora probaremos la unicidad de  $f_j$ . Supongamos que existen dos funciones estrictamente crecientes,  $f_j^{(1)}$  y  $f_j^{(2)}$ , tales que  $f_j^{(1)}(X_j) \sim \mathcal{N}(0, 1)$  y  $f_j^{(2)}(X_j) \sim \mathcal{N}(0, 1)$ . Entonces,  $\forall t \in \mathbb{R}$  tenemos que,

$$\begin{aligned} F_j \left[ \left( f_j^{(1)} \right)^{-1} (t) \right] &= \mathbb{P} \left[ X_j \leq \left( f_j^{(1)} \right)^{-1} (t) \right] = \mathbb{P} \left[ f_j^{(1)} (X_j) \leq t \right] = \Phi(t) \\ &= \mathbb{P} \left[ f_j^{(2)} (X_j) \leq t \right] = \mathbb{P} \left[ X_j \leq \left( f_j^{(2)} \right)^{-1} (t) \right] = F_j \left[ \left( f_j^{(2)} \right)^{-1} (t) \right], \end{aligned}$$

y del hecho de que  $F_j$  es estrictamente creciente se sigue que  $\left( f_j^{(1)} \right)^{-1} (t) = \left( f_j^{(2)} \right)^{-1} (t)$  para todo  $t$ , lo que implica que  $f_j^{(1)} = f_j^{(2)}$ . Ahora notemos que  $f_j = \Phi^{-1} (F_j)$  es una función estrictamente creciente que transforma  $X_j$  en una variable aleatoria normal estándar, de donde se sigue la conclusión. □

Esta demostración se extrajo de [Mai and Zou \(2015\)](#) (Lema 1), sin embargo nos pareció interesante presentarla porque determina explícitamente al conjunto de funciones en la Definición 3.3.1 sobre la distribución NPN, así toda vez que tengamos un modelo de cópula gaussiana sabremos exactamente cual es el v.a. latente con distribución gaussiana.

### A.2. NPN y cópula gaussiana

A continuación se formula una prueba para la equivalencia entre un modelo NPN y uno de cópula gaussiana.

**Demostración.** Primero probemos que si  $\mathbf{X} \sim NPN_p (\Sigma^0, f^0)$ , entonces  $C_{\mathbf{X}}$  es una cópula gaussiana como en (3.3.2).

A partir del Lema 3.3.1 se sigue que  $f_j^0(X_j) = \Phi^{-1}(F_j(X_j))$  con  $j = 1, \dots, p$ , son las únicas transformaciones estrictamente crecientes que admite el modelo NPN, entonces si  $f^0 = \{f_j^0\}_{j=1}^p$  por el Lema 2.5.1 se sigue que  $C_{\mathbf{X}} = C_{f^0(\mathbf{X})}$ , pero  $f^0(\mathbf{X}) \sim \mathcal{N}_p(\mathbf{0}, \Sigma^0)$ , por lo tanto

$$C_{\mathbf{X}}(u_1, \dots, u_p) = C_{f^0(\mathbf{X})}(u_1, \dots, u_p) = \Phi_{\Sigma^0}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_p)),$$

como se buscaba probar.

Ahora, probaremos que un modelo de cópula gaussiana implica distribución NPN.

Sea  $\mathbf{X} = (X_1, \dots, X_p)^T$  un vector aleatorio tal que,

$$C_{\mathbf{X}}(u_1, \dots, u_p) = \Phi_{\Sigma^0}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_p)).$$

Con base en el Lema 3.3.1 sabemos que el conjunto de transformaciones  $f^0 = \{f_j^0(X_j) = \phi^{-1}(F_j(X_j))\}_{j=1}^p$  es tal que  $f_j^0(X_j) \sim \mathcal{N}(0, 1)$ , además  $f^0$  es único.  $f^0$  se compone de transformaciones estrictamente crecientes, entonces por el Lema 2.5.1 la cópula de las variables transformadas es la misma que la del vector original, es decir,

$$C_{f^0(\mathbf{X})}(u_1, \dots, u_p) = \Phi_{0, \Sigma^0}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_p)),$$

y como cada  $f_j^0(X_j) \sim \mathcal{N}(0, 1)$  se concluye que  $f^0(\mathbf{X}) \sim \mathcal{N}_p(\mathbf{0}, \Sigma^0)$ , en otras palabras,  $\mathbf{X} \sim NPN_p(\Sigma^0, f^0)$ .

Quedando así probada la doble implicación.

□

La motivación de presentar una propuesta de prueba al Lema 3.3.2 es el hecho de que no encontramos en la literatura una demostración clara a esta equivalencia (c.f. [Liu et al. \(2009\)](#), [Liu et al. \(2012a\)](#) y [Han and Liu \(2014\)](#)).

## Apéndice B

# Condicionales de cópula gaussiana

En este apéndice proponemos una demostración a la Proposición 3.3.2, misma que asegura que las condicionales de cópula gaussiana son *cópulas gaussianas también* con una matriz de correlaciones parciales. Con lo anterior se prueba en paralelo que la cópula gaussiana cumple el *supuesto simplificador*.

### B.1. Condicionales de cópula gaussiana

**Demostración.** Procederemos vía inducción matemática sobre la cardinalidad de los conjuntos involucrados en la proposición. Primero probaremos que se cumple para los subconjuntos  $A_k$  y  $B_k$  tales que  $\text{card}(A_k) = p - k$  y  $\text{card}(B_k) = k$ , con  $k = 1, \dots, p - 2$ .

Por practicidad de notación y sin pérdida de generalidad pediremos orden a los índices en los subconjuntos  $A_k$  y  $B_k$  de forma tal que,

$$A_k = (k + 1) : p, \quad B_k = 1 : k.$$

#### Base de inducción ( $k = 1$ ).

p.d.  $C_{A_1|B_1}$  es gaussiana con parámetro  $\Sigma^{A_1|B_1}$  independientemente del valor del condicionante.

Comencemos por analizar la distribución conjunta condicional. Notemos que  $\mathbf{X}_{A_1} = (X_2, \dots, X_p)^T$ , así que tomemos  $x_i \in \text{Ran } X_i$ ,  $i = 2, \dots, p$ . Por otro lado  $\mathbf{X}_{B_1} = X_1$ , así que tomemos  $\mathbf{v}_1 \in \text{Ran } X_1$ .

$$\begin{aligned} H_{A_1|B_1}(x_2, \dots, x_p | \mathbf{v}_1) &= \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_p} \frac{f_{1:p}(\mathbf{v}_1, z_2, \dots, z_p)}{f_1(\mathbf{v}_1)} dz_2 \dots dz_p \\ &= \frac{1}{f_1(\mathbf{v}_1)} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_p} f_{1:p}(\mathbf{v}_1, z_2, \dots, z_p) dz_2 \dots dz_p \\ &= \frac{1}{f_1(\mathbf{v}_1)} \cdot \frac{\partial}{\partial \mathbf{v}_1} H_{1:p}(\mathbf{v}_1, x_2, \dots, x_p) \\ &= \frac{1}{f_1(\mathbf{v}_1)} \cdot \frac{\partial}{\partial F_1(\mathbf{v}_1)} \mathcal{C}_{1:p}(F_1(\mathbf{v}_1), F_2(x_2), \dots, F_p(x_p)) f_1(\mathbf{v}_1) \\ &= \frac{\partial}{\partial F_1(\mathbf{v}_1)} \mathcal{C}_{1:p}(F_1(\mathbf{v}_1), F_2(x_2), \dots, F_p(x_p)), \end{aligned} \tag{B.1.1}$$

donde  $f_{1:p}$  es la función de densidad conjunta de  $\mathbf{X}$ .

Denotemos con  $b_i = \Phi^{-1}(u_i)$ ,  $i = 1, \dots, p$ . Recordemos que  $\Phi$  y  $\phi$  son la FDA y densidad de una distribución normal estándar univariada. Analicemos la expresión en (B.1.1),

$$\begin{aligned}
\frac{\partial}{\partial u_1} \mathcal{C}_{1:p}(u_1, \dots, u_p) &= \frac{\partial}{\partial u_1} \int_{-\infty}^{\Phi^{-1}(u_1)} \cdots \int_{-\infty}^{\Phi^{-1}(u_p)} g_{1:p}(z_1, \dots, z_p; \Sigma) dz_1 \cdots dz_p \\
&= \frac{\partial}{\partial u_1} \int_{-\infty}^{b_1} \cdots \int_{-\infty}^{b_p} g_{1:p}(z_1, \dots, z_p; \Sigma) dz_1 \cdots dz_p \\
&= \frac{\partial b_1}{\partial u_1} \cdot \frac{\partial}{\partial b_1} \int_{-\infty}^{b_1} \cdots \int_{-\infty}^{b_p} g_{1:p}(z_1, \dots, z_p; \Sigma) dz_1 \cdots dz_p \\
&= \frac{1}{\phi(b_1)} \int_{-\infty}^{b_2} \cdots \int_{-\infty}^{b_p} \frac{\partial}{\partial b_1} \int_{-\infty}^{b_1} g_{1:p}(z_1, \dots, z_p; \Sigma) dz_1 \cdots dz_p \\
&= \frac{1}{\phi(b_1)} \int_{-\infty}^{b_2} \cdots \int_{-\infty}^{b_p} g_{1:p}(b_1, z_2, \dots, z_p; \Sigma) dz_2 \cdots dz_p \\
&= \frac{g_1(b_1; \sigma_1^2 = 1)}{\phi(b_1)} \int_{-\infty}^{b_2} \cdots \int_{-\infty}^{b_p} \frac{g_{1:p}(b_1, z_2, \dots, z_p; \Sigma)}{g_1(b_1; \sigma_1^2 = 1)} dz_2 \cdots dz_p \\
&= \frac{\phi(b_1)}{\phi(b_1)} \int_{-\infty}^{b_2} \cdots \int_{-\infty}^{b_p} g_{2:p|1}(z_2, \dots, z_p | b_1; \Sigma) dz_2 \cdots dz_p \\
&= \int_{-\infty}^{\frac{b_2 - \rho_{21} b_1}{\sqrt{1 - \rho_{21}^2}}} \cdots \int_{-\infty}^{\frac{b_p - \rho_{p1} b_1}{\sqrt{1 - \rho_{p1}^2}}} g_{2:p}(z_2, \dots, z_p; \Sigma^{2:p|1}) dz_2 \cdots dz_p \\
&= \Phi_{\Sigma^{2:p|1}} \left( \frac{b_2 - \rho_{21} b_1}{\sqrt{1 - \rho_{21}^2}}, \dots, \frac{b_p - \rho_{p1} b_1}{\sqrt{1 - \rho_{p1}^2}} \right). \tag{B.1.2}
\end{aligned}$$

A continuación se dará justificación a algunas de las igualdades obtenidas en el desarrollo anterior.

- Sea  $(Y_1, \dots, Y_p)^T \sim \mathcal{N}_p(\mathbf{0}, \Sigma)$  (un vector genérico Gaussiano), con  $\Sigma$  una matriz de correlaciones t.q.  $\text{diag}(\Sigma) = \mathbf{1}$ . Tomemos la siguiente partición de  $\Sigma$ ,

$$\Sigma = \begin{bmatrix} \Sigma_1 & \Sigma_* \\ \Sigma_*^T & \Sigma_{2:p} \end{bmatrix} \quad \text{con tamaños} \quad \begin{bmatrix} 1 \times 1 & 1 \times (p-1) \\ (p-1) \times 1 & (p-1) \times (p-1) \end{bmatrix}.$$

Entonces  $(Y_2, \dots, Y_p)^T | \{Y_1 = b_1\} \sim \mathcal{N}_{p-1}(\mu^{2:p|1}(b_1), P^{2:p|1}(b_1))$ , donde

$$\mu^{2:p|1}(b_1) = \Sigma_*^T \Sigma_1^{-1} b_1 = \Sigma_*^T b_1. \tag{B.1.3}$$

$$P^{2:p|1}(b_1) = \Sigma_{2:p} - \Sigma_*^T \Sigma_1^{-1} \Sigma_* = \Sigma_{2:p} - \Sigma_*^T \Sigma_*. \tag{B.1.4}$$

De (B.1.3) se sigue que,

$$\mathbb{E}[Y_{j+1} | Y_1 = b_1] = (\mu^{2:p|1}(b_1))_j = \rho_{j+1,1} b_1 \quad j = 1, \dots, p-1. \tag{B.1.5}$$

De (B.1.4) se sigue que,

$$\text{Var}[Y_{j+1} | Y_1 = b_1] = (P^{2:p|1}(b_1))_{jj} = 1 - \rho_{j+1,1}^2 \quad j = 1, \dots, p-1. \tag{B.1.6}$$

Por lo tanto, siguiendo (B.1.5) y (B.1.6) tenemos que,

$$\left\{ \left( \frac{Y_2 - \rho_{21} b_1}{\sqrt{1 - \rho_{21}^2}}, \dots, \frac{Y_p - \rho_{p1} b_1}{\sqrt{1 - \rho_{p1}^2}} \right) \middle| Y_1 = b_1 \right\} \sim \mathcal{N}_{p-1}(\mathbf{0}, \Sigma^{2:p|1}),$$

donde  $\Sigma^{2:p|1}$  es una matriz de correlaciones parciales con respecto a  $B_1$ , i.e.

$$\left(\Sigma^{2:p|1}\right)_{ij} = \rho_{i+1,j+1|1} = \rho_{i+1,j+1} - \rho_{i+1,1}\rho_{1,j+1}, \quad i, j \in \{1, \dots, p-1\},$$

tal que  $\text{diag}(\Sigma^{2:p|1}) = \mathbf{1}$ .

Regresando a la prueba, sustituyendo (B.1.2) en (B.1.1) tenemos que

$$H_{A_1|B_1}(x_2, \dots, x_p | \mathbf{v}_1) = \Phi_{\Sigma^{2:p|1}} \left( \frac{\Phi^{-1}(F_2(x_2)) - \rho_{21}\Phi^{-1}(F_1(\mathbf{v}_1))}{\sqrt{1-\rho_{21}^2}}, \dots, \frac{\Phi^{-1}(F_p(x_p)) - \rho_{p1}\Phi^{-1}(F_1(\mathbf{v}_1))}{\sqrt{1-\rho_{p1}^2}} \right). \quad (\text{B.1.7})$$

Sean  $u_2, \dots, u_p \in \mathbb{I}$ , vía (2.4.2) extraeremos la cópula condicional de la distribución condicional, i.e.

$$\mathcal{C}_{A_1|B_1}(u_2, \dots, u_p; \mathbf{v}_1) = H_{A_1|B_1}(F_{2|1}^{-1}(u_2; \mathbf{v}_1), \dots, F_{p|1}^{-1}(u_p; \mathbf{v}_1) | \mathbf{v}_1). \quad (\text{B.1.8})$$

De (B.1.8) observamos que necesitamos conocer las expresiones de  $F_{j|1}^{-1}$ , con  $j = 2, \dots, p$ , entonces recurriremos a la fórmula de Joe (1996), Proposición 4.2.1. Sea  $x \in \text{Ran}(X_j | \{X_1 = \mathbf{v}_1\})$ ,  $j = 2, \dots, p$ , tenemos que

$$\begin{aligned} F_{j|1}(x | \mathbf{v}_1) &= \frac{\partial \mathcal{C}_{j1}(F_j(x), F_1(\mathbf{v}_1))}{\partial F_1(\mathbf{v}_1)} \\ &= \Phi \left( \frac{\Phi^{-1}(F_j(x)) - \rho_{j1}\Phi^{-1}(F_1(\mathbf{v}_1))}{\sqrt{1-\rho_{j1}^2}} \right), \end{aligned} \quad (\text{B.1.9})$$

(B.1.9) se sigue como caso particular de (B.1.7).

Ahora obtengamos  $F_{j|1}^{-1}$  para  $j = 2, \dots, p$ . Sea  $u \in \mathbb{I}$  t.q.  $F_{j|1}(x | \mathbf{v}_1) = u$ , se sigue que,

$$\begin{aligned} \Phi \left( \frac{\Phi^{-1}(F_j(x)) - \rho_{j1}\Phi^{-1}(F_1(\mathbf{v}_1))}{\sqrt{1-\rho_{j1}^2}} \right) &= u \\ \frac{\Phi^{-1}(F_j(x)) - \rho_{j1}\Phi^{-1}(F_1(\mathbf{v}_1))}{\sqrt{1-\rho_{j1}^2}} &= \Phi^{-1}(u) \\ \Phi^{-1}(F_j(x)) &= \Phi^{-1}(u)\sqrt{1-\rho_{j1}^2} + \rho_{j1}\Phi^{-1}(F_1(\mathbf{v}_1)) \\ F_j(x) &= \Phi \left[ \Phi^{-1}(u)\sqrt{1-\rho_{j1}^2} + \rho_{j1}\Phi^{-1}(F_1(\mathbf{v}_1)) \right] \\ x &= F_j^{-1} \left\{ \Phi \left[ \Phi^{-1}(u)\sqrt{1-\rho_{j1}^2} + \rho_{j1}\Phi^{-1}(F_1(\mathbf{v}_1)) \right] \right\}, \end{aligned}$$

por lo tanto,

$$F_{j|1}^{-1}(u; \mathbf{v}_1) = F_j^{-1} \left\{ \Phi \left[ \Phi^{-1}(u)\sqrt{1-\rho_{j1}^2} + \rho_{j1}\Phi^{-1}(F_1(\mathbf{v}_1)) \right] \right\}. \quad (\text{B.1.10})$$

Analicemos el argumento  $j$  de  $H_{A_1|B_1}(\cdot, \dots, \cdot | \mathbf{v}_1)$ , (B.1.7), evaluado en  $F_{j|1}^{-1}(u_j; \mathbf{v}_1)$ ,

$$\frac{\Phi^{-1} \left( F_j \left( F_{j|1}^{-1}(u_j; \mathbf{v}_1) \right) \right) - \rho_{j1}\Phi^{-1}(F_1(\mathbf{v}_1))}{\sqrt{1-\rho_{j1}^2}} = \Phi^{-1}(u_j), \quad (\text{B.1.11})$$

(B.1.11) se sigue de (B.1.10).

Finalmente, con base en (B.1.7), (B.1.8) y (B.1.11), se puede concluir que,

$$\mathcal{C}_{A_1|B_1}(u_2, \dots, u_p; \mathbf{v}_1) = \Phi_{\Sigma^{2:p|1}}(\Phi^{-1}(u_2), \dots, \Phi^{-1}(u_p)) = \mathcal{C}_{A_1|B_1}(u_2, \dots, u_p),$$

es decir  $\mathcal{C}_{A_1|B_1}$  es cópula gaussiana de dimensión  $p-1$  con parámetro  $\Sigma^{2:p|1} = \Sigma^{A_1|B_1}$  y además éste no depende del valor del condicionante.

**Hipótesis de inducción** (se supone válido para  $k$ , con  $k \in \{2, \dots, p-3\}$ ).

$\mathcal{C}_{A_k|B_k}$  es cópula gaussiana con parámetro  $\Sigma^{A_k|B_k}$  independiente del valor del condicionante.

**Prueba de inducción** (se prueba para  $k+1$ ).

p.d.  $\mathcal{C}_{A_{k+1}|B_{k+1}}$  es cópula gaussiana con parámetro  $\Sigma^{A_{k+1}|B_{k+1}}$  independiente del valor del condicionante.

Comencemos por analizar la distribución conjunta condicional al paso  $k+1$ . Notemos que  $\mathbf{X}_{A_{k+1}} = (X_{k+2}, \dots, X_p)$ , así que tomemos  $x_i \in \text{Ran } \mathbf{X}_i$ ,  $i = k+2, \dots, p$ . Además  $\mathbf{X}_{B_{k+1}} = (X_1, \dots, X_{k+1})^T$ , así que tomemos  $\mathbf{v}_{k+1} = (v_1, \dots, v_{k+1}) \in \text{Ran } \mathbf{X}_{B_{k+1}}$ . También denotaremos con  $\mathbf{v}_k = (v_1, \dots, v_k)$  a los primeros  $k$  elementos de  $\mathbf{v}_{k+1}$ .

$$\begin{aligned} H_{A_{k+1}|B_{k+1}}(x_{k+2}, \dots, x_p | \mathbf{v}_{k+1}) &= \int_{-\infty}^{x_{k+2}} \dots \int_{-\infty}^{x_p} f_{k+2:p|1:k+1}(z_{k+2}, \dots, z_p | \mathbf{v}_{k+1}) dz_{k+2} \dots dz_p \\ &= \int_{-\infty}^{x_{k+2}} \dots \int_{-\infty}^{x_p} \frac{f_{1:p}(\mathbf{v}_{k+1}, z_{k+2}, \dots, z_p)}{f_{1:k+1}(\mathbf{v}_{k+1})} dz_{k+2} \dots dz_p \\ &= \int_{-\infty}^{x_{k+2}} \dots \int_{-\infty}^{x_p} \frac{f_{1:p}(\mathbf{v}_{k+1}, z_{k+2}, \dots, z_p)}{f_{1:k+1}(\mathbf{v}_{k+1})} \cdot \left( \frac{f_{1:k}(\mathbf{v}_k)}{f_{1:k}(\mathbf{v}_k)} \right) dz_{k+2} \dots dz_p \\ &= \left( \frac{f_{1:k}(\mathbf{v}_k)}{f_{1:k+1}(\mathbf{v}_{k+1})} \right) \int_{-\infty}^{x_{k+2}} \dots \int_{-\infty}^{x_p} \frac{f_{1:p}(\mathbf{v}_k, v_{k+1}, z_{k+2}, \dots, z_p)}{f_{1:k}(\mathbf{v}_k)} dz_{k+2} \dots dz_p \\ &= \left( \frac{f_{1:k}(\mathbf{v}_k)}{f_{1:k+1}(\mathbf{v}_k, v_{k+1})} \right) \int_{-\infty}^{x_{k+2}} \dots \int_{-\infty}^{x_p} f_{k+1:p|1:k}(v_{k+1}, z_{k+2}, \dots, z_p | \mathbf{v}_k) dz_{k+2} \dots dz_p \\ &= \left( \frac{1}{f_{k+1|1:k}(v_{k+1} | \mathbf{v}_k)} \right) \frac{\partial H_{k+1:p|1:k}(v_{k+1}, z_{k+2}, \dots, z_p | \mathbf{v}_k)}{\partial v_{k+1}} \\ &= \left( \frac{1}{f_{k+1|1:k}(v_{k+1} | \mathbf{v}_k)} \right) \frac{\partial H_{A_k|B_k}(v_{k+1}, z_{k+2}, \dots, z_p | \mathbf{v}_k)}{\partial v_{k+1}} \\ &= \left( \frac{f_{k+1|1:k}(v_{k+1} | \mathbf{v}_k)}{f_{k+1|1:k}(v_{k+1} | \mathbf{v}_k)} \right) \times \\ &\quad \frac{\partial \mathcal{C}_{A_k|B_k}(F_{k+1|B_k}(v_{k+1} | \mathbf{v}_k), F_{k+2|B_k}(z_{k+2} | \mathbf{v}_k), \dots, F_{p|B_k}(z_p | \mathbf{v}_k))}{\partial F_{k+1|B_k}(v_{k+1} | \mathbf{v}_k)} \\ &= \frac{\partial \mathcal{C}_{A_k|B_k}(F_{k+1|B_k}(v_{k+1} | \mathbf{v}_k), F_{k+2|B_k}(z_{k+2} | \mathbf{v}_k), \dots, F_{p|B_k}(z_p | \mathbf{v}_k))}{\partial F_{k+1|B_k}(v_{k+1} | \mathbf{v}_k)}, \end{aligned} \quad (\text{B.1.12})$$

donde por hipótesis  $\mathcal{C}_{A_k|B_k}$  es gaussiana con parámetro  $\Sigma^{A_k|B_k}$  independiente del valor del condicionante.

Denotemos con  $b_i = \Phi^{-1}(u_i)$ ,  $i = k+1, \dots, p$ . De manera totalmente análoga a como se obtuvo la igualdad (B.1.2) se obtiene una igualdad para la parcial de la cópula en (B.1.12).

$$\frac{\partial}{\partial u_{k+1}} \mathcal{C}_{A_k|B_k}(u_{k+1}, \dots, u_p) = \Phi_{\Sigma^{A_k|B_k}} \left( \frac{b_{k+2} - \rho_{k+2,k+1|B_k} b_{k+1}}{\sqrt{1 - \rho_{k+2,k+1|B_k}^2}}, \dots, \frac{b_p - \rho_{p,k+1|B_k} b_{k+1}}{\sqrt{1 - \rho_{p,k+1|B_k}^2}} \right). \quad (\text{B.1.13})$$

La justificación de (B.1.13) es la misma que para (B.1.2), tomando un v.a.  $\mathbf{Y}^* = (Y_{k+1}, \dots, Y_p)^T \sim \mathcal{N}_p(\mathbf{0}, \Sigma^{A_k|B_k})$ , calculando la distribución de  $(Y_{k+2}, \dots, Y_p)^T | \{Y_{k+1} = b_{k+1}\}$  y posteriormente estandarizando este último v.a.

Por practicidad denotaremos con  $(*) = H_{A_{k+1}|B_{k+1}}(x_{k+2}, \dots, x_p | \mathbf{v}_{k+1})$  a la función de distribución conjunta al paso  $k+1$ . Sustituyendo (B.1.13) en (B.1.12) tenemos que,

$$(*) = \Phi_{\Sigma^{A_{k+1}|B_{k+1}}} \left( \frac{\Phi^{-1}(F_{k+2|B_k}(x_{k+2} | \mathbf{v}_k)) - \rho_{k+2, k+1|B_k} \Phi^{-1}(F_{k+1|B_k}(v_{k+1} | \mathbf{v}_k))}{\sqrt{1 - \rho_{k+2, k+1|B_k}^2}}, \dots, \frac{\Phi^{-1}(F_{p|B_k}(x_p | \mathbf{v}_k)) - \rho_{p, k+1|B_k} \Phi^{-1}(F_{k+1|B_k}(v_{k+1} | \mathbf{v}_k))}{\sqrt{1 - \rho_{p, k+1|B_k}^2}} \right). \quad (\text{B.1.14})$$

Ahora, sean  $u_{k+2}, \dots, u_p \in \mathbb{I}$ , vía (2.4.2) extraeremos la cópula condicional de la distribución condicional,

$$\mathcal{C}_{A_{k+1}|B_{k+1}}(u_{k+2}, \dots, u_p; \mathbf{v}_{k+1}) = H_{A_{k+1}|B_{k+1}}(F_{k+2|B_{k+1}}^{-1}(u_{k+2}; \mathbf{v}_{k+1}), \dots, F_{p|B_{k+1}}^{-1}(u_p; \mathbf{v}_{k+1}) | \mathbf{v}_{k+1}). \quad (\text{B.1.15})$$

De (B.1.15) observamos que necesitamos conocer las expresiones de  $F_{j|B_{k+1}}^{-1}$ , con  $j = k+2, \dots, p$ . Nuevamente recurriremos a la fórmula de Joe. Sea  $x \in \text{Ran}(X_j | \{X_{B_{k+1}} = \mathbf{v}_{k+1}\})$ ,  $j = k+2, \dots, p$ , tenemos que

$$\begin{aligned} F_{j|B_{k+1}}(x | \mathbf{v}_{k+1}) &= \frac{\partial \mathcal{C}_{j, k+1|B_k}(F_{j|B_k}(x | \mathbf{v}_k), F_{k+1|B_k}(v_{k+1} | \mathbf{v}_k); \mathbf{v}_k)}{\partial F_{k+1|B_k}(v_{k+1} | \mathbf{v}_k)} \\ &= \frac{\partial \mathcal{C}_{j, k+1|B_k}(F_{j|B_k}(x | \mathbf{v}_k), F_{k+1|B_k}(v_{k+1} | \mathbf{v}_k))}{\partial F_{k+1|B_k}(v_{k+1} | \mathbf{v}_k)} \\ &= \Phi \left( \frac{\Phi^{-1}(F_{j|B_k}(x | \mathbf{v}_k)) - \rho_{j, k+1|B_k} \Phi^{-1}(F_{k+1|B_k}(v_{k+1} | \mathbf{v}_k))}{\sqrt{1 - \rho_{j, k+1|B_k}^2}} \right), \end{aligned} \quad (\text{B.1.16})$$

la igualdad en (B.1.16) se sigue como caso particular de (B.1.13).

Sea  $u \in \mathbb{I}$ , de (B.1.16) se sigue que

$$F_{j|B_{k+1}}^{-1}(u; \mathbf{v}_{k+1}) = F_{j|B_k}^{-1} \left\{ \Phi \left[ \Phi^{-1}(u) \sqrt{1 - \rho_{j, k+1|B_k}^2} + \rho_{j, k+1|B_k} \Phi^{-1}(F_{k+1|B_k}(v_{k+1} | \mathbf{v}_k)) \right]; \mathbf{v}_k \right\}. \quad (\text{B.1.17})$$

Analizando el argumento  $j$  de  $\Phi_{\Sigma^{A_{k+1}|B_{k+1}}}$  con respecto a  $H_{A_{k+1}|B_{k+1}}(\cdot, \dots, \cdot | \mathbf{v}_{k+1})$  en (B.1.14) evaluado en  $F_{j|B_{k+1}}^{-1}(u_j)$ ,  $j = k+2, \dots, p$ , tenemos que

$$\frac{\Phi^{-1}(F_{j|B_k}(F_{j|B_{k+1}}^{-1}(u_j; \mathbf{v}_{k+1}) | \mathbf{v}_k)) - \rho_{j, k+1|B_k} \Phi^{-1}(F_{k+1|B_k}(v_{k+1} | \mathbf{v}_k))}{\sqrt{1 - \rho_{j, k+1|B_k}^2}} = \Phi^{-1}(u_j), \quad (\text{B.1.18})$$

la igualdad se sigue de (B.1.17).

Finalmente, con base en (B.1.14), (B.1.15) y (B.1.18), se puede concluir que

$$\mathcal{C}_{A_{k+1}|B_{k+1}}(u_{k+2}, \dots, u_p; \mathbf{v}_{k+1}) = \Phi_{\Sigma^{A_{k+1}|B_{k+1}}}(\Phi^{-1}(u_{k+2}), \dots, \Phi^{-1}(u_p)) = \mathcal{C}_{A_{k+1}|B_{k+1}}(u_{k+2}, \dots, u_p),$$

es decir  $\mathcal{C}_{A_{k+1}|B_{k+1}}$  es cópula gaussiana de dimensión  $p - (k+1)$  con parámetro  $\Sigma^{A_{k+1}|B_{k+1}} = \Sigma^{k+2:p|1:k}$  y además éste no depende del valor del condicionante. □

La demostración es análoga para la Proposición 3.3.3 con respecto a las cópulas condicionales de cópula  $\mathcal{T}$ .





## Apéndice C

# Apéndice C: prueba para la fórmula de Joe

Este apéndice plantea una demostración a la fórmula de [Joe \(1996\)](#), Proposición 4.2.1, alineada a la notación incluida en el documento.

**Demostración.** Sea  $x \in \overline{\mathbb{R}}$ . Sin pérdida de generalidad, tomaremos  $A = 1 : m$  con  $1 \leq m < p$ ,  $w_j = m$  y  $k = m + 1$ . Notemos que  $A_{-j} = 1 : m - 1$ . Se sigue que,

$$\begin{aligned}
 F_{k|A}(x|\mathbf{v}) &= F_{m+1|1:m}(x|\mathbf{v}) = \int_{-\infty}^x f_{m+1|1:m}(u|\mathbf{v}) du \\
 &= \int_{-\infty}^x \frac{f_{1:m+1}(\mathbf{v}, u)}{f_{1:m}(\mathbf{v})} du \\
 &= \int_{-\infty}^x \frac{f_{1:m+1}(\mathbf{v}_{-j}, \mathbf{v}_j, u)}{f_{1:m}(\mathbf{v})} du \\
 &= \int_{-\infty}^x \frac{f_{1:m+1}(\mathbf{v}_{-j}, \mathbf{v}_j, u)}{f_{1:m}(\mathbf{v})} \left( \frac{f_{1:m-1}(\mathbf{v}_{-j})}{f_{1:m-1}(\mathbf{v}_{-j})} \right) du \\
 &= \left( \frac{f_{1:m-1}(\mathbf{v}_{-j})}{f_{1:m}(\mathbf{v})} \right) \cdot \int_{-\infty}^x \frac{f_{1:m+1}(\mathbf{v}_{-j}, \mathbf{v}_j, u)}{f_{1:m-1}(\mathbf{v}_{-j})} du \\
 &= \left( \frac{f_{1:m-1}(\mathbf{v}_{-j})}{f_{1:m}(\mathbf{v})} \right) \cdot \int_{-\infty}^x f_{m:m+1}(\mathbf{v}_j, u|\mathbf{v}_{-j}) du \\
 &= \left( \frac{f_{1:m-1}(\mathbf{v}_{-j})}{f_{1:m}(\mathbf{v})} \right) \cdot \frac{\partial H_{m:m+1}(\mathbf{v}_j, x|\mathbf{v}_{-j})}{\partial \mathbf{v}_j} \\
 &= \left( \frac{f_{1:m-1}(\mathbf{v}_{-j})}{f_{1:m}(\mathbf{v})} \right) \cdot \frac{\partial \mathcal{C}_{m:m+1}(F_{m|1:m-1}(\mathbf{v}_j|\mathbf{v}_{-j}), F_{m+1|1:m-1}(x|\mathbf{v}_{-j}); \mathbf{v}_{-j})}{\partial F_{m|1:m-1}(\mathbf{v}_j|\mathbf{v}_{-j})} \times \\
 &\hspace{15em} f_{m|1:m-1}(\mathbf{v}_j|\mathbf{v}_{-j}) \\
 &= \left( \frac{f_{1:m-1}(\mathbf{v}_{-j})}{f_{1:m}(\mathbf{v})} \right) \cdot \frac{\partial \mathcal{C}_{m:m+1}(F_{m|1:m-1}(\mathbf{v}_j|\mathbf{v}_{-j}), F_{m+1|1:m-1}(x|\mathbf{v}_{-j}); \mathbf{v}_{-j})}{\partial F_{m|1:m-1}(\mathbf{v}_j|\mathbf{v}_{-j})} \times \\
 &\hspace{15em} \left( \frac{f_{1:m}(\mathbf{v})}{f_{1:m-1}(\mathbf{v}_{-j})} \right) \\
 &= \frac{\partial \mathcal{C}_{m:m+1}(F_{m|1:m-1}(\mathbf{v}_j|\mathbf{v}_{-j}), F_{m+1|1:m-1}(x|\mathbf{v}_{-j}); \mathbf{v}_{-j})}{\partial F_{m|1:m-1}(\mathbf{v}_j|\mathbf{v}_{-j})} \\
 &= \frac{\partial \mathcal{C}_{w_j k|A_{-j}}(F_{w_j|A_{-j}}(\mathbf{v}_j|\mathbf{v}_{-j}), F_{k|A_{-j}}(x|\mathbf{v}_{-j}); \mathbf{v}_{-j})}{\partial F_{w_j|A_{-j}}(\mathbf{v}_j|\mathbf{v}_{-j})}.
 \end{aligned}$$

□



# Bibliografía

- Aas, K. and Berg, D. (2009). Models for construction of multivariate dependence—a comparison study. *The European Journal of Finance*, 15(7-8):639–659.
- Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and economics*, 44(2):182–198.
- Abdous, B., Genest, C., and Rémillard, B. (2005). Dependence properties of meta-elliptical distributions. In *Statistical modeling and analysis for complex data problems*, pages 1–15. Springer.
- Bacchus, F. and Grove, A. (1995). Graphical models for preference and utility. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pages 3–10. Morgan Kaufmann Publishers Inc.
- Bedford, T. and Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial intelligence*, 32(1-4):245–268.
- Bedford, T. and Cooke, R. M. (2002). Vines: A new graphical model for dependent random variables. *Annals of Statistics*, pages 1031–1068.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 289–300.
- Benjamini, Y. and Liu, W. (1999). A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence. *Journal of Statistical Planning and Inference*, 82(1):163–170.
- Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, pages 1165–1188.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 192–236.
- Borgelt, C., Steinbrecher, M., and Kruse, R. R. (2009). *Graphical models: representations for learning, reasoning and data mining*. John Wiley & Sons.
- Bühlmann, P. and Van De Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications*. Springer.
- Bullmore, E. T. and Bassett, D. S. (2011). Brain graphs: graphical models of the human brain connectome. *Annual review of clinical psychology*, 7:113–140.
- Cai, T., Liu, W., and Luo, X. (2011). A constrained l1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607.
- Casella, G. and Berger, R. L. (2002). *Statistical inference*, volume 2. Duxbury Pacific Grove, CA.
- Castillo, E., Gutierrez, J. M., and Hadi, A. S. (2012). *Expert systems and probabilistic network models*. Springer Science & Business Media.
- Czado, C. (2010). Pair-copula constructions of multivariate copulas. In *Copula theory and its applications*, pages 93–109. Springer.

- Demarta, S. and McNeil, A. J. (2005). The t copula and related copulas. *International statistical review*, 73(1):111–129.
- Dempster, A. P. (1972). Covariance selection. *Biometrics*, pages 157–175.
- Donoho, D. L. et al. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. *AMS Math Challenges Lecture*, pages 1–32.
- Drton, M. and Perlman, M. D. (2007). Multiple testing and error control in gaussian graphical model selection. *Statistical Science*, pages 430–449.
- Drton, M. and Perlman, M. D. (2008). A sinful approach to gaussian graphical model selection. *Journal of Statistical Planning and Inference*, 138(4):1179–1200.
- Durante, F. and Sempi, C. (2010). Copula theory: an introduction. In *Copula theory and its applications*, pages 3–31. Springer.
- Edwards, D. (2012). *Introduction to graphical modelling*. Springer Science & Business Media.
- Efron, B. (2010). *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*, volume 1. Cambridge University Press.
- Embrechts, P., Lindskog, F., and McNeil, A. (2001). Modelling dependence with copulas. *Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich*.
- Embrechts, P., McNeil, A., and Straumann, D. (1999). Correlation: pitfalls and alternatives. *RISK-LONDON-RISK MAGAZINE LIMITED-*, 12:69–71.
- Erdely, A. and Diaz-Viera, M. (2010). Nonparametric and semiparametric bivariate modeling of petrophysical porosity-permeability dependence from well log data. In *Copula Theory and Its Applications*, pages 267–278. Springer.
- Fan, J., Feng, Y., and Xia, L. (2015). A conditional dependence measure with applications to undirected graphical models. *arXiv preprint arXiv:1501.01617*.
- Fang, H.-B., Fang, K.-T., and Kotz, S. (2002). The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis*, 82(1):1–16.
- Fischer, M., Köck, C., Schlüter, S., and Weigert, F. (2009). An empirical analysis of multivariate copula models. *Quantitative Finance*, 9(7):839–854.
- Frahm, G., Junker, M., and Szimayer, A. (2003). Elliptical copulas: applicability and limitations. *Statistics & Probability Letters*, 63(3):275–286.
- Frey, R. and McNeil, A. J. (2003). Dependent defaults in models of portfolio credit risk. *Journal of Risk*, 6:59–92.
- Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.
- Genest, C. and Rémillard, B. (2004). Test of independence and randomness based on the empirical copula process. *Test*, 13(2):335–369.
- Gijbels, I., Veraverbeke, N., and Omelka, M. (2011). Conditional copulas, association measures and their applications. *Computational Statistics & Data Analysis*, 55(5):1919–1932.
- Guo, W. and Rao, M. B. (2008). On control of the false discovery rate under no assumption of dependency. *Journal of Statistical Planning and Inference*, 138(10):3176–3188.
- Haff, I. H., Aas, K., and Frigessi, A. (2010). On the simplified pair-copula construction - simply useful or too simplistic? *Journal of Multivariate Analysis*, 101(5):1296–1310.
- Hammersley, J. M. and Clifford, P. (1971). Markov fields on finite graphs and lattices.

- Han, F. and Liu, H. (2014). High dimensional semiparametric scale-invariant principal component analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(10):2016–2032.
- Han, F., Zhao, T., and Liu, H. (2013). Coda: High dimensional copula discriminant analysis. *The Journal of Machine Learning Research*, 14(1):629–671.
- Hoeffding, W. (1940). *Massstabinvariante korrelationstheorie*. Teubner.
- Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.
- Hu, J. (2010). Dependence structures in chinese and us financial markets: a time-varying conditional copula approach. *Applied Financial Economics*, 20(7):561–583.
- Huang, J.-J., Lee, K.-J., Liang, H., and Lin, W.-F. (2009). Estimating value at risk of portfolio by conditional copula-garch method. *Insurance: Mathematics and economics*, 45(3):315–324.
- Janssen, P., Swanepoel, J., and Veraverbeke, N. (2015). Bernstein estimation for a copula derivative with application to conditional distribution and regression functionals. *TEST*, pages 1–24.
- Jaworski, P., Durante, F., Hardle, W. K., and Rychlik, T. (2010). *Copula theory and its applications*. Springer.
- Joe, H. (1996). Families of  $m$ -variate distributions with given margins and  $m(m-1)/2$  bivariate dependence parameters. *Lecture Notes-Monograph Series*, pages 120–141.
- Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. CRC Press.
- Joe, H. (2014). *Dependence modeling with copulas*. CRC Press.
- Jordan, M. I. (2004). Graphical models. *Statistical Science*, pages 140–155.
- Kauermann, G., Schellhase, C., and Ruppert, D. (2013). Flexible copula density estimation with penalized hierarchical b-splines. *Scandinavian Journal of Statistics*, 40(4):685–705.
- Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 419–430.
- Killiches, M., Kraus, D., and Czado, C. (2015). Model distances for vine copulas in high dimensions with application to testing the simplifying assumption. *arXiv preprint arXiv:1510.03671*.
- Koller, D. and Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.
- Kotz, S. and Nadarajah, S. (2004). *Multivariate  $t$ -distributions and their applications*. Cambridge University Press.
- Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association*, 53(284):814–861.
- Kurowicka, D. and Cooke, R. M. (2006). *Uncertainty analysis with high dimensional dependence modelling*. John Wiley & Sons.
- Lauritzen, S. L. (1996). *Graphical models*. Oxford University Press.
- Liu, H., Han, F., Yuan, M., Lafferty, J., Wasserman, L., et al. (2012a). High-dimensional semiparametric gaussian copula graphical models. *The Annals of Statistics*, 40(4):2293–2326.
- Liu, H., Han, F., and Zhang, C.-h. (2012b). Transelliptical graphical models. In *Advances in Neural Information Processing Systems*, pages 809–817.
- Liu, H., Lafferty, J., and Wasserman, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *The Journal of Machine Learning Research*, 10:2295–2328.
- Mai, Q. and Zou, H. (2015). Sparse semiparametric discriminant analysis. *Journal of Multivariate Analysis*, 135:175–188.

- Mardia, K. V. (1962). Multivariate pareto distributions. *The Annals of Mathematical Statistics*, pages 1008–1015.
- Mari, D. D. and Kotz, S. (2001). *Correlation and dependence*, volume 518. World Scientific.
- McNeil, A. J. (2008). Sampling nested archimedean copulas. *Journal of Statistical Computation and Simulation*, 78(6):567–581.
- Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, pages 1436–1462.
- Mesfioui, M. and Quessy, J.-F. (2008). Dependence structure of conditional archimedean copulas. *Journal of Multivariate Analysis*, 99(3):372–385.
- Min, A. and Czado, C. (2011). Bayesian model selection for d-vine pair-copula constructions. *Canadian Journal of Statistics*, 39(2):239–258.
- Murphy, K. P. (2006). Undirected graphical models.
- Murray, I. and Ghahramani, Z. (2004). Bayesian learning in undirected graphical models: approximate mcmc algorithms. In *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, pages 392–399. AUAI Press.
- Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.
- Nikoloulopoulos, A. K., Joe, H., and Li, H. (2012). Vine copulas with asymmetric tail dependence and applications to financial return data. *Computational Statistics & Data Analysis*, 56(11):3659–3673.
- Palaro, H. P. and Hotta, L. K. (2006). Using conditional copula to estimate value at risk. *Journal of Data Science*, 4:93–115.
- Patton, A. J. (2001a). Estimation of copula models for time series of possibly different lengths. *U of California, Econ. Disc. Paper*, (2001-17).
- Patton, A. J. (2001b). Modelling time-varying exchange rate dependence using the conditional copula.
- Patton, A. J. (2002). *Applications of copula theory in financial econometrics*. PhD thesis, University of California, San Diego.
- Patton, A. J. (2006). Modelling asymmetric exchange rate dependence\*. *International economic review*, 47(2):527–556.
- Pircalabelu, E. (2015). New methodologies for model selection and estimation of probabilistic graphical models. *status: accepted*.
- Pourahmadi, M. (2013). *High-Dimensional Covariance Estimation: With High-Dimensional Data*. John Wiley & Sons.
- Radhakrishna Rao, C. (1981). A lemma on g-inverse of a matrix and computation of correlation coefficients in the singular case. *Communications in Statistics-Theory and Methods*, 10(1):1–10.
- Ripley, B. D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 172–212.
- Robins, G., Pattison, P., Kalish, Y., and Lusher, D. (2007). An introduction to exponential random graph ( $p^*$ ) models for social networks. *Social networks*, 29(2):173–191.
- Rodriguez, J. C. (2007). Measuring financial contagion: A copula approach. *Journal of Empirical Finance*, 14(3):401–423.
- Roth, M. (2012). On the multivariate t distribution.
- Ruschendorf, L. (1976). Asymptotic distributions of multivariate rank order statistics. *The Annals of Statistics*, pages 912–923.
- Salvador, R., Suckling, J., Schwarzbauer, C., and Bullmore, E. (2005). Undirected graphs of frequency-dependent functional connectivity in whole brain networks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1457):937–946.

- Sancetta, A. and Satchell, S. (2004). The bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric Theory*, 20(03):535–562.
- Sarkar, S. K. (2002). Some results on false discovery rate in stepwise multiple testing procedures. *Annals of statistics*, pages 239–257.
- Savu, C. and Tiede, M. (2010). Hierarchies of archimedean copulas. *Quantitative Finance*, 10(3):295–304.
- Schweizer, B. and Wolff, E. F. (1981). On nonparametric measures of dependence for random variables. *The annals of statistics*, pages 879–885.
- Shen, X., Zhu, Y., and Song, L. (2008). Linear b-spline copulas with applications to nonparametric estimation of copulas. *Computational Statistics & Data Analysis*, 52(7):3806–3819.
- Siburg, K. F. and Stoimenov, P. A. (2008). Gluing copulas. *Communications in Statistics. Theory and Methods*, 37(19):3124–3134.
- Sklar, M. (1959). *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8.
- Stöber, J., Joe, H., and Czado, C. (2013). Simplified pair copula constructions - limitations and extensions. *Journal of Multivariate Analysis*, 119:101–118.
- Studeny, M. (2006). *Probabilistic conditional independence structures*. Springer Science & Business Media.
- Takahasi, K. (1965). Note on the multivariate burr's distribution. *Annals of the Institute of Statistical Mathematics*, 17(1):257–260.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.
- Valdés-Sosa, P. A., Sánchez-Bornot, J. M., Lage-Castellanos, A., Vega-Hernández, M., Bosch-Bayard, J., Melie-García, L., and Canales-Rodríguez, E. (2005). Estimating brain functional connectivity with sparse multivariate autoregression. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1457):969–981.
- Varoquaux, G., Gramfort, A., Poline, J.-B., and Thirion, B. (2010). Brain covariance selection: better individual functional connectivity models using population prior. In *Advances in Neural Information Processing Systems*, pages 2334–2342.
- Veraverbeke, N., Omelka, M., and Gijbels, I. (2011). Estimation of a conditional copula and association measures. *Scandinavian Journal of Statistics*, 38(4):766–780.
- Wainwright, M. J. and Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1-2):1–305.
- Wasserman, S. and Pattison, P. (1996). Logit models and logistic regressions for social networks: I. an introduction to markov graphs andp. *Psychometrika*, 61(3):401–425.
- Whelan, N. et al. (2004). Sampling from archimedean copulas. *Quantitative finance*, 4(3):339–352.
- Yuan, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *The Journal of Machine Learning Research*, 11:2261–2286.
- Zhang, K., Peters, J., Janzing, D., and Schölkopf, B. (2012). Kernel-based conditional independence test and application in causal discovery. *arXiv preprint arXiv:1202.3775*.