



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

INSTITUTO DE GEOFÍSICA

MODELOS BASADOS EN CÓPULAS PARA LA  
SIMULACIÓN ESTOCÁSTICA CONJUNTA DE  
PROPIEDADES DE REDES DE FRACTURAS  
DISCRETAS EN MEDIOS POROSOS  
FRACTURADOS

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

**Doctor en Ciencias de la Terra**

PRESENTA:

**Francisco Mendoza Torres**

DIRECTOR DE TESIS:

Dr. Martín Alberto Díaz Viera

**UNAM**  
POSGRADO

Ciudad Universitaria, Cd. Mx., 2017



# Resumen

*Sin acento*

La motivación principal de este trabajo fue desarrollar un enfoque metodológico para la caracterización, modelado y simulación de redes de fracturas discretas en 2D que tome en cuenta las relaciones de dependencias entre las propiedades de las fracturas. Para la modelación de las dependencias se utilizó la teoría de cópulas con el enfoque no paramétrico de cópulas de Bernstein. Dicho enfoque permitió que se considerara de manera explícita las relaciones orientación-longitud y orientación-longitud-apertura en la modelación, ya que la orientación, al ser un dato direccional, impone nuevas restricciones a la metodología convencional. La importancia radica en que tales relaciones pueden tener consecuencias relevantes en las propiedades de percolación y consecuentemente en las propiedades de flujo y transporte del sistema en estudio.

El enfoque de redes de fracturas discretas consiste en aplicar un método de simulación estocástica booleano, también conocido como método de simulación basado en objetos, donde las fracturas se representan como objetos geométricos simplificados (segmentos de línea en 2D y polígonos en 3D). De esta manera, las fracturas son consideradas explícitamente y pueden ser incluidas en los modelos de flujo y transporte cuando dichas fracturas sean relevantes.

Dentro de la modelación geológica-petrofísica de yacimientos, y desde el punto de vista metodológico, se describen detalladamente la modelación de cada uno de los elementos de una red de fracturas discretas tanto con la metodología estándar como con la metodología propuesta. También se hizo énfasis en mostrar los diagramas de flujo dentro de cada etapa de la modelación, el flujo de trabajo general, así como en mostrar su relación con los modelos numéricos y computacionales.

Esta metodología se aplicó en dos redes de fracturas generadas sintéticamente, pero respetando argumentos geológicos tomados de la literatura. En el primer caso de estudio solamente se modela la relación dirección-longitud y en el otro se modela el trío dirección-longitud-apertura. En ambos ejemplos se mostró que el enfoque común en donde no se toma en cuenta la dependencia genera DFNs que distan mucho de los datos reales cuando hay dependencia. En particular, se muestra que con el enfoque convencional se puede reproducir el comportamiento univariado de cada propiedad de fractura, pero el comportamiento bi- o trivariado es totalmente diferente al modelado con la teoría de cópulas, el cual sí reproduce la estructura de dependencia de los datos.

*→ aclarar siglas*



# Abstract

(DFN)

The main motivation of this work was to develop a methodological approach for the characterization, modeling and simulation of discrete fractures networks in 2D that considers the dependence structure between the properties of the fracture properties. For modeling of dependence, copula theory was used with the non-parametric approach of Bernstein's copulas. This approach allowed explicit consideration of the orientation-length and orientation-length-aperture relationships in the modeling, since the orientation, being a directional data, imposes new restrictions to the conventional methodology. Such relationships can have relevant consequences on the percolation properties and consequently on the flow and transport properties of the system being studied.

The approach of discrete fracture networks consists in applying a Boolean stochastic simulation method, also known as an object-based simulation method, where fractures are represented as simplified geometric objects (2D line segments and 3D polygons). In this way, fractures are considered explicitly and can be included in the flow and transport models when these fractures are relevant.

Within the geological-petrophysical reservoir modeling, and from a methodological point of view, it is described in detail the modeling of each of the elements of a discrete fracture network with both the standard methodology and the proposed methodology. Emphasis was also placed on showing flowcharts within each stage of modeling as well as showing their relationship with numerical and computational models.

This methodology was applied in two fracture networks generated synthetically but respecting geological arguments taken from the literature. In the first case study, only the direction-length relationship is modeled and in the other, the triad direction-length-aperture is modeled. In both examples, it was shown that the common approach where the dependency is not considered generates DFNs that are very far from the actual data when there is dependence. In particular, it is shown that with the conventional approach the univariate behavior of each fracture property can be reproduced, but the bi- or trivariate behavior is totally different from that of the copula theory, which actually reproduces the dependence structure of the data.



# Índice general

	Página
<b>Resumen</b>	<b>I</b>
<b>Abstract</b>	<b>III</b>
<b>Índice general</b>	<b>VII</b>
<b>Lista de figuras</b>	<b>XII</b>
<b>Lista de tablas</b>	<b>XIII</b>
<b>1. Introducción</b>	<b>1</b>
<b>2. Modelos de fracturas en el marco de la modelación geológico-petrofísica de yacimientos</b>	<b>5</b>
2.1. Modelación geológico-petrofísica de yacimientos . . . . .	5
2.1.1. Modelo geológico . . . . .	6
2.1.1.1. Geometría y arquitectura del yacimiento . . . . .	6
2.1.1.2. Marco estratigráfico . . . . .	7
2.1.1.3. Modelo <del>Litológico</del> <i>minúscula</i> . . . . .	9
2.1.1.4. Modelado de heterogeneidades . . . . .	11
2.1.2. Modelo petrofísico . . . . .	12
2.1.2.1. Estimación de propiedades petrofísicas . . . . .	12
2.1.2.2. Metodología de simulación de propiedades petrofísicas y técnicas geoestadísticas para modelar variables continuas <i>sin acento</i> . . . . .	14
2.2. Modelos de fracturas . . . . .	16
2.2.1. Método Gaussiano truncado utilizando simulación indicador . . . . .	16
2.2.2. Modelo con geoestadística multi-punto . . . . .	18
2.2.3. Modelos booleanos . . . . .	19

<b>3. Modelos booleanos basados en objetos estocásticos</b>	<b>23</b>
3.1. Procesos puntuales . . . . .	23
3.2. Construcción del modelo booleano . . . . .	25
3.3. El funcional que evita del modelo booleano . . . . .	26
3.4. Propiedades de estabilidad . . . . .	27
3.5. Simulación del modelo booleano . . . . .	29
3.6. El caso estacionario . . . . .	34
<b>4. Estadística de datos orientados</b>	<b>37</b>
4.1. Variables aleatorias . . . . .	37
4.2. Variables orientadas . . . . .	41
<b>5. Modelación de dependencias mediante la teoría de cópulas.</b>	<b>45</b>
5.1. Generalidades . . . . .	45
5.2. Cópulas bivariadas de Bernstein . . . . .	51
5.3. Cópulas bivariadas para vectores aleatorios con componentes orientadas . . . . .	56
5.4. Simulación conjunta de variables aleatorias: caso bivariado . . . . .	61
5.5. Simulación de conjunta de variables aleatorias: caso multivariado . . . . .	63
5.6. Vine cópulas: Caso trivariado . . . . .	63
5.7. Cópulas de Bernstein: caso general multivariado . . . . .	64
5.8. Cópulas multivariadas para vectores aleatorios con componentes orientadas . . . . .	66
<b>6. Metodología de simulación de redes de fracturas discretas</b>	<b>69</b>
6.1. Análisis, modelado y simulación de variables aleatorias . . . . .	69
6.2. Análisis, modelado y simulación de la estructura de dependencia entre variables aleatorias . . . . .	71
6.3. Diagrama de flujo para el análisis, modelado y simulación de redes de fracturas discretas tomando en cuenta su estructura de dependencia . . . . .	74
<b>7. Casos de estudio</b>	<b>77</b>
7.1. Caso bivariado orientación-longitud . . . . .	77
7.1.1. Descripción de los datos . . . . .	77
7.1.2. La metodología estándar . . . . .	83
7.1.3. El enfoque con la cópula de Bernstein . . . . .	87
7.2. Caso de estudio trivariado: orientación-longitud-apertura. . . . .	89
7.3. Resultados y discusión . . . . .	93
<b>8. Conclusiones y trabajo futuro</b>	<b>95</b>
<b>A. Teoría de la aproximación de funciones</b>	<b>99</b>
A.1. Curvas de Bézier . . . . .	100

<i>Índice general</i>	VII
<b>B. Software</b>	<b>105</b>
B.1. Introducción . . . . .	105
B.2. Validación del software . . . . .	105
<b>Bibliografía</b>	<b>118</b>



# Índice de figuras

2.1. a) fracturas; b) fracturas digitalizadas; c) Fracturas pixeladas; d) centroides de fracturas. . . . .	17
2.2. a) Variograma; b) Simulación. . . . .	18
2.3. Simulaciones basadas en la geoestadística multi-punto. . . . .	19
2.4. Modelado de una fractura. Izquierda: fractura real; derecha: su modelo de placas paralelas. . . . .	20
3.1. Ejemplos de procesos puntuales. a) Caso Homogéneo, b) Intensidad no uniforme c) Proceso de conglomerados. . . . .	24
3.2. a) un ejemplo simple de modelo booleano cuyos objetos son círculos; b) Segmentos lineales. . . . .	26
4.1. Ejemplo de función de distribución de probabilidad. . . . .	38
4.2. Función de distribución empírica. Los rectángulos en negro con líneas de guiones se muestra el correspondiente arreglo computacional para almacenar los valores de $\hat{F}_n$ . . . . .	39
4.3. Motivación de la estadística de datos orientados. ¿Cuál es la media? ¿Y la desviación estándar? . . . . .	41
5.1. Donde $i, j, k \in \{1, \dots, n\}$ . Tonos más oscuros representan valores de $n\hat{C}_n$ más grandes. Cada cambio en tono representa una diferencia de uno (por la función indicador). . . . .	48
5.2. Representación gráfica de la cópula empírica y su relación con su almacenaje computacional en una matriz. El centro de cada círculo representa una pseudo-observación $\mathbf{u}_i = (u_i, v_i)$ . La malla regular de líneas de guiones representa el arreglo en la memoria, $n\mathbf{C}$ , que guarda los valores enteros $C_{ij} := n\hat{C}_n(u_i, v_j)$ . Donde $i$ denota el rengón y $j$ la columna. Nótese que las $u_i$ están ordenadas crecientemente pero no existe un orden en las $v_i$ . Alternativaemnte se pudo haber ordenado por $v_i$ y dejar las $u_i$ sin orden. . . . .	49

5.3. Ejemplo de la aplicación de la teoría de cópulas de Bernstein (Erdely y Díaz-Viera, 2012). . . . .	51
5.4. Representación esquemática de los pesos $p_{ij}$ de la densidad de cópula. . . . .	56
5.5. Corrección para que la cópula de Bernstein cumpla con la condición de periodicidad en la primera variable (Ecuación 5.30). . . . .	58
5.6. Densidad marginal estimada (top), función de distribución (medio) y gráficos de cuantiles (abajo) para dirección del viento (izquierda) y para la precipitación pluvial (derecha). Tomada de Carnicero <i>et al.</i> (2013). . . . .	59
5.7. Función de densidad conjunta estimada (utilizando cópulas de Bernstein) de la variable circular-lineal ( $\Theta, X$ ) . La cual describe la dirección del viento y la precipitación pluvial. . . . .	60
5.8. Densidad marginal estimada (top) y función de distribución (abajo) para la dirección de dos boyas. . . . .	61
5.9. Histogramas de un ejemplo de simulación no paramétrica. 473 datos (izquierda) y 100 datos simulados (derecha). . . . .	62
5.10. Diagrama de árbol que muestra cómo se van agregando la condición de periodicidad a las variables aleatorias para la cópula de Bernstein multivariada. Aquí se muestra el caso para 3 variables aleatorias periódicas. Nótese que también muestra la cantidad de sumandos de los elementos de las esquinas, que para este caso es $2^3$ . . . . .	67
6.1. Diagrama de flujo para obtener una simulación de una variable aleatoria. En este caso puede ser la orientación o la longitud. Compárese con la Figura 4.1. . . . .	70
6.2. Diagrama de flujo para obtener las pseudo-observaciones. Compárese con la Figura 4.1. . . . .	71
6.3. Obtención del diagrama de pseudo-observaciones con el cual se estudia la dependencia entre dos variables aleatorias. Las funciones de distribución marginales que se muestran son continuas pero pueden ser no-continuas o una combinación de ambas. . . . .	72
6.4. Diagrama de dispersión con múltiples zonas remarcadas a fin de hacer un análisis exploratorio preliminar. . . . .	73
6.5. Diagrama de dispersión en el cual se han resaltado dos zonas con dependencias más simples. . . . .	73
6.6. Modelación de la cópula a partir de las pseudo-observaciones. . . . .	74
6.7. Modelación bivariada de propiedades de objetos booleanos. . . . .	75
6.8. Diagrama de flujo para obtener ubicaciones espaciales a través de un proceso puntual. Desde el punto de vista computacional, la parte superior del diagrama indica crear un arreglo vacío que almacene las coordenadas de todos los vértices $\mathbf{x}_i$ . . . . .	75
6.9. Flujo de trabajo para la caracterización, modelado y simulación de redes de fracturas discretas en medios porosos fracturados. . . . .	76
7.1. Representación gráfica de la red de fracturas de los datos sintéticos. Aproximadamente el 10 % de las fracturas son de color negro. . . . .	79

7.2.	Roseta de los datos a analizar. $f_0$ está en la dirección N-S, $f_{90}$ en la dirección perpendicular (E-O). . . . .	80
7.3.	Histograma de las longitudes de fractura. La longitud se distribuye lognorm(meanlog = 3.00, sdlog = 0.73) . . . . .	81
7.4.	Función de distribución acumulativa y su respectivo modelo ajustado mediante la función cuantil de Bernstein-Kantorovich. Izquierda: dirección; Derecha: longitud. . . . .	82
7.5.	Diagrama de dispersión de la longitud vs dirección (izquierda), y su correspondiente gráfico de pseudo-observaciones (derecha). . . . .	83
7.6.	Roseta de orientaciones de una simulación. Compárese con Figura 7.2 . . . . .	84
7.7.	Histograma de longitudes de una simulación. Compárese con la Figura 7.3. . . . .	85
7.8.	Funciones de distribución acumulativa empírica de los datos y sus respectivos modelos. Izquierda: Direcciones; Derecha: longitudes. . . . .	86
7.9.	Diagramas de dispersión de la simulación. Izquierda: los datos; derecha: las pseudo-observaciones. . . . .	86
7.10.	DFN de una simulación de los datos. Compárese con Figura 7.1. . . . .	87
7.11.	Diagrama de dispersión y gráfico de pseudo-observaciones de la simulación que toma en cuenta la modelación de la estructura de dependencia. Compárese con la Figura 7.5 y con la Figura 7.9. . . . .	88
7.12.	DFN de la simulación basada en las cónulas de Bernstein. Compárese con Figura 7.1 y con Figura 7.12. . . . .	88
7.13.	Sistema de fracturas a analizar. . . . .	89
7.14.	Matrix de diagramas de dispersión con las correspondientes distributiones marginales en las entradas de la diagonal. . . . .	90
7.15.	Diagramas de dispersión de las pseudo-observaciones. . . . .	91
7.16.	Distribuciones marginales en la diagonal dentro de la matriz de simulaciones. . . . .	92
7.17.	Gráfico de pseudo-observaciones simuladas. . . . .	92
A.1.	Aproximación vs interpolación (línea continua). . . . .	99
A.2.	Interpolación de Lagrange. Observe las oscilaciones en la curva polinomial aún cuando no se observa tal oscilación en los puntos de control. . . . .	100
A.3.	Curva de Bézier para el caso de puntos equiespaciados a lo largo del eje $x$ . . . . .	101
A.4.	Sean $P$ y $Q$ dos puntos en el conjunto $S$ . (a) el segmento rectilíneo $\overline{PQ}$ yace completamente dentro del conjunto convexo $S$ ; (b) ejemplo que muestra un conjunto no convexo, en el cual parte del segmento rectilíneo yace fuera del conjunto. . . . .	102
A.5.	Líneas que intersectan la curva de Bézier y sus puntos de control. . . . .	103
B.1.	(izquierda) Roseta de orientaciones, (derecha) histograma de precipitación. . . . .	106
B.2.	Figuras equivalentes a la Fig. 2 del artículo.(recuadros superiores) Densidad marginal estimada, (recuadros inferiores) funciones de distribución acumulada, para la dirección del viento (izquierda) y para la precipitación pluvial (derecha) . . . . .	106

B.3. Figura equivalentes a la Fig. 4 del artículo. Diagrama de dispersión de precipitación y dirección del viento (theta). . . . .	107
B.4. Densidad de cópula. . . . .	107
B.5. Función de densidad de probabilidad empírica y ajustada con el enfoque no paramétrico utilizando polinomios de Bernstein-Kantorovich. En gris se muestran los datos observados y la línea es el ajuste. . . . .	108
B.6. Histogramas de ejemplo de simulación no paramétrica utilizando polinomios de Bernstein-Kantorovich para modelar la función de distribución de probabilidad. 473 datos (izquierda) y 100 datos simulados (derecha). . . . .	109
B.7. Tiempo de cómputo de simulaciones de la cópula de Bernstein en función del grado del polinomio. . . . .	109
B.8. Media (recuadro izquierdo) y varianza (recuadro derecho) en función del grado del polinomio de Bernstein de datos simulados. . . . .	110

# Índice de tablas

2.1. Valores de porcentajes de píxeles negros para los resultados en la Figura 2.3. . . . .	20
4.1. Funciones de densidad de probabilidad para datos orientados. $0 \leq \theta < 2\pi$ , $0 \leq \phi < \pi/2$ , $0 \leq \kappa < \infty$ . Ver ec. 3.5.17 de Mardia y Jupp (2000), ec. 2.2.6 en la sección 2.2.4 de Jammalamadaka y SenGupta (2001) y ec. 4.22 de Fisher <i>et al.</i> (1993). . . . .	43
6.1. Símbolos básicos para construir diagramas de flujo. . . . .	69
7.1. Parámetros de la combinación de dos distribuciones de von Mises de donde fueron muestreadas las 400 direcciones de la red de fractura a modelar. . . . .	79
7.2. Estadística circular de las direcciones de los datos sintéticos. . . . .	80
7.3. Estadígrafos de las longitudes de los datos sintéticos. . . . .	81
7.4. Estadígrafos circulares de las direcciones de una simulación de los datos. Compárese con Tabla 7.2. . . . .	84
7.5. Estadígrafos de las longitudes de una simulación de los datos. Compárese con la Tabla 7.3. . . . .	85



## Introducción

La modelación de redes de fracturas es de suma importancia en los medios porosos fracturados, en particular dentro de la modelación geológica-petrofísica. En un contexto dinámico las redes de fracturas pueden afectar el comportamiento del flujo y almacenamiento de fluidos subterráneos. En un contexto estático, también son importantes porque ayudan a explicar parte de la historia geológica de la zona de estudio o forman vetas o corredores de vetas en las que se almacenan minerales. Como Nelson (2001) lo describe: inicialmente hay que considerar que el yacimiento es fracturado.

Debido a que los sistemas fracturas muestran una amplia complejidad (Adler y Thovert, 1999),  
y que, por razones económicas, el volumen de roca muestreada es típicamente una muy pequeña fracción del yacimiento (Chilès y Delfiner, 2012, p. 1), el enfoque de modelación es probabilístico ya que permite cuantificar la incertidumbre de los parámetros medidos o estimados. Es decir, cada propiedad involucrada es usualmente modelada mediante una variable aleatoria.

Gracias al poder computacional que existe actualmente, el modelo de redes de fracturas Discretas (DFN, por sus siglas en inglés) se ha vuelto realizable. Este enfoque estocástico consiste en aplicar un método de simulación basada en objetos, el cual también se conoce como método booleano (Cacas *et al.*, 2001; Chilès y Delfiner, 2012; Stoyan *et al.*, 1987), en donde las fracturas son representadas como objetos geométricos simplificados. Por ejemplo, las fracturas en 2D se representan usualmente como segmentos rectilíneos cuya orientación y longitud representan las características de mismo nombre de las fracturas. Algunas propiedades de fracturas como la porosidad y la apertura no se modelan geométricamente pero sí como atributos de las fracturas.

La metodología estándar supone independencia entre las variables aleatorias que modelan cada propiedad, ya que cada propiedad de fractura se modela de manera independiente (Adler y Thovert, 1999; Bourbiaux *et al.*, 2002; Elmo *et al.*, 2014; Gringarten, 1997; Zellou *et al.*, 2003). Se ha demostrado (Mendoza-Torres *et al.*, 2017) que cuando existe dependencia, las redes de fracturas simuladas pueden tener un ~~comportamiento~~ totalmente diferente al de los datos, incluso aún obedeciendo las mismas distribuciones individuales (marginales). Por ejemplo, cuando se espera que las fracturas más largas estén alineadas en cierta dirección y el enfoque independiente produce

*sin ac en +*

tales fracturas en la dirección perpendicular.

La dependencia estadística entre las propiedades de las fracturas usualmente es no-lineal y compleja. Por lo tanto, las técnicas estadísticas tradicionales, que generalmente se basan en suposiciones de linearidad, son muy restrictivas para modelar dichas redes.

Algunos autores (Balankin *et al.*, 2001; Olson, 2007) reportan que las longitudes y aperturas de fracturas siguen distribuciones de probabilidad con colas pesadas, lo que a su vez provoca diagramas de dispersión complicados, no solamente entre las variables, sino también entre otras variables que no necesariamente tienen colas pesadas.

En particular, la dependencia entre longitud, orientación y apertura, tanto por pares como de manera trivariada, es no lineal y generalmente no se modela. Esto también debido a la falta de modelos para la dependencia entre datos orientados y datos convencionales como la longitud. Cabe notar, que *estas* relaciones de dependencia son muy importantes dentro de un contexto de flujo y transporte ya que pueden afectar severamente las propiedades de percolación.

Un enfoque ampliamente utilizado para modelar las relaciones de dependencia es el enfoque de regresión lineal. Debido a sus limitaciones, muchas veces ajustan modelos de manera artificial al transformar alguna de las variables involucradas o ambas. Decimos de manera artificial por dos aspectos principales: 1) no es claro que la regresión satisfaga los requisitos de la teoría de regresión lineal; 2) el sesgo que se obtiene al transformar los datos (Box, 1971; Miller, 1984; Seber y Wild, 2003); y 3) el esfuerzo que conlleva comparar transformaciones y evaluar los supuestos de la regresión.

*Intimamente* con la regresión se tienen los coeficientes de correlación. Éstos son estadígrafos que intentan explicar la dependencia entre los datos. En ciencias de la tierra, comúnmente se tienen estructuras de dependencia entre los datos que los coeficientes de correlación no pueden explicar. Por otro lado, se sabe que dos conjuntos de datos pueden tener el mismo coeficiente de correlación y, por el contrario, tener estructura de dependencia diferente (Chernih *et al.*, 2007; Embrechts *et al.*, 1999; Kat, 2003; King, 1986), es decir, pueden ser *espurios*. ?? *sin ac en +*

Como solución a éstas limitantes, la teoría de cópulas ha mostrado modelar de manera flexible y general una gran gama de estructuras de dependencia. Una de las ventajas de *esta* teoría es poder separar la estructura de dependencia de las funciones de distribución marginales univariadas (Sklar, 1959).  \*

Una función cópula válida es la aproximación de la cópula empírica (Berghaus *et al.*, 2017; Bucher y Volgushev, 2011; Carnicero *et al.*, 2013; Deheuvels, 1979; Fermanian *et al.*, 2004; Radulović *et al.*, 2017) mediante los polinomios de Bernstein (Sancetta y Satchell, 2004). Éste enfoque no paramétrico es más fácil de usar que otros en los que se utilizan cópulas paramétricas, no depende de la forma de la dependencia y la reproduce de manera adecuada.

Sin embargo, una de sus limitantes es no poder modelar de manera natural datos orientados como lo son las direcciones de fracturas. Carnicero *et al.* (2013) adaptaron *este* modelo para incluir datos periódicos. La teoría de cópulas ya ha tenido aplicaciones dentro de Ciencias de la Tierra, por ejemplo en geoestadística (Bárdossy, 2006; Díaz-Viera y Casar-González, 2005a; Haslauer *et al.*, 2010; Kazianka, 2012; Kazianka y Pilz, 2010, 2011) o en valores extremos (Salvadori *et al.*, 2007). En particular, las cópulas de Bernstein se han utilizado para datos no periódicos (Erdely y

*sin ac-en +*

Díaz-Viera, 2009; Hernández-Maldonado *et al.*, 2012b,c), pero ésta tesis es el primer proyecto en considerar explícitamente la función de relación entre dirección de fracturas y longitud y apertura de la misma.

El desarrollo de éste trabajo produjo una metodología sistemática para la simulación estocástica conjunta de las propiedades geométricas con dependencias complejas de los objetos que representan fracturas discretas mediante la modelación de su función de distribución de probabilidad conjunta usando cópulas.

En el Capítulo 2 se muestra el papel de las redes de fracturas dentro de la modelación geológica-petrofísica de yacimientos así mismo se muestran algunos enfoques de modelación de redes de fracturas, entre ellos el modelo adoptado en esta tesis. Los fundamentos matemáticos de parte de la geometría estocástica concerniente a los modelos booleanos se muestra en el Capítulo 3: procesos puntuales y modelos booleanos. Estos modelos tienen objetos que a su vez tienen propiedades. La orientación de las fracturas es una de ellas para la cual se utiliza la estadística de datos orientados explicada en el Capítulo 4. En el siguiente capítulo se muestra a detalle el tema principal de éste trabajo. Se muestran las bases de la teoría de cópulas, el caso de cópulas de Bernstein y su adaptación para incluir datos orientados. También se muestran algunos resultados multivariados así como el enfoque de cópulas de Vine para el caso trivariado. Cabe mencionar que a lo largo de éste capítulo se relacionan los modelos matemáticos, sus algoritmos numéricos, su almacenaje en memoria computacional y el nombre del código que contiene dicha implementación. El Capítulo 6 es una síntesis en flujos de trabajo del capítulo anterior, así como también se proporciona una metodología general para simulación estocástica de redes de fracturas discretas. En el Capítulo 7 se aplica la metodología y los programas computacionales en dos casos de estudio: uno bivariado y otro trivariado. Aquí se comparan los resultados con la metodología estándar. En el Capítulo 8 se muestran las conclusiones y trabajo futuro. Por último, la teoría de aproximación se dejó en el Apéndice A ya que aunque importante para distinguirla de la interpolación, no es totalmente necesaria dentro del cuerpo de la tesis.

De esta manera, en este trabajo se enriquece la metodología de la modelación de redes de fracturas discretas al considerar la dependencia que existe entre las propiedades intrínsecas de los modelos booleanos, en particular, en la modelación de redes de fracturas discretas. Se brindan diagramas de flujo de trabajo en cada parte de la modelación así como un flujo de trabajo general. Esta metodología se aplicó a dos casos de estudio que reprodujeron la estructura de dependencia de los datos; en contraparte, la metodología estándar produjo resultados bivariados muy diferentes a los esperados.



# Modelos de fracturas en el marco de la modelación geológico-petrofísica de yacimientos

## 2.1. Modelación geológico-petrofísica de yacimientos

La caracterización estática de yacimientos se basa en el empleo de técnicas geoestadísticas, las cuales permiten obtener modelos numéricos que representan el modelo geológico y el modelo petrofísico de un yacimiento. Las técnicas geoestadísticas facilitan combinar información de manera integral y analizar la influencia de los diversos parámetros que intervienen en el modelo. El objetivo al aplicar técnicas geoestadísticas es obtener un modelo numérico confiable de los yacimientos, además que, por tratarse de técnicas probabilísticas, es factible realizar un análisis del grado de incertidumbre asociado a los modelos.

En el proceso de caracterizar un yacimiento, se parte del supuesto de que los datos conocidos nunca son suficientes, ni exhaustivos para describir un yacimiento de una manera completa y exacta. Bajo este enfoque, los métodos geoestadísticos representan una herramienta práctica mediante la cual es posible realizar distribuciones de facies y de parámetros petrofísicos consistentes con los datos conocidos y con el conocimiento geológico de los procesos que formaron los yacimientos. En muchas ocasiones no se comprende el potencial de la modelación geoestadística, esto se debe a que en ocasiones prevalece la idea de que al aplicar los modelos estocásticos se está empleando el azar, como si se lanzara un dado para decidir si en una determinada ubicación hay un tipo u otro de facies. En realidad, la modelación estocástica se basa en la estadística de la información conocida y en un modelo de correlación espacial de la información, además de que permite evaluar el grado de incertidumbre que se tiene acerca del modelo realizado, la cual es el reflejo de la complejidad geológica-petrofísica del yacimiento.

Es importante señalar que la MGPy parte de un modelo geológico conceptual, el cual incluye el ambiente en el cual se realizó el depósito de sedimentos y un marco estratigráfico producto de la sedimentación y la posterior formación de las rocas que constituyen el yacimiento. El trabajo de formular un modelo geológico involucra disciplinas tales como: geología, estratigrafía de secuen-

*clarar siglas*

cias, sedimentología, interpretación de registros de pozos, bioestratigrafía, estudios geológicos de análogos de afloramientos e interpretación sísmica. El modelo sedimentológico depositacional del yacimiento debe de proporcionar una evaluación semi-cuantitativa de los parámetros geométricos respecto de la extensión y distribución de la facies, así como la relación que guardan entre sí. Dicha información es básica para el proceso de modelación estocástica del yacimiento.

En tema de geología estructural se debe de tener identificadas y definidas las características estructurales del yacimiento como son las fallas regionales, locales y las fracturas, este trabajo se apoya principalmente en la interpretación de la información sísmica, pero también en la información de pozos, el estudio de núcleos extraídos de los pozos y en trabajos de geología de campo en análogos de yacimientos que pueden aportar información útil o corroborar la información que se interpreta de la sísmica.

Con respecto de la información petrofísica, *esta* se obtiene de las diferentes curvas de los registros geofísicos de pozo y por lo general se realiza un proceso de calibración con la información petrofísica obtenida de los núcleos mediante pruebas de laboratorio. La integración de datos dinámicos como son pruebas de presión, pruebas de trazadores y datos de producción, ofrecen información a gran escala relativa al flujo y en ocasiones puede ser incluida en el modelo geológico conceptual.

### **2.1.1. Modelo geológico**

La definición del modelo geológico constituye una de las fases más importantes de la caracterización de un yacimiento debido al volumen de trabajo que involucra y por el impacto que tiene en el resultado. Consta de las siguientes etapas:

1. Geometría y arquitectura del yacimiento
2. Marco estratigráfico
3. Modelo litológico
4. Modelado de heterogeneidades

#### **2.1.1.1. Geometría y arquitectura del yacimiento**

Consiste en establecer e identificar la arquitectura básica del yacimiento, es decir se deben de definir las fronteras del depósito. La arquitectura del yacimiento involucra identificar e interpretar las principales fallas, así como las superficies que representan la cima y la base del yacimiento. Más aún de debe de identificar las fallas que pueden actuar como unidades de flujo o aquellas que constituyen barreras impermeables. Mientras que las características estructurales como las fallas regionales son definidas de manera determinística usando datos sísmicos, las características de las fallas locales y/o fracturas pueden ser simuladas usando modelos estocásticos. Los datos sísmicos, proporcionan información sobre la distribución lateral de los cuerpos geológicos, son imágenes en tiempo, amplitud o impedancia, y pueden ser usada para el cálculo de las funciones de distribución espacial de las facies y también pueden ser integrados en el proceso de simulación de propiedades

petrofísicas como variables secundarias, siempre y cuando posean una buena correlación entre la característica petrofísica y el atributo sísmico.

### Definición de la arquitectura del yacimiento

Consiste en identificar el marco geométrico básico de la trampa de hidrocarburos, lo cual incluye definir las fronteras del mismo, en particular el mapa estructural de la parte superior del yacimiento o cima estructural. El mapa estructural de la parte superior o cima del yacimiento se define a partir de: Datos sísmicos apilados en 2D o 3D que producen mapas en tiempo de la estructura del yacimiento y luego se convierten a profundidad mediante un modelo de velocidad de la unidad geológica involucrada. Datos de pozos que son usados cuando se carece de la sísmica o ésta es de mala calidad. En este caso puede ser alta la incertidumbre entre pozos, no obstante, si se tiene una malla densa de pozos puede incluso dar mejores resultados que la información sísmica.

### Modelado de Fallas

Las fallas son identificadas en base a tres tipos principales de información:

Evidencia geológica: Detectados mediante la inconsistencia en el esquema de correlación estratigráfica.

Evidencia en pozos: Ausencia o repetición de espesores. Cuando faltan secuencias o espesores se puede relacionar con fallas normales, mientras que cuando se repiten secuencias o espesores se relacionan con fallas inversas. No obstante, los pozos verticales tienen poca probabilidad de cruzar una falla comparados con los desviados y por lo tanto se usan más para validar localmente la interpretación sísmica.

Datos sísmicos: Las fallas pueden ser detectadas a partir de discontinuidades en los patrones de reflexión luego de haberse eliminado el ruido y migrado la señal (corregida la posición en espacio). Cuando existen datos sísmicos en 3D de buena calidad se puede hacer una descripción estructural precisa del yacimiento. El grado de detalle depende del tamaño de las características estructurales que deseamos identificar y que tienen un impacto en el flujo. La información sísmica sola no es suficiente para establecer un patrón estructural que sea relevante para el flujo. Se requiere de otras fuentes de información para integrar la interpretación geofísica, como es la información dinámica: las pruebas de pozos, los trazadores, las pruebas de inyección y la información de la producción de hidrocarburos.

#### 2.2.1.3 Procedimiento para la construcción del marco estructural en 3D

En términos generales se siguen los siguientes pasos:

Definición de las fallas principales: Las fallas principales son las que limitan los bloques más grandes del yacimiento. Construcción de las superficies geológicas: Dentro de cada bloque se modelan los principales horizontes geológicos (superior, inferior, y los eventos mayores correlacionables) mediante superficies interpoladas a partir de la información disponible. Modelación de las fallas menores: Las que cortan y desplazan los principales horizontes geológicos.

### 2.1.1.2. Marco estratigráfico

Esta parte se encarga de definir el marco o estructura interna del yacimiento, el proceso considera la identificación y definición de las superficies que delimitan a las principales unidades de

flujo del yacimiento a partir de la correlación de unidades geológicas en los pozos. Otra parte importante es la definición y construcción de la malla estratigráfica, cuyo objetivo principal implica la definición de la geometría interna estratigráfica, en el sentido vertical, la cual debe de representar la arquitectura de las unidades que conforman el yacimiento. El conocimiento geológico es sintetizado mediante funciones de distribución de las facies (curvas de proporción vertical, curvas de probabilidad de ocurrencia y variogramas). El modelo geológico conceptual del yacimiento nos provee de un medio adicional para inferir la longitud de correlación de las facies (rango o alcance del variograma), las dimensiones y los espesores promedio de las unidades geológicas.

Definir el marco estratigráfico y la correlación de unidades geológicas involucra un considerable número de disciplinas y técnicas tales como: sísmica de exploración, sedimentología, estratigrafía de secuencias, interpretación de registros geofísicos de pozos, petrofísica, bioestratigrafía, estudios geológicos de afloramientos, entre otras.

### **Estratigrafía de secuencias**

Esta puede ser definida como el estudio de facies genéticamente relacionadas dentro de un marco crono-estratigráfico. El principio básico detrás de esta definición es que la deposición de patrones de sedimentación fue controlada por cambios en el nivel relativo del mar y este a su vez esta controlados por los movimientos eustáticos, la tectónica y la velocidad de sedimentación. Hay varias razones por las que la estratigrafía secuencia puede ser considerada como una herramienta ideal para la caracterización integral de yacimientos y estas son:

La aplicación de la estratigrafía secuencia a la escala de yacimiento proporciona un marco estratigráfico detallado que pueden reducir el riesgo de correlaciones erróneas entre las diferentes unidades genéticas.

La estratigrafía de secuencias puede ser estudiada a diferentes escalas y en este sentido tiene aplicación en estudios de fractalidad, esto permite la utilización y la integración de los datos tomados a diferentes escalas y con distintas herramientas.

Su aplicación permite la predicción de la presencia, la continuidad y la extensión de las facies de un yacimiento e incluso fuera de las áreas desarrolladas.

Sus principios se pueden aplicar tanto a YAA como a YNF.

### **Construcción de la malla estratigráfica**

Desde el punto de vista estratigráfico la cuestión principal es la definición de la geometría interna para la arquitectura de las unidades geológicas. Existen cuatro posibilidades básicas:

**Capas paralelas:** Estas se construyen en forma paralela siguiendo una superficie, que puede ser la base o la cima, las cuales se verán truncadas una vez que alcanzan la siguiente superficie de referencia. Se define un espesor constante para las celdas. Se consideran paralelas siguiendo la base y se truncan con la cima, o paralelas siguiendo la cima y se truncan contra la base.

**Capas proporcionales:** La suma de las capas será constante, e independiente del espesor de la unidad. Solo se especifica el número de capas a construir.

**Capas por fracciones:** Es una manera de construir capas proporcionales, pero se puede especificar espesores diferentes relativos entre capas. Por ejemplo 1, 2, 1, este código indica que al generar 3 capas la capa de en medio será dos veces el espesor de la capa inferior y superior.

Capas combinadas: Combinaciones entre ellas. Por ejemplo, las capas son paralelas siguiendo la cima y usando una superficie para guiar la manera como las capas serán colocadas.

### 2.1.1.3. Modelo Litológico

La distribución de facies en el yacimiento representa una herramienta poderosa para guiar la distribución de propiedades petrofísicas del mismo, ya que en la mayoría de los casos estos dos aspectos están íntimamente relacionados. Así el modelo de distribución de facies implica la definición de facies, lo cual está basado en el modelo geológico conceptual y en el modelo estratigráfico-sedimentológico del yacimiento. El concepto de facies es particularmente adecuado para los estudios integrales de yacimientos ya que pueden ser consideradas como el volumen elemental práctico del yacimiento y pueden representar el bloque básico para la construcción de modelos geológicos en 3-D. Las facies constituyen el elemento básico para manejar y transferir la información geológica a través de las diferentes etapas de un proceso de modelación.

La clasificación de facies puede simplificarse mediante su agrupación para formar litotipos y/o de clases petrofísicas. Una alternativa práctica y simplista se reduce a la definición de dos tipos de facies: la almacenadora y la que no almacenadora. Cuando se tiene suficiente información y de buena calidad, es posible identificar un número mayor de facies, en ocasiones se puede intentar un enfoque más sofisticado basado en el tratamiento estadístico multivariado de los datos. El proceso de definir facies se realiza, por lo general, de la siguiente manera: (1) Se parte de los pozos que cuentan con núcleos y que a su vez que tengan datos de registros de pozo más completos y confiables, además de estar ubicados en áreas representativas del yacimiento. (2) Las facies se definen e identifican en los registros de pozo, y se ajustan con la información de núcleos, finalmente se agrupan en un número reducido que se denominan litotipos. Para esto último es posible usar técnicas como el análisis de agrupamiento como cluster analysis y/o el método de componentes principales. Cuando las propiedades petrofísicas del yacimiento no están directamente controladas por las facies o litotipos, se puede usar un enfoque alternativo en términos de clases petrofísicas (agrupamiento por propiedades petrofísicas). (3) El esquema final de clasificación es extendido al resto de los pozos.

Uno de los intereses principales en cómo están definidas las facies, tiene que ver en cómo construir distribuciones realistas en 3-D de dichas facies de manera que puedan ser usadas posteriormente durante la modelación del yacimiento. Las facies deben poseer un control significativo sobre parámetros tales como la porosidad, la permeabilidad, o la saturación de agua, de otra manera, la modelación de la distribución de las facies sería de poco beneficio, ya que la incertidumbre no se reduciría y los modelos resultantes no tendrían un adecuado poder predictivo.

El modelo litológico es construido integrando:

Modelo sedimentológico conceptual (representación conceptual del yacimiento)

Clasificación en tipos de facies (litotipos) o de clases petrofísicas.

Distribución de facies (litotipos) o de clases Petrofísicas.

#### Modelo sedimentológico conceptual

El modelo sedimentológico depositacional forma parte del modelo geológico conceptual, este

modelo suministra información cuantitativa o semi-cuantitativa de los parámetros geométricos relativos a dimensiones y espesores de cuerpos sedimentarios que serán útiles para el proceso de modelación estocástica. Esto es que, en el caso de obtener variogramas poco confiables debido a la carencia de datos, se debe optar por el diseño de variogramas de acuerdo con modelo sedimentológico depositacional. La definición del modelo sedimentológico depositacional implica identificar el marco sedimentológico, el cual se refiere a sedimentación marina, fluvial, deltaica, turbidítica, etc., y al proceso depositacional, como puede ser por corrientes de alta o baja energía, flujos de escombros, corrientes de turbidez o bien una depositación o precipitación de origen químico.

### **Definición y clasificación de facies**

En cuanto a la definición, descripción y clasificación de litofacies, estas se realizan comúnmente a partir de las muestras de núcleos y su objetivo es la clasificación de la roca del yacimiento desde el punto de vista litológico, depositacional y/o petrofísico.

Las facies pueden ser consideradas como las unidades básicas de la modelación geológica.

Se distinguen varios tipos de facies como pueden ser:

Litofacies (definidas en núcleos o muestras de campo) Electrofacies (en registros geofísicos de pozo) Sismofacies (en información de sísmica de exploración) Litotipos (grupos de facies) Clases petrofísicas (agrupamiento por propiedades petrofísicas)

### **Definición y clasificación de facies**

El concepto de facies es particularmente adecuado para estudios integrales de yacimientos ya que pueden ser consideradas como el volumen elemental práctico del yacimiento y representan el bloque básico para la construcción de modelos geológicos en 3-D. Las facies pueden considerarse como una herramienta para transferir la información geológica a través de las diferentes etapas de un estudio.

En el estudio más simple se reduce a la definición de dos tipos de facies: la almacenadora y la que no almacenadora. Cuando se tiene suficiente información y de buena calidad, es posible identificar un número mayor de facies, en ocasiones se puede intentar un enfoque más sofisticado basado en el tratamiento estadístico multivariado de los datos. La clasificación de facies puede simplificarse mediante su agrupación para formar litotipos y/o clases petrofísicas. El proceso de definir facies se realiza, por lo general, de la siguiente manera:

Se parte de los pozos que cuentan con núcleos y que a su vez que tengan datos de registros de pozo más completos y confiables, además de estar ubicados en áreas representativas del yacimiento. Las facies se definen e identifican en los registros de pozo, y se ajustan y calibran con la información de núcleos, finalmente se agrupan en un número reducido que se denominan litotipos. Para esto último es posible usar técnicas como el análisis de agrupamiento como cluster analysis y/o el método de componentes principales. Cuando las propiedades petrofísicas del yacimiento no están directamente controladas por las facies o litotipos, se puede usar un enfoque alternativo en términos de clases petrofísicas (agrupamiento por propiedades petrofísicas).

El esquema final de clasificación (litotipos o clases petrofísicas) es extendido al resto de los pozos. Esto puede realizarse usando técnicas de reconocimiento de patrones como redes neuronales.

### **Modelado de la distribución de facies**

Uno de los intereses principales en cómo están definidas las facies, tiene que ver en cómo

construir distribuciones realistas en 3-D de dichas facies de manera que puedan ser usadas posteriormente durante la modelación del yacimiento. Las facies deben poseer un control significativo sobre parámetros tales como la porosidad, la permeabilidad, o la saturación de agua, de otra manera, la modelación de la distribución de las facies sería de poco beneficio, ya que la incertidumbre no se reduciría y los modelos resultantes no tendrían un adecuado poder predictivo.

En lo que respecta a los métodos de simulación estocástica para la modelación de facies, las técnicas empleadas son de dos tipos principalmente:

Basados en Celdas (o continuos):

Las facies se convierten a variables categóricas y se simula mediante técnicas de simulación secuencial combinado con técnicas de estimación de kriging indicador.

Se considera a la variable a ser simulada como una realización de una función aleatoria continua cuya distribución (usualmente Gaussiana) es caracterizada con diferentes umbrales (valores de corte), los cuales identifican a diferentes facies o diferentes rangos de propiedades petrofísicas.

Estos métodos trabajan mejor en presencia de asociaciones de facies que varían suavemente, como es frecuentemente en el caso de los yacimientos arenos-arcillosos. No se hace ninguna suposición acerca de la forma de los cuerpos sedimentarios.

Basados en Objetos (o booleanos):

Generan distribuciones espaciales de cuerpos sedimentarios los cuales son obtenidos mediante la superposición de geometrías simplificadas como: láminas, discos o sinusoides, típicamente simuladas dentro de una facies arcillosa de fondo.

Los parámetros de los objetos (orientación, sinuosidad, longitud, ancho, etc.) pueden ser estimados sobre la base del modelo sedimentológico, la información sísmica, los análogos de yacimiento y las interpretaciones de las pruebas de presión en pozos.

En algunos ambientes depositacionales, especialmente los fluviales con meandros o los de tipo turbidíticos, en donde los canales de arena son medios propicios para formar un yacimiento, estos modelos pueden producir imágenes muy realistas de la arquitectura de dichas facies.

#### 2.1.1.4. Modelado de heterogeneidades

Las heterogeneidades del yacimiento son características geológicas de escala menor que no puede atrapar la malla estratigráfica y que pueden no ser significativas desde el punto de vista estrictamente estático de la caracterización del yacimiento pero que tienen un impacto significativo en el flujo de los hidrocarburos, tal es el caso de las fracturas y los vúrgulos en los yacimientos carbonatados. La relación entre la heterogeneidad del yacimiento y los parámetros dinámicos del campo es uno de los puntos claves de un estudio integral y pueden ser un factor determinante en el grado de detalle y la precisión que puede ser alcanzado en la MGPy. La integración de datos dinámicos como son pruebas de pozos, trazadores y datos de producción mediante las técnicas estocásticas de simulación representa aún un reto de gran importancia, ya que éstos ofrecen información relativa al flujo y a una escala mayor, y que podría ser esencial para la construcción de modelos de yacimientos confiables.

## 2.1.2. Modelo petrofísico

La comprensión de la estrecha relación existente entre la red de poros, las propiedades de la roca y el flujo constituye la parte medular del estudio de los yacimientos, así las propiedades petrofísicas son las que en última instancia controlan el flujo y transporte de fluidos en el yacimiento. Las rocas que forman parte del yacimiento son usualmente interpretadas como un medio poroso, donde el flujo de fluidos tiene lugar a través de una red interconectada de espacios de poro. Las características y las propiedades de esta red porosa, son las propiedades petrofísicas, que están íntimamente relacionadas con la distribución original del tamaño de los granos de la roca del yacimiento, así como de los procesos posteriores de diagénesis que ha sufrido ésta a lo largo de su historia geológica.

En presenta una breve explicación de los procedimientos de estimación e interpretación de las principales propiedades petrofísicas: porosidad, saturación de agua y permeabilidad, a nivel de pozo usando mediciones en núcleos y registros de pozos; posteriormente se discute la metodología de simulación de propiedades petrofísicas y se mencionarán las principalmente técnicas geoestadísticas más utilizadas para modelar variables continuas y como que integran la información a escala de pozo con la información sísmica con el fin de derivar los modelos 3D de las propiedades petrofísicas.

### 2.1.2.1. Estimación de propiedades petrofísicas

La porosidad se define como la relación entre el volumen de espacio de poro y el volumen bruto de la roca del yacimiento. Es un parámetro sin dimensión y puede ser expresado en fracción o por ciento. Desde el punto de vista del proceso responsable de la formación de la porosidad se clasifica fundamentalmente en dos tipos: (1) primaria: es la porosidad original después de la deposición de los sedimentos y su compactación inicial y (2) secundaria: debida a los esfuerzos tectónicos (fracturas, estilolitas) y a la circulación del agua subterránea (disolución, recristalización, lixiviación, dolomitización, etc.). La porosidad secundaria es más frecuente en rocas carbonatadas debido a su relativa fragilidad solubilidad. Otra clasificación de la porosidad tiene que ver con la conectividad de los poros: (1) total: incluye toda la porosidad y (2) efectiva: es la porosidad interconectada.

#### Porosidad

La porosidad ( $\phi$ ) se obtiene de estudios realizados a núcleo, debe de tenerse en consideración para su evaluación, la precisión de la prueba y la representatividad de los núcleos. La otra fuente de evaluación de la porosidad son los registros geofísicos de pozo, mediante los cuales es posible calcular la porosidad utilizando diferentes curvas y baja diversas metodologías. La integración de mediciones de núcleo y registros, es un procedimiento general para la estimación de la porosidad y depende del tipo de datos disponibles y del tipo de yacimiento bajo estudio, por lo general incluye una revisión de la información con que se cuenta y después de establecer la correspondencia en profundidad de los registros con los núcleos, se comparan los valores de registros y núcleos mediante un gráfico de dispersión.

Mayormente la porosidad no representa gran dificultad excepto en los casos de litología compleja, ya que la mayoría de los métodos dependen del conocimiento de la misma o en los YNF,

donde la porosidad secundaria (debido a las fracturas) representa un porcentaje significativo de la porosidad total.

### Saturación del Agua

La saturación de agua ( $S_w$ ), el espacio de poros de las rocas del yacimiento está lleno de fluidos, normalmente hidrocarburos y agua. La determinación de las condiciones de saturación del medio poroso es una de las tareas importantes en un estudio de yacimiento, ya que afecta al cálculo de volumen de hidrocarburos, a la mecánica de los fluidos y al desempeño de producción esperado de un campo. Las saturaciones de los fluidos puede ser determinada en núcleos de dos maneras: mediante la medición de la cantidad de fluidos extraída o a través de mediciones de la presión capilar. La práctica común para evaluar la saturación del agua es mediante la interpretación de las curvas de resistividad de los registros de pozo. Se considera que la conductividad eléctrica de la formación se debe a la presencia de agua en los poros, ya que la matriz porosa y los hidrocarburos son aislantes perfectos. La interpretación consiste en comparar la resistividad medida de la formación con su resistividad teórica si sólo contuviera agua. En caso de que sea mayor la primera entonces se infiere la presencia de hidrocarburos.

### Permeabilidad

La permeabilidad ( $k$ ) se define como la capacidad de la formación rocosa de conducir fluidos. Es la propiedad petrofísica más importante de un yacimiento, pero también es la más difícil de medir. Desde el punto de vista del concepto de integración se requiere de un entendimiento profundo de las implicaciones estáticas y dinámicas de la permeabilidad. Existen muchas técnicas que ofrecen una información directa o indirecta acerca de la permeabilidad, pero cada una de ellas aporta una pieza particular del mosaico, por lo que la integración no es una cuestión trivial. Se debe distinguir entre permeabilidad absoluta y efectiva, la escala y el tipo de medición, las condiciones ambientales que afectan a los datos, etc.

Una característica peculiar de la permeabilidad es que es una cantidad direccional expresada mediante un tensor y que con frecuencia muestra anisotropía. Primariamente depende de las características de la textura de la roca, las cuales son el resultado de los procesos de deposición. Los arreglos de los granos que constituyen el sedimento tienen un fuerte impacto en las características del flujo. La existencia de anisotropía en la permeabilidad se debe a la orientación y la alineación de los granos, la presencia de arcillas u horizontes limosos, o a elementos con características laminadas como estilolitas o fracturas.

El análisis en núcleos permite la medición directa de la permeabilidad absoluta en el laboratorio, bajo diferentes condiciones experimentales. En general, las muestras que pueden ser analizadas pueden tener diferentes tamaños y volúmenes. Los volúmenes más pequeños son muestras recuperadas de núcleos laterales, donde la dimensión típica de un tapón es  $<1$  pulgada de largo. Un poco más grande, de 1 a  $1\frac{1}{2}$  pulgadas, son los tapones tradicionales tomadas de núcleos convencionales de rotación. Muestras de diámetro completo de hasta varias pulgadas de largo, pueden ser también analizadas en el caso de formaciones muy heterogéneas.

El modo más común de estimar perfiles de permeabilidad en pozos es mediante algún predictor de la permeabilidad, típicamente en forma de una ecuación empírica. Esto normalmente requiere de un conjunto de datos para la calibración, el cual está constituido por uno o más pozos claves

en donde la información sea completa en términos de datos de registros y núcleos. Por mucho el estimador de la permeabilidad más usado es la relación de porosidad-permeabilidad, el cual ha sido ampliamente reconocido que la mayoría de las rocas de los yacimientos muestran en una escala semilogarítmica una relación razonablemente lineal entre estas dos propiedades, lo cual permite la estimación de la permeabilidad cuando está disponible el perfil de la porosidad. Otros estimadores que tradicionalmente se han usado son las regresiones lineales múltiples y las ecuaciones empíricas existentes. Es importante hacer notar que el objetivo de todas estas técnicas es la estimación de la permeabilidad absoluta en condiciones *in situ* ya que los parámetros que se usan provienen de mediciones de registros de pozo.

Existe la necesidad de establecer modelos de dependencias entre propiedades petrofísicas, como es el caso de las relaciones de porosidad-permeabilidad. Como se ha mencionado los modelos usualmente aplicados se basan en relaciones funcionales empíricas y/o técnicas de regresión que no permiten representar toda la variabilidad del comportamiento observado de dichas propiedades, subestimando su varianza y valores extremos, que mayormente constituyen una parte fundamental de la información. Más aún, el enfoque de regresión muchas veces está restringido a suponer dependencias lineales entre las propiedades, lo cual en la práctica puede ser poco frecuente y limitado a ciertos casos particulares.

Es sabido que para realizar simulaciones estocásticas de más de una variable se requiere del conocimiento de su función de distribución de probabilidad multivariada, lo cual es prácticamente imposible de conocer; lo que mayormente se hace es ajustar los datos a uno de los pocos modelos de distribución de probabilidad establecidos, lo cual puede ser muy forzado y desvirtuar el comportamiento real de los datos. Una manera novedosa de representar distribuciones de probabilidad empíricas de datos multivariados dependientes es mediante unas funciones denominadas cónulas. Las cónulas son funciones que describen dependencias entre variables y representan la herramienta matemática apropiada para crear distribuciones de datos multivariados correlacionados.

#### **2.1.2.2. Metodología de simulación de propiedades petrofísicas y técnicas geoestadísticas para modelar variables ~~continuas~~ *in acreto***

Los valores de las propiedades petrofísicas son asignados dentro de una facies de manera que reproduzcan las características estadísticas representativas (histograma, variograma, correlación con otra variable) de la propiedad para dicha facies. Las propiedades petrofísicas tales como porosidad y permeabilidad son modeladas dentro de cada facies y capa del yacimiento. A las arcillas y a las facies no-netas se le pueden asignar valores bajos de manera arbitraria. Sin embargo, las propiedades petrofísicas dentro de la mayoría de las facies deben ser asignadas de manera que reproduzcan el histograma, el variograma y la correlación con respecto a variables secundarias. Las técnicas de simulación Gaussianas son las más empleadas para este propósito.

Las propiedades petrofísicas que usualmente se modelan son porosidad y permeabilidad. La porosidad a diferentes escalas se comporta en general de manera lineal, por lo general se modela la porosidad efectiva y no la total. Como se ha mencionado anteriormente la permeabilidad no es una propiedad intrínseca de la roca, depende de las condiciones de frontera fuera del volumen de

la medición y puede variar en varios órdenes de magnitud a diferentes escalas. Para su modelación se debe considerar: que los datos deben ser corregidos a las condiciones de fluido y presión del yacimiento; la anisotropía horizontal respecto a la vertical es un factor importante, por lo que los componentes direccionales principales de la permeabilidad son modelados o se modela de manera global y luego se aplica la relación de anisotropía vertical-horizontal obtenida de las facies.

Las principales propiedades petrofísicas, porosidad y permeabilidad se modelan dentro de cada facies y unidad del yacimiento, ya que las capas del yacimiento pueden ser distintas debido a su deposición en diferentes ambientes o edades geológicas. Las propiedades del yacimiento dentro de la misma facies en diferentes capas pueden tener características similares, sin embargo, deben ser modeladas separadamente debido a que pertenecen a sistemas estratigráficos diferentes. Mientras que las propiedades dentro de diferentes facies en una misma capa son significativamente diferentes y no están relacionadas, las propiedades dentro de diferentes facies pueden ser modeladas de manera independiente cuando las facies no están relacionadas. Las propiedades en diferentes facies pueden estar relacionadas unas con otras, es decir, puede existir una correlación de las propiedades petrofísicas a través de la frontera de las facies.

Para modelar propiedades petrofísicas existen varios enfoques y es recomendable hacer uso de variables secundarias o variables de apoyo, así la impedancia sísmica está correlacionada negativamente con la porosidad, la permeabilidad está correlacionada con la porosidad, la saturación de agua está correlacionada con la permeabilidad y los datos de producción ofrecen una fuente externa de información.

Respecto a las técnicas es posible partir de técnicas de estimación tipo cokriging en donde se hace uso de variables secundarias (atributos sísmicos correlacionados con propiedades petrofísicas), o bien utilizar métodos de simulación estocástica basados en celdas de tipo Gaussiano como son Bandas Rotantes y Simulación Secuencial Gaussiana (SGS), siendo este último el más utilizado y más aun empleando un enfoque de cosimulación con aplicación de variables secundarias. Típicamente la modelación se realiza de manera secuencial, por lo que los datos secundarios disponibles en cada etapa son redundantes o un solo tipo ofrece la mayor información, por lo general, se procede modelando primero la porosidad usando la sísmica como variable secundaria y posteriormente se modela la permeabilidad usando la porosidad como variable secundaria.

Generar realizaciones estocásticas de las propiedades petrofísicas de un yacimiento con un adecuado nivel de correlación espacial en los valores extremos de una propiedad como la permeabilidad, es uno de los retos más difíciles en el modelado geoestadístico de yacimientos. El método más comúnmente aplicado para la generación de realizaciones estocásticas de datos petrofísicos, como ya se mencionó es la simulación secuencial Gaussiana (SGS), la cual da por resultado una disminución de la longitud de correlación de los valores en las colas de la distribución, así la SGS es incapaz de simular grandes regiones conectadas o zonas de muy alta o muy baja permeabilidad. De acuerdo con la investigación realizada, la simulación fractal basada en el Movimiento Fraccional de Lévy (fLm) se muestra como una alternativa a los enfoques Gaussianos para modelar propiedades petrofísicas en formaciones geológicas altamente heterogéneas.

## 2.2. Modelos de fracturas

Modelar y simular numéricamente flujo de fluidos a través un medio poroso fracturado es complicado debido a la heterogeneidad y anisotropía creada por la compleja distribución de las fracturas, las cuales se presentan a diferentes escalas y cuyo espesor es demasiado pequeño en comparación con las dimensiones de la matriz (Romano-Perez, 2016). Además, el flujo de fluidos comprende dos medios, la matriz y las fracturas, con propiedades drásticamente diferentes. Para representar medios porosos fracturados han sido propuestos varios modelos conceptuales, los cuales pueden ser clasificados dentro de tres enfoques principales:

- Enfoque continuo.
- Enfoque discreto.
- Enfoque híbrido.

Estos enfoques están basados en consideraciones totalmente diferentes, por lo que cada uno de ellos tiene distintas limitaciones y se aplican bajo ciertas condiciones.

Dentro del contexto estático, es decir, sin considerar explícitamente el flujo, se han propuesto varios modelos los cuales se explican a continuación.

En las geociencias, en general es difícil o costoso un muestreo exhaustivo del sistema de fracturas que se quiere estudiar lo que conlleva a cierta limitación de datos. Por estas razones, el enfoque para atacar dicho problema es vía modelos estocásticos, y por lo tanto modelos Booleanos (o modelos basados en objetos) están también incluidos. Los métodos reportados en la literatura para la simulación de redes de fractura incluyen también enfoques basados en la geometría estocástica, la estadística multi-punto y una combinación de geoestadística para la densidad de la fractura y el modelado basado en objetos para geometrías de fractura. Dentro de ellos se incluye el conocido enfoque de redes de fracturas discretas el cual es un caso particular de los modelos Booleanos. Para mayor detalle sobre la siguiente discusión se puede consultar en el trabajo de Dowd *et al.* (2007).

### 2.2.1. Método Gaussiano truncado utilizando simulación indicador

El objetivo de este método consiste en reproducir mediante simulación Gaussiana truncada el sistema de fracturas al reproducir la proporción de píxeles de la imagen de fracturas. La imagen es binaria (Figura 2.1) con colores negros para representar la fractura y blancos para representar el fondo. Esta imagen se puede ver como una representación indicador del conjunto de datos. Un pixel negro en  $x$  corresponde a  $\mathbb{1}(x) = 1$ , lo que significa que al menos una fractura pasa por el pixel centrado en  $x$ .

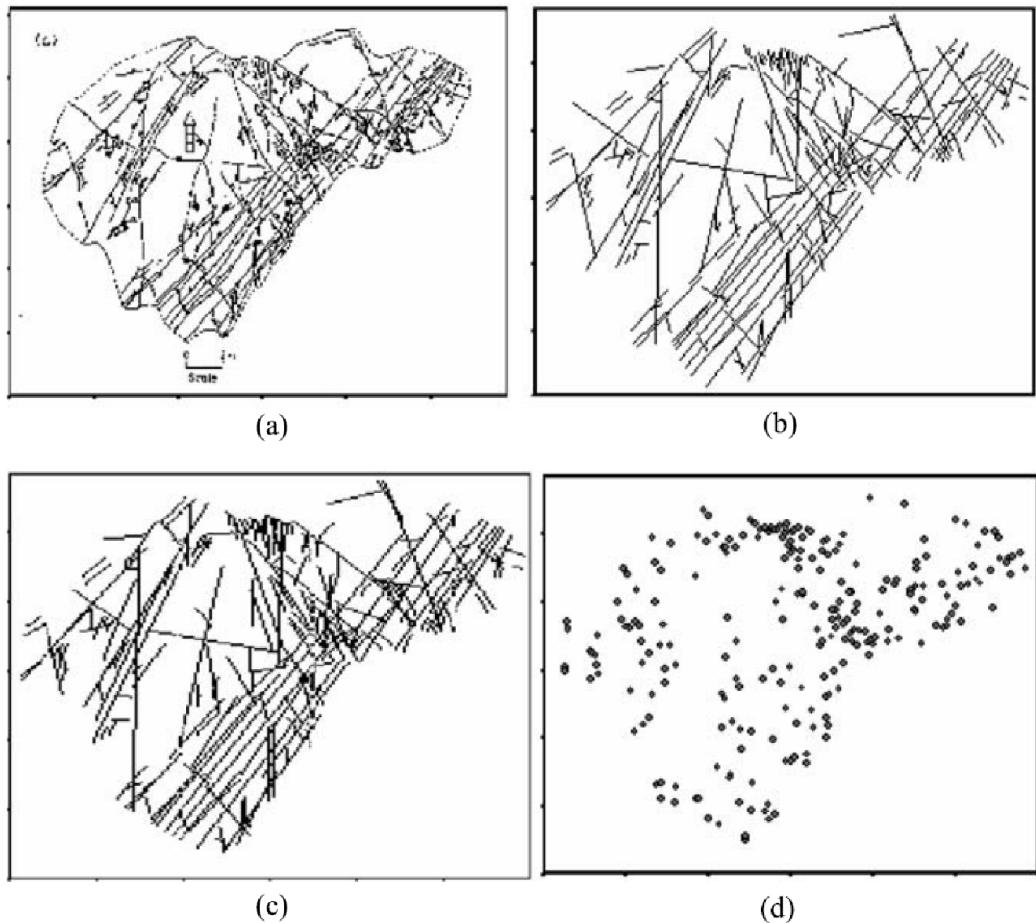


Figura 2.1: a) fracturas; b) fracturas digitalizadas; c) Fracturas pixeladas; d) centroides de fracturas.

La función indicador se define como

$$\mathbb{1}(x) = \begin{cases} 0, & \text{si } x \in F; \\ 1, & \text{en otro caso.} \end{cases} \quad (2.1)$$

Donde es el conjunto aleatorio de las fracturas "pixeladas". Por lo tanto, la simulación (no condicionada) puede ser realizada con el método Gaussiano truncado, el cual a su vez utiliza variogramas. Los resultados de la modelación con este método se muestran en la Figura 2.2.

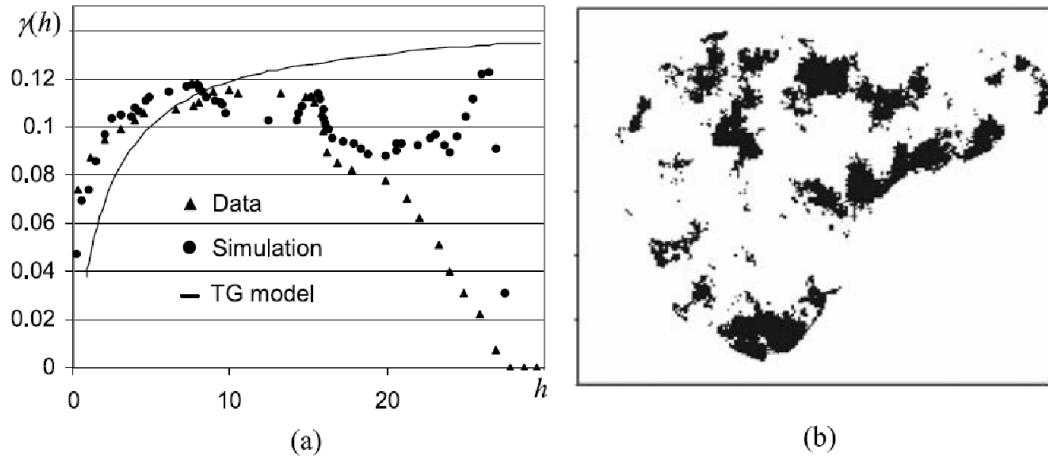


Figura 2.2: a) Variograma; b) Simulación.

Aunque el modelo de variograma es reproducido razonablemente bien, la imagen simulada difiere considerablemente de la imagen original. Este ejemplo muestra que reproducir el primer orden y órdenes superiores de un campo aleatorio no necesariamente reproduce un proceso aleatorio.

### 2.2.2. Modelo con geoestadística multi-punto

Utilizando el contexto de la geoestadística convencional, el uso de sólo el primer y segundo momento es claramente insuficiente para simular las complejas estructuras de una red de fracturas. Utilizando las estadísticas de orden superior debería mejorar la simulación, aunque la reconstrucción de todas las propiedades estocásticas del campo aleatorio no se puede garantizar. Inferencia de orden superior en modelos estadísticos espaciales es difícil, ya que se basan en las configuraciones espaciales en lugar de las distancias por pares y orientaciones simples usados en las formas tradicionales de la geoestadística. Un enfoque consiste en utilizar la probabilidad condicional para una configuración particular para inferir experimentalmente el modelo espacial para describir el campo aleatorio.

Para una variable indicador, la probabilidad condicional es equivalente a la esperanza condicional de la configuración. La probabilidad condicional de que  $\mathbb{1} = 1$ , dado  $n$  eventos  $\mathbb{1}(x) = 0$  ó 1 para  $i = 1, \dots, n$  se puede expresar como la suma de términos de la siguiente manera:

$$Prob \left\{ \mathbb{1}(x) = \mathbb{1}(x_i), \quad i = \{1, \dots, n\} = \lambda_0 + \sum_{i=1}^n \lambda_i^{(1)} \mathbb{1}(x_i) + \sum_{i=1}^n \sum_{j>i}^n \lambda_{ij}^{(2)} \mathbb{1}(x_i) \mathbb{1}(x_j) + \dots \right\} \quad (2.2)$$

Donde  $\lambda_{..}^{(0)}$  son los  $2^n + 1$  coeficientes obtenidos por un sistema de Kriging simples extendidos. Los primeros  $n + 1$  términos se pueden resolver por Kriging (geoestadística a dos puntos) y los otros términos describen el enfoque multi-punto de orden 3 al  $n$ .

Cuando se ha inferido la probabilidad condicional de la Ecuación 2.2,  $\mathbb{1}(x)$  puede ser simulada mediante muestreo Monte Carlo. Generalmente, la inferencia se basa en realizaciones de la variable aleatoria, a las cuales se les conoce como imágenes de entrenamiento. Estas imágenes de entrenamiento pueden ser datos o postulados en base a la experiencia. La simulación secuencial de  $\mathbb{1}(x)$  depende del número  $n$  de datos indicadores,  $\mathbb{1}(x_i)$ , más cercanos a cada nodo  $x$ .

En la Figura 2.3 se muestran los resultados para diferentes valores de  $n$ . Estas simulaciones muestran la importancia de éste parámetro en la estadística multi-punto. La selección ~~de~~ depende lo complejo de las estructuras a simular. En el ejemplo que aquí se muestra se puede concluir que valores  $n \leq 12$  son inadecuados para describir el sistema de fracturas en cuestión.

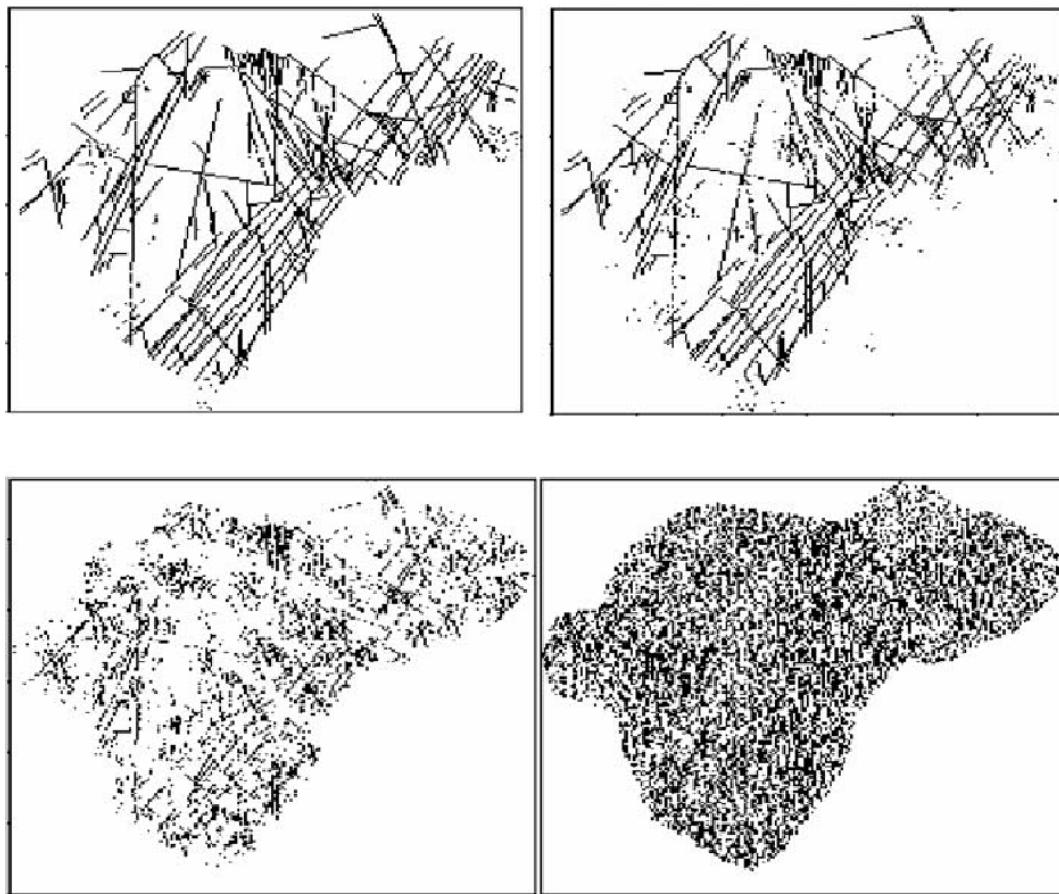


Figura 2.3: Simulaciones basadas en la geoestadística multi-punto.

### 2.2.3. Modelos booleanos

Debido a la complejidad de las fracturas y más aún de los sistemas de fracturas, no es viable modelarlas con demasiada precisión. Por ejemplo, el simple hecho de intentar modelar las caras

	Pixeles negros (%)	Posición en la Figura 2.3
20	19.63	Superior izquierda
16	19.26	Superior derecha
12	20.88	Inferior izquierda
8	43.97	Inferior derecha

Tabla 2.1: Valores de porcentajes de píxeles negros para los resultados en la Figura 2.3.

de las fracturas es computacionalmente lento. Si el objetivo buscado es la simulación de flujo, éste último puede tomar mucho más tiempo si se considera la rugosidad de la fractura. En muchos de los casos, la rugosidad de estas superficies puede ser auto-afín, es decir tener propiedades fractales (Adler y Thovert, 1999, ch. 4).

Cuando se desea modelar un sistema de fracturas en algún yacimiento, el enfoque del párrafo anterior se torna imposible de implementar. Un enfoque más viable es simplificar la geometría de las fracturas. Por ejemplo, se puede suponer que las caras son planos (Figura 2.4), modelo conocido como de placas paralelas (Dietrich *et al.*, 2005).

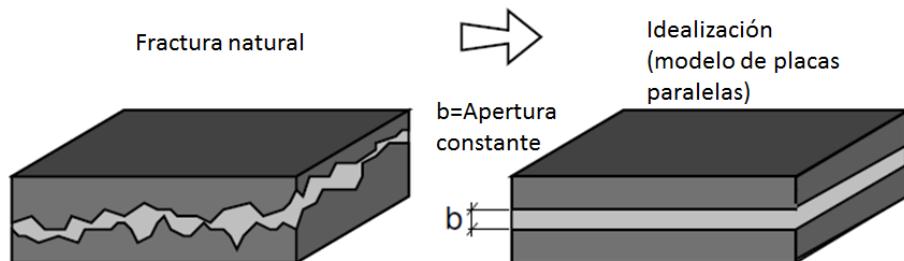


Figura 2.4: Modelado de una fractura. Izquierda: fractura real; derecha: su modelo de placas paralelas.

Con el enfoque de Dietrich *et al.* (2005), se supone que las fracturas son discretas y simplificadas. Los modelos más usuales para representar sistemas de fracturas en medios porosos en general y en particular en la modelación de Yacimientos Naturalmente Fracturados (YNF) son los modelos de Redes de Fracturas Discretas (DFN). Este tipo de modelos nos proporcionan el número de fracturas en determinado volumen con longitudes y orientaciones dadas. Estos modelos consideran a una fractura como un ente geométrico simplificado, usualmente segmentos rectilíneos, paralelogramos y elipses en 2D, y polígonos, paralelepípedos y elipsoides en 3D. Por lo que una fractura discreta puede ser caracterizada por su longitud, apertura, orientación, etc.

El método estocástico de simulación de fracturas discretas está basado en el modelo booleano, el cual a su vez se construye a partir de procesos puntuales de Poisson que generan los centros geométricos de los objetos que representan a las fracturas. Aquí por objeto se está considerando a los entes geométricos que representan a las fracturas. Los objetos no son determinísticos, son aleatorios, es decir éstos son generados estocásticamente, por lo que sus características geométricas

están dadas por funciones de distribución de probabilidad.

Dado que este es el enfoque computacionalmente más viable para la modelación de fracturas en un yacimiento, es el que se adoptó en este trabajo doctoral. En el siguiente capítulo se explica más la teoría de los modelos booleanos. Por lo pronto diremos que la idea intuitiva es la de objetos simplificados aleatoriamente colocados en el espacio. La palabra aleatoria se utiliza en un sentido probabilístico y no significa que estén uniformemente distribuidos.



# Capítulo 3

## Modelos booleanos basados en objetos estocásticos

Los modelos booleanos, aunque simples en la idea intuitiva detrás de la matemáticas requieren otros ingredientes aún más esenciales. En particular, a continuación se presenta una breve reseña de la teoría de procesos puntuales.

### 3.1. Procesos puntuales

Un proceso puntual aleatorio es un modelo estocástico (Lantuéjoul, 2002) en el cual cada realización está formada por un número finito ó contable de puntos en  $\mathbb{R}^d$ .

Aunque los elementos de un proceso puntual son muy básicos, estos procesos pueden formar estructuras complicadas. La Figura 3.1 muestra algunos ejemplos, pero existe una gran variedad de modelos. Además, en muchos fenómenos la simplificación de eventos espaciales a puntos es bastante adecuada. Por ejemplo, para modelar árboles en los bosques, epicentros de terremotos, estrellas en las galaxias, entre otros. Una razón más de la importancia de los procesos puntuales, es que ellos sirven de ingredientes para modelos más complicados como en los que se enfoca esta tesis.

En un proceso puntual *localmente finito* el número de puntos dentro de cualquier conjunto compacto (cerrado y acotado) es casi seguramente finito. En este caso es conveniente describir las propiedades estadísticas del proceso puntual mediante un conteo del número de puntos dentro de subconjuntos del espacio. Sea  $N(B)$  el número aleatorio de puntos dentro del conjunto  $B$ , denominado el funcional que cuenta.  $B$  es un elemento de  $\mathcal{B}_0$ , el conjunto de los Boreelianos acotados.

La distribución espacial de un **proceso homogéneo de Poisson** con intensidad  $\theta$  está caracterizado por las siguientes propiedades:

1. Si  $A \in \mathcal{B}_0$ , entonces el número de puntos dentro de  $A$ ,  $N(A)$ , es una variable aleatoria de

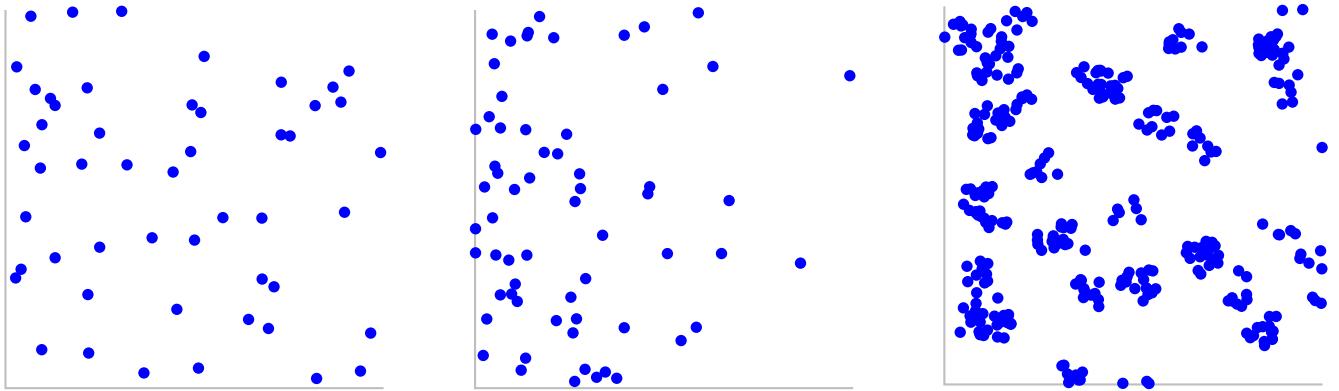


Figura 3.1: Ejemplos de procesos puntuales. a) Caso Homogéneo, b) Intensidad no uniforme c) Proceso de conglomerados.

Poisson con parámetro  $\lambda\|A\|$  donde  $\|A\|$  es el  $d$ -volumen de  $A$

$$P\{N(A) = n\} = \exp\{-\lambda\|A\|\} \frac{[\lambda\|A\|]^n}{n!}$$

2. Si  $(A_i, i \in I)$  es una familia finita de elementos disjuntos de  $\mathcal{B}_0$ , entonces las variables aleatorias  $N(A_i), i \in I$  son mutuamente independientes.

Dado que el valor medio de la distribución de Poisson es igual a su parámetro  $\lambda$ , la media del número de puntos sobre  $A$  es igual a  $\lambda\|A\|$ . En particular,  $\lambda$  es la media del número de puntos por unidad de  $d$ -volumen. Si se requiere simular una cantidad exacta de puntos, entonces  $\|A\|/n = <\lambda>$  da un valor (medio  $<\cdot>$ ) aproximado de la intensidad. Si embargo esta intensidad calculada no es la única porque existe un conjunto no-numerable de valores de intensidad que pueden producir el mismo  $n$ .

El siguiente teorema, cuya demostración se puede encontrar en (Lantuéjoul, 2002, p. 121), muestra que los puntos de un proceso puntual homogéneo de Poisson son independientes y con ubicación totalmente aleatoria.

Sea  $A \in \mathcal{B}_0$ . Si  $N(A) = n$ , los  $n$  puntos son independientes y uniformemente distribuidos en  $A$ .

Este teorema permite un algoritmo simple para simular un proceso puntual de Poisson en una región  $A$ .

1. Inicialice el conjunto  $X = \emptyset$ ,
2. Obtenga cuántos puntos van a ser simulados al muestrear  $n \sim \text{Poisson}(\theta\|A\|)$ ,
3. si  $n > 0$ , para  $i = 1, \dots, n$  genere  $x_i \sim \text{Unif}(A)$ ,
4. Los puntos  $x_i \in \mathbb{R}^d$  pertenecientes a la realización del proceso puntual son  $X = \{x_1, \dots, x_n\}$ .

*en que software?*

Computacionalmente,  $X$  puede ser un arreglo bidimensional de  $n \times d$ ,  $n$  renglones por  $d$  columnas. Aunque este algoritmo es computacionalmente rápido, se puede hacer más eficiente si se toma en cuenta el orden de este arreglo ( $n \times d$  ó  $d \times n$ ); dependiendo del paradigma en la memoria: row-major vs column-major (Karniadakis y Kirby, 2003; Matloff, 2016). Este algoritmo está computacionalmente implementado en la función `spatstat::rpoint`. Si ya se sabe el valor de  $n$ , entonces la función que simula los puntos se llama `spatstat::runifpoint`.

Los mejores ejemplos para la metodología mostrada en esta tesis se obtuvieron con procesos puntuales homogéneos en los cuales  $\lambda$  es constante. Otros modelos comunes son cuando  $\lambda$  varía espacialmente (Proceso puntual de Poisson no homogéneo) y cuando la intensidad espacial es una función aleatoria (Proceso de Cox).

## 3.2. Construcción del modelo booleano

Se necesitan dos ingredientes básicos e independientes para la construcción de un modelo booleano en  $\mathbb{R}^d$ :

1. Un conjunto de *semillas*, esto es, un proceso puntual de Poisson  $\mathcal{P}$  con función de intensidad  $\theta = (\theta(x), x \in \mathbb{R}^d)$ .
2. Una familia  $(A(x), x \in \mathbb{R}^d)$  de subconjuntos de  $\mathbb{R}^d$  compactos, no vacíos, aleatorios e independientes. Al subconjunto  $A(x)$  se le llama el *objeto implantado en x* y su funcional que pega se denota por  $\mathcal{T}_x$ .

De este modo, un modelo booleano  $X$  es la *unión de todos los objetos implantados en las semillas Poisson*

$$X = \bigcup_{x \in \mathcal{P}} A(x)$$

La apariencia de una realización del modelo depende en gran medida de la forma que tengan los objetos. Éstos pueden ser, por ejemplo, segmentos de recta, discos, polígonos, etcétera. En la Figura 3.2 se muestran tres realizaciones de modelos booleanos. En el primer caso se muestra un conjunto de círculos –los objetos– implantados sobre semillas de Poisson homogéneamente distribuidas. En el segundo caso se muestra un ejemplo similar pero los objetos tienen ciertas propiedades las cuales se visualizan mediante sus radios y colores. En el tercer caso se muestra otra realización en el cual los objetos son segmentos rectilíneos.



Figura 3.2: a) un ejemplo simple de modelo booleano cuyos objetos son círculos; b) Segmentos lineales.

Ya que un modelo booleano es posiblemente la unión de una infinidad de objetos, no existe garantía de que sea cerrado. Este será ciertamente el caso si para cada punto  $x \in \mathbb{R}^d$  existe una vecindad a la cual sólo le pega una cantidad finita de objetos (casi seguramente), y en particular si la población de objetos ( $A(x), x \in \mathcal{P}$ ) es de orden finito. Específicamente, sea  $N(K)$  el número de objetos que le pegan al compacto  $K$ . Entonces

$$N(K) = \sum_{x \in \mathcal{P}} \mathbb{1}_{K \cap A(x) \neq \emptyset}.$$

Un cálculo sencillo muestra que la media de  $N(K)$  es

$$\vartheta(K) = \int_{\mathbb{R}^d} \theta(x) \mathcal{T}_x(K) dx \leq \infty.$$

Se dice que un modelo booleano es de **orden finito** si  $\vartheta(K) < \infty \forall K \in \mathcal{K}$ . A lo largo de este texto se supondrá que los modelos booleanos considerados son de orden finito.

### 3.3. El funcional que evita del modelo booleano

Comencemos determinando la distribución del número de objetos que le pegan al subconjunto compacto  $K$ .

Se tendrá que  $N(K)$  es una variable aleatoria Poisson con media  $\vartheta(K)$ : Para cada  $D \in \mathcal{K}$  consideraremos el número  $N_D(K)$  de objetos implantados en  $D$  que le pegan a  $K$ .

$$N_D(K) = \sum_{x \in \mathcal{P} \cap D} \mathbb{1}_{K \cap A(x) \neq \emptyset}.$$

Por definición, el número de puntos en  $\mathcal{P} \cap D$  sigue una distribución Poisson con media

$$\theta(D) = \int_D \theta(x) dx.$$

Supongamos que este número es igual a  $n$ . De acuerdo a las propiedades de un proceso puntual de Poisson estos puntos están distribuidos de manera uniforme e independiente sobre  $D$ , con función de densidad de probabilidad  $\frac{\theta(\cdot)}{\theta(D)}$ . Más aún, un objeto implantado en  $x$  le pega a  $D$  con probabilidad  $\mathcal{T}_x(K)$  y lo evita con la probabilidad complementaria  $1 - \mathcal{T}_x(K)$ . En consecuencia, la función generadora de  $N_D(K)$  es

$$\mathbb{E} \{ s^{N_D(K)} \} = \sum_{n=0}^{\infty} e^{-\theta(D)} \frac{\theta(D)^n}{n!} \left( \int_D \frac{\theta(x)}{\theta(D)} [(1 + (s-1)\mathcal{T}_x(K)] dx \right)^n,$$

donde  $0 \leq s \leq 1$ , y al sumar se tiene que

$$\mathbb{E} \{ s^{N_D(K)} \} = \exp \left\{ (s-1) \int_D \theta(x) \mathcal{T}_x(K) dx \right\}.$$

Para extender este resultado a todo el espacio, sea  $(D_n, n \in \mathbb{N})$  una sucesión creciente de compactos que cubren a  $\mathbb{R}^d$ . Entonces  $(N_{D_n}(K), n \in \mathbb{N})$  también es una sucesión creciente y converge casi seguramente a  $N(K)$ . En consecuencia

$$\mathbb{E} \{ s^{N(K)} \} = \lim_{n \rightarrow \infty} \mathbb{E} \{ s^{N_{D_n}(K)} \} = \exp \left\{ (s-1) \int_{\mathbb{R}^d} \theta(x) \mathcal{T}_x(K) dx \right\}.$$

Esta última expresión se reconoce como la función generadora de una distribución Poisson con media  $\vartheta(K)$ .

Ya que  $\mathbb{P}\{X \cap K = \emptyset\} = \mathbb{P}\{N(K) = 0\}$ , deducimos que el funcional que evita del modelo booleano es

$$\mathbb{P}\{X \cap K = \emptyset\} = e^{-\vartheta(K)}, \quad K \in \mathcal{K}.$$

### 3.4. Propiedades de estabilidad

Para poder explorar las propiedades de estabilidad del modelo booleano es necesario echar mano de conceptos básicos de morfología, a saber, el concepto de *dilatación*.

Dados  $x \in \mathbb{R}^d$  y  $A \subset \mathbb{R}^d$ ,  $A_x$  denota al conjunto obtenido al trasladar  $A$  por  $x$ . Si  $B$  es un subconjunto de  $\mathbb{R}^d$  entonces la *suma de Minkowski* de  $A$  y  $B$  es

$$A \oplus B = \bigcup_{y \in B} A_y.$$

Decimos entonces que el punto  $x$  pertenece al conjunto  $A$  *dilatado* por  $B$  si  $B_x$  le pega a  $A$  (es decir, si  $B_x \cap A \neq \emptyset$ ).

Por otra parte, necesitaremos también el concepto de  $i$ -planos, los cuales son espacios afines de dimensión  $i$ . [Un espacio afín es un espacio vectorial sin origen].

~~Ahora sí estamos listos~~ para enunciar las propiedades básicas de estabilidad de un modelo booleano:

- La unión de dos modelos booleanos independientes es un modelo booleano.
- Un modelo booleano dilatado por un subconjunto compacto no vacío de  $\mathbb{R}^d$  es un modelo booleano.
- La intersección de un modelo booleano con un subconjunto compacto de  $\mathbb{R}^d$  es un modelo booleano.
- La intersección de un modelo booleano y un  $i$ -plano es un modelo booleano.

A continuación, se proporcionan argumentos para justificar estas afirmaciones:

Sean  $X'$  y  $X''$  dos modelos booleanos independientes con funciones de identidad  $\theta'$  y  $\theta''$ , y funcionales que pegan  $\mathcal{T}'$  y  $\mathcal{T}''$ . Debido a la independencia se tiene que

$$\begin{aligned}\mathbb{P}\{(X' \cup X'') \cap K = \emptyset\} &= \mathbb{P}\{X' \cap K = \emptyset\}\mathbb{P}\{X'' \cap K = \emptyset\} \\ &= \exp\left\{-\int_{\mathbb{R}^d} \theta'(x)\mathcal{T}'_x(K) + \theta''(x)\mathcal{T}''_x(K)dx\right\} \\ &= \exp\left\{-\int_{\mathbb{R}^d} \theta(x)\mathcal{T}_x(K)dx\right\}\end{aligned}$$

donde  $\theta = \theta' + \theta''$  y  $\mathcal{T}_x$  queda definido como

$$\mathcal{T}_x(K) = \frac{\theta''(x)}{\theta'(x) + \theta''(x)}\mathcal{T}'_x(K) + \frac{\theta'(x)}{\theta'(x) + \theta''(x)}\mathcal{T}''_x(K), \quad K \in \mathcal{K}.$$

Este es el funcional que pega de un objeto aleatorio igual a  $A'(x)$  con probabilidad  $\frac{\theta'(x)}{\theta'(x) + \theta''(x)}$ , e igual a  $A''(x)$  con probabilidad complementaria  $\frac{\theta''(x)}{\theta'(x) + \theta''(x)}$ . Este hecho prueba la primera propiedad.

Ahora bien, sea  $D$  un subconjunto de  $\mathbb{R}$  compacto y no vacío. La segunda propiedad es consecuencia directa de la distributividad de la suma de Minkowski sobre la unión de conjuntos:

$$X \oplus D = \bigcup_{x \in \mathcal{P}} A(x) \oplus D.$$

De este modo, la dilatación de un modelo booleano vuelve a ser la unión de objetos aleatorios implantados en las semillas determinadas por  $\mathcal{P}$ , lo cual determina un nuevo modelo booleano. La diferencia es que en lugar de tener los objetos aleatorios originales  $A(x)$ , se tienen los objetos aleatorios dilatados  $A(x) \oplus D$ .

En cuanto a la tercera propiedad, podemos escribir simplemente  $X \cap D = \bigcup_{x \in \mathcal{P}} A(x) \cap D$ . Sin embargo, esta fórmula es difícil de interpretar ya que los objetos  $A(x) \cap D$  serán casi seguramente vacíos. En cambio, consideraremos la siguiente expresión:

$$\begin{aligned}
\mathbb{P}\{(X \cap D) \cap K = \emptyset\} &= \mathbb{P}\{X \cap (D \cap K) = \emptyset\} \\
&= \exp\left\{-\int_{\mathbb{R}}^d \theta(x) \mathcal{T}_x(D \cap K) dx\right\} \\
&= \exp\left\{-\int_{\mathbb{R}}^d \mathcal{T}_x(D) \frac{\theta(x) \mathcal{T}_x(D \cap K)}{\mathcal{T}_x(D)} dx\right\}
\end{aligned}$$

Pero  $\frac{\theta(x) \mathcal{T}_x(D \cap K)}{\mathcal{T}_x(D)}$  es el funcional que pega de  $A(x) \cap D$ , ya que este objeto es no vacío. En consecuencia,  $X \cap D$  es un modelo booleano con función de intensidad  $\theta \mathcal{T}(D)$  y funcional que pega  $\frac{\theta \mathcal{T}(D \cap K)}{\mathcal{T}(D)}$ .

Finalmente, sea  $P$  un  $i$ -plano de  $\mathbb{R}^d$ . La última propiedad se puede deducir de la tercera tomando una sucesión creciente de compactos no vacíos ( $D_n, n \in \mathbb{N}$ ) que converja a  $P$ .

El modelo booleano es un caso especial de los modelos basados en objetos. Una primera extensión consiste en reemplazar el proceso puntual de Poisson que especifica la localización de los objetos a través de un proceso espacial de nacimiento y muerte (Preston, 1977; Stoyan et al., 1987). También es posible permitir dependencia entre los objetos. En ese caso, un objeto se inserta o se elimina dependiendo de una función de los objetos ya presentes (y no solamente el número de éstos). Esto lleva al concepto de un proceso de vida y muerte de objetos en el espacio. Un ejemplo típico es el modelo de interacción dos a dos considerado por Sylverseen y Omre (1994). La primera versión de este modelo incluyó un número fijo de objetos, lo cual permitía la simulación a través de un muestreo de Gibbs. Esta restricción fue eliminada en un artículo posterior (1997). Es posible extender el modelo de muchas otras maneras, como el proceso de saltos de Markov de objetos.

### 3.5. Simulación del modelo booleano

El problema entre manos es simular el modelo booleano en  $D$ , sujeto a las condiciones de que dos subconjuntos finitos  $C_0$  y  $C_1$  deben estar contenidos en  $X^c$  y  $X$ , respectivamente. Este problema es importante en la industria petrolera. Los ingenieros petroleros requieren conocer la geometría del yacimiento para poder ejecutar programas de simulación de flujos. Un algoritmo desarrollado por Haldersen en 1983 fue notablemente mejorado por Chessa en 1995. Consiste en simular de manera independiente los cuerpos arenosos (los objetos) que intersecan los pozos y aquellos que no. Esta perspectiva dicotómica es posible gracias a las propiedades de independencia del proceso Poisson. La dificultad de esta estrategia es que la distribución de un objeto que interseca al pozo depende no solamente de dónde esté implantado, sino también del número y la localización de los pozos que interseca. Usualmente, el problema se vuelve intratable en cuanto este número es mayor a uno.

El algoritmo iterativo descrito a continuación fue diseñado a partir de comunicaciones privadas entre Matheron y Lantuéjoul en 1990. Posteriormente Gedler (1991) llevó a cabo el trabajo preliminar bajo la supervisión de Lantuéjoul. Ya que este algoritmo de simulación condicional es

una versión modificada de un algoritmo de simulación no condicional (mediante la restricción del kernel de transición), se presenta primero el algoritmo de simulación no condicional.

Se discutió anteriormente que  $X \cap D$  es un modelo booleano con función de intensidad  $\theta(\cdot)\mathcal{T}_x(D)$ . De manera acorde, el número de objetos de  $X \cap D$  es de distribución Poisson con media

$$\vartheta(D) = \int_{\mathbb{R}^d} \theta(x)\mathcal{T}_x(D)dx.$$

El funcional que pega de un objeto de  $X \cap D$  implantado en  $x$  es  $\frac{\mathcal{T}_x(D \cap \cdot)}{\mathcal{T}_x}(D)$ .

Un *objeto típico* de  $X \cap D$  es un objeto seleccionado uniformemente entre los objetos de  $X \cap D$ . Un objeto típico de  $X \cap D$  se implanta de acuerdo a la función de densidad  $\theta(\cdot)\mathcal{T}(D)$ . Su funcional que pega es

$$\mathcal{T}(K) = \int_{\mathbb{R}^d} \frac{\theta(x)\mathcal{T}_x(D)}{\vartheta(D)} \frac{\mathcal{T}_x(D \cap K)}{\mathcal{T}_x(D)} dx = \frac{1}{\vartheta(D)} \int_{\mathbb{R}^d} \theta(x)\mathcal{T}_x(D \cap K)dx$$

Notemos que en particular  $\mathcal{T}(D) = 1$  y  $\mathcal{T}(K) = 0$  si  $K$  es disjunto de  $D$ .

Por otra parte,  $X \cap D$  tiene la misma distribución que una unión de  $N$  objetos típicos independientes, donde  $N$  tiene distribución Poisson con media  $\vartheta(D)$ . Este hecho se prueba del siguiente modo: Sea  $Y$  tal unión. Su funcional que evita es

$$\begin{aligned} \mathbb{P}\{Y \cap K = \emptyset\} &= \sum_{n=0}^{\infty} e^{-\vartheta(D)} \frac{\vartheta(D)^n}{n!} [1 - \mathcal{T}(K)]^n \\ &= \exp\{-\vartheta(D)\mathcal{T}(K)\} \\ &= \exp\left\{-\int_{\mathbb{R}^d} \theta(x)\mathcal{T}_x(D \cap K)dx\right\} \end{aligned}$$

el cual resulta ser exactamente el funcional que evita de  $X \cap D$ .

En consecuencia, una simulación no condicional del modelo booleano se puede obtener utilizando el siguiente algoritmo:

#### **Algoritmo para el modelo booleano:**

1. Hacer  $X = \emptyset$ .
2. Generar  $N \sim \text{Poisson}(\vartheta(D))$ .
3. Si  $N = 0$ , regresar  $X$ .
4. Generar un objeto típico  $A \sim \mathcal{T}$ .
5. Hacer  $X = X \cup A$ ,  $N = N - 1$  e ir al paso 3.

Este algoritmo requiere de la simulación de un objeto típico. El algoritmo estándar es:

#### **Algoritmo para el objeto típico**

1. Generar  $X \sim \theta(\cdot)\mathcal{T}(D)$ .
2. Generar  $A \sim \frac{\mathcal{T}_x(D \cap \cdot)}{\mathcal{T}_x(D)}$ .

3. Regresar  $A$ .

Si resulta difícil simular un objeto implantado en  $x$  y pegarle a  $D$  directamente, es posible emplear un algoritmo alternativo de rechazo.

#### **Algoritmo para el objeto típico (método de rechazo)**

1. Generar  $X \sim \theta(\cdot)\mathcal{T}(D)$ .
2. Generar  $A \sim \mathcal{T}_x$ .
3. Si  $A \cap D = \emptyset$ , ir al paso 2.
4. Regresar  $A \cap D$ .

Finalmente, el primer algoritmo se modifica ligeramente para volverlo iterativo. Esto no presenta dificultad alguna ya que algoritmo de Metrópolis puede ser utilizado para simular una distribución Poisson. El siguiente algoritmo simula iterativamente un modelo Poisson. Específicamente, simula la población de objetos que constituye el modelo booleano. En este algoritmo el número de objetos de la población  $\Phi$  se denota por  $\#\Phi$ .

#### **Algoritmo para objetos de un modelo booleano**

1. Hacer  $\Phi = \emptyset$ .
2. Generar una variable aleatoria uniforme  $U$  que tome los valores 1, -1 y 0, con probabilidades:

$$p_1 = \frac{\vartheta(D)}{\vartheta(D) + \#\Phi + 1}, \quad p_{-1} = \frac{\#\Phi}{\vartheta(D) + \#\Phi}, \quad p_0 = 1 - p_1 - p_{-1}.$$

3.
  - a) Si  $U = 1$ , entonces generar  $A \sim \mathcal{T}$  y hacer  $\Phi = \Phi \cup A$ .
  - b) Si  $U = -1$ , entonces generar  $A \sim Unif(\Phi)$  y hacer  $\Phi = \Phi \setminus A$ .
4. Ir al paso 2.

#### **Simulación condicional**

Empezaremos con dos comentarios. En primer lugar, la cadena de Markov del segundo algoritmo presentado es reversible (hereda esta propiedad del algoritmo de Metrópolis). En segundo lugar, la restricción de esta cadena de Markov al conjunto  $\Omega_c$  de todas las poblaciones de objetos que respetan las condiciones en  $C_0$  y  $C_1$  es irreducible. Esto se deriva del hecho de que  $\Omega_c$  es estable bajo concatenación, es decir que si  $\Phi, \Psi \in \Omega_c$ , entonces la población  $\Phi + \Psi$  determinada por todos los objetos de  $\Phi$  y todos los objetos de  $\Psi$  también es un elemento de  $\Omega_c$ .

En consecuencia, se puede aplicar la técnica de restricción del kernel de transición<sup>1</sup> y el algoritmo en cuestión puede volverse condicional al pedir que las condiciones sean satisfechas en cada paso de la iteración. A continuación, se presenta el algoritmo condicional:

#### **Algoritmo condicional para el modelo booleano**

---

<sup>1</sup>Esta técnica se discute en la cuarta sección del octavo capítulo del mismo libro. A grandes rasgos, consiste en prohibir excursiones fuera de  $\Omega_c$ . Se toma  $x \in \Omega_c$  y se simula  $y$  hasta asegurar que  $y \in \Omega_c$ . Luego se hace  $x = y$ . Esta idea fue propuesta por Lantuéjoul en 1996.

1. Generar  $\Phi \in \Omega_c$ .
2. Generar una variable aleatoria  $U$  que tome los valores  $1, -1$  y  $0$ , con probabilidades:

$$p_1 = \frac{\vartheta(D)}{\vartheta(D) + \#\Phi + 1}, \quad p_{-1} = \frac{\#\Phi}{\vartheta(D) + \#\Phi}, \quad p_0 = 1 - p_1 - p_{-1}.$$

3. a) Si  $U = 1$ , entonces generar  $A \sim \mathcal{T}$ . Si  $A \cap C_0 \neq \emptyset$  entonces hacer  $\Phi = \Phi \cup A$ .
- b) Si  $U = -1$ , entonces generar  $A \sim \text{Unif}(\Phi)$ . Si  $C_1 \subset \Phi \setminus A$ , entonces hacer  $\Phi = \Phi \setminus A$ .
4. Ir al paso 2.

Por supuesto, este algoritmo debe iniciar con una población permitida  $\Phi \in \Omega_c$ . Una manera en que se puede obtener es simulando una sucesión de objetos típicos independientes. Cada vez que un objeto le pega a  $C_0$  es descartado automáticamente. El procedimiento se continúa hasta que los objetos que quedan cubran completamente  $C_1$ . Para evitar iniciar con demasiados objetos se recomienda conservar sólo los objetos que son los primeros en cubrir puntos de  $C_1$ .

En consecuencia, se tiene el siguiente algoritmo:

#### **Algoritmo para la inicialización del modelo booleano condicional**

1. Hacer  $\Phi = \emptyset$  y  $C = C_1$ .
2. Generar  $A \sim \mathcal{T}$ .
3. Si  $A \cap C_0 \neq \emptyset$  o si  $A \cap C = \emptyset$ , ir al paso 2.
4. Hacer  $\Phi = \Phi \cup A$  y  $C \setminus A$ .
5. Si  $C \neq \emptyset$ , ir al paso 2.
6. Regresar  $\Phi$ .

Sea  $N_g$  el número de objetos generados para completar este procedimiento. En el caso en que  $C_1 \neq \emptyset$  un cálculo sencillo muestra que el valor medio de  $N_g$  es

$$\mathbb{E}\{N_g\} = \sum_{\emptyset \neq C \subset C_1} (-1)^{|C|+1} \frac{1}{\mathcal{T}(C_0 \cup C) - \mathcal{T}(C_0)}.$$

Este valor medio es finito si y sólo si  $\mathbb{P}\{C_0 \subset (X \cap D)^c, C_1 \subset X \cap D\} > 0$ .

La afirmación anterior se explica del siguiente modo: Tenemos que  $\mathbb{E}\{N_g\} < \infty$  si y sólo si  $\mathcal{T}(C_0 \cup \{c\}) - \mathcal{T}(C_0) > 0$  para cada  $c \in C_1$ . Para aprovechar estas desigualdades expresaremos  $X \cap D$  como la unión de  $\#C_1$  modelos booleanos independientes e idénticamente distribuidos  $(X(c), c \in C_1)$ . Su función de identidad común es  $\frac{\vartheta(\cdot)\mathcal{T}(D)}{\#C_1}$ . Notemos que  $X \cap D$  contiene a  $C_1$  en cuanto  $X(c)$  contiene a  $c$ . En consecuencia

$$\begin{aligned} \mathbb{P}\{C_0 \subset (X \cap D)^c, C_1 \subset X \cap D\} &\geq \prod_{c \in C_1} \mathbb{P}\{C_0 \cap X(c) = \emptyset, C_1 \cap X(c)\} \\ &= \exp\left\{-\frac{\vartheta(D)}{\#C_1}\mathcal{T}(C_0)\right\} - \exp\left\{-\frac{\vartheta(D)}{\#C_1}\mathcal{T}(C_0 \cup \{c\})\right\} > 0. \end{aligned}$$

La elección del criterio para detener el algoritmo depende de su tasa de convergencia. Una posibilidad es comparar el funcional que evita  $Q_n$  del conjunto aleatorio  $X_n$  producido en la enésima iteración y el funcional que evita  $Q_\infty$  que se quiere simular. Sin entrar en detalles es posible usar el argumento dado por Lantuéjoul (1997), el cual aprovecha el hecho de que el número de objetos en  $X_n$  evoluciona de acuerdo a una cadena de Markov con un kernel de transición de Jacobi compacto. Denotando por  $\psi$  al eigenvector normado asociado con el eigenvalor  $\lambda$  con el mayor módulo estrictamente mayor a 1, se puede determinar que

$$|Q_n(K) - Q_\infty(K)| \leq 2\lambda^n \mathbb{E}\{\psi(\#\Phi_0)\},$$

donde  $\Phi_0$  denota a la población inicial de objetos. Es posible obtener el valor de  $\lambda$  mediante la técnica de integración de rango<sup>2</sup>. Explícitamente, se obtiene el valor de  $\lambda = 0.993$ , el cual corresponde a un rango de integración de aproximadamente 320 iteraciones. Kendall y Thunnes diseñaron un algoritmo exacto de simulación condicional para el caso en que los objetos son acotados.

Es posible aplicar el algoritmo del modelo booleano condicional cuando se tienen restricciones más generales que  $C_0$  y  $C_1$ . Este algoritmo funciona siempre y cuando el conjunto de estados permisibles  $\Omega_c$  sea tal que  $\mathbb{P}\{\Omega_c\} > 0$  y que la cadena de Markov restringida a  $\Omega_c$  sea irreducible. Este hecho es útil para ingenieros petroleros que necesitan incorporar a sus simulaciones más información que sólo datos del pozo (datos dinámicos, datos sísmicos o incluso interpretaciones geológicas).

El modelo booleano es un caso especial de los modelos basados en objetos. Una primera extensión consiste en reemplazar el proceso puntual de Poisson que especifica la localización de los objetos a través de un proceso espacial de nacimiento y muerte (Preston, 1977; Stoyan et al., 1987). También es posible permitir dependencia entre los objetos. En ese caso, un objeto se inserta o se elimina dependiendo de una función de los objetos ya presentes (y no solamente el número de éstos). Esto lleva al concepto de un proceso de vida y muerte de objetos en el espacio. Un ejemplo típico es el modelo de interacción dos a dos considerado por Sylverseen y Omre (1994). La primera versión de este modelo incluyó un número fijo de objetos, lo cual permitía la simulación a través de un muestreo de Gibbs. Esta restricción fue eliminada en un artículo posterior (1997). Es posible extender el modelo de muchas otras maneras, como el proceso de saltos de Markov de objetos.

---

<sup>2</sup>La cuarta sección del noveno capítulo del libro discute la determinación empírica de la tasa de convergencia. La idea inicial consiste en estimar la tasa mediante una simulación para cierto número de iteraciones, permitiendo un período de calentamiento. Sin embargo, dada la interpretación de la tasa se necesita gran precisión al estimarla, y este planteamiento inicial no permite estimar el número de iteraciones necesario para obtener una precisión aceptable. Lantuéjoul propone utilizar el *rango integral* para salvar este obstáculo. Al este es una herramienta simple y poderosa que permite cuantificar las fluctuaciones estadísticas de un modelo estocástico. Se define utilizando herramientas variográficas y se discute a detalle en el cuarto capítulo.

### 3.6. El caso estacionario

Ahora se considera el caso en que:

1. La función de intensidad  $\theta$  es constante.
2. Todos los objetos son idénticamente distribuidos salvo por traslaciones. Dicho de otro modo, sea  $A$  un objeto implantado en el origen. Entonces  $A(x)$  tiene el mismo funcional que pega que  $A$  trasladado por  $x$ :

$$\mathbb{P}\{A(x) \cap K \neq \infty\} = \mathbb{P}\{A_x \cap K \neq \emptyset\} = \mathbb{P}\{A \cap K_{-x} \neq \emptyset\} = \mathcal{T}(K_{-x}).$$

Utilizando estas nuevas hipótesis el funcional que evita de  $X$  se convierte en

$$\mathbb{P}\{X \cap K = \emptyset\} = \exp \left\{ -\theta \int_{\mathbb{R}^d} \mathcal{T}(K_{-x}) dx \right\}.$$

Pero

$$\begin{aligned} \int_{\mathbb{R}^d} \mathcal{T}(K_{-x}) dx &= \mathbb{E} \left\{ \int_{\mathbb{R}^d} \mathbb{1}_{A \cap K_{-x} \neq \emptyset} dx \right\} \\ &= \mathbb{E} \left\{ \int_{\mathbb{R}^d} \mathbb{1}_{-x \in A \oplus K} dx \right\}, \\ &= \mathbb{E} \{|A \oplus K|\} \end{aligned}$$

y finalmente se tiene que el funcional que evita de un modelo booleano estacionario con intensidad  $\theta$  y objeto  $A$  es

$$\mathbb{P}\{X \cap K = \emptyset\} = e^{-\theta \mathbb{E}\{|A \oplus K|\}}, \quad K \in \mathcal{K}.$$

Consideremos ahora algunos casos particulares de  $K$ :

- Si  $K = \{x\}$  consiste de un solo punto, entonces obtenemos la probabilidad de que  $x$  no pertenezca a ningún objeto, la cual queda determinada por  $e^{-\theta \mathbb{E}\{|A|\}}$ .
- Si  $K = \{x, x + h\}$  es un par de puntos, entonces obtenemos la función de distribución bivariada de  $X^c$ . Ya que  $|A \oplus K| = |A \cup A_h| = 2|A| - |A \cap A_h|$ , es conveniente introducir el covariograma geométrico  $G_A$  de  $A$ :

Sea  $f$  la función indicadora de algún subconjunto  $X$  de  $\mathbb{R}^d$ . Suponiendo que  $f$  es integrable, se tiene que el volumen  $|X|$  de  $X$  es finito.  $X_h$  denotará al conjunto que se obtiene de trasladar  $X$  por  $h$ . Entonces podemos definir el siguiente mapeo:

$$G(h) = \int_{\mathbb{R}^d} \mathbb{1}_X(x) \mathbb{1}_X(x + h) dx.$$

Si desarrollamos la expresión anterior tenemos que

$$G(h) = \int_{\mathbb{R}^d} \mathbb{1}_X(x) \mathbb{1}_{X-h}(x) dx = |X \cap X-h|.$$

Tenemos que  $G(h) = G(-h)$ , y entonces definimos el covariograma geométrico de  $X$  como el mapeo  $G(h) = |X \cap X-h|$ ,  $h \in \mathbb{R}^d$ .

Utilizando el covariograma  $G_A$  tenemos

$$\mathbb{P}\{x \in X^c, x+h \in X^c\} = q^2 e^{\theta G_A(h)}.$$

De esta expresión derivamos lo siguiente:

$$\mathbb{P}\{x \in X, x+h \in X^c\} = q - q^2 e^{\theta G_A(h)},$$

$$\mathbb{P}\{x \in X, x+h \in X\} = 1 - 2q + q^2 e^{\theta G_A(h)}.$$

- Si  $K = [x, x+h]$  es un segmento de recta, entonces  $|A \oplus K|$  es tratable si y sólo si  $A$  es casi seguramente convexo. Sean  $|h|$  y  $\alpha$  el módulo y la dirección de  $K$ , entonces tenemos que  $|A \oplus K| = |A| + |h||A_\alpha|$  donde  $A_\alpha$  denota al volumen ( $d-1$ -dimensional) de  $A$  proyectado sobre el hiperplano ortogonal a  $\alpha$ . En consecuencia

$$\mathbb{P}\{[x, x+h] \subset X^c\} = q e^{-\theta|h|\mathbb{E}\{|A_\alpha|\}}.$$

Más aún, si suponemos que  $A$  es isotrópico (es decir, que su funcional que pega es invariante bajo rotaciones) entonces la fórmula de Cauchy<sup>3</sup> puede aplicarse y la fórmula anterior se simplifica a

$$\mathbb{P}\{[x, x+h] \subset X^c\} = q e^{-\theta|h|\frac{\omega_{d-1}}{d\omega_d} \mathbb{E}\{|\partial A|\}}.$$

Esta probabilidad se comporta como una función exponencial de módulo  $h$ .

---

<sup>3</sup>Esta fórmula representa el valor medio de los funcionales de Minkowski. Estos funcionales surgen en el campo de la estereología al estudiar funciones sobre los conjuntos convexos de  $\mathbb{R}^d$ . Es deseable que esta familia de funciones respete la convexidad, y los funcionales de Minkowski la generan. Están definidas salvo alguna constante multiplicativa, y convencionalmente se utiliza por constante de normalización el volumen de una esfera unitaria en  $\mathbb{R}^d$ , a saber

$$\omega_d = \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+1)}.$$

Los funcionales de Minkowski se denotan por  $W_i$ , donde  $i$  se denomina el grado del funcional, y varios de ellos tienen interpretaciones muy sencillas. Por ejemplo  $W_0(K) = |K|$ ,  $W_1(K) = \frac{|\partial K|}{d}$ , donde  $|\partial K|$  denota al volumen ( $d-1$ -dimensional de la frontera de  $K$ .

- Si  $K = B(x, r)$ , una bola de radio  $r$  con centro en  $x$ , también es posible realizar cálculos explícitos en el caso en los objetos son casi seguramente convexos. En este caso se puede aplicar la fórmula de Steiner<sup>4</sup>:

$$\mathbb{E}\{|A \oplus K|\} = \sum_{i=0}^d \binom{d}{i} \mathbb{E}\{W_i(A)\}r^i$$

y obtenemos

$$\mathbb{P}\{B(x, r) \subset X^c\} = \exp \left\{ -\theta \sum_{i=0}^d \binom{d}{i} \mathbb{E}\{W_i(A)\}r^i \right\}.$$

Estas fórmulas son muy útiles para comprobar la compatibilidad de un conjunto de datos experimentales con el modelo booleano, así como para inferir estadísticamente sus parámetros.

---

<sup>4</sup>Si  $B$  es un convexo, definimos a  $\dot{B}$  como el conjunto obtenido de  $B$  después de una rotación aleatoria uniforme. La fórmula de Steiner proporciona la media de los funcionales de Minkowski de  $K \oplus \dot{B}$ :

$$\mathbb{E}\{W_i(K \oplus \dot{B})\} = \sum_{j=0}^{d-i} \binom{d-i}{j} W_{i+j}(K)W_{d-j}(B)$$

## Capítulo

# 4

## Estadística de datos orientados

Antes de revisar el tema principal de éste capítulo es conveniente revisar algunos temas de probabilidad que son comunes o equivalentes.

### 4.1. Variables aleatorias

Una variable aleatoria es una función en donde el dominio es un espacio de posibles opciones  $S$  y la imagen son los números reales  $\mathbb{R}$  (Casella y Berger, 2002). El término *aleatorio* en la definición hace referencia a la ley probabilidad y no significa que tenga la misma probabilidad en cualquier intervalo de su imagen. A largo de este escrito se entenderá de esta manera.

*Notación:* Las variables aleatorias se denotarán con letras mayúsculas,  $X$  por ejemplo. Por otro lado, las realizaciones/observaciones de esta variable aleatoria (o su imagen) serán denotadas con las correspondientes letras minúsculas,  $x$  para el ejemplo anterior. ~~Como ejemplo más específico tenemos el siguiente:  $X$ , la estatura de los mexicanos, puede tomar el valor  $x = 1.7 \text{ m}$ .~~

El objeto de estudio de este proyecto doctoral fueron fenómenos que se modelaron con variables aleatorias continuas, es decir, variables aleatorias cuya imagen podía tomar cualquier valor dentro de cierto intervalo no nulo de los reales.

Las funciones aleatorias son completamente caracterizadas por su función de distribución de probabilidad  $F$ , también conocidas como función de distribución acumulativa (CDF, por sus siglas en inglés):

$$F_X(x) = P(X \leq x), \quad \forall x. \quad (4.1)$$

*O igual*

donde  $P$  denota la probabilidad de que la variable aleatoria  $X$  sea menor que el valor  $x$ . ~~Por ejemplo, la probabilidad de que la estatura,  $X$ , sea menor que  $x = 1.5 \text{ m}$ .~~

Para caracterizar completamente una variable aleatoria continua  $X$ , alternativamente a la función de distribución  $F$ , se tiene la función de densidad de probabilidad  $f$ :

$$F_X(x) = \int_{-\infty}^x f_X(t)dt \quad \forall x. \quad (4.2)$$

Si  $f_X(x)$  es continua, entonces, por el teorema fundamental del cálculo:

$$\frac{d}{dx} F_X(x) = f_X(x) \quad (4.3)$$

El histograma es un estimador de la función de densidad  $f_X$ .

Para mostrar las características de una CDF usemos la diferencia hacia adelante de primer orden:

$$\Delta F := F(x + d) - F(x) \quad (4.4)$$

Una función de distribución  $F$  satisface las siguientes condiciones:

- $F(x)$  es una función no-decreciente de  $x$ . Es decir,  $\Delta F \geq 0$ .
- $F(x)$  continua por la derecha. Es decir, para cada número  $x_0$ ,  $\lim_{x \rightarrow x_0^+} F(x) = F(x_0)$ .
- $\lim_{x \rightarrow -\infty} F(x) = 0$ , y  $\lim_{x \rightarrow +\infty} F(x) = 1$ .

Nótese que no hay significado probabilístico en estas definiciones, dicha interpretación se obtiene a partir de la definición en la Ecuación 4.1. Gráficamente, todas estas características se pueden observar en la Figura 4.1.

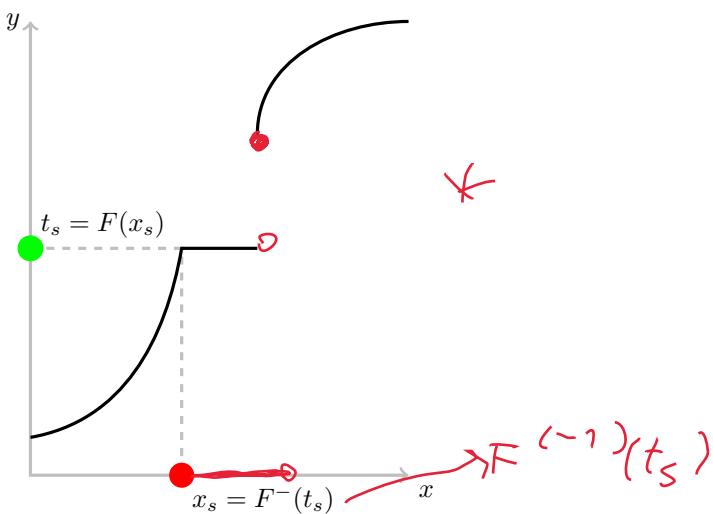


Figura 4.1: Ejemplo de función de distribución de probabilidad.

Sea  $S := \{x_1, \dots, x_n\}$  el conjunto formado por observaciones (valores medidos o muestreados) de la variable aleatoria  $X$ . Entonces se pueden estimar los valores de la función de distribución  $F$  a partir de su función de distribución empírica.

$$\hat{F}_n(x) = \frac{1}{n} \sum_{k=1}^n \mathbb{1}(x_k \leq x) \quad (4.5)$$

donde  $\mathbb{1}$  es la función indicador (Ecuación 2.1) y  $n$  es la cantidad de datos.

Esta función empírica es la función no paramétrica aproximante más conocida de una función de distribución, pero también existen otras como la de Weibull plotting position ó estándar (Salvadori *et al.*, 2007, p. 10). No se debe confundir con la función de densidad de Weibull, la cual es continua. En este trabajo utilizaremos la Ecuación 4.5 debido a su simplicidad y a su similitud con la cópula empírica que se verá más adelante.

A partir de ahora, en lo que resta de esta subsección, supondremos que los datos observados  $x_i$  están ordenados en orden creciente ( $x_i < x_{i+1}$ ) y no hay datos duplicados, es decir, que los datos son iguales a sus estadísticos de orden  $x_{(i)}$ .

El gráfico de la función de distribución empírica (Figura 4.2) es constante a tramos con saltos de magnitud  $1/n$  cada  $x_i$ . Si definimos las **pseudo-observaciones** como  $u_i := \hat{F}_n(x_i)$ , interpretada como una probabilidad en el intervalo  $[0, 1]$ ,  $u_i$  es una cantidad adimensional.

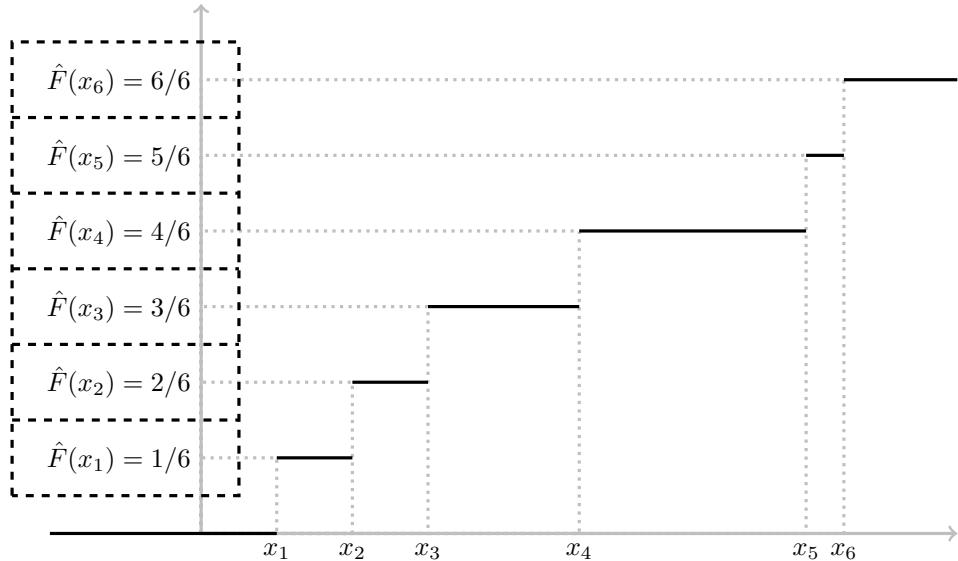


Figura 4.2: Función de distribución empírica. Los rectángulos en negro con líneas de guiones se muestran el correspondiente arreglo computacional para almacenar los valores de  $\hat{F}_n$ .

Desde el punto de vista computacional, para obtener valores de esta función empírica solamente basta con almacenar los valores de  $\hat{F}_n(x_i)$ , para toda  $i = 1, \dots, n$ .

$$[\hat{F}_n(x_1) \mid \hat{F}_n(x_2) \mid \hat{F}_n(x_3) \mid \dots \mid \hat{F}_n(x_{n-1}) \mid \hat{F}_n(x_n)]$$

De esta manera para saber el valor de  $\hat{F}_n$  en cualquier otro valor  $x$ , solamente hay que compararlo contra todas las observaciones y saber cual  $x_i$  es el inmediato inferior. Ésto es lo que significa

~~X~~ la ~~ecuación~~ Ecuación 4.5: para obtener un valor  $\hat{F}(x)$ ,  $x$  se compara con cada observación  $x_i$ . Nótese que ya no se requieren los valores  $x_i$  ya que están implícitos en vector de valores de  $\hat{F}_n$ , solamente es necesario saber cuál es el índice  $i$  para poder acceder al arreglo. Para lenguajes de programación de bajo nivel como C/C++ o Fortran, esto se pudiera implementar en un arreglo de enteros sin signo, lo que reduce el uso de memoria comparado con un arreglo de números reales (valores flotantes). Los índices de la Figura 4.2 corresponden uno-a-uno con arreglos de lenguajes de programación con offset 1 como R o matlab, pero no para arreglos de lenguajes de programación de offset 0 como python o C++. Aún más, pudiera ahorrarse guardar tal arreglo ya que el cálculo de  $1/i$  es computacionalmente muy rápido.

La implementación computacional de la función de distribución empírica está en la función

\* ~~stats::ecdf.~~ *en R*

Como puede verse en la Figura 4.2, la función de distribución empírica no es continua, pero existen muchos modelos que sí lo son. Por ejemplo, se encuentran la función normal, la lognormal, ley de potencias, Weibull, exponencial, entre otras. Estos ejemplos, salvo la distribución normal, tienen asimetría positiva, y son comúnmente utilizadas en el modelado de las longitudes de fracturas (Bonnet *et al.*, 2001; Bour y Davy, 1997; Gudmundsson *et al.*, 2011). Aunque estas distribuciones dependen de ciertos parámetros (la media y la varianza en el caso de la distribución normal), también existen funciones de distribución no paramétricas. Otras variables que pueden ser modeladas con dichas funciones son la apertura, la porosidad y la permeabilidad.

En muchas ocasiones uno de los objetivos buscados mediante el análisis y modelado de los datos es la simulación, ya que permite cuantificar la incertidumbre del fenómeno. Un algoritmo de simulación denominado de la transformada inversa requiere la función cuantil. Ésta es una función asociada con una variable aleatoria  $X$ , la cual se define como la inversa generalizada  $F^- : [0, 1] \rightarrow \bar{\mathbb{R}} = [-\infty, \infty]$  de la función de distribución  $F$  (Embrechts y Hofert, 2013):

$$F^-(y) = \inf\{x \in \mathbb{R} : F(x) \geq y\} \quad (4.6)$$

con la convención que  $\inf \emptyset = \infty$ . Esta inversa puede visualizarse en la Figura 4.1. Si  $F$  es estrictamente creciente, entonces  $F^- = F^{-1}$ , la inversa generalizada es igual a la inversa usual.

Dicha función cuantil, puede o no existir en forma explícita independientemente si la función  $F$  existe en forma explícita. Por ejemplo, se tiene una fórmula algebraica para la función de densidad de la distribución normal pero no se tiene una para calcular la función de distribución o la función cuantil, para estos casos se tiene que recurrir a métodos numéricos. Un ejemplo parecido es el encontrado con la función de densidad de von Mises (Veáse Sección 4.2).

Un ejemplo opuesto se tiene con los polinomios de Bernstein-Kantorovich (Muñoz-Pérez y Fernández-Palacín, 1987), que representan de forma semi-explicativa la función cuantil, pero para obtener su correspondiente función de distribución se tiene que recurrir a métodos numéricos (Dahlquist y Björck, 2008; Quarteroni *et al.*, 2006). Esta función no paramétrica y continua es un caso particular de una curva de Bézier (ver Apéndice A). La implementación computacional de esta función puede ser utilizada mediante la función `lmomco::dat2bernqua` (Asquith, 2017).

Encontrar numéricamente la inversa de una función  $F$  para un valor  $y$  en su imagen, es equi-

valente a encontrar la raíz  $x$  de la función  $G(x, y) = F(x) - y = 0$  (wikipediaRootFinding2016, 2016). Una implementación computacional para encontrar dicha raíz se puede encontrar en el software R con la función `stats::uniroot`. Otra función que podría ser más rápida es utilizando una versión paralelizada del método de bisección para encontrar dicha raíz (Karniadakis y Kirby, 2003; Miranker, 1971; Nijmeijer, 2015; Overflow, 2017). También se puede aprovechar el hecho de que la función  $G$ , debido a que es una traslación de  $F$ , es monótona (Exchange, 2017; MathOverflow, 2017).

## 4.2. Variables orientadas

Los *datos direccionales* (también conocidos como *datos orientados*) surgen de varias maneras. En particular, cuando se trata de una dimensión (una sola variable aleatoria) se les llama *datos circulares* y esta variable se encuentra definida sobre el círculo unitario,  $\mathbb{S}^1 = [0, 2\pi]$ . Nótese que ésta es una variable aleatoria con periodo  $2\pi$ . Algunos ejemplos de variables aleatorias periódicas son la dirección de una brújula (como dirección del viento, dirección del vuelo, dirección de rumbo, etc.), la posición de las manecillas del reloj, la temperatura diurna, o la anual, etc.

 En la caracterización ~~de~~ de fracturas o fallas geológicas, se cuenta con datos cuya característica es tener dos componentes de este estilo. Estas componentes son el *echado* y *azimut*; ambas asociadas a un solo dato y ambas medidas en ángulos. Estas dos componentes generan un vector (o dato) en 3D, por lo que la herramienta que se debería considerar para el análisis estadístico son los *datos esféricos* Fisher *et al.* (1993); Jammalamadaka y SenGupta (2001).

En el caso univariado, la estadística de datos orientados surge de manera natural si se intenta calcular la media y la desviación estándar del conjunto de datos que se muestra en la Figura 1. Nótese que, si se calcula el promedio usual, el valor estimado para la media sería de  $180^\circ = (10+350)/2$ , la cual es una dirección diametralmente opuesta a la sugerida por los datos ( $0^\circ$  = Este geográfico). También nótese que si se procede con este camino la varianza correspondiente también discrepa mucho de lo que se observa en el *histograma circular* de la izquierda de la figura.

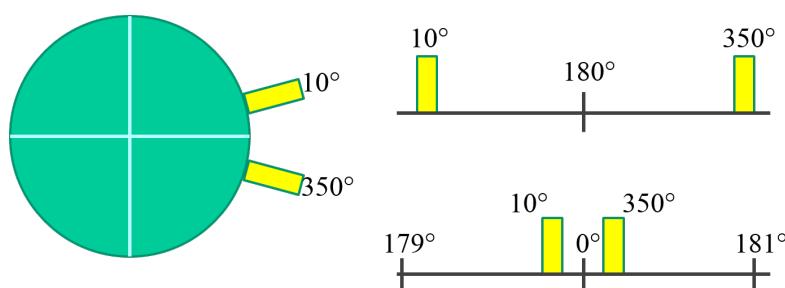


Figura 4.3: Motivación de la estadística de datos orientados. ¿Cuál es la media? ¿Y la desviación estandar?

La solución a este problema lo proporciona la estadística de datos orientados al hacer uso del cálculo vectorial. Con este enfoque los estadígrafos básicos para datos circulares se definen de la

siguiente manera (Mardia y Jupp, 2000, ch. 2) y están implementados en el paquete **circular** del software estadístico *R*:

Supongamos que se tienen  $n$  datos, si cada orientación  $\theta_i$  se puede hacer corresponder con un vector unitario  $\mathbf{x}_i = (\cos \theta_i, \sin \theta_i)$ ; entonces la dirección media  $\bar{\theta}$  es la dirección del vector resultante  $\mathbf{x}_1 + \dots + \mathbf{x}_n$ . Las coordenadas cartesianas del centro de masa son  $(\bar{C}, \bar{S})$ , donde

$$\bar{C} = \frac{1}{n} \sum_{j=1}^n \cos \theta_j, \quad \bar{S} = \frac{1}{n} \sum_{j=1}^n \sin \theta_j \quad (4.7)$$

Por lo tanto  $\bar{\theta}$  es la solución a las ecuaciones

$$\bar{C} = \bar{R} \cos \bar{\theta}, \quad \bar{S} = \bar{R} \sin \bar{\theta} \quad (4.8)$$

Donde la longitud resultante media  $\bar{R}$  está dada por

$$\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2} \quad (4.9)$$

Observe que  $0 \leq \bar{R} \leq 1$ . Si las direcciones  $\theta_1, \dots, \theta_n$  son muy cercanas entre sí, entonces  $\bar{R}$  será cercano a 1, pero si está muy dispersas entonces  $\bar{R}$  será cercano a 0. Por lo tanto  $\bar{R}$  es una medida de *concentración* de los datos. Nótese que en la *estadística convencional* (o de *datos no orientados*), la varianza nos proporciona una medida de *dispersión*, opuesto a la interpretación de  $\bar{R}$ . Análogo a la varianza convencional y otros estadígrafos pueden encontrarse en el capítulo 2 de Mardia y Jupp (2000). Para datos esféricos o datos direccionales en  $\mathbb{R}^p$ , consulte el capítulo 9 del mismo libro o el libro de Fisher *et al.* (1993).

Supongamos que se ha especificado una dirección (origen) y orientación (sentido de giro) inicial. Dos ejemplos de tal especificación son: 1) un plano cartesiano matemático ( $X^+ = 0, Y^+ = 90^\circ$ ), y 2) uno geográfico ( $N=0^\circ, E=90^\circ$ ). Entonces la función de distribución  $F$  para un ángulo aleatorio  $\theta$  se define como la función en  $\mathbb{R}$  dada por

$$F(x) = \Pr(0 < \theta \leq x)$$

$$0 \leq x \leq 2\pi$$

$$\begin{aligned} &y \\ F(x + 2\pi) - F(x) &= 1 \quad -\infty < x < \infty \end{aligned}$$

La ecuación significa que cualquier arco de longitud  $2\pi$  tiene probabilidad 1 (ya que tal arco es la totalidad de la circunferencia del círculo).

A diferencia de distribuciones en los reales,  $\mathbb{R}$ ,

$$F(x) = \infty, \quad F(x) = -\infty$$

Por definición,

$$F(0) = 0, \quad F(2\pi) = 1$$

Para una variable circular, una función  $f$  es la densidad de probabilidad de una función de distribución totalmente (absolutely) continua si y sólo si

- $f(\theta) \geq 0$  casi por doquier en  $(-\infty, \infty)$ ,
- $f(\theta + 2k\pi) = f(\theta)$  casi por doquier en  $(-\infty, \infty)$  con  $k \in \mathbb{Z}$  (i.e.,  $f$  es periódica),
- $\int_0^{2\pi} f(\theta) d\theta = 1$ .

Las funciones de densidad uni y bidimensional de probabilidad paramétricas para datos orientados más comunes son:

von Mises	$f(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \mu)}$
Fisher	$f(\theta, \phi; \alpha, \beta, \kappa) = \frac{1}{4\pi \sinh(\kappa)} \exp[\kappa[\sin \theta \sin \alpha \cos(\phi - \beta) + \cos \theta \cos \alpha]] \sin \theta$

Tabla 4.1: Funciones de densidad de probabilidad para datos orientados.  $0 \leq \theta < 2\pi$ ,  $0 \leq \phi < \pi/2$ ,  $0 \leq \kappa < \infty$ . Ver ec. 3.5.17 de Mardia y Jupp (2000), ec. 2.2.6 en la sección 2.2.4 de Jammalamadaka y SenGupta (2001) y ec. 4.22 de Fisher *et al.* (1993).

La función de densidad de von-Mises es el análogo a la función de densidad de probabilidad normal ya que también consta de dos parámetros, es unimodal y simétrica.



# Capítulo 5

## Modelación de dependencias mediante la teoría de cópulas.

La metodología moderna para modelar dependencias estocásticas es la teoría de cópulas. Ésta ha demostrado más generalidad y versatilidad que el enfoque de regresión clásico (lineal, por ejemplo) para modelar dependencias.

Aunque la teoría de cópulas no es una teoría tan nueva (Sklar, 1959), se han dado pocas aplicaciones prácticas. Dentro de la industria petrolera se han hecho algunas aplicaciones para simular conjuntamente propiedades petrofísicas usando cópulas (Díaz-Viera, 2012; Díaz-Viera y Anguiano-Rojas, 2006; Díaz-Viera y Casar-González, 2005a,b; Díaz-Viera *et al.*, 2006; Erdely y Díaz-Viera, 2009, 2011, 2012; Erdely *et al.*, 2012; Hernández-Maldonado *et al.*, 2012a,b,c; Hernández-Maldonado y Díaz-Viera, 2014) pero aún no se han hechos trabajos respecto a las propiedades de fractura.

Desde un punto de vista, las cópulas son funciones que unen o “acoplan” funciones de distribución multivariada a sus funciones de distribución marginal unidimensional. No existen modelos clásicos que combinen de manera general datos orientados con datos no orientados por lo que con este trabajo se propondrán modelos que de manera adecuada tomen en cuenta la dependencia con la cópula, y a la vez respeten y reproduzcan las marginales.

### 5.1. Generalidades

\* *¿*Qué relación existe entre una función de distribución multivariada,  $H$ , y sus funciones de distribución marginal,  $F$  y  $G$ ? La solución a este problema la dio Sklar (1959) al crear una nueva clase de funciones a las que llamó cópulas.

Para poder abordar adecuadamente la definición de cópulas y el teorema de Sklar (1959) observemos primero algunas definiciones.

Sean  $S_1, S_2 \subset \mathbb{R}$  y  $H : S_1 \times S_2 \rightarrow \mathbb{R}$ . Si denotamos al ínfimo como “ínf”, y suponemos que

\*  $a_1 = \inf S_1$ ,  $a_2 = \inf S_2$ . Decimos que  $H$  está *aterrizada* si  $H(x, a_2) = 0 = H(a_1, y)$  para todo  $(x, y) \in S_1 \times S_2$ . De igual manera, si denotamos al supremo como “sup”, y suponemos que  $b_1 = \sup S_1$  y  $b_2 = \sup S_2$ . Decimos que una función  $H$  tiene *márgenes*, y que los *márgenes de  $H$*  son las funciones  $F$  y  $G$  dadas por

$$\text{Dom } F = S_1 \text{ y } F(x) = H(x, b_2) \text{ para todo } x \in S_1$$

$$\text{Dom } G = S_2 \text{ y } G(y) = H(b_1, y) \text{ para todo } y \in S_2$$

Donde “Dom” denota el dominio de la función correspondiente.

Sean  $S_1, S_2 \subset \mathbb{R}$ , y  $B = [x_1, y_1] \times [x_2, y_2]$  un rectángulo tal que todos sus vértices están en  $S_1 \times S_2$ . Una función  $H : S_1 \times S_2 \rightarrow \mathbb{R}$  es bi-creciente si ...

*márgenes*

?

**Definición. Subcópula.** Una subcópula bidimensional (o 2-subcópula) es una función  $C'$  con las siguientes características:

1.  $\text{Dom} C' = S_1 \times S_2$  donde  $S_1$  y  $S_2$  son subconjuntos de  $\mathbb{I} = [0, 1]$  que contienen al cero y a 1,
2.  $C'$  está aterrizada y es bi-creciente.
3. Para cada  $u \in S_1$ ,  $C'(u, 1) = u$ ; y para cada  $v \in S_2$ ,  $C'(1, v) = v$ .

**Definición. Cúpula.** Una cúpula bidimensional (ó 2-cúpula) es una 2-subcópula  $C$  tal que  $\text{Dom } C = \mathbb{I}^2$ .

**Teorema de Sklar (1959).** Sea  $H$  una función de distribución con funciones *márgenes*  $F$  y  $G$ . Entonces existe una cúpula  $C$  tal que para todo  $x, y \in \mathbb{R}$ ,

$$H(x, y) = C(F(x), G(y)) \quad (5.1)$$

Si  $F$  y  $G$  son continuas, entonces  $C$  es única, de otra manera  $C$  está determinada de manera única en  $\text{Ran } F \times \text{Ran } G$ . Si  $C$  es una cúpula y  $F$  y  $G$  son funciones de distribución, entonces la función  $H$  definida por la ecuación (5) es una función de distribución conjunta con *márgenes*  $F$  y  $G$ .

En la teoría clásica de cópulas se suponen modelos paramétricos tanto para  $H$  como para las *márgenes*  $F$  y  $G$ . Por ejemplo,  $F$  pudiera ser *normal* con parámetros  $\mu$  y  $\sigma^2$ ; y  $G$  *exponencial* con parámetro  $\lambda$ , mientras  $H$  la distribución normal bivariada con sus correspondientes parámetros.

Los modelos paramétricos son importantes en varios aspectos. Por ejemplo, sus parámetros proveen una interpretación rápida del fenómeno en estudio. Existen todo un catálogo de modelos (?) que explican una gran variedad de fenómenos. Cuando se tiene poca confianza en los datos ó se cuenta con modelos hipotéticos, se pueden probar ciertos modelos paramétricos. Generalmente, son más fáciles de implementar computacionalmente que los modelos no paramétricos y más rápidos.

Como complemento a los modelos paramétricos se tienen los modelos no-paramétricos. Estos últimos son realmente útiles cuando la estructura de dependencia es más compleja que relaciones monótonas entre variables únicamente (Joe, 2014). En el caso de existir *márgenes* con formas

complicadas (no simétricas, multimodales, etc.), entonces el enfoque paramétrico puede ser aplicado a la distribución multivariada directamente.

El ejemplo más simple de una cópula no paramétrica es la cópula empírica, que se puede definir de la siguiente manera (Berghaus *et al.*, 2017; Bucher y Volgushev, 2011; Carnicer *et al.*, 2013; Deheuvels, 1979; Fermanian *et al.*, 2004; Radulović *et al.*, 2017). Dada una muestra,  $\{(x_1, y_1), \dots, (x_n, y_n)\}$ , de la función de distribución conjunta del vector aleatorio  $(X, Y)$ ; con sus correspondientes pseudo-observaciones obtenidas mediante la Ecuación 4.5. Entonces, la función de distribución cópula empírica se define como,

$$\hat{C}_n(u, v) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(u_i \leq u, v_i \leq v) \quad (5.2)$$

Para  $1 \leq i \leq n$ ,  $1 \leq j \leq n$  donde  $\mathbb{1}(a \leq b, c \leq d)$  si  $a \leq b$  y  $c \leq d$  y cero en otro caso, con  $a, b, c, d \in \mathbb{R}$ .

La Ecuación 5.2 se puede reescribir de la forma de la Ecuación 5.3:

$$n\hat{C}_n(u, v) = \sum_{i=1}^n \mathbb{1}(u_i \leq u, v_i \leq v) \quad (5.3)$$

De esta manera  $n\hat{C}_n$ , de manera análoga a la forma empírica univariada (Ecuación 4.5), puede ser almacenado como un arreglo bidimensional de enteros sin signo. *Ésto* lleva a evitar errores de redondeo que se puedan propagar ya que para  $n$  grandes, digamos  $n = 1 \times 10^6$ , lo normal en estos tiempos,  $1/n = 1 \times 10^{-6} = 0.000001$ . Además, también hay un ahorro considerable en la memoria de la computadora. Como ventaja adicional, ayuda a obtener visualizaciones más sencillas.

Para encontrar cualquier valor  $\hat{C}_n(u, v)$  simplemente se compara  $(u, v)$  con los valores  $(u_i, v_i)$  siguiendo la función indicador y de forma acumulativa 2-creciente (hacia la derecha y hacia la arriba).

*sin acento*

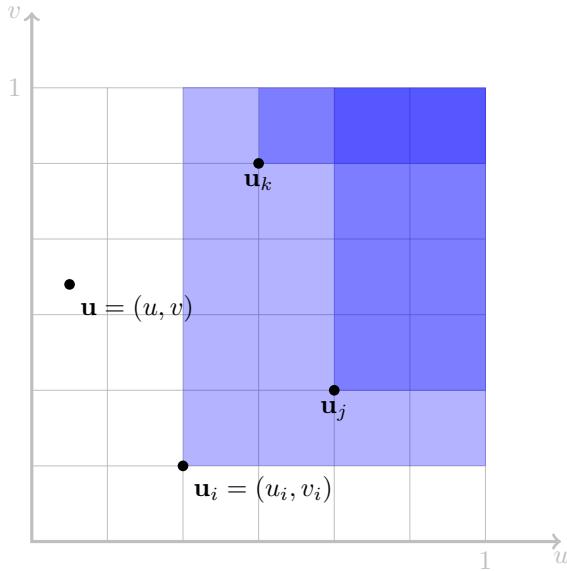


Figura 5.1: Donde  $i, j, k \in \{1, \dots, n\}$ . Tonos más oscuros representan valores de  $n\hat{C}_n$  más grandes. Cada cambio en tono representa una diferencia de uno (por la función indicador).

*muchos espacios*

Esta forma permite un diagrama más sencillo que relaciona la ubicación en el gráfico de pseudo-observaciones con su correspondiente arreglo en memoria de la computadora. Supongamos que los datos  $\{(x_i, y_i)\}$  están ordenados por valores de  $x$  crecientes, es decir,  $u_i = u_{(i)}$ . Esto no implica que  $v_i = v_{(i)}$ . Además, suponemos que no hay valores repetidos.

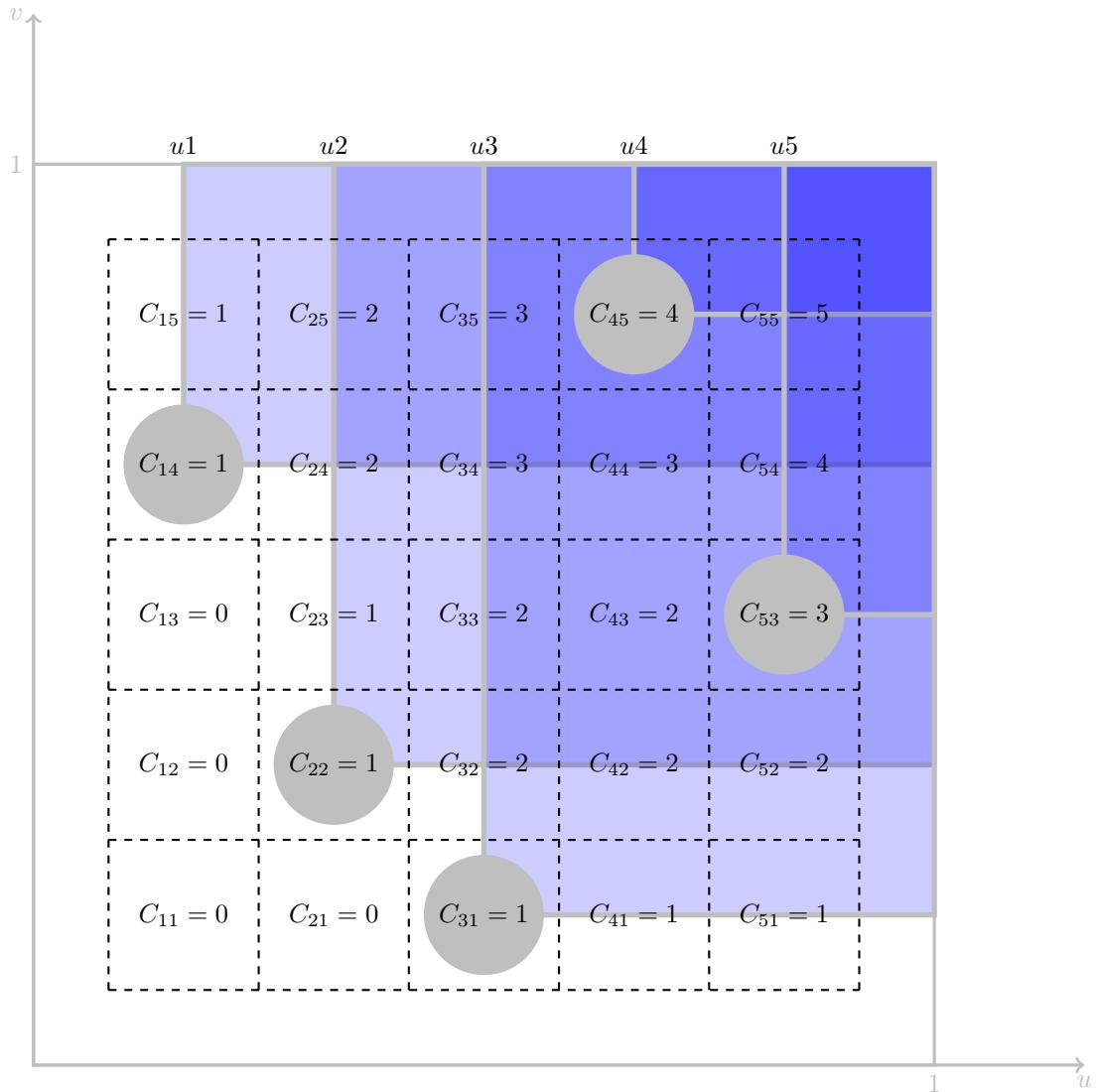


Figura 5.2: Representación gráfica de la cópula empírica y su relación con su almacenaje computacional en una matriz. El centro de cada círculo representa una pseudo-observación  $\mathbf{u}_i = (u_i, v_i)$ . La malla regular de líneas de guiones representa el arreglo en la memoria,  $n\mathbf{C}$ , que guarda los valores enteros  $C_{ij} := n\hat{C}_n(u_i, v_j)$ . Donde  $i$  denota el renglón y  $j$  la columna. Nótese que las  $u_i$  están ordenadas crecientemente pero no existe un orden en las  $v_i$ . Alternativamente se pudo haber ordenado por  $v_i$  y dejar las  $u_i$  sin orden.

Para ligar la Figura 5.2 con su correspondiente representación matricial  $n\mathbf{C}$  (Ecuación 5.4), se referirá a la primera mediante palabras como horizontal/vertical, ó paralelo al eje  $u/v$ ; mientras que para referirse al correspondiente arreglo/matriz computacional se nombrarán renglones y columnas. Por ejemplo, el segundo renglón de la matriz  $n\mathbf{C}$ ,  $n\mathbf{C}(2, \cdot)$ , corresponde al rectángulo vertical de la Figura 5.2 centrado en  $u_2$ , paralelo al eje  $v$ .

sin acento

$$n\mathbf{C} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} \\ C_{21} & C_{22} & C_{23} & C_{24} & C_{25} \\ C_{31} & C_{32} & C_{33} & C_{34} & C_{35} \\ C_{41} & C_{42} & C_{43} & C_{44} & C_{45} \\ C_{51} & C_{52} & C_{53} & C_{54} & C_{55} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 & 2 \\ 1 & 2 & 2 & 3 & 3 \\ 1 & 2 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix} \quad (5.4)$$

Veámos el arreglo en la Figura 5.2 por columnas, es decir, por rectángulos verticales de ancho  $\Delta u_i$  y centrados en cada una de las pseudo-observaciones. Nótese que, si fijamos un renglón, digamos el renglón  $k$  que corresponde a  $u_k$ , los valores de  $n\mathbf{C}$  satisfacen la siguiente relación de recurrencia:

$$n\mathbf{C}(k, \cdot) = n\mathbf{C}(k - 1, \cdot) + \mathbf{v}_k \quad (5.5)$$

donde el vector  $\mathbf{v}_k := (0, 0, \dots, 0, 1, 1, \dots, 1)$  contiene su primer 1 en la entrada  $rank(v_k)$ , el rango de  $v_k$ . Si  $n\mathbf{C}$  fue inicializada con puros 0, entonces se puede ahorrar memoria omitiendo la creación de un nuevo arreglo  $\mathbf{v}_k$  al utilizar  $n\mathbf{C}(k, \cdot) := \mathbf{v}_k$ . Alternativamente, se puede proceder, primero por columnas y luego por renglones para obtener el mismo resultado. La preferencia dependerá del paradigma del arreglo en memoria del lenguaje de programación utilizado: *row major* ó *column major*. También se puede hacer más eficiente al observar que el último renglón y la última columna tienen entradas correspondientes a 1, 2, 3, 4, 5, para este caso. En el caso general tendrán entradas 1, 2, ...,  $n$ . La implementación computacional de esta matriz se encuentra en la función `berNSTEINcop::empCopM`. La función `berNSTEINcop::genmat.copem` implementa un algoritmo similar pero no igual ya que compara los valores de  $v_i$  en lugar de tomar el rango.

Para calcular el valor de la cópula empírica en cualquier posición  $(u, v)$  solamente basta con saber los máximos índices  $i$  y  $j$  que cumplen  $(u_i, v_j) \leq (u, v)$  elemento-a-elemento. Y  $\hat{C}_n(u, v) = n\mathbf{C}(i, j)/n$  (ver la Figura 5.1).

Nótese que la Ecuación 5.2 se puede escribir de la siguiente forma más compacta y más parecida a su contraparte univariada (Ecuación 4.5):

$$\hat{C}_n(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(\mathbf{u}_i \leq \mathbf{u}) \quad (5.6)$$

donde  $\mathbf{u}_i := (u_i, v_i)$ ,  $\mathbf{u} := (u, v)$  y la relación de orden,  $\leq$ , se efectúa componente a componente.

Se puede ver que la cópula empírica no es continua. Un modelo continuo de estructura de dependencia asociada dos variables aleatorias continuas e independientes  $X$  y  $Y$  es:

$$C_{XY}(u, v) = \Pi(u, v) := uv \quad (5.7)$$

Aunque la cópula empírica no es continua, ella es útil para obtener versiones continuas de la cópula.

## 5.2. Cópulas bivariadas de Bernstein

Una versión continua de la cópula empírica se puede obtener mediante un ajuste polinomial. Una base del espacio de polinomios de grado  $n$  son los polinomios de Bernstein (veáse la Sección A.1):

$$B(w|k, n) := \binom{k}{n} w^k (1-w)^{n-k} \quad (5.8)$$

donde  $w \in [0, 1]$ ,  $k \in \{0, 1, \dots, n\}$ .

Muchos lenguajes de programación tienen implementada eficientemente ésta función, conocida en estadística como la función de masa binomial. Por ejemplo, en R está implementada bajo el nombre `dbinom`. Estos polinomios pueden ser utilizados para aproximar y suavizar la cópula empírica (Ecuación 5.2), lo que nos lleva a la cópula empírica de Bernstein propuesta por Sancetta y Satchell (2004):

$$C_B(u, v) = \sum_{i=0}^n \sum_{j=0}^n \hat{C}_n \left( \frac{i}{n}, \frac{j}{n} \right) B(u|i, n) B(v|j, n) \quad (5.9)$$

En ésta definición, la cópula empírica está definida incluso para valores  $u = 0$ , mientras que en la definición de la función acumulativa empírica (Ecuación 4.5) la mínima pseudo-observación tomaba el valor  $u_1 = 1/n$ . Este inconveniente se resuelve añadiendo  $u_0 = 0$  (Sancetta y Satchell, 2004) al vector de pseudo-observaciones empíricas. Un ejemplo de la aplicación de esta teoría en ciencias de la Tierra se muestra en la Figura 5.3.

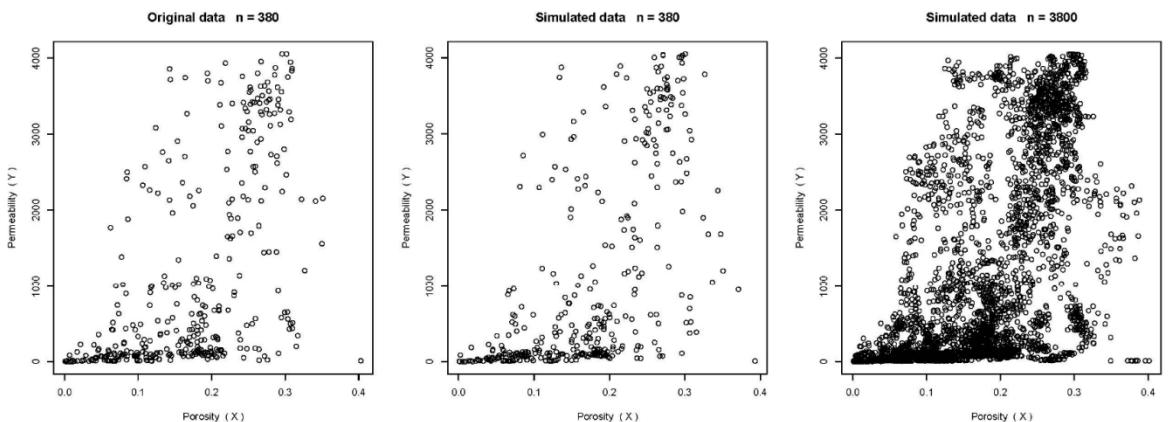


Figura 5.3: Ejemplo de la aplicación de la teoría de cópulas de Bernstein (Erdely y Díaz-Viera, 2012).

Para obtener  $C_B(u, v)$  se puede utilizar el algoritmo de de'Casteljau ó se puede hacer mediante el uso del producto tensorial. Dado que hay suficientes trabajos describiendo el primer enfoque (Goldman, 2002; Mann, 2006), expliquemos el enfoque mediante tensores (Renteln, 2014) ya que permite una representación más fácil para su implementación computacional.

$$C_B(u, v) = \sum \mathbf{A} \quad (5.10)$$

donde hemos definido la notación  $\sum \mathbf{A}$  como la suma de todos los elementos del arreglo  $\mathbf{A}$ , la cual está definida de la siguiente manera:

$$\mathbf{A} = \begin{pmatrix} \hat{C}(0, 0)B(u|0, n)B(v|0, n) & \hat{C}(0, 1)B(u|0, n)B(v|1, n) & \cdots & \hat{C}(0, n)B(u|0, n)B(v|n, n) \\ \hat{C}(1, 0)B(u|1, n)B(v|0, n) & \hat{C}(1, 1)B(u|1, n)B(v|1, n) & \cdots & \hat{C}(1, n)B(u|1, n)B(v|n, n) \\ \hat{C}(2, 0)B(u|2, n)B(v|0, n) & \hat{C}(2, 1)B(u|2, n)B(v|1, n) & \cdots & \hat{C}(2, n)B(u|2, n)B(v|n, n) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{C}(n, 0)B(u|n, n)B(v|0, n) & \hat{C}(n, 1)B(u|n, n)B(v|1, n) & \cdots & \hat{C}(n, n)B(u|n, n)B(v|n, n) \end{pmatrix} \quad (5.11)$$

también definimos  $\hat{C}(i, j) := \hat{C}_n(i/n, j/n)$ .

Existen varios factores repetidos que se pueden omitir si calculamos los siguientes vectores columna:

$$\mathbf{B}(u|n) := (B(u|0, n), B(u|1, n), \dots, B(u|n, n))^T \quad (5.12)$$

$$\mathbf{B}(v|n) := (B(v|0, n), B(v|1, n), \dots, B(v|n, n))^T \quad (5.13)$$

y haciendo el producto exterior (o producto tensorial) de dos vectores  $\mathbf{a}$  y  $\mathbf{b}$ , denotado mediante  $\mathbf{a} \otimes \mathbf{b}$ , el cual produce un matriz  $\mathbf{w}$  con coordenadas  $w_{ij} = a_i b_j$ . Si los vectores son vectores columna,  $\mathbf{a} \otimes \mathbf{b}$  es equivalente a la multiplicación matricial  $\mathbf{ab}^T$ . Un ejemplo particular se tiene si  $\mathbf{a} = (1, 2)^T \in \mathbb{R}^2$  y  $\mathbf{b} = (3, 4, 5)^T \in \mathbb{R}^3$ :

$$\mathbf{a} \otimes \mathbf{b} = \mathbf{ab}^T = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \begin{pmatrix} b_1 & b_2 & b_3 \end{pmatrix} = \begin{pmatrix} a_1 b_1 & a_1 b_2 & a_1 b_3 \\ a_2 b_1 & a_2 b_2 & a_2 b_3 \end{pmatrix} \quad (5.14)$$

$$= \begin{pmatrix} 1 \\ 2 \end{pmatrix} \begin{pmatrix} 3 & 4 & 5 \end{pmatrix} = \begin{pmatrix} 1 \cdot 3 & 1 \cdot 4 & 1 \cdot 5 \\ 2 \cdot 3 & 2 \cdot 4 & 2 \cdot 5 \end{pmatrix} = \begin{pmatrix} 3 & 4 & 5 \\ 6 & 8 & 10 \end{pmatrix} \quad (5.15)$$

Por lo tanto,  $\mathbf{A}$  se puede reescribir de la siguiente manera:

$$\mathbf{A} = \hat{\mathbf{C}} \odot [\mathbf{B}(u|n) \otimes \mathbf{B}(v|n)] \quad (5.16)$$

donde  $\odot$  es el producto elemento-a-elemento de las matrices (también conocido como el producto Hadamard o de Schur), y  $\hat{\mathbf{C}}$  es la matriz cópula empírica (Ecuación 5.4) cuyos elementos son  $\hat{C}_n(u_i, v_j)$ . Nótese que tanto  $\mathbf{B}(u|n) \otimes \mathbf{B}(v|n)$  como  $\hat{\mathbf{C}}$  pertenecen a  $\mathbb{M}^{(n+1) \times (n+1)}$ , las matrices de  $(n+1) \times (n+1)$ ,  $n+1$  renglones por  $n+1$  columnas.

De esta forma, el código computacional se puede hacer eficiente y fácil de leer:

$$C_B(u, v) = \sum \left\{ \hat{\mathbf{C}} \odot [\mathbf{B}(u|n) \otimes \mathbf{B}(v|n)] \right\} \quad (5.17)$$

Más adelante, en la sección de simulación, se requiere  $\partial C_B / \partial u$ , la que a su vez requiere las diferencias de primer orden con respecto a  $u$  de la cópula empírica,  $\Delta_u \hat{C}_n$ :

$$\Delta_u \hat{C}_n \left( \frac{i}{n}, \frac{j}{n} \right) = \hat{C}_n \left( \frac{i+1}{n}, \frac{j}{n} \right) - \hat{C}_n \left( \frac{i}{n}, \frac{j}{n} \right) \quad (5.18)$$

Para entender la representación matricial de la Ecuación 5.18, a continuación se muestra un arreglo en el que cada entrada tiene la forma  $C_{(i+1)j} - C_{ij}$ .

$$\Delta_u \hat{\mathbf{C}} = \begin{pmatrix} C_{10} - C_{00} & C_{11} - C_{01} & C_{12} - C_{02} & \cdots & C_{1n} - C_{0n} \\ C_{20} - C_{10} & C_{21} - C_{11} & C_{22} - C_{12} & \cdots & C_{2n} - C_{1n} \\ C_{30} - C_{20} & C_{31} - C_{21} & C_{32} - C_{22} & \cdots & C_{3n} - C_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{n0} - C_{(n-1)0} & C_{n1} - C_{(n-1)1} & C_{n2} - C_{(n-1)2} & \cdots & C_{nn} - C_{(n-1)n} \end{pmatrix} \quad (5.19)$$

Este arreglo, se obtiene directamente del arreglo de la cópula empírica (Ecuación 5.4) extendida mediante  $C_{0i} = C_{i0} = 0$ .

$$\hat{\mathbf{C}} = \begin{pmatrix} C_{00} & C_{01} & C_{02} & C_{03} & \cdots & C_{0(n-1)} & C_{0n} \\ C_{10} & C_{11} & C_{12} & C_{13} & \cdots & C_{1(n-1)} & C_{1n} \\ C_{20} & C_{21} & C_{22} & C_{23} & \cdots & C_{2(n-1)} & C_{2n} \\ C_{30} & C_{31} & C_{32} & C_{33} & \cdots & C_{3(n-1)} & C_{3n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{(n-1)0} & C_{(n-1)1} & C_{(n-1)2} & C_{(n-1)3} & \cdots & C_{(n-1)(n-1)} & C_{(n-1)n} \\ C_{n0} & C_{n1} & C_{n2} & C_{n3} & \cdots & C_{n(n-1)} & C_{nn} \end{pmatrix} \quad (5.20)$$

Nótese que mientras  $\hat{\mathbf{C}} \in \mathbb{M}^{(n+1) \times (n+1)}$ ,  $\Delta_u \hat{\mathbf{C}} \in \mathbb{M}^{n \times (n+1)}$ . Al igual que el arreglo de la cópula empírica, para ahorrar memoria y evitar errores de redondeo que se puedan propagar a través de los cálculos, el arreglo  $\Delta_u \hat{\mathbf{C}}$  se puede guardar como  $\Delta_u n \hat{\mathbf{C}}$  (Ecuación 5.4).

Por otro lado, también es útil tener a la mano  $\Delta_v \hat{\mathbf{C}}$ , la diferencia finita de primer orden de la cónula empírica con respecto a la segunda variable. En este caso se tiene que  $\Delta_u \hat{\mathbf{C}} \in \mathbb{M}^{(n+1) \times n}$  y su arreglo tiene la forma de la Ecuación 5.21.

$$\Delta_v \hat{\mathbf{C}} = \begin{pmatrix} C_{01} - C_{00} & C_{02} - C_{01} & C_{03} - C_{02} & \cdots & C_{0n} - C_{0(n-1)} \\ C_{11} - C_{10} & C_{12} - C_{11} & C_{13} - C_{12} & \cdots & C_{1n} - C_{1(n-1)} \\ C_{21} - C_{20} & C_{22} - C_{21} & C_{23} - C_{22} & \cdots & C_{2n} - C_{2(n-1)} \\ C_{31} - C_{30} & C_{32} - C_{31} & C_{33} - C_{32} & \cdots & C_{3n} - C_{3(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{n1} - C_{n0} & C_{n2} - C_{n1} & C_{n3} - C_{n2} & \cdots & C_{nn} - C_{n(n-1)} \end{pmatrix} \quad (5.21)$$

Ahora, estamos preparados para definir las derivadas de la cónula de Bernstein (Phillips, 2003; Sancetta y Satchell, 2004):

$$\partial C_B / \partial u = \sum_{j=0}^n \left\{ B(v|j, n) n \left[ \sum_{i=0}^{n-1} \Delta_u \hat{C}_n \left( \frac{i}{n}, \frac{j}{n} \right) B(u|i, n-1) \right] \right\} \quad (5.22)$$

$$= n \sum_{i=0}^{n-1} \sum_{j=0}^n B(u|i, n-1) \Delta_u \hat{C}_n \left( \frac{i}{n}, \frac{j}{n} \right) B(v|j, n) \quad (5.23)$$

Los sumandos de esta ecuación, al igual que para la cónula de Bernstein (Ecuación 5.4), se pueden escribir como elementos de la matriz de la Ecuación 5.24:

$$\begin{pmatrix} B(u|0, n-1) \Delta_u \hat{C}_n(0, 0) B(v|0, n) & \cdots & B(u|0, n-1) \Delta_u \hat{C}_n(0, 1) B(v|n, n) \\ B(u|1, n-1) \Delta_u \hat{C}_n(\frac{1}{n}, 0) B(v|0, n) & \cdots & B(u|1, n-1) \Delta_u \hat{C}_n(\frac{1}{n}, 1) B(v|n, n) \\ B(u|2, n-1) \Delta_u \hat{C}_n(\frac{2}{n}, 0) B(v|0, n) & \cdots & B(u|2, n-1) \Delta_u \hat{C}_n(\frac{2}{n}, 1) B(v|n, n) \\ \vdots & \ddots & \vdots \\ B(u|n-1, n-1) \Delta_u \hat{C}_n(\frac{n-1}{n}, 0) B(v|0, n) & \cdots & B(u|n-1, n-1) \Delta_u \hat{C}_n(\frac{n-1}{n}, 1) B(v|n, n) \end{pmatrix} \quad (5.24)$$

Reutilizando la Ecuación 5.13 y la Ecuación 5.19, tenemos que

$$\frac{\partial C_B}{\partial u}(u, v) = n \sum \left\{ \Delta_u \hat{\mathbf{C}} \odot [\mathbf{B}(u|n-1) \otimes \mathbf{B}(v|n)] \right\} \quad (5.25)$$

De manera similar tenemos:

$$\frac{\partial C_B}{\partial v}(u, v) = n \sum \left\{ \Delta_v \hat{\mathbf{C}} \odot [\mathbf{B}(u|n) \otimes \mathbf{B}(v|n-1)] \right\} \quad (5.26)$$

La función de densidad de cópula  $c$  ( $c$  minúscula) es la segunda derivada parcial con resperco a cada una de las variables. En el caso de la copúa de Bernstein, ésta se puede escribir como (Carnicero *et al.*, 2013):

$$\hat{c}_B(u, v) = \sum_{i=0}^n \sum_{j=0}^n p_{ij} \beta(u|i, n-i+1) \beta(v|j, n-j+1) \quad (5.27)$$

Donde

$$\beta(x|a, b) = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1}$$

Y donde los pesos son

$$p_{ij} = \frac{1}{n} \sum_{k=1}^n \mathbb{1} \left( \frac{i-1}{n} < u_k \leq \frac{i}{n}, \frac{j-1}{n} < v_k \leq \frac{j}{n} \right) \quad (5.28)$$

Para  $i, j = 1, \dots, n$ .

Estos pesos son equivalentes al estimador del histograma en  $[0, 1]^2$  de la cópula (para más detalle de dichos estimadores veáse Scott (1992)); es decir,  $p_{ij}$  es  $1/n$  veces el número de puntos  $(u_i, v_i)$  dentro del rectángulo delimitado por el rectángulo  $(i-1, i]/k \times (j-1, j]/k$  (veáse Figura 5.4).

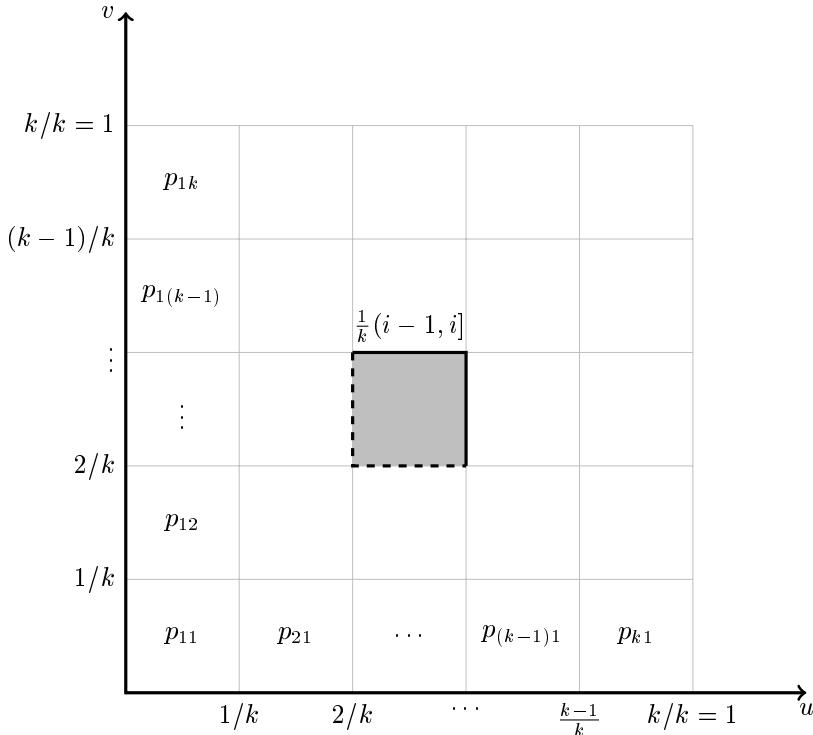


Figura 5.4: Representación esquemática de los pesos  $p_{ij}$  de la densidad de cópula.

Sancetta y Satchell (2004) y Phillips (2003) demuestran que el sesgo de los polinomios Bernstein de orden  $k$  es del orden de  $d/k + \mathcal{O}(1/k)$  para alguna constante finita  $d$ . Ellos también proporcionan valores óptimos para  $k$ .

Recientemente el uso de splines penalizados se ha incorporado a la teoría de cópulas no paramétricas Kauermann y Schellhase (2013). En dicho trabajo se muestra las ventajas que tienen los splines penalizados sobre otros enfoques no paramétricos como el de Bernstein, wavelets y kernel. Kauermann y Schellhase (2013) aproximan la copula densidad,  $c$ ,

$$c(u_1, \dots, u_p) \approx \sum_{\mathbf{k} \in \mathcal{K}} b_{\mathbf{k}} \phi_{\mathbf{k}}(u_1, \dots, u_p)$$

Donde  $b_{\mathbf{k}}$  son los coeficientes y  $\phi_{\mathbf{k}}$  son splines penalizados.

### 5.3. Cópulas bivariadas para vectores aleatorios con componentes orientadas

Johnson y Wehrly (1978); Wehrly y Johnson (1980) mostraron que la función de distribución conjunta de una variable  $X$ , y una circular  $\Theta$ , se puede expresar como

$$f_{\Theta,X}(\theta, x) = 2\pi g(2\pi [F_\Theta(\theta) + F_X(x)]) f_\Theta(\theta) f_X(x)$$

Donde  $g(\cdot)$  es una función de densidad de una variable aleatoria circular. Por lo tanto, la copula para  $\Theta$  y  $X$  debe satisfacer:

$$c(F_\Theta(\theta), F_X(x)) = 2\pi g(2\pi [F_\Theta(\theta) + F_X(x)])$$

Una de las aplicaciones de esta teoría se ha hecho en la estimación de ozono en la ciudad de México Fernández-Durán (2004) usando sumas trigonométricas no negativas para modelar la función de densidad de variables aleatorias circulares y cópulas circulares-lineales para modelar la dependencia conjunta. Nótese que el enfoque de Fernández-Durán es un enfoque paramétrico.

Otro caso de cópulas no paramétricas de datos que incluyen datos circulares es la adaptación de Carnicero *et al.* (2013) a las cópulas de Bernstein de Sancetta y Satchell (2004). Dicha adaptación consiste en modelos propuestos para tomar en cuenta la periodicidad de los datos circulares.

La función de densidad de probabilidad  $f_{\Theta,X}(\cdot, \cdot)$  de un vector aleatorio  $(\Theta, X)$  es una función no negativa definida en la superficie de un cilindro  $\mathbb{S}^1 \times \mathbb{R}$ . Esta densidad tiene periodo  $2\pi$  en la componente circular:

$$f_{\Theta,X}(\theta, x) = f_{\Theta,X}(\theta + 2k, x) \quad (5.29)$$

para  $k \in \mathbb{Z}, -\infty < \theta < \infty$ ,

Y su integral es uno sobre un intervalo de longitud  $2\pi$  por los reales. Es decir, se integra a uno en todo su dominio.

Como consecuencia del teorema de Sklar (1959),

$$f_{\Theta,X}(\theta, x) = c(F_\Theta(\theta), F_X(x)) f_\Theta(\theta) f_X(x)$$

donde  $c(\cdot, \cdot)$  es la densidad cópula.

Debido a la condición de periodicidad para la función de distribución conjunta (Ecuación 5.29, es necesario que la cópula satisfaga:

$$c(0, v) = c(1, v) \quad \forall v \in [0, 1] \quad (5.30)$$

Las cópulas de Bernstein no cumplen esta condición, para la cual es necesario que los pesos de la Ecuación 5.28 satisfagan  $p_{1j_2} = p_{kj_2}$  para  $j_2 = 1, \dots, k$ . Para remediarlo, Carnicero *et al.* (2013) proponen,

$$\tilde{p}_{1j_2} = \tilde{p}_{kj_2} = \frac{p_{1j_2} + p_{kj_2}}{2} \quad j_2 = 1, \dots, k \quad (5.31)$$

$$\tilde{p}_{j_1j_2} = p_{j_1j_2} \quad j_1 \neq 1, k \quad (5.32)$$

Partiendo de la Figura 5.4, la corrección de la Ecuación 5.31 equivale a promediar los vectores de pesos  $p[1, :]$  y  $p[k, :]$ . De manera visual, dicha corrección se obtiene al promediar las dos celdas del mismo tono de gris en la Figura 5.5.

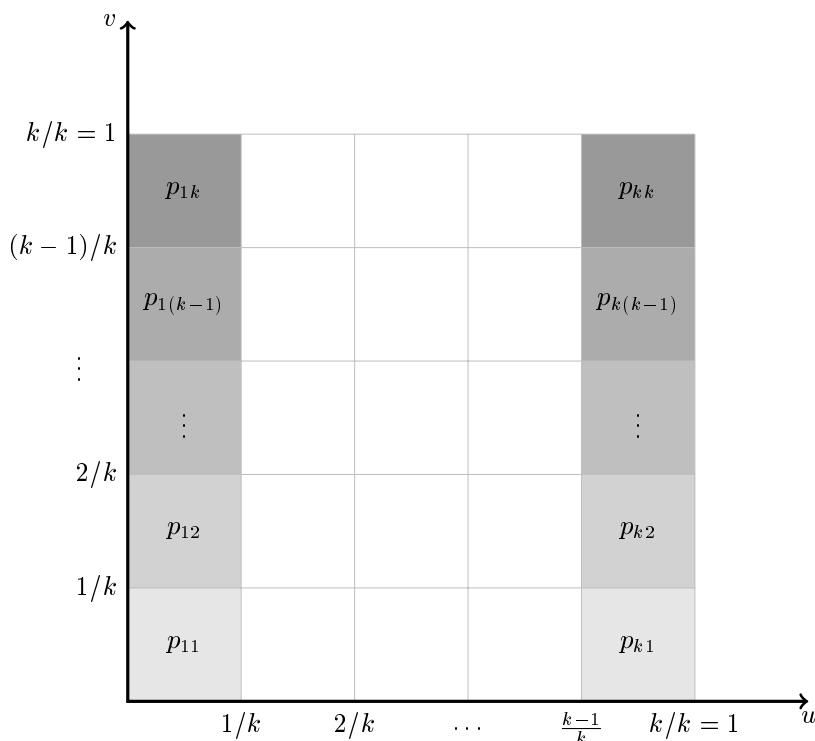


Figura 5.5: Corrección para que la cópula de Bernstein cumpla con la condición de periodicidad en la primera variable (Ecuación 5.30).

Para validar el software desarrollado en R se reprodujeron los resultados del artículo de Carnicer et al. (2013) en el que se muestra la aplicación de esta metodología a datos de dirección del viento y precipitación pluvial. Para tal objetivo, se estimaron las funciones de densidad de probabilidad univariada para ambos datos. Para la dirección del viento la función de densidad univariada fue una combinación de tres distribuciones de von Mises y para los datos de precipitación la función de densidad se compone de una combinación de tres exponenciales (Figura 5.6).

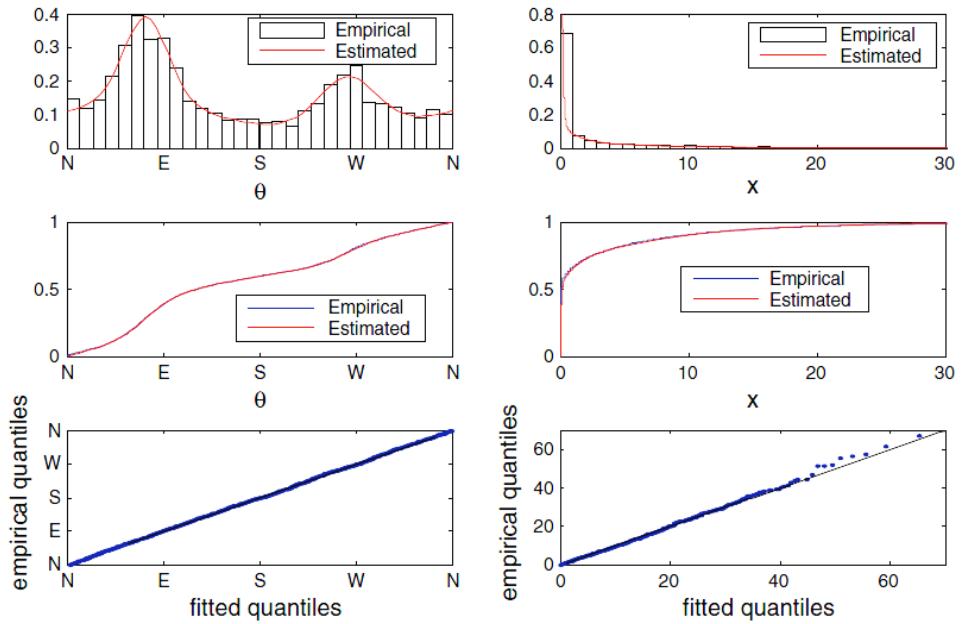


Figura 5.6: Densidad marginal estimada (top), función de distribución (medio) y gráficos de cuantiles (abajo) para dirección del viento (izquierda) y para la precipitación pluvial (derecha). Tomada de Carnicero *et al.* (2013).

mucho espacio !!

\_\_\_\_\_

Para estimar la función de densidad conjunta se utiliza la metodología mostrada en esta sección en lo que respecta a la parte bivariada. En la parte univariada aquí proponemos que sea mediante los polonimos de Bernstein-Kantorovich, pero para reproducir los resultados de Carnicero *et al.* (2013) se utilizaron los modelos que ellos ajustaron. El resultado se muestra en la Figura 5.7 y en la Figura 5.6. Nótese que las marginales son claramente respetadas y la función de densidad es periódica con respecto a la variable aleatoria orientada.

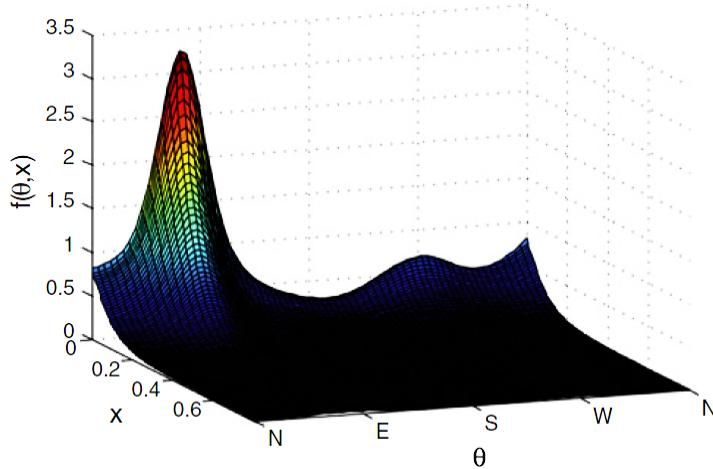


Figura 5.7: Función de densidad conjunta estimada (utilizando cópulas de Bernstein) de la variable circular-lineal  $(\Theta, X)$ . La cual describe la dirección del viento y la precipitación pluvial.

Nótese que, aunque la corrección propuesta por Carnicero *et al.* (2013) permite una cópula continua, no asegura diferenciabilidad, es decir, no asegura que la cópula sea suave en las fronteras. Una opción más versátil que los polinomios de Bernstein que permite modelar curvas y sus derivadas al mismo tiempo son los splines. Dado que la suavidad se pide en la densidad, el enfoque de cópulas con splines propuesto por Schellhase y Kauermann (2012) o el de Ferreira (2008 directional logspline) puede ser útil para una aproximación más natural en el caso de vectores aleatorios con componentes circulares.

Hasta éste momento solamente se ha estudiado el caso con una sola componente circular, pero más es posible. En el caso bivariado, se puede hacer que ambas componentes sean circulares.

La función de densidad de probabilidad  $f_{\Theta_1, \Theta_2}(\cdot, \cdot)$  de un vector aleatorio circular-circular  $(\Theta_1, \Theta_2)$  es una función no negativa definida en la superficie de un toro,  $\mathbb{S}^1 \times \mathbb{S}^1$ . Esta densidad tiene periodo  $2\pi$  en ambas componentes:

$$f_{\Theta_1, \Theta_2}(\theta_1, \theta_2) = f_{\Theta_1, \Theta_2}(\theta_1 + 2j\pi, \theta_2 + 2k\pi), \quad j, k \in \mathbb{Z}, -\infty < \theta_1, \theta_2 < \infty \quad (5.33)$$

Y su integral es uno sobre el producto de dos intervalos de longitud  $2\pi$ . Es decir, es uno al integrar sobre todo su dominio.

Debido a la condición (Ecuación 5.33) para la función de distribución conjunta, es necesario que la cópula para datos circulares-circulares, además de la Ecuación 5.30, satisfaga:

$$c(u, 0) = c(u, 1) \quad \forall u \in [0, 1] \quad (5.34)$$

En el caso de las cópulas de Bernstein, los pesos son similares a la Ecuación 5.28 salvo que las dos variables aleatorias correspondientes son circulares. Por lo tanto los pesos tienen que satisfacer  $p_{1j_2} = p_{kj_2}$  y  $p_{j_11} = p_{j_1k}$  para  $j_1, j_2 = 1, \dots, k$ . Y la corrección de periodicidad, además de aplicarse a

la primera variable  $u$ , posteriormente tiene que aplicarse a la segunda variable  $v$ . Lo que produce los siguientes pesos:

$$\tilde{p}_{1j_2} = \tilde{p}_{kj_2} = \frac{p_{1j_2} + p_{kj_2}}{2}$$

$$\tilde{p}_{j_11} = \tilde{p}_{j_1k} = \frac{p_{j_11} + p_{j_1k}}{2}$$

$$\tilde{p}_{11} = \tilde{p}_{1k} = \tilde{p}_{k1} = \tilde{p}_{kk} = \frac{p_{11} + p_{1k} + p_{k1} + p_{kk}}{4}$$

$$\text{Y } \tilde{p}_{j_1j_2} = p_{j_1j_2} \text{ para } j_1 \neq 1, k \text{ y } j_2 \neq 1, k$$

Lo que garantiza que la cópula se continua y periódica en ambas variables.

Para ilustrar la metodología, Carnicero *et al.* (2013) muestra la aplicación de esta metodología a datos de dirección dos boyas en el mar. Para tal objetivo, se estimaron las funciones de densidad de probabilidad univariada para ambos datos. En ambas variables, la función de densidad univariada fue una combinación de tres distribuciones de von Mises (Figura 5.8).

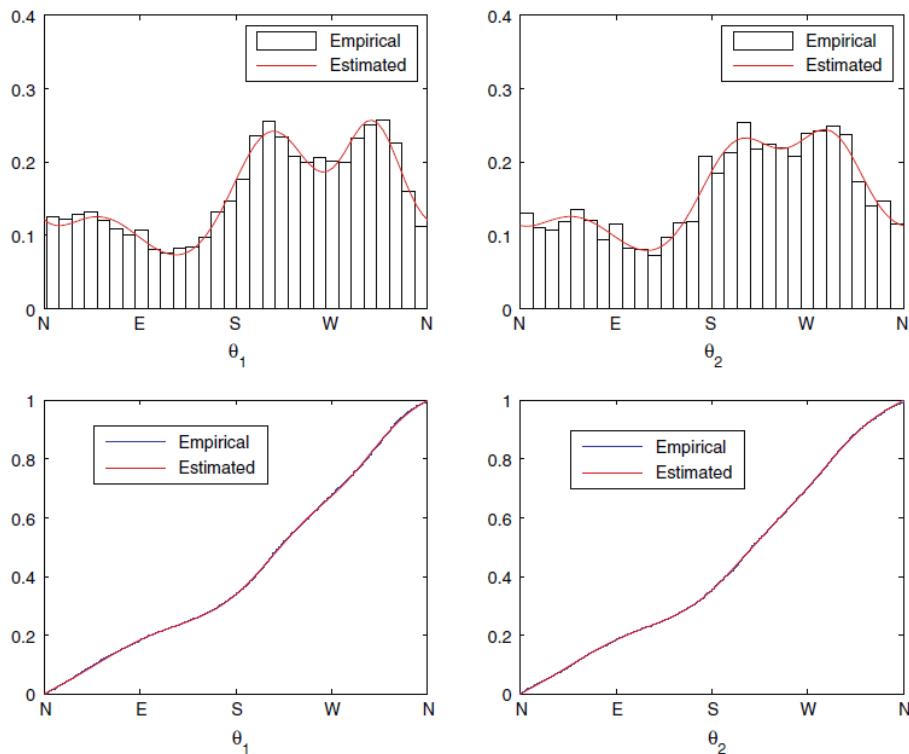


Figura 5.8: Densidad marginal estimada (top) y función de distribución (abajo) para la dirección de dos boyas.

## 5.4. Simulación conjunta de variables aleatorias: caso bivariado

Dado un modelo de dependencia (cópula), uno de los objetivos perseguidos al modelar la función de densidad de probabilidad conjunta es simular valores de las variables aleatorias. Para

simular datos a partir de la cópula de se utiliza una variante del método de distribución condicional (Nelsen, 2006) en el cual se utiliza un enfoque no paramétrico (Erdely y Díaz-Viera, 2009, 2012):

1. Genere dos variables aleatorias  $u$  y  $t$  que sean independientes y continuas en  $(0, 1)$ .
2. Obtenga  $v = c_u^{(-1)}(t)$ . Donde  $t = c_u(v) = \frac{\partial \tilde{C}(u,v)}{\partial u}$  y  $c_u^{(-1)}$  es la inversa generalizada (Ecación 4.6) de  $c_u$ .
3. Los datos simulados,  $(x, y)$ , se obtienen utilizando funciones cuantiles no paramétricas de  $X$  y  $Y$  respectivamente. Estas funciones  $\tilde{Q}$  y  $\tilde{R}$  son estimadas mediante polinomios de Bernstein-Kantorovich (Muñoz-Pérez y Fernández-Palacín, 1987). Por lo tanto  $(x, y) = (\tilde{Q}(u), \tilde{R}(v))$ .

Atendiendo al punto 1 de esta metodología es necesario simular valores con una distribución uniforme en el intervalo  $[0, 1]$ , ya que estos son requeridos por la cópula. Esto se puede conseguir con la función `stats::runif`. Recuérdese que la cópula bivariada tiene dominio  $[0, 1]^2$ . Para el segundo paso se requieren a) la derivada parcial de la cópula Cópula; b) obtener una función inversa. En el caso de la cópula de Bernstein la derivada está dada por la Ecación 5.23. Al terminar este segundo paso se cuenta con dos valores  $u_i$  y  $v_i$  que están distribuidos uniformemente pero que conservan la misma dependencia que los datos originales. Por último, estos valores son los argumentos de las funciones cuantiles para obtener  $x_i$  y  $y_i$ . La implementación computacional de estas funciones cuantiles se encuentra en el paquete `1momco` bajo el nombre de `dat2bernqua`. La Figura 5.9 muestra un ejemplo de un conjunto de datos a los cuales se les modeló su función de distribución con los polinomios de Bernstein-Kantorovich. Por lo tanto  $(x_i, y_i) = (\tilde{Q}(u_i), \tilde{R}(v_i))$ .

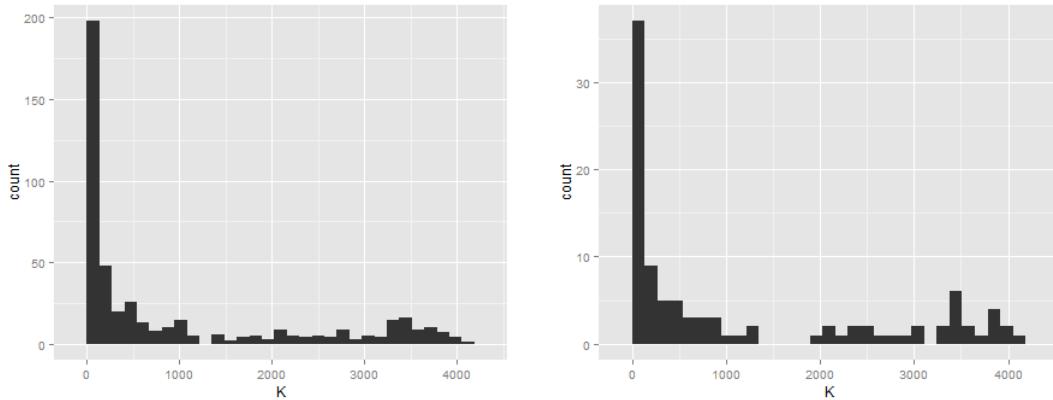


Figura 5.9: Histogramas de un ejemplo de simulación no paramétrica. 473 datos (izquierda) y 100 datos simulados (derecha).

## 5.5. Simulación de conjunta de variables aleatorias: caso multivariado

Uno de los objetivos al modelar vectores aleatorios es generar muestras estadísticamente equivalentes a los datos observados para así poder cuantificar incertidumbre. Para lograr este cometido, se ilustra el algoritmo de simulación.

El algoritmo de simulación se simplifica si todas las  $d$  marginales unidimensionales se distribuyen uniformemente lo cual se logra obteniendo la función de distribución acumulada de los datos ya que  $F_i(x_i) \sim \text{Uniforme}(0, 1)$ . La generalidad del algoritmo se recupera al mapear cada marginal  $u_i := F_i(x_i)$  en su variable correspondiente  $x_i$  mediante el método de la transformada inversa (Robert y Casella, 2009, sec. 2.1.2, pp. 44) utilizando la inversa generalizada (Ecuación 4.6).

Para obtener una muestra simulada se procede con el siguiente algoritmo recursivo (Mai y Scherer, 2012, sec. 5.3.2, pp. 208; Dißmann, 2010):

$$\text{inicio: } W_i \stackrel{i.i.d.}{\sim} \text{Uniforme}(0, 1), \quad \forall i \in \{1, \dots, d\} \quad (5.35)$$

$$\text{posteriormente: } U_1 := W_1 \quad (5.36)$$

$$U_2 := F_{2|1}^-(W_2|U_1) \quad (5.37)$$

$$U_3 := F_{3|2,1}^-(W_3|U_2, U_1) \quad (5.38)$$

$$U_d := F_{d|d-1, \dots, 1}^-(W_d|U_{d-1}, \dots, U_1) \quad (5.39)$$

En el caso de teoría de cópulas cada  $F_{i|i-1, \dots, 1}^-$  es la inversa generalizada de la Ecuación 5.44.

## 5.6. Vine cópulas: Caso trivariado

Es bien sabido que el análisis y visualización bivariado de datos es más fácil que el caso multivariado. Esto también se cumple en el caso de la modelación de dependencias mediante cópulas. Un enfoque que permite estudiar dependencias con  $d \geq 3$  mediante descomposiciones bivariadas es el conocido como Vine copulas (también conocido como construcciones de cópulas mediante pares).

El siguiente caso trivariado ilustra cómo se efectúa la descomposición en pares a partir de la función de densidad conjunta  $f_{123}$  (Díaz-Viera, 2015):

$$f_{123} = f_{23|1} \cdot f_1 \quad (5.40)$$

$$= c_{23|1}(F_{2|1}, F_{3|1}) \cdot f_{2|1} \cdot f_{3|1} \cdot f_1 \quad (5.41)$$

$$= c_{23|1}(F_{2|1}, F_{3|1}) \cdot \frac{f_{12}}{f_1} \cdot \frac{f_{13}}{f_1} \cdot f_1 \quad (5.42)$$

$$= c_{23|1}(F_{2|1}, F_{3|1}) \cdot c_{12}(F_1, F_2) \cdot c_{13}(F_1, F_3) \cdot f_1 \cdot f_2 \cdot f_3 \quad (5.43)$$

Nótese que para la misma densidad  $f_{123}$  existen otras dos posibles descomposiciones dependiendo de la variable condicionante.

Cada uno de los argumentos de las cópulas bivariadas de la descomposición Vine (ecuación 5.43) está dado como (Aas *et al.*, 2009; Dißmann, 2010; Mai y Scherer, 2012):

$$F(x|\mathbf{x}) = \frac{\partial C_{X,X_j|\mathbf{x}_{-j}}(F(x|\mathbf{x}_j), F(x_j|\mathbf{x}_{-j}))}{\partial F(x_j|\mathbf{x}_{-j})} \quad (5.44)$$

donde  $\mathbf{x} \subset \{x_1, \dots, x_d\}$ ,  $\mathbf{x}^c := \{x_1, \dots, x_d\} - \mathbf{x}$ ,  $x \notin \mathbf{x}$  y  $x_j \notin \mathbf{x}_j \subset \mathbf{x}$ .

Nótese que en la ecuación (5.44) se utilizan mayúsculas para denotar variables aleatorias mientras que las minúsculas denotan observaciones (realizaciones). La notación también distingue entre escalares ( $x$  ó  $X$ ) y vectores ( $\mathbf{x}$ ).

Para el caso trivariado se tiene (Mai y Scherer, 2012, pp. 209, example 5.7)

$$F(u_3|u_1, u_2) = \frac{\partial C_{3,1|2}(F(u_3|u_2), F(u_1|u_2))}{\partial F(u_1|u_2)} \quad (5.45)$$

Un caso particular de las cópulas de Vine se tiene haciendo el *supuesto de simplificación* (Gijbels *et al.*, 2011; Joe, 2014)

$$C_{3,1|2} = C_{3,1} \quad (5.46)$$

Un caso más particular pero frecuente se tiene si  $U_3 \perp U_1|U_2$  (ver 5.7):

$$C_{3,1|2}(u, w) = uw \quad (5.47)$$

Utilizando 5.47 en 5.45 se tiene:

$$\frac{\partial C_{3,1|2}}{\partial F(u_1|u_2)} = F(u_3|u_2) \quad (5.48)$$

Lo que reduce el tiempo de cómputo para la simulación.

## 5.7. Cópulas de Bernstein: caso general multivariado

En el caso multivariado, para modelar dependencias, una aproximación de una función  $d$ -dimensional utilizando los polinomios de Bernstein toma la siguiente forma:

$$G(u_1, \dots, u_d) = \sum_{v_1=0}^{m_1} \cdots \sum_{v_d=0}^{m_d} \alpha \left( \frac{v_1}{m_1}, \dots, \frac{v_d}{m_d} \right) \prod_{i=1}^d B(u_i|v_i, m_i) \quad (5.49)$$

Si definimos los arreglos ordenados (vectores)  $n$ -dimensionales

$$\mathbf{m} := (m_1, \dots, m_n) \quad (5.50)$$

$$\mathbf{u} := (u_1, \dots, u_n) \quad (5.51)$$

$$\mathbf{v} := (v_1, \dots, v_n) \quad (5.52)$$

$$\mathbf{0} := (0, \dots, 0) \quad (5.53)$$

entonces podemos definir

$$\sum_{\mathbf{v}=\mathbf{0}}^{\mathbf{m}} := \sum_{v_1=0}^{m_1} \cdots \sum_{v_d=0}^{m_d} \quad (5.54)$$

y por lo tanto la ecuación (5.49) se puede expresar como

$$G(\mathbf{u}) = \sum_{\mathbf{v}=\mathbf{0}}^{\mathbf{m}} \alpha \left( \frac{v_1}{m_1}, \dots, \frac{v_d}{m_d} \right) \prod_{i=1}^d B(u_i | v_i, m_i) \quad (5.55)$$

El caso más sencillo se obtiene cuando  $n = 1$ :

$$G(u) = \sum_{v=0}^m \alpha \left( \frac{v}{m} \right) B(u | v, m) \quad (5.56)$$

Dado que los polinomios satisfacen de manera natural muchas de las propiedades de las funciones de distribución acumulada, los polinomios de Bernstein han demostrado ser una buena aproximación de la cópula empírica, que en el caso bivariado toma la siguiente forma (Deheuvels, 1979):

Para las simulaciones es importante tener la derivada de (5.49), lo cual se puede obtener haciendo uso de 5.57:

$$\frac{\partial G}{\partial u_l} = \frac{\partial}{\partial u_l} \left[ \sum_{v_1=0}^{m_1} \cdots \sum_{v_l=0}^{m_l} \cdots \sum_{v_d=0}^{m_d} \alpha \left( \frac{v_1}{m_1}, \dots, \frac{v_l}{m_l}, \dots, \frac{v_d}{m_d} \right) \prod_{i=1}^d B(u_i | v_i, m_i) \right] \quad (5.57)$$

$$= \sum_{v_1=0}^{m_1} \cdots \sum_{v_d=0}^{m_d} \prod_{i \neq l} B(u_i | v_i, m_i) \frac{\partial}{\partial u_l} \left[ \sum_{v_l=0}^{m_l} \alpha \left( \frac{v_1}{m_1}, \dots, \frac{v_d}{m_d} \right) B(u_l | v_l, m_l) \right] \quad (5.58)$$

$$= \sum_{v_1=0}^{m_1} \cdots \sum_{v_d=0}^{m_d} \prod_{i \neq l} B(u_i | v_i, m_i) \frac{\partial}{\partial u_l} \left[ m_l \sum_{v_l=0}^{m_l-1} \Delta_l \alpha \left( \frac{v_1}{m_1}, \dots, \frac{v_d}{m_d} \right) B(u_l | v_l, m_l) \right] \quad (5.59)$$

El operador  $k$ -dimensional diferencia finita hacia adelante se define como

$$\Delta_{1,\dots,k} \alpha \left( \frac{v_1}{m}, \dots, \frac{v_k}{m} \right) = \sum_{l_1=0}^1 \cdots \sum_{l_k=0}^1 (-1)^{k+l_1+\dots+l_k} \alpha \left( \frac{v_1 + l_1}{m}, \dots, \frac{v_k + l_k}{m} \right) \quad (5.60)$$

En el caso univariado  $k = 1$ :

$$\Delta_1 \alpha \left( \frac{i}{m} \right) = \sum_{l=0}^1 (-1)^{1+1} \alpha \left( \frac{i+l}{m} \right) \quad (5.61)$$

$$= \alpha \left( \frac{i+1}{m} \right) - \alpha \left( \frac{i}{m} \right) \quad (5.62)$$

En el caso multivariado haciendo la diferencia finita de una sola variable:

$$\Delta_r \alpha \left( \frac{i_1}{m}, \dots, \frac{i_d}{m} \right) = \sum_{l=0}^1 (-d)^{d+1} \alpha \left( \frac{i_r + l}{m} \right) \quad (5.63)$$

$$= \alpha \left( \frac{i_r + 1}{m} \right) - \alpha \left( \frac{i_r}{m} \right) \quad (5.64)$$

La diferencia finita de la cópula empírica bivariada con respecto a la primera variable se reduce a la Ecuación 5.18.

## 5.8. Cópulas multivariadas para vectores aleatorios con componentes orientadas

Nótese que en el trabajo que se ha mostrado hasta ahora sobre datos periódicos solamente se ha hablado de cópulas bivariadas. Para el caso  $m$ -variado no se tiene que modelo de cópulas dentro, del contexto de cópulas de Bernstein, que tome en cuenta cualquier cantidad de variables aleatorias periódicas. La cópula densidad  $m$ -variada de Bernstein (Sancetta & Satchell, 2004, sec. 4.1) es una generalización de la ecuación (9)

$$\hat{c}_B(u_1, u_2, \dots, u_m) = \sum_{j_1=1}^k \sum_{j_2=1}^k \dots \sum_{j_m=1}^k p_{j_1, j_2, \dots, j_m} \prod_{i=1}^m \beta(u_i | j_i, k - j_i + 1)$$

Para imponer las condiciones de periodicidad sobre la variable aleatoria  $\mathcal{L}_1$ -ésima,  $\mathcal{L}_1 \in \{1, 2, \dots, m\}$  se agregan condiciones similares a (16)

$$\tilde{p}_{j_1 \dots 1 \dots j_m} = \tilde{p}_{j_1 \dots k \dots j_m} = \frac{p_{j_1 \dots 1 \dots j_m} + p_{j_1 \dots k \dots j_m}}{2}$$

Para  $j_i = 1, \dots, k$ , con  $i \in \{1, 2, \dots, \mathcal{L}_1 - 1, \mathcal{L}_1 + 1, \dots, m\}$ , i.e.,  $i \neq \mathcal{L}_1$

$$\tilde{p}_{j_1 \dots j_{\mathcal{L}_1} \dots j_m} = p_{j_1 \dots j_{\mathcal{L}_1} \dots j_m}$$

Para  $j_{\mathcal{L}_1} \neq 1, k$ .

Se puede demostrar que la cópula converge uniformemente.

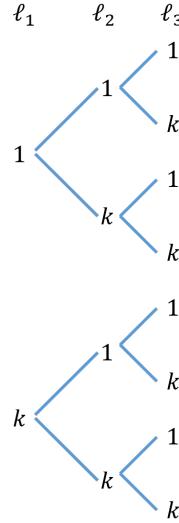


Figura 5.10: Diagrama de árbol que muestra cómo se van agregando la condición de periodicidad a las variables aleatorias para la cópula de Bernstein multivariada. Aquí se muestra el caso para 3 variables aleatorias periódicas. Nótese que también muestra la cantidad de sumandos de los elementos de las esquinas, que para este caso es  $2^3$ .

Como es de esperarse, la ecuación (22) se reduce a la ecuación (16) en el caso bivariado cuando  $\mathcal{L}_1 = 1$ .

Nótese que si aplicamos nuevamente las ecuaciones (24) a las  $\tilde{p}$ , pero esta vez a la variable aleatoria  $\mathcal{L}_2$ -ésima, con  $\mathcal{L}_1 \neq \mathcal{L}_2$ , obtenemos el caso de dos variables aleatorias periódicas

$$\tilde{p}_{j_1 \dots j_{\mathcal{L}_1} \dots 1 \dots j_m} = \tilde{p}_{j_1 \dots j_{\mathcal{L}_1} \dots k \dots j_m} = \frac{\tilde{p}_{j_1 \dots j_{\mathcal{L}_1} \dots 1 \dots j_m} + \tilde{p}_{j_1 \dots j_{\mathcal{L}_1} \dots k \dots j_m}}{2}$$

Para  $j_i = 1, \dots, k$ , con  $i \in \{1, 2, \dots, \mathcal{L}_2 - 1, \mathcal{L}_2 + 1, \dots, m\}$ , i.e.,  $i \neq \mathcal{L}_2$

$$\tilde{p}_{j_1 \dots j_{\mathcal{L}_1} \dots j_{\mathcal{L}_2} \dots j_m} = \tilde{p}_{j_1 \dots j_{\mathcal{L}_1} \dots j_{\mathcal{L}_2} \dots j_m}$$

Para  $j_{\mathcal{L}_2} \neq 1, k$ .

Que para  $j_{\mathcal{L}_1} = 1, k$  y  $j_{\mathcal{L}_2} = 1, k$ , es decir, en las esquinas frontera.

$$\tilde{p}_{j_1 \dots 1 \dots 1 \dots j_m} = \tilde{p}_{j_1 \dots 1 \dots k \dots j_m} = \tilde{p}_{j_1 \dots k \dots 1 \dots j_m} = \tilde{p}_{j_1 \dots k \dots k \dots j_m} \quad (5.65)$$

$$= \frac{p_{j_1 \dots 1 \dots 1 \dots j_m} + p_{j_1 \dots k \dots 1 \dots j_m}}{2} + \frac{p_{j_1 \dots 1 \dots k \dots j_m} + p_{j_1 \dots k \dots k \dots j_m}}{2} \quad (5.66)$$

$$= \frac{p_{j_1 \dots 1 \dots 1 \dots j_m} + p_{j_1 \dots k \dots 1 \dots j_m} + p_{j_1 \dots 1 \dots k \dots j_m} + p_{j_1 \dots k \dots k \dots j_m}}{2^2} \quad (5.67)$$

Nótese que para el caso particular de la cópula bivariada la ecuación (23) se reduce a la ecuación (20).

De esta manera se podría continuar con el algoritmo para dar la periodicidad a las variables aleatorias requeridas.



## Capítulo

# 6

## Metodología de simulación de redes de fracturas discretas

En este capítulo se muestra de manera esquemática los ingredientes principales de los capítulos anteriores. Para ello, es necesario explicar la manera en que se desarrollarán tales ideas principales.

Un diagrama de flujo (flowchart) es una representación gráfica que resalta los elementos de un programa computacional o una metodología, las relaciones entre ellos, así como su correspondiente orden de ejecución (flujo lógico). Para mayor información sobre diagramas de flujo su puede consultar los libros de Venit y Drake (2014) y Farrell (2014).

Tabla 6.1: Símbolos básicos para construir diagramas de flujo.

Símbolo	Nombre	Descripción
	Proceso	Representa cualquier proceso, cálculo ó función.
	Input/output	Argumento de entrada o salida.
	Decisión	Representa una decisión binaria, usualmente Sí o No.
	Flecha	Indica la dirección de flujo de control.

Además, algunos símbolos de la Tabla 6.1 rellenos con color verde ( ) representan los inicios de cada diagrama de flujo, mientras que en color rojo ( ) los finales de los mismos. En el caso de diagramas verticales, el mismo se lee de arriba hacia abajo.

### 6.1. Análisis, modelado y simulación de variables aleatorias

El enfoque que se utilizará en esta sección es el convencional en ciencias de la tierra en donde, teniendo datos, se procede a caracterizarlos, es decir, a entender su comportamiento en la mayor medida posible. Posteriormente se puede hacer el modelado teniendo en cuenta el paso anterior

y posteriormente pueden hacer simulaciones estadísticamente equivalentes con los datos. Estos pasos, para el caso univariado, se puede ver reflejado en el diagrama de flujo de la Figura 6.1, yendo de lo rojo hacia lo verde.

La caracterización se basa principalmente en análisis exploratorio de los datos  $x_s$  (en el globo rojo en la Figura 6.1). En esta etapa del análisis estadístico se analizan individualmente cada una de las propiedades de las fracturas, así como de la red de fracturas. El resultado esperado para esta etapa es entender el fenómeno subyacente a los datos, para así poder modelar o sugerir modelos de distribución de probabilidad para cada una de las propiedades de fractura.

Para lograr este objetivo se hace uso de estadígrafos y de gráficas. Por ejemplo, la media proporciona una medida de tendencia central, mientras que la desviación estándar proporciona información sobre qué tan dispersos se encuentran los datos con respecto a la media obtenida. Esto es válido tanto para las longitudes como para las orientaciones, claro, cada variable con su teoría correspondiente. El análisis exploratorio siempre debe de combinar los estadígrafos y los gráficos. Para las longitudes se puede utilizar los gráficos de caja y los histogramas principalmente. Para las orientaciones es útil utilizar las rosetas de orientaciones, pero un histograma resalta más cualquier moda sutil que se escape a la vista en una roseta.

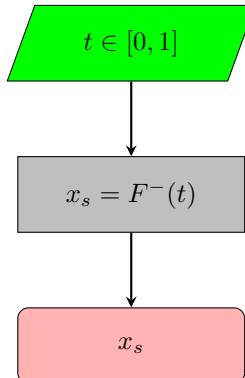


Figura 6.1: Diagrama de flujo para obtener una simulación de una variable aleatoria. En este caso puede ser la orientación o la longitud. Compárese con la Figura 4.1.

En la siguiente etapa, en color gris en la Figura 6.1, se procede a modelar la función de probabilidad para los datos. Aquí se puede proceder con el enfoque común de hacer pruebas de hipótesis para varios modelos propuestos o se puede hacer con el enfoque de los cuantiles de Bernstein-Kantorovich utilizado en este trabajo doctoral.

Para hacer la simulación, hay que recorrer la Figura 6.1 en sentido contrario en el que lo hicimos para la caracterización y el modelado. Ahora, el algoritmo de simulación muestra que debemos empezar con un valor simulado  $t$  obtenido de una distribución  $Uniforme(0, 1)$ . La función cuantil  $F^-(t)$  nos proporcionará un valor simulado. De esta manera podemos obtener las simulaciones necesarias para estimar la incertidumbre.

## 6.2. Análisis, modelado y simulación de la estructura de dependencia entre variables aleatorias

Igual de importante que estudiar individualmente las propiedades de fractura es estudiar la relación que existen entre tales propiedades. Se recomienda estudiar, primero, las relaciones por pares, es decir de manera bivariada. Dos ejemplos de relaciones bivariadas serían (1) longitud y apertura, y (2) orientación y longitud. Posteriormente se puede estudiar de manera trivariada o multivariada pero en estos casos es difícil obtener una idea visual de la estructura de dependencia.

El gráfico más útil para estudiar las dependencias entre variables aleatorias es el gráfico o diagrama de pseudo-observaciones. Dicho gráfico se forma con las pseudo-observaciones obtenidas a partir de la función de distribución empírica (Ecuación 4.5). Para obtenerlas se presenta el diagrama de flujo en la Figura 6.2. Nótese que en el diagrama se muestra la función de distribución de probabilidad  $F(\cdot)$  de manera general, pero en este trabajo se utiliza  $F(x) = \hat{F}_n(x)$ . Los puntos de cada pseudo-observación  $(u_i, v_i)$ ,  $i \in \{1, \dots, n\}$  formados con las variables  $X$  y  $Y$  se grafican en el plano  $uv$  (Figura 6.3) para obtener un diagrama de dispersión de las pseudo-observaciones.

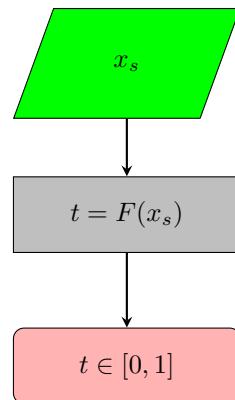


Figura 6.2: Diagrama de flujo para obtener las pseudo-observaciones. Compárese con la Figura 4.1.

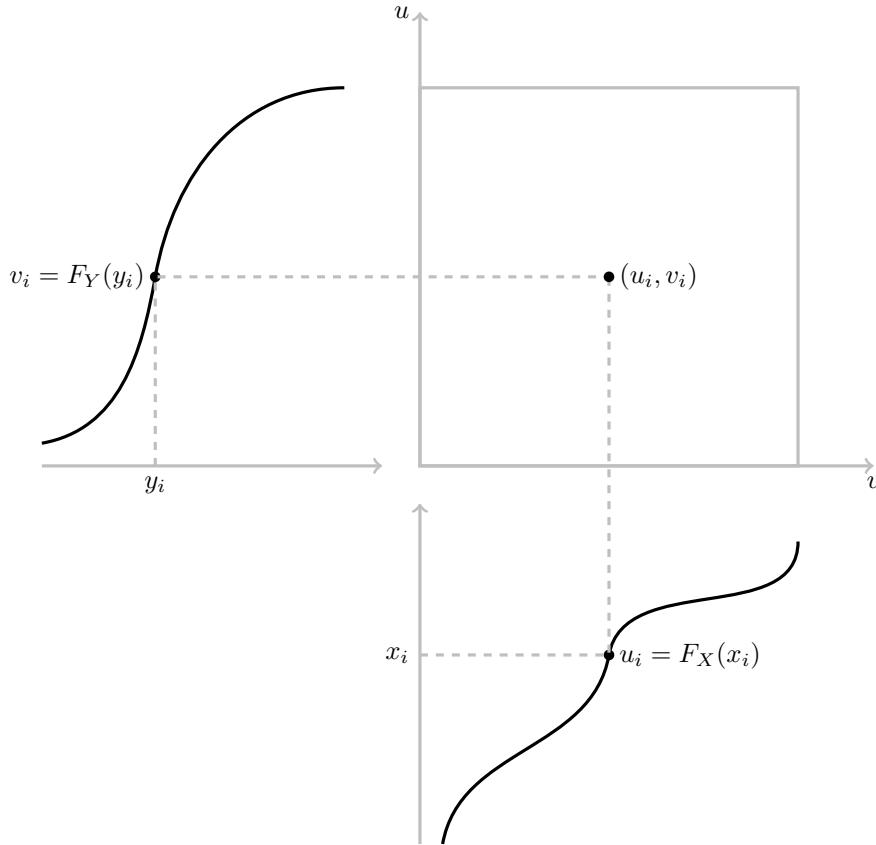


Figura 6.3: Obtención del diagrama de pseudo-observaciones con el cual se estudia la dependencia entre dos variables aleatorias. Las funciones de distribución marginales que se muestran son continuas pero pueden ser no-continuas o una combinación de ambas.

Dentro del contexto de la teoría de cópulas el comportamiento conjunto queda totalmente determinado por la cópula, y de manera casi directa por las pseudo-observaciones. Una manera de facilitar el análisis del gráfico de pseudo-observaciones es agregando una sub-capa tipo mosaico o pixeles. Un ejemplo muy sencillo se muestra en el recuadro derecho de la Figura 6.5. Se propone, como primer paso, que la sub-capa sea determinada por los cuartiles y los valores máximos y mínimos de los datos. En particular se propone que los cuartiles se calculen a partir de los modelos de los datos ya que de esta manera se pueden comparar diversas simulaciones sobre un mismo mosaico. En el ejemplo de la Figura 6.5 se tiene que  $q_{1X} = F_X^{-}(0.25)$ ,  $M = F_Y^{-}(0.5)$ .

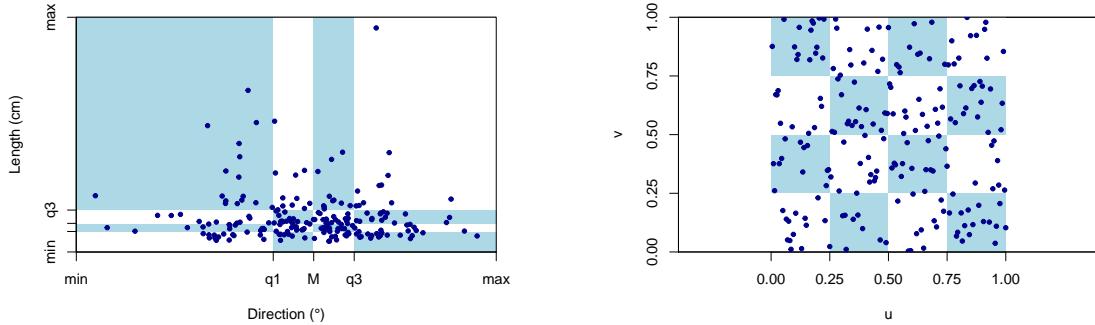


Figura 6.4: Diagrama de dispersión con múltiples zonas remarcadas a fin de hacer un análisis exploratorio preliminar.

Aunque la cópula tiene la información de la estructura de dependencia, conviene graficar también el diagrama de dispersión de los datos ya que ellos, a diferencia de las pseudo-observaciones, están en el rango real. Nótese que los cuadros del gráfico de pseudo-observaciones se extienden o encogen en el diagrama de dispersión, incluso de forma diferente dependiendo el mosaico/pixel (intervalo bidimensional). En la Figura 6.4 se muestra otro mosaico para otro conjunto de datos debido a que se quiere resaltar que la estructura de dependencia general se puede descomponer en dos estructuras más simples.

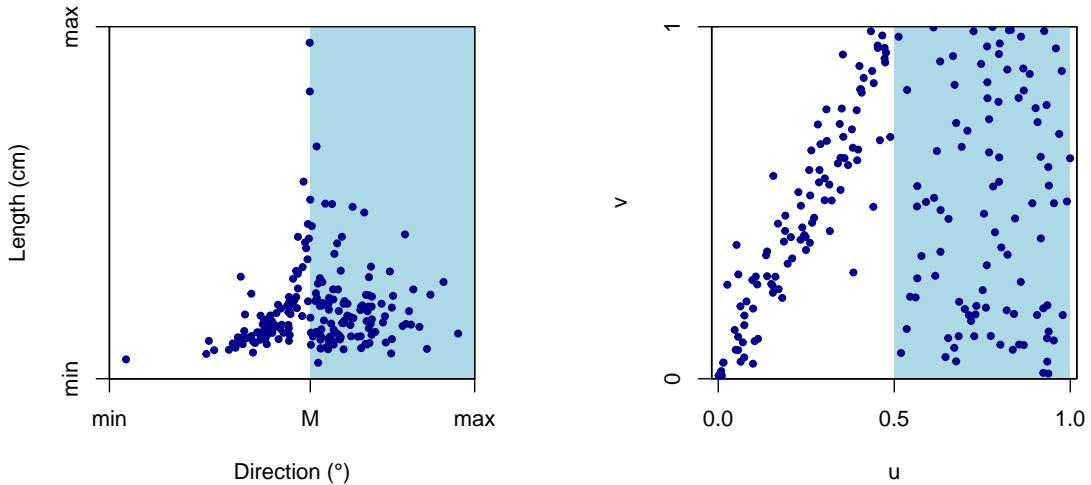


Figura 6.5: Diagrama de dispersión en el cual se han resaltado dos zonas con dependencias más simples.

De manera cuantitativa, se pueden estimar coeficientes de correlación y/o hacer pruebas de hipótesis sobre la independencia de los datos. También es útil agregar estos valores al gráfico

combinado (de dispersión y pseudo-observaciones) ya sea en el gráfico o en el pie de figura.

Para la modelación de la cópula bivariada, dentro del contexto de las cópulas de Bernstein, se procede a un diagrama de flujo similar al de la Figura 6.1. Se analizan los pares de pseudo-observaciones obtenidas mediante la Figura 6.2. Éstas son inspeccionadas visualmente en un diagrama de dispersión, y cuantitativamente con estadígrafos (coeficientes de correlación ó asociación) para determinar si existe o no alguna estructura de dependencia que se deba tomar en cuenta. Si hay dependencia, entonces se modela la cópula empírica  $\hat{C}_n$  con la cual se ajusta la cópula estimada  $C$ .

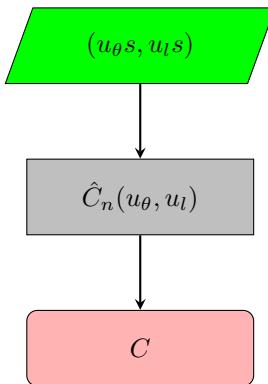


Figura 6.6: Modelación de la cópula a partir de las pseudo-observaciones.

### 6.3. Diagrama de flujo para el análisis, modelado y simulación de redes de fracturas discretas tomando en cuenta su estructura de dependencia

Resulta ilustrativo el diagrama de flujo de la Figura 6.7 en la modelación del vector aleatorio bivariado que representa las propiedades de los objetos booleanos en el cual se parte a partir de los datos observados  $\theta_i$  y  $l_i$ . Se han escogido éstos símbolos para hacer la analogía con la orientación y la longitud de fracturas, pero el diagrama de flujo es válido para cuquier par de variables aleatorias continuas.

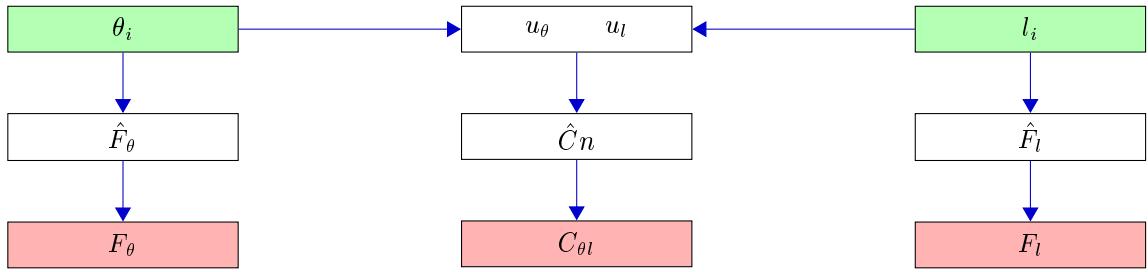


Figura 6.7: Modelación bivariada de propiedades de objetos booleanos.

Nótese que de arriba hacia abajo, se procede con la caracterización de los datos hasta llegar al modelo, en lo parte inferior, tanto de manera univariada como bivariada. En la parte univariada en los extremos izquierdo y derecho del diagrama son los casos particulares del inverso de la metodología en la Figura 6.1. Si el diagrama en cuestión se recorre en sentido contrario a las líneas de flujo, entonces se obtiene el algoritmo de simulación bivariado.

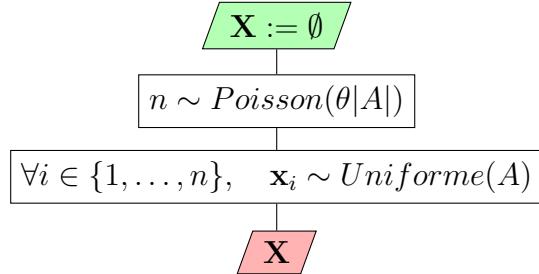


Figura 6.8: Diagrama de flujo para obtener ubicaciones espaciales a través de un proceso puntual. Desde el punto de vista computacional, la parte superior del diagrama indica crear un arreglo vacío que almacene las coordenadas de todos los vértices  $x_i$ .

Estos dos últimos diagramas de flujo forman la esencia del algoritmo de simulación de redes de fracturas. Con el algoritmo de la Figura 6.7 se generan los objetos, es decir la orientación, longitud, etc. de las fracturas; mientras que con el diagrama de la Figura 6.8 se obtienen sus ubicaciones espaciales.

El flujo de trabajo estándar (Figura 6.9) puede ser siendo utilizado en la parte que corresponde al cluster análisis para determinar familias, el tamaño de las familias, la dirección y su dispersión, ya que este análisis proporciona información valiosa para el entendimiento del sistema geológico, sin embargo, la modelación resulta más compatible con los datos si se hace de manera total, ya que no se crean los artificios generados al segmentar bruscamente las familias mediante un valor de corte. Además, entre mayor sea la cantidad de datos, mejor es la modelación de funciones de probabilidad.

Con respecto a las características multivariadas en color azul, es claro que sí hay que hacer un análisis basado en la teoría de cópulas a como se mostró en la sección pasada. Ya que el enfoque individualista no puede producir relaciones de dependencia que se pueden observar en los sistemas de fracturas reales. Éste es el aporte principal al conocimiento que se hace con este trabajo de investigación, en particular la posibilidad para modelar datos orientados y otras propiedades como longitud o apertura.

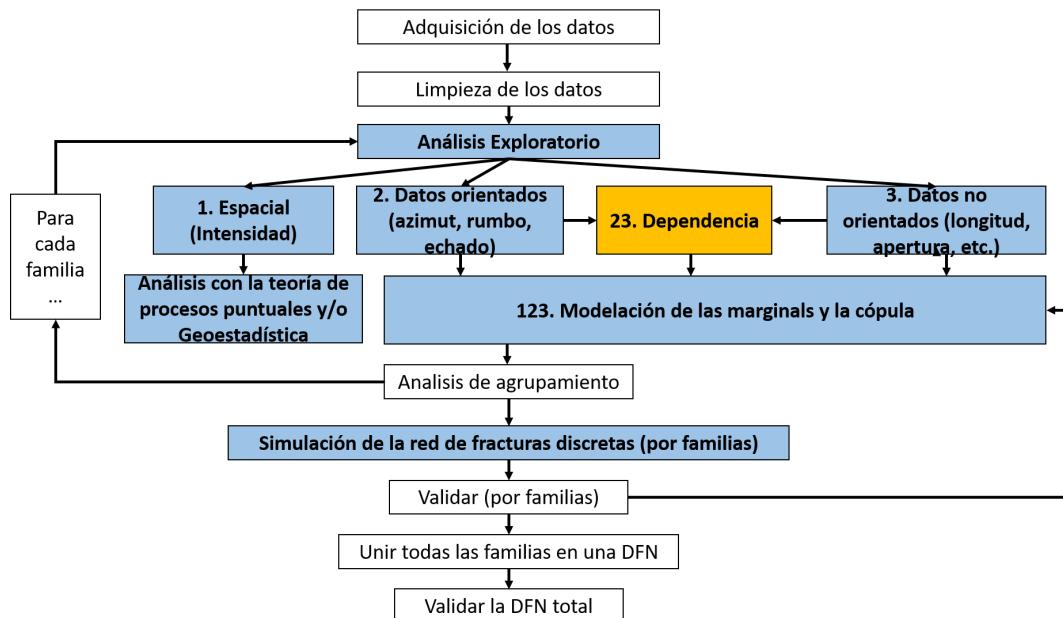


Figura 6.9: Flujo de trabajo para la caracterización, modelado y simulación de redes de fracturas discretas en medios porosos fracturados.

El flujo de trabajo (Figura 6.9) comienza en la parte superior con la adquisición de los datos, posteriormente hay que limpiarlos para poder hacer un análisis exploratorio, el cual se realiza de manera univariada (intensidad, orientación, longitud, apertura, ...) y de manera bivariada se estudian las propiedades mediante las pseudo-observaciones. Como resultado hasta este momento se pudo haber identificado grupos/clusters, los cuales se pueden separar para comenzar el ciclo desde al análisis exploratorio para cada uno de los grupos. Posteriormente se ajustan modelos a cada una de las variables y de manera conjunta con cópulas. Estos modelos permiten las simulaciones de fracturas por familias, las cuales tienen que ser validadas, nuevamente, mediante otro análisis exploratorio para verificar que cumple con las mismas propiedades estadísticas que los datos. Ahora ya se pueden integrar todas las familias y para finalmente validar la familia global.

Cabe mencionar que el análisis de dependencia también puede ayudar para clasificar familias que se distingan de manera bivariada una de otra. Es decir, también se puede integrar el análisis de dependencia a cada una de las familias obtenida con el análisis de clusters.

# Casos de estudio

A continuación, se aplican los resultados de los capítulos anteriores a dos casos particulares. El primer caso fue tomado del artículo de Mendoza-Torres *et al.* (2017). En él se muestra la aplicación a una red de fracturas considerando solamente la dependencia entre la orientación y la longitud. La aplicación trivariada (orientación-longitud-apertura) se muestra en el segundo caso.

## 7.1. Caso bivariado orientación-longitud

### 7.1.1. Descripción de los datos

Aunque la metodología es válida para cualquier sistema de fracturas, en esta sección se presenta una red de fracturas cuyo comportamiento es frecuentemente encontrado dentro de la corteza terrestre. Debido a la falta de datos reales, se generó un conjunto de datos tratando de reproducir cierto conocimiento geológico:

*Fracura de extensión:* Fractura formada por la extensión perpendicular a las paredes de la fractura. La magnitud de la extensión puede ser minúscula como en el caso de *juntas*, o puede ser tan grande como las *venas*

---

Fossen (2010)

las juntas de cizalla comúnmente ocurren en *conjuntos conjugados*... las juntas de cizalla ocurren en conjuntos conjugados oblicuos en donde las juntas de extensión aparecen como juntas longitudinales y transversas que un par ortogonal

---

Singhal y Gupta (2010)

Aún más, algunos sistemas de fracturas tienen relaciones específicas de orientación y longitud:

Los mineros se refieren a estas fracturas en direcciones características como *cruceros de carbón*: un crucero constituido por face cleats, largas fracturas dominantes (juntas sistemáticas) que pueden extenderse por varios metros de manera horizontal; y por, butt cleats, otro conjunto de fracturas más cortas escasamente desarrolladas que terminan en las face cleats en ángulos rectos (juntas transversales).

---

Davis *et al.* (2011)

A partir de este conocimiento de la geología estructural, el conjunto de datos de orientación y longitud a estudiar debe satisfacer:

- Tener dos familias  $f_{90}$  y  $f_0$ , siendo esta última la familia dominante, es decir, con fracturas más largas.
- las longitudes de la familia  $f_{90}$  son menores que la familia conjugada.
- La familia en la dirección Este ( $f_{90}$ ), con fracturas más pequeñas, debe tener más fracturas que la familia dominante.

Nos referiremos al conjunto de datos a estudiar como *los datos*, o los *datos sintéticos*, y serán considerados como si fueran reales. Estos 400 datos se componen de las coordenadas cartesianas de los centros de fracturas  $\vec{x} = (x, y) \in [0, 400] \times [0, 400]$ . Debido a razones de visualización y comparación, estos centros de fracturas se obtuvieron con un proceso homogéneo de Poisson con intensidad  $\theta = 1$ . Las direcciones de rumbo de las fracturas fueron muestradas de una combinación de dos distribuciones de von Mises con parámetros mostrados en la Tabla 7.1. Estos ángulos se encuentran en el rango  $[0, 180]$  en un sistema de coordenadas geográfico (Norte =  $0^\circ$ , Este =  $90^\circ$ ). Las longitudes fueron muestradas mediante una distribución lognormal. Aunque generalmente las fracturas muestran un comportamiento auto-afín por lo que se observan de manera similar a varias escalas; atendiendo a las citas geológica de Davis *et al.* (2011) en este trabajo, utilizaremos cm para las longitudes de fracturas.

Con respecto a la estructura de dependencia, los datos anteriores se muestrearon pegando dos cópulas (gluing copulas) de Placket (copBasic::PLACKETTcop). Para pseudo-observaciones en el intervalo  $u_\theta = [0, 0, 5]$ , se utilizó una dependencia aproximadamente lineal y directa (parámetro de placket = 0.06); para las pseudo-observaciones en el intervalo  $u_\theta = [0.5, 1]$  se utilizó una dependencia aproximadamente lineal e inversa (parámetro de placket = 20). Esta estructura de dependencia se puede observar en la Figura 7.5.

Como resultado, la red de fracturas discretas en ser analizadas y modelada con la metodología estándar y conjunta se muestra en la Figura 7.1. En esta figura, para analizar la relación de dependencia entre dirección y longitud, se colorearon en negro las fracturas con longitudes mayores a 51 cm mientras que las fracturas menores en azul claro. Este corte se basó en el  $90^\circ$  percentil del modelo de longitudes. Es decir, aproximadamente el 10 % del total de fracturas son negras.

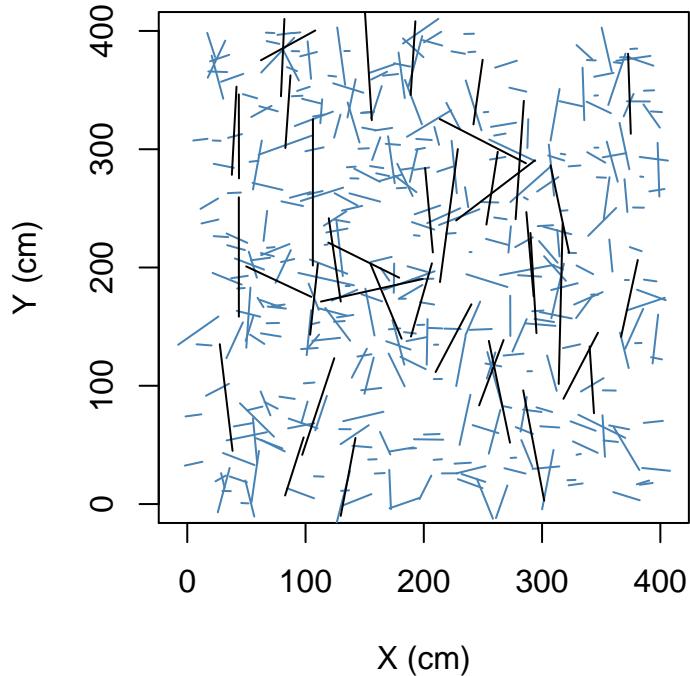


Figura 7.1: Representación gráfica de la red de fracturas de los datos sintéticos. Aproximadamente el 10 % de las fracturas son de color negro.

Las orientaciones fueron generadas usando la función `rmixedvonmises` del paquete `circular` (Agostinelli y Lund, 2013) con los parámetros de la Tabla 7.1. En la Tabla 7.2 se muestran algunos de sus estadígrafos, calculados con el mismo paquete. La media circular ( $89.7^\circ$ ) es muy similar a la mediana ( $88.5^\circ$ ). Ambos estadígrafos muy parecidos a la familia f90. El valor muy pequeño de asimetría es un indicador de que la distribución es altamente simétrica. Estos resultados también se pueden visualizar en la Figura 7.2, la cual muestra que la familia f90 tiene una población mayor que la familia en la dirección N-S. Esta última observación en total acuerdo con los requerimientos del modelo.

familia	proporción	$\mu(\circ)$	$\kappa$
f0	0.3	0	10
f90	0.7	90	10

Tabla 7.1: Parámetros de la combinación de dos distribuciones de von Mises de donde fueron muestreadas las 400 direcciones de la red de fractura a modelar.

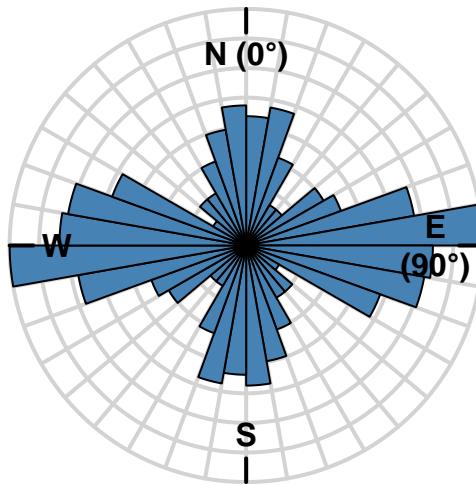


Figura 7.2: Roseta de los datos a analizar.  $f_0$  está en la dirección N-S,  $f_{90}$  en la dirección perpendicular (E-O).

Estadígrafo	valor
Mediana	88.458°
$\mu$	89.728°
Desviación estándar	44.063
Asimetría	0.045

Tabla 7.2: Estadística circular de las direcciones de los datos sintéticos.

Las longitudes fueron generadas utilizando la función `rlnorm` del paquete `stats`. Los parámetros de la distribución lognormal en dicha función son  $meanlog = 3$ ,  $sdlog = 0.73$ , la media y la desviación estándar en la escala logarítmica. Los estadígrafos en la Tabla 7.3 muestran la diferencia entre los valores de la media y la mediana. Esto es una muestra de alta asimetría también en la misma tabla. El histograma además muestra una pequeña moda alrededor de los 70 cm.

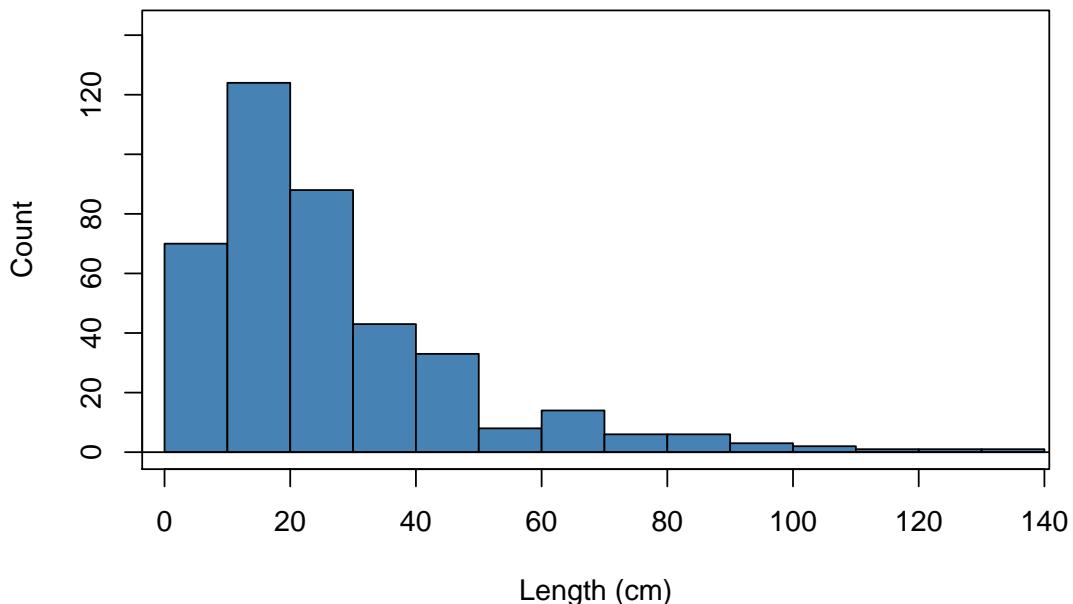


Figura 7.3: Histograma de las longitudes de fractura. La longitud se distribuye lognorm( $\text{meanlog} = 3.00$ ,  $\text{sdlog} = 0.73$ )

Estadígrafo	valor
Mediana	20.389 cm
Media	26.186 cm
Desviación estandar	20.894 cm
Asimetría	1.914

Tabla 7.3: Estadígrafos de las longitudes de los datos sintéticos.

La Figura 7.4 muestra las funciones de distribución acumulativa de la dirección y de la longitud. Aunque estos datos tienen un comportamiento marcadamente diferente, la función cuantil de Bernstein-Kantorovich muestra representar bastante bien ambos comportamientos. Para los datos orientados, en este trabajo se prefiere la flexibilidad de dicha función cuantil en vez de ajustar artificialmente una función como la von Mises.

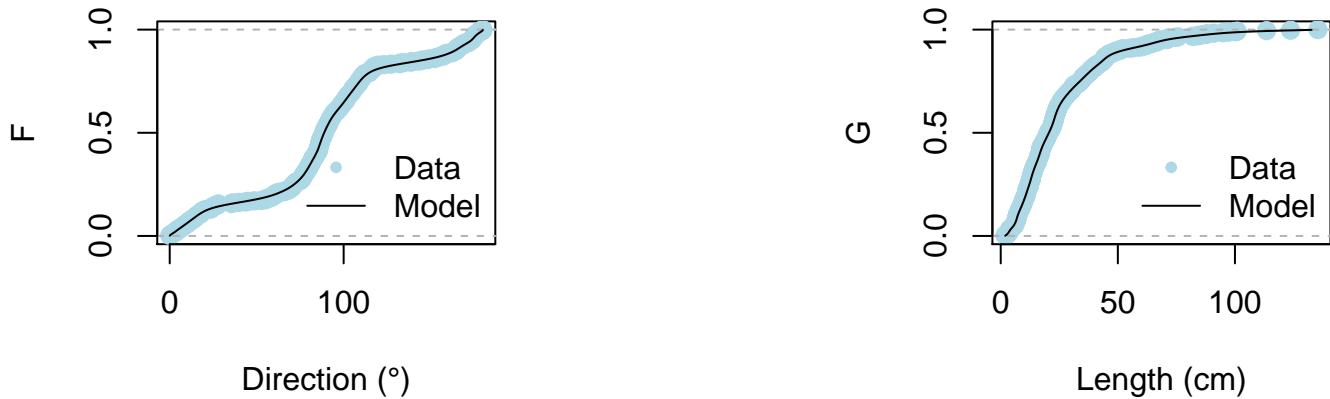


Figura 7.4: Función de distribución acumulativa y su respectivo modelo ajustado mediante la función cuantil de Bernstein-Kantorovich. Izquierda: dirección; Derecha: longitud.

El diagrama de dispersión (Figura 7.5) dirección-longitud y su correspondiente gráfico de pseudo-observaciones  $uv$  se muestran para entender el comportamiento bivariado, es decir, para entender la estructura de dependencia subyacente. La diferencia entre ambos gráficos reside en que el diagrama de dispersión de los datos vive en el producto cartesiano de los rangos de las dos variables aleatorias mientras que el gráfico de pseudo-observaciones vive en el cuadrado unitario ( $[0, 1] \times [0, 1]$ ).

Dado que el análisis de dependencia utilizando la teoría de cópula no es estándar fuera del ámbito estadístico, expliquemos brevemente cómo se efectúa. Los valores en los ejes del gráfico de pseudo-observaciones Figura 7.5 corresponden a las probabilidades (acumulativas) de los cuantiles (ejes en el diagrama de dispersión). Por ejemplo, en el eje horizontal del diagrama de dispersión (izquierda de Figura 7.5), el ángulo  $90^\circ$  (la media) corresponde a  $u = 0.5$  en el gráfico de pseudo-observaciones; mientras que el tercer cuatil en el diagrama de dispersión corresponde a  $u = 0.75$ . El ángulo  $0^\circ$  corresponde a  $u = 0$ . De manera similar para el eje vertical.

El gráfico de pseudo-observaciones permite entender de manera visual la estructura de dependencia. De esta manera se puede estudiar el efecto que tiene la estructura de dependencia sobre el diagrama de dispersión.

El coeficiente de correlación de rango para estos datos ( $\rho_M = 0.635$ ) es un indicador de que la orientación no es independiente de la longitud. Éste número fue calculado con los códigos que se pueden encontrar en el trabajo de Tu (2015).

Un indicador más confiable para determinar independencia se debe basar en la teoría de cópulas. La función `indepTest` del paquete `copula` permite hacer una prueba de independencia basada en cópulas. Para los datos en uso se obtuvo un estadístico de la prueba con valor 0.408 y un p-valor de  $5 \times 10^{-4}$ , lo que indica que la independencia es rechazada. Por lo tanto, sí es útil continuar con el modelado de la estructura de dependencia.

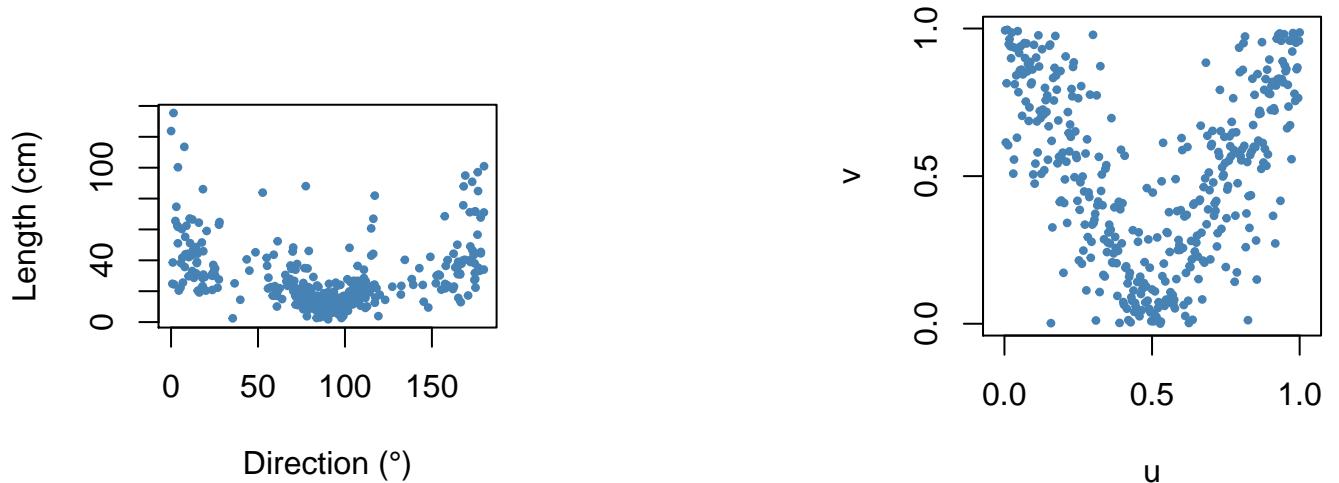


Figura 7.5: Diagrama de dispersión de la longitud vs dirección (izquierda), y su correspondiente gráfico de pseudo-observaciones (derecha).

En el diagrama de dispersión claramente se muestra una moda en la dirección N-S pero la otra familia es casi indistinguible. Es decir, la familia f0 resalta más que la familia f90. En contraste, en la roseta de direcciones sucede lo casi contrario. Ahí sí se pueden visualizar claramente ambas modas, pero la que más resalta es la de la familia en la dirección E-O. En el diagrama de dispersión se puede observar que la probabilidad de obtener fracturas con longitudes menores a, digamos, 18cm, dentro de la familia N-S es casi cero. Un fenómeno inverso se puede observar en la familia E-O: muchas fracturas pequeñas y casi ninguna grande.

Este fenómeno se debe a la estructura de dependencia mostrada en el gráfico de pseudo-observaciones. Nótese también que hay muchas más fracturas pequeñas (menores a 51 cm, por ejemplo) que fracturas largas a lo largo del rango de direcciones,  $[0, 180^\circ]$ .

### 7.1.2. La metodología estándar

Para el siguiente análisis es conveniente distinguir entre las palabras 'sintético' y simulado. Cuando usemos el primero será para referirnos al conjunto de datos de la sección pasada. Nótese pue para un conjunto de datos pueden existir múltiples simulaciones.

En esta sección se muestran resultados utilizando una metodología cuasi-estándar. El término *cuasi* se debe a que no se hace clasificación de fracturas alguna ya que con la función cuantil de Bernstein-Kantorovich se puede modelar directamente datos multimodales. De esta manera, las funciones marginales univariadas son modeladas de manera idéntica que en la sección siguiente. Esto nos permite ser más objetivos en la comparación del análisis bivariado, que en la metodología estándar se hace simulando cada variable de manera independiente.

Los modelos marginales univariados utilizados serán los mismos que en la Figura 7.4. Para facilitar la comparación, los anchos de clase son los mismos tanto en la roseta de direcciones de los datos sintéticos (Figura 7.2) como en la de los datos simulados (Figura 7.6). De manera similar

para los histogramas de los datos sintético y los simulados. De manera bivariada, los rangos tanto en el eje vertical como en el horizontal fueron fijados. También en la parte espacial, los centros de las fracturas son exactamente los mismos.

La roseta de las simulaciones independientes (Figura 7.6) parece más suavizada, es decir, comparando con la roseta de los datos (Figura 7.2), hay menos contraste entre las modas y los valles. A diferencia de este comentario, las ubicaciones de dichas modas y valles son satisfactoriamente reproducidas. Incluso, una ligera asimetría alrededor de  $90^\circ$  es reproducida. De manera cuantitativa, los estadígrafos circulares (Tabla 7.4) también muestra lo bueno que es el ajuste. Sin embargo, el aumento de desviación estándar remarca la sospecha del efecto de suavizado.

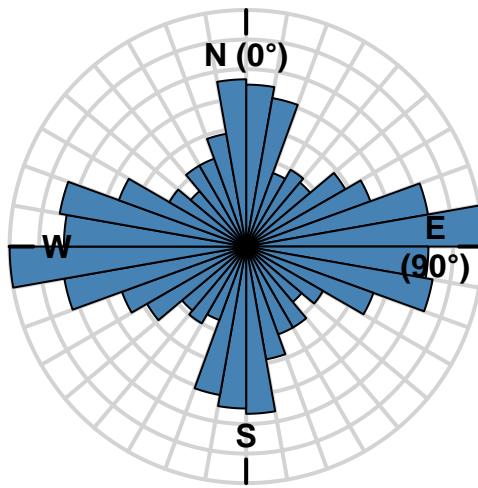


Figura 7.6: Roseta de orientaciones de una simulación. Compárese con Figura 7.2

Estadígrafo	valor
Mediana	$83.930^\circ$
$\mu$	$85.431^\circ$
Desviación estándar	48.920
Asimetría	0.166

Tabla 7.4: Estadígrafos circulares de las direcciones de una simulación de los datos. Compárese con Tabla 7.2.

Las longitudes, simuladas independientemente de las orientaciones, también fueron reproducidos. Nótese por ejemplo que el valle observado a aproximadamente 55cm se reproduce en el histograma (Figura 7.7). La moda sutil alrededor de los 70 cm también aparece en esta simulación. Por el lado cuantitativo, la media y la mediana son distintos entre sí pero muy parecidos sus correspondientes en los datos sintéticos. Aquí también se muestra un aumento en la desviación estándar.

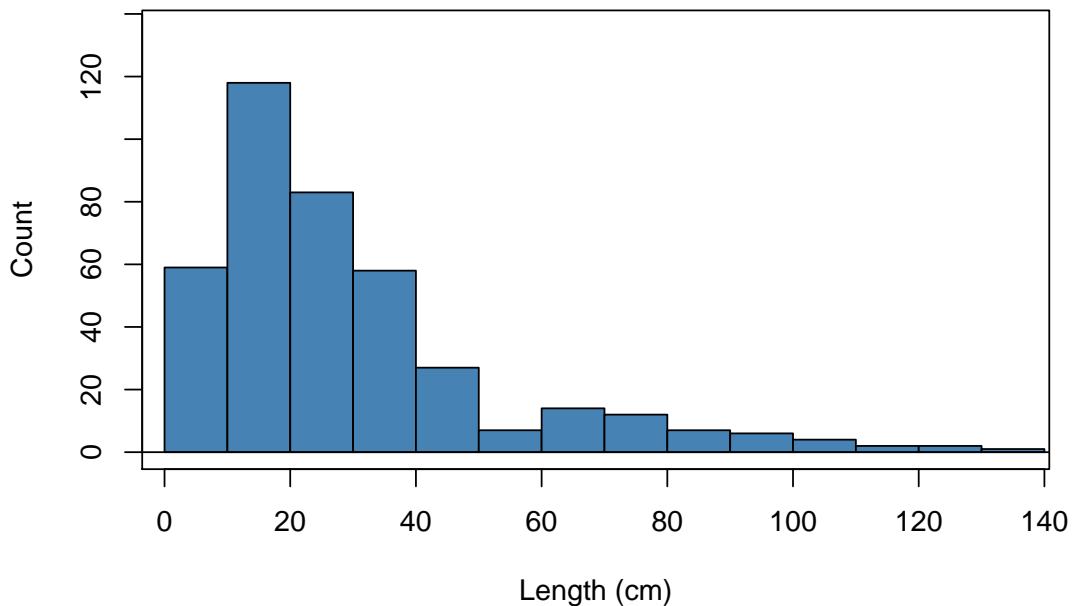


Figura 7.7: Histograma de longitudes de una simulación. Compárese con la Figura 7.3.

Estadígrafo	valor
Mediana	22.877 cm
Media	29.400 cm
Desviación estandar	23.930 cm
Asimetría	1.752

Tabla 7.5: Estadígrafos de las longitudes de una simulación de los datos. Compárese con la Tabla 7.3.

Esta coherencia entre la simulación y los datos también se observa en la cercanía de la función de distribución acumulativa empírica con la de los modelos (Figura 7.8). Éste es el resultado de la cualidad no-paramétrica de la función cuantil de Bernstein-Kantorovich.

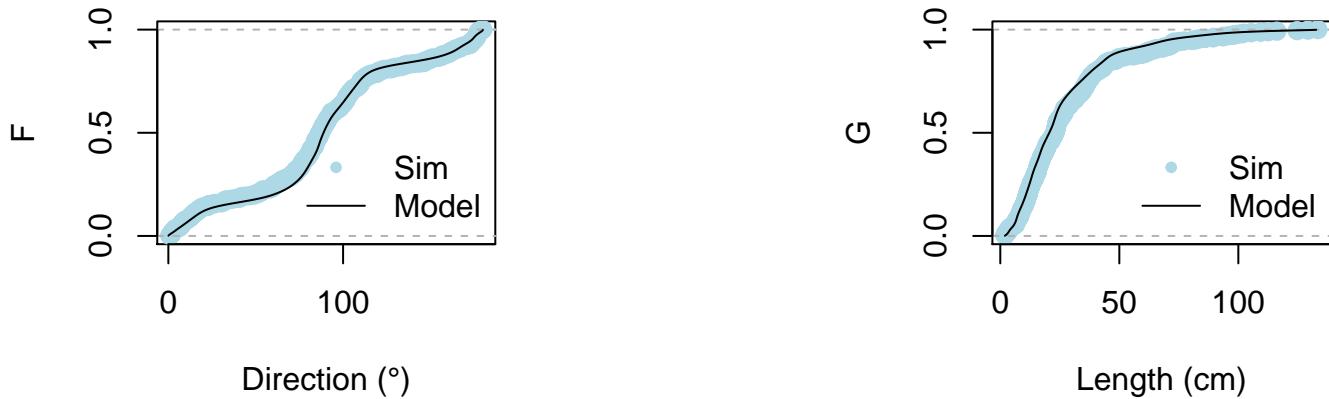


Figura 7.8: Funciones de distribución acumulativa empírica de los datos y sus respectivos modelos. Izquierda: Direcciones; Derecha: longitudes.

Nótese sin embargo que en el caso bivariado (Figura 7.9) el modelado no es tan satisfactorio como en el caso univariado. Tanto en el diagrama de dispersión como en el gráfico de pseudo-observaciones los puntos están más dispersos, sin forma preferencial alguna. Cuantitativamente, el coeficiente de correlación es claramente diferente ( $\rho_M = 0.016$ ). Ni es diagram de dispersión ni el gráfico de pseudo-observaciones muestra la estructura de los datos que se quiso modelar. La razón: la simulación no tomó en cuenta ningún dato bivariado. La metodología estándar siempre produce resultados como éste gráfico de pseudo-observaciones: puntos distribuidos independientemente (Figura 7.9).

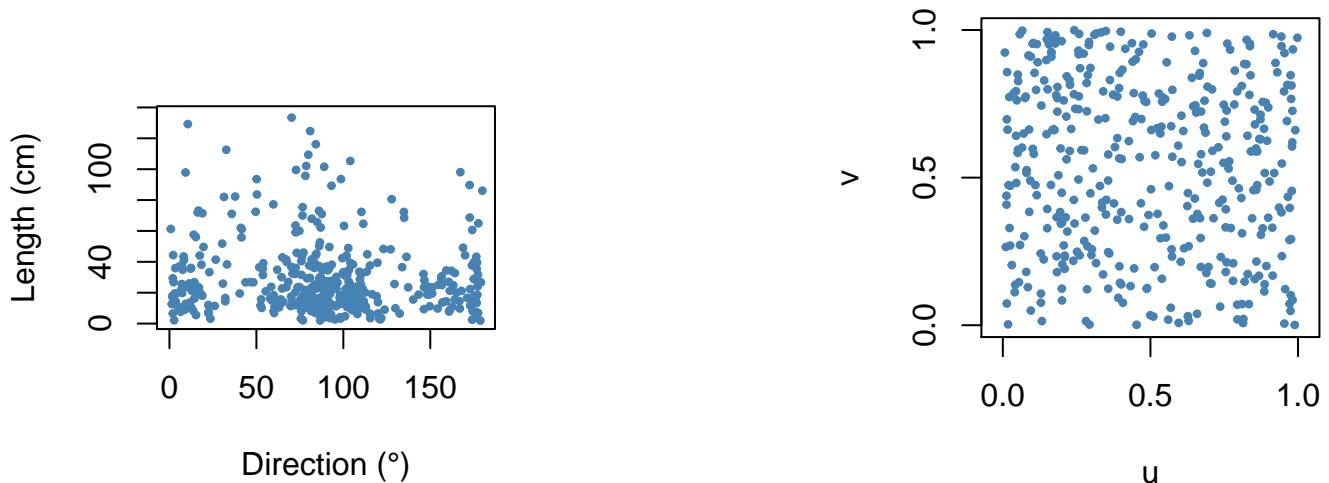


Figura 7.9: Diagramas de dispersión de la simulación. Izquierda: los datos; derecha: las pseudo-observaciones.

La DFN resultante paga el precio al mostrar, por ejemplo, que el 10 % de las fracturas más largas no están preferencialmente en la dirección N-S como los datos originales. Al contrario, ahora las fracturas más largas se encuentran en la dirección E-W.

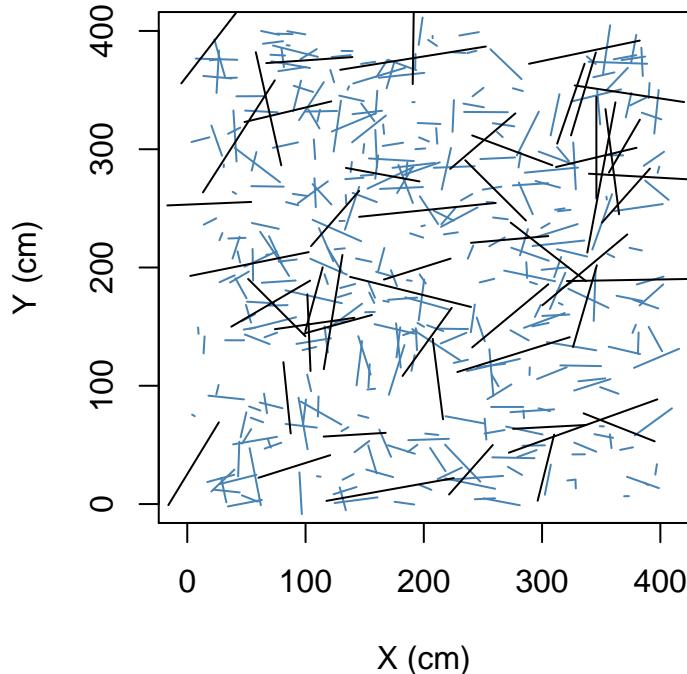


Figura 7.10: DFN de una simulación de los datos. Compárese con Figura 7.1.

### 7.1.3. El enfoque con la cópula de Bernstein

La metodología en esta sección sí incluye la estructura de dependencia a través de la cópula de Bernstein de las pseudo-observaciones en la Figura 7.5. Con este modelo de cópula, se simularon nuevas pseudo-observaciones (izquierda de la Figura 7.11) utilizando los pasos 1 y 2 del algoritmo de simulación bivariado. Después, utilizando el paso 3, se obtuvieron los valores simulados de orientación y longitud (derecha de la Figura 7.11) mediante las funciones cuantiles mostradas en la Figura 7.4.

La teoría de cópulas permitió tener las mismas marginales que en el caso independiente, es decir, la roseta de direcciones, el histograma de longitudes y los estadígrafos son exactamente los mismos que en el caso independiente. La diferencia radica en la estructura de dependencia. La Figura 7.11 muestra los diagramas de dispersión. Nótese el contraste en aspecto visual de esta simulación comparada con los datos reales. Cuantitativamente, el coeficiente de correlación de rango ( $\rho_M = 0.571$ ) también es coherente con el de los datos.

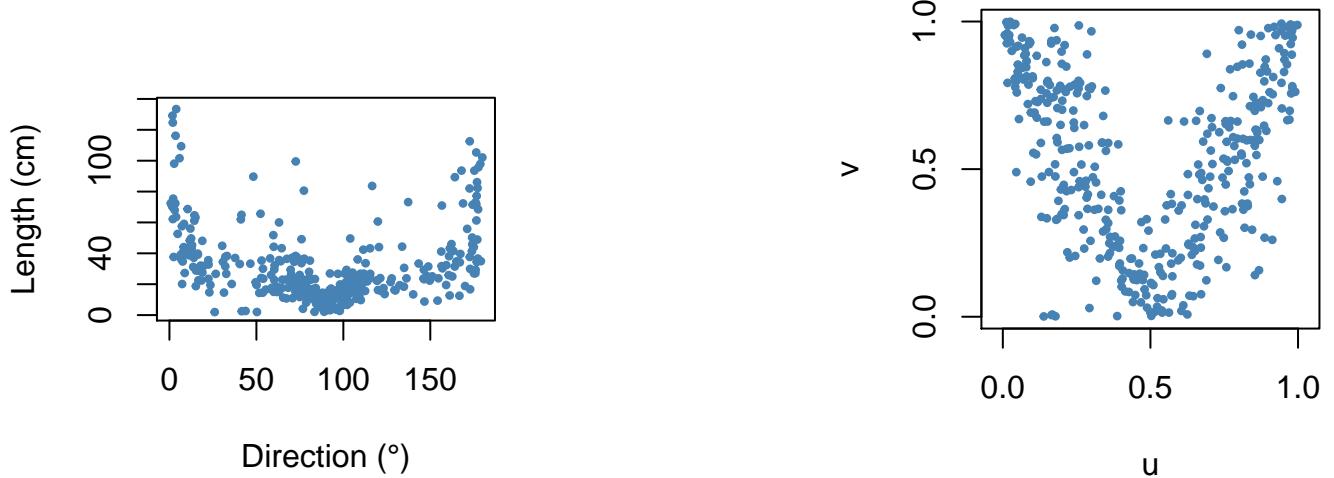


Figura 7.11: Diagrama de dispersión y gráfico de pseudo-observaciones de la simulación que toma en cuenta la modelación de la estructura de dependencia. Compárese con la Figura 7.5 y con la Figura 7.9.

La red de fracturas discretas correspondiente, al igual que en los datos sintéticos, muestra la preferencia de las fracturas más largas por la dirección N-S. Sin embargo, sin faltar a la realidad, todavía existe la posibilidad de obtener fracturas en otras direcciones.

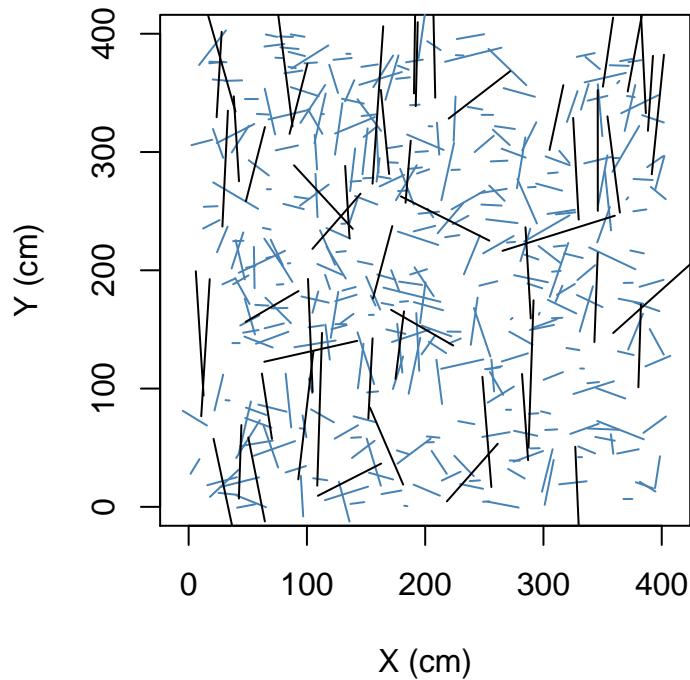


Figura 7.12: DFN de la simulación basada en las cópulas de Bernstein. Compárese con Figura 7.1 y con Figura 7.12.

## 7.2. Caso de estudio trivariado: orientación-longitud-apertura.

Para mostrar la utilidad de esta teoría se generó computacionalmente un sistema de fracturas (Figura 7.13) al cual se le dio una estructura de dependencia compleja en el que las fracturas en la dirección vertical son más pequeñas que las que están en dirección horizontal. Además, la relación longitud-apertura es monótona, aunque no lineal (Figura 7.15). Por otro lado, la ubicación espacial fue simulada mediante un proceso puntual de Poisson con intensidad constante ya que un proceso de Poisson con intensidad variable no permitió una mejor visualización de la red de fracturas. Los parámetros de los modelos se muestran a continuación:

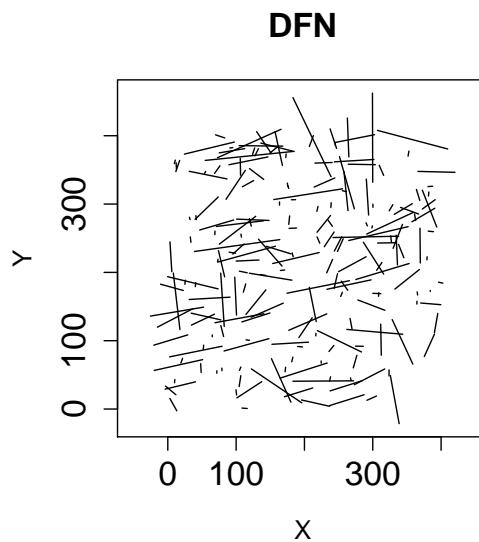


Figura 7.13: Sistema de fracturas a analizar.

- $x - \text{poisproc}(I = 0.5)$
- $\theta - \text{vonmises}(\mu = 90, 180; \kappa = 10, 10)$
- $L - \text{lnorm}(\mu = 4, 2; \sigma = 0.4, 0.6)$
- $b - \text{lnorm}(\mu = \exp(-4); \sigma = \exp(1))$

Nótese que tanto los datos de orientación como los datos de longitud fueron simulados cada uno como una combinación de las distribuciones de probabilidad correspondientes, lo que resulta en distribuciones bimodales tanto para los ángulos como para las longitudes. La relación de dependencia entre estas dos variables (Figura 7.15) parece fácil ya que es una unión de dependencia lineal y uniforme, sin embargo, en la práctica es difícil modelarla cuando se dan de esta manera.

Por fines de comparación con simulaciones de este mismo sistema de fracturas, a continuación, se muestran algunos estadígrafos del análisis exploratorio de los datos. Las direcciones de las fracturas están medidas en un rango de  $0^\circ$  a  $180^\circ$  sobre un sistema de coordenadas geográfico. Los estadígrafos de la variable direccional  $\Theta$  fueron calculados mediante la estadística de datos orientados (Fisher, 1995; Jammalamadaka y SenGupta, 2001; Mardia y Jupp, 2000).

Statistics	$L$	$\theta$
$\mu$	34.6203	36.3756
median	24.1315	36.5734
$\sigma$	30.3200	59.5328
skewness	0.8456	-0.2369

Nótese que estos estadígrafos no representan adecuadamente los modelos ya que no son modelos univariados. Sin embargo, se espera que la metodología presentada sí logre reproducir cada uno de los ingredientes para generar esta red de fracturas, así como los estadígrafos.

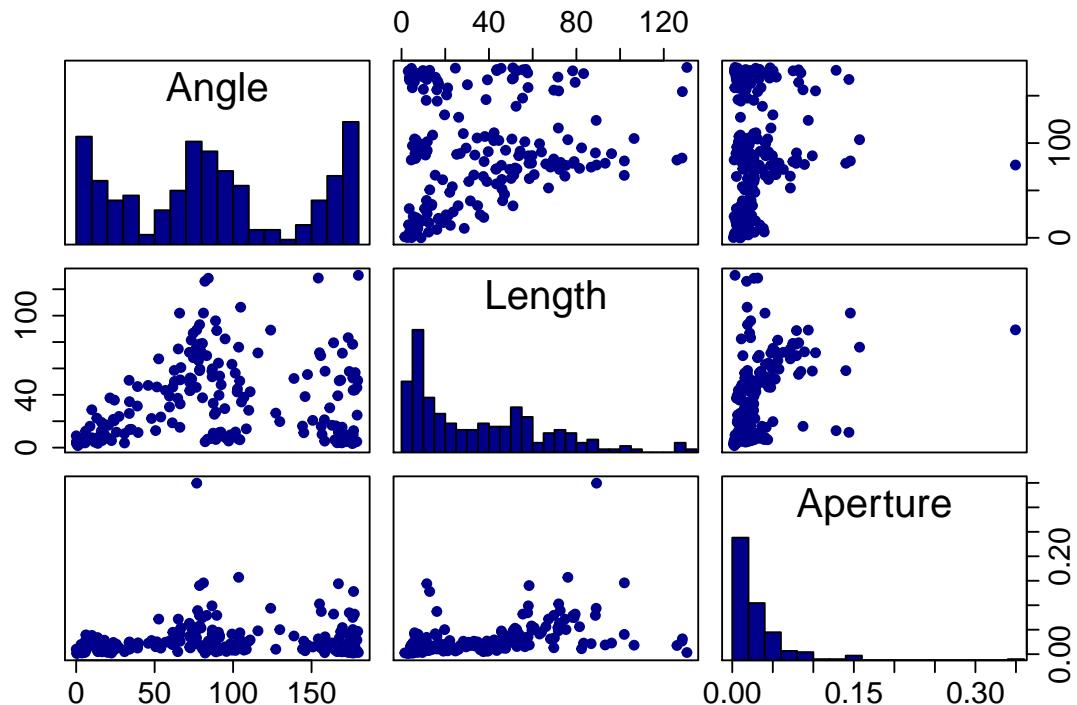


Figura 7.14: Matrix de diagramas de dispersión con las correspondientes distribuciones marginales en las entradas de la diagonal.

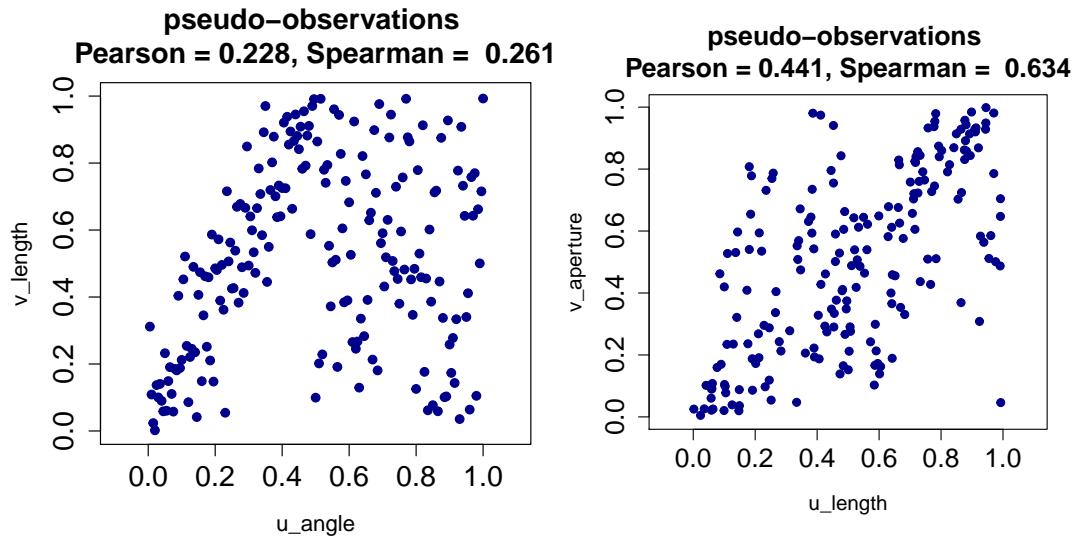


Figura 7.15: Diagramas de dispersión de las pseudo-observaciones.

Con los datos de la Figura 7.13 se modeló de manera separada: a) la distribución marginal de las orientaciones, b) la distribución marginal de las longitudes, c) la distribución marginal de las aperturas, d) la cópula de la relación orientación-longitud, e) la cópula longitud-apertura. Este último estos últimos dos incisos se llevaron a cabo suponiendo independencia condicional.

Uno de los objetivos buscados al modelar vectores aleatorios es simular realizaciones del mismo proceso subyacente. A continuación, se muestra el resultado de una simulación.

Statistics	$L$	$\theta$
$\mu$	38.1409	51.8984
median	32.6340	55.5260
$\sigma$	31.0912	55.5536
skewness	0.7451	0.2844

Nótese que los estadígrafos son congruentes con los de los datos pero en el caso de las medidas de tendencia central de las orientaciones no tanto, quizás porque es una distribución bimodal o quizás sea solamente para esta realización en particular. Sin embargo, los histogramas se ven muy congruentes ya que sí respetan las modas y valles, incluso para los datos de orientación. Las simulaciones de longitud y de apertura son muy congruentes con los datos tanto en estadígrafos como en el histograma. Se respeta el comportamiento bimodal para los datos de longitud y la forma con asimetría positiva para los datos de apertura.

De manera bivariada la congruencia también es satisfactoria, tanto en el diagrama de dispersión como en el gráfico de pseudo-observaciones, respetándose las zonas de nula, baja y alta densidad de datos. Dentro de esos gráficos también se muestran algunos estadígrafos bivariados (coeficientes de correlación) que confirman la bondad de la metodología mostrada.

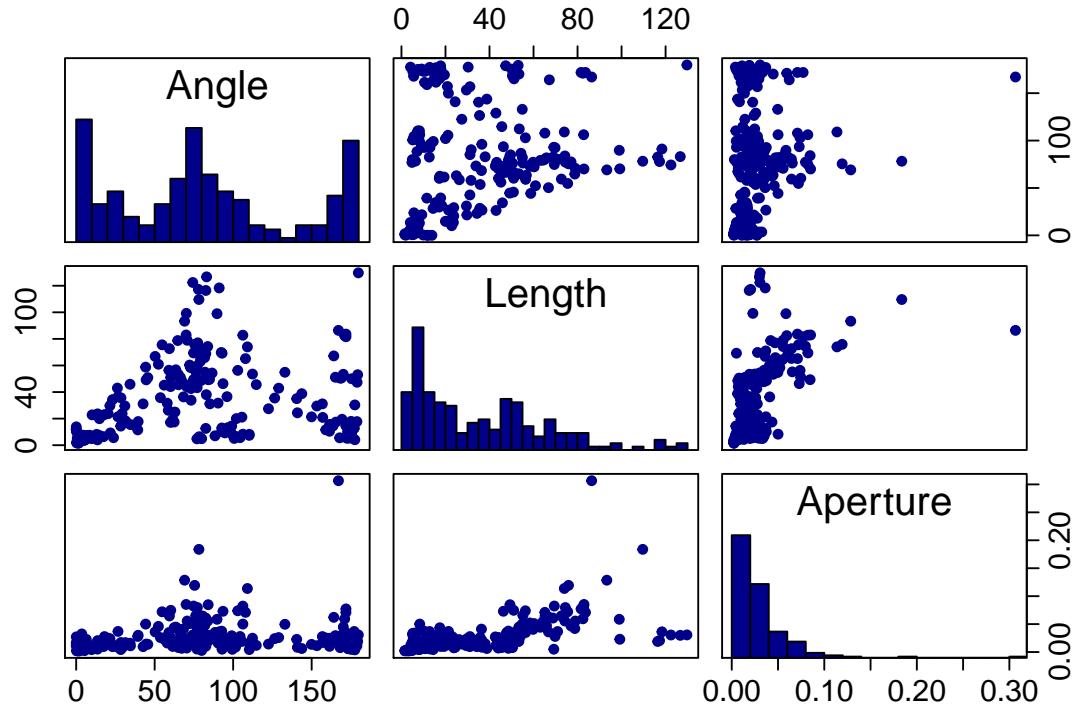


Figura 7.16: Distribuciones marginales en la diagonal dentro de la matriz de simulaciones.

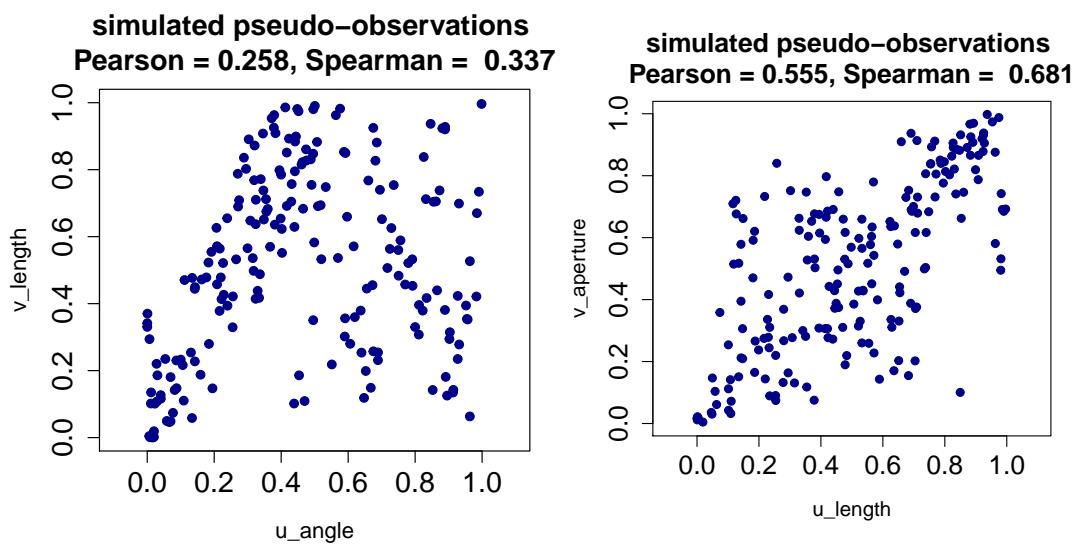


Figura 7.17: Gráfico de pseudo-observaciones simuladas.

### 7.3. Resultados y discusión

Por construcción, el enfoque de Bernstein-Kantorovich para modelar la función cuantil univariada es no paramétrico, pero se observaron muy buenos resultados para modelar todas las variables, incluyendo la dirección de las fracturas. Esto debido a que el problema de la periodicidad se compensa en la modelación al tener muchos datos en esta frontera. Las simulaciones con éste enfoque univariado reproducen incluso modas sutiles en los datos.

En la parte bivariada, los diagramas de pseudo-observaciones muestran una correcta correspondencia entre los datos y las simulaciones con las cópulas de Bernstein: la mayoría de las fracturas pequeñas están alineadas en la dirección E-O y las fracturas más grandes en la dirección perpendicular. Los resultados también se aprecian en los diagramas de dispersión.

Por el contrario, el diagrama de pseudo-observaciones utilizando la metodología estándar no imita la dependencia de los datos. Aún más grave, la mayoría de las fracturas grandes apuntan en la dirección de las fracturas pequeñas. Incluso hay fracturas grandes en las zonas de probabilidad donde no debería haber y también fracturas pequeñas en zonas de casi nula probabilidad de ocurrencia. Por ejemplo, se observan fracturas muy pequeñas en la dirección N-S, pero en los datos originales no las hay.

Se observó una buena correspondencia tanto en la DFN de los datos como de las simulaciones con la metodología propuesta tomando en cuenta la estructura de dependencia. En general, todas las variables involucradas fueron reproducidas, tanto de manera gráfica como en sus estadígrafos, además de marginal como de manera conjunta.

Las rosetas y los histogramas para cada una de las simulaciones parece más suavizado que su correspondiente de los datos originales. Tal suavizado también se reflejó de manera cuantitativa en una disminución de la desviación estándar. Investigando se encontró (Phillips, 2003, p. 251) que este fenómeno se debe a que la convergencia uniforme de los polinomios de Bernstein es muy lenta, o cual se traduce en un suavizado de la curva.

Debido a que la relación orientación-longitud de fracturas es muy importante en un contexto geológico y dinámico, además de para obtener propiedades de percolación de un yacimiento naturalmente fracturado dado, esta metodología es de mucha utilidad cuando se observan dependencias entre tales variables.

Este trabajo se ha limitado a desarrollar algoritmos, metodología y ejemplos de aplicación en los casos bivariado y trivariado con cópulas de Vine. Sin embargo, se han mostrado algunas fórmulas para incluir más de tres variables. En este sentido, Wei y Scheffer (2012) mostraron cómo extender la metodología dentro del enfoque de cópulas de Vine.

Aunque se podría haber utilizado modelos geoestadísticos para la intensidad de fractura se decidió utilizar procesos puntuales con intensidad constante ya que después de varios intentos, el objeto de estudio de este trabajo (la dependencia entre las propiedades de las fracturas) se analizaba mejor visualmente con un proceso puntual homogéneo.



## Conclusiones y trabajo futuro

Se estableció una metodología sistemática para la simulación estocástica conjunta de las propiedades geométricas con dependencias complejas de los objetos que representan fracturas discretas mediante la modelación de su función de distribución de probabilidad conjunta usando cópulas, en particular, usando las cópulas no paramétricas de Bernstein.

A diferencia de las restricciones de los modelos de regresión lineal entre las variables, el enfoque de la cópula de Bernstein permite investigar estadísticamente estructuras de dependencia complejas de manera muy flexible. Este enfoque es dirigido por los datos. En particular, la condición periódica de Carnicero *et al.* extiende el enfoque para incluir variables tales como la dirección de la fractura.

El uso de las cópulas para modelar la estructura de dependencia permite evitar el sesgo producido al transformar las variables aleatorias involucradas en simulaciones, por ejemplo la transformación logarítmica de las longitudes de fractura.

Los enfoques no paramétricos utilizados para las distribuciones marginales y la cópula permitieron una muy buena coincidencia de la simulación de la red de fracturas discretas, incluso en presencia de asimetría en la distribución. Esto permite estimar, a través del análisis de simulaciones, propiedades de percolación más realistas de los medios porosos fracturados. Otra ventaja con este enfoque no paramétrico es la facilidad de uso, ya que no se requiere una prueba bondad de ajuste.

Basándonos en la teoría de cópulas, la geometría estocástica, y la modelación geológica-petrofísica, se establecieron flujos de trabajo para la metodología propuesta para analizar, modelar y simular redes de fracturas discretas. Se mostraron flujos de trabajo generales y particulares. Como resultado, a partir de un conjunto de datos de fracturas, se pueden obtener simulaciones de redes de fracturas discretas tomando en cuenta su estructura de dependencia. De manera más general, esta metodología se puede aplicar a los modelos booleanos.

Como resultado colateral se ha creado un software en R que permite implementar la metodología paso a paso junto con ~~esta~~ tesis. Cabe mencionar que como parte del flujo de trabajo se ha incluido el análisis exploratorio de los datos ya que a veces suele omitirse. En esta etapa es donde se entiende

el comportamiento de los datos, que a su vez es un descriptor del fenómeno subyacente. Dicha comprensión permite validar las simulaciones.

En lugar de usar un estadígrafo para estudiar la dependencia de variables aleatorias, tiene mucho potencial utilizar una función (modelo de dependencia), la cópula, para modelar dicha dependencia. Como ejemplo se mostró un caso con dependencia combinando una dependencia lineal en un rango hasta la mediana y para el segundo rango de los datos se modeló de manera independiente. Aunque esto se puede hacer por separado la cópula puede modelar ambos casos al mismo tiempo. Para la relación longitud-apertura se utilizó una dependencia cuasi-monótona pero a la vez compleja. De esta manera se ha mostrado la versatilidad de la cópula para modelar dependencias comúnmente utilizadas, así como dependencias no frecuentemente mostradas en la literatura.

También se ha mostrado cómo la metodología bivariada se puede llevar al caso tri-variado de variables aleatorias con el enfoque de Vine copulas. En particular cuando hay independencia condicional, en el cual se mostró que la longitud es el enlace entre orientación y apertura. La independencia observada se mostró entre la orientación y la apertura.

~~\* Cabe mencionar que durante el desarrollo de ésta tesis se detectaron varias áreas que pueden enriquecer los resultados. A continuación se mencionan, ya que pueden ser relevantes tanto para incluir más fenómenos como para reducir el tiempo de cómputo.~~

Como trabajo futuro se pueden explorar otras definiciones de la función de distribución empírica como la de Weibull. Éstas tienen la ventaja de que las simulaciones no estarían restringidas a valores entre el mínimo y el máximo de los datos.

Una de las desventajas de las cópulas de Bernstein es el tiempo de cómputo, el cual también se podría reducir si se trabaja en un algoritmo que calcule la inversa de manera que considere las propiedades (por ejemplo, la monotonía) de las funciones de distribución univariada. Tal algoritmo podría ser el de bisección de manera que tome en cuenta el cómputo en paralelo.

Se ha demostrado que las cópulas de Bernstein no reproducen dependencia en las colas, por lo que los resultados de esta tesis también se pueden enriquecer con los resultados de dicho artículo.

Las cópulas de Bernstein reproducen muy bien los datos, lo que puede llevar a un sobreajuste, es decir, las simulaciones solamente reproducen los datos utilizados y puede no reproducir otros datos agregados en campañas de adquisición posteriores. Se puede buscar métodos que permitan reducir dicho sobreajuste ~~quizás algún método de validación cruzada.~~

Solamente se trabajó en el enfoque con la cópula de Bernstein ya que reproduce una gran cantidad de estructuras de dependencias complejas. También queda comparar y combinar la metodología utilizando cópulas paramétricas como las arquimedianas ya que ése tipo de cópulas tienen estructuras de dependencia monótonas o una combinación de éste tipo de dependencias, lo cual hace más fácil la interpretación del fenómeno en estudio.

El caso trivariado directo, sin cópulas de Vine, también se podría implementar. Para obtener simulaciones, se requiere obtener la inversa de una función de tres variables, lo cual podría tomar demasiado tiempo computacional. En el dado caso de existir datos orientados, también se tendría que implementar el cálculo del histograma multivariado.

Se puede explorar los resultados con el enfoque de B-splines para modelar, no la cópula  $C$

como en este trabajo, sino la función de densidad de cópula  $c$  como en el trabajo de Schellhase y Kauermann (2012).



## Teoría de la aproximación de funciones

En los cursos básicos de geometría analítica y de cálculo diferencial se encuentra frecuentemente el problema de encontrar la recta que pasa por dos puntos dados (**puntos de control**). Este es quizás el ejemplo más sencillo de interpolación: ajustar la curva que pasa por los puntos dados.

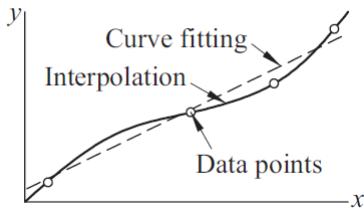


Figura A.1: Aproximación vs interpolación (línea continua).

Para el caso de datos experimentales, no se tiene la total certeza de la exactitud de los mismos por lo que una curva que se *aproxime* bien podría modelar el fenómeno. Además, se ha demostrado que en ciertos casos el interpolar puede resultar en comportamientos que difieren del esperado en el fenómeno, por ejemplo, la curva puede oscilar demasiado y no reproducir adecuadamente las formas deseadas (Figura A.2). Se ha observado que en el caso de interpolaciones polinomiales, estas oscilaciones incrementan con la cantidad de datos, equivalentemente, con el grado del polinomio. Por lo tanto, no importa cuántos puntos de control se agreguen, no hay garantía de que los interpolantes polinomiales converjan a la curva o superficie que se quiere representar.

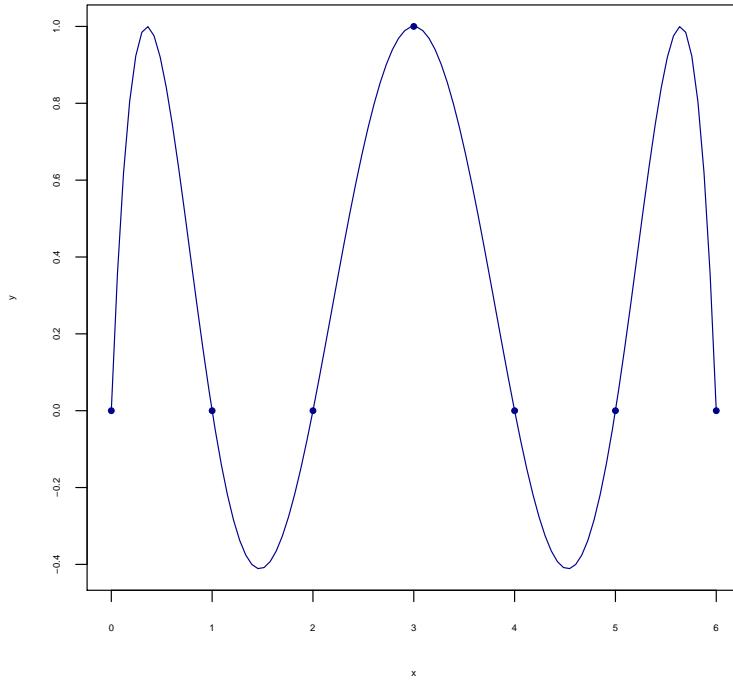


Figura A.2: Interpolación de Lagrange. Observe las oscilaciones en la curva polinomial aún cuando no se observa tal oscilación en los puntos de control.

## A.1. Curvas de Bézier

Un ejemplo de aproximación se forma al utilizar los polinomios de Bernstein, los cuales se definen de la siguiente manera (de Villiers, 2012; Goldman, 2002; Mann, 2006; Phillips, 2003):

$$B_i^n(u) = \binom{n}{i} u^i (1-u)^{n-i} \quad (\text{A.1})$$

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

El conjunto de estos polinomios para una  $n$  dada forma una base del espacio de polinomios de grado  $n$ .

De esta manera, el teorema de Weierstrass establece que cualquier función continua se puede aproximar uniformemente mediante polinomios, en particular mediante las curvas de Bézier:

$$B(u) = \sum_{i=0}^n P_i B_i^n(u) \quad (\text{A.2})$$

donde  $P_i \in \mathbb{R}^n$  son puntos de control tomados de la función a aproximar.

Así, se puede aproximar cualquier curva en  $\mathbb{R}^n$ . En el caso de puntos de control equiespaciados (Figura A.3) de la forma  $(i/n, y_i)$ , las curvas de Bézier presentan funciones de la forma  $y = f(x)$ :

$$B(u) = \sum_{i=0}^n (i/n, y_i) B_i^n(u) \quad (\text{A.3})$$

$$= \left( u, \sum_{i=0}^n y_i B_i^n(u) \right) \quad (\text{A.4})$$

Lo cual se puede re-escribir de la siguiente manera:

$$B(u) = \sum_{i=0}^n f\left(\frac{i}{n}\right) B_i^n(u) \quad (\text{A.5})$$

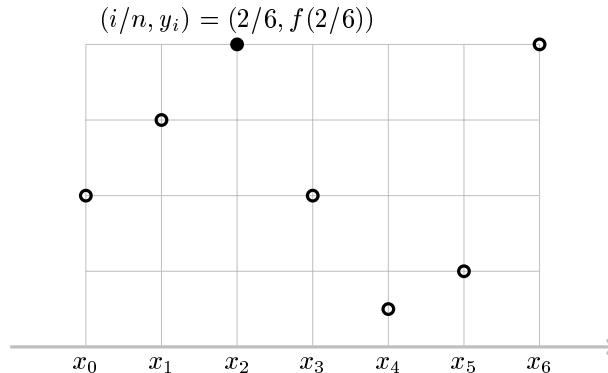


Figura A.3: Curva de Bézier para el caso de puntos equiespaciados a lo largo del eje  $x$ .

Los polinomios de Bernstein-Kantorovich son un ejemplo de aplicación de esta última ecuación en la que se approxima la función cuantil haciendo  $f(i/n) = (x_{(i)} + x_{(i+1)})/2$  (Muñoz-Pérez y Fernández-Palacín, 1987). Su implementación computacional se encuentra con el nombre `dat2bernqua` dentro del paquete `lmomco` del software estadístico R.

Estas aproximaciones tienen muchas propiedades que la hacen populares en varios ámbitos como en CAGD (Computer aided geometric design) y en probabilidad con las cópulas de Bernstein.

Por ejemplo, si  $f(x)$  es *monótona*, también lo es su correspondiente curva de Bézier (Phillips, 2003, Theorem 7.1.2, p. 251). Ejemplo de funciones monótonamente crecientes son las funciones de distribución de probabilidad.

*Invarianza afín.* Se dice que una curva tiene invarianza afín si sus puntos de control viven en un espacio afín. Esta característica garantiza que la curva sea independiente del sistema de

coordenadas, y por lo tanto, una transformación afín de la curva solamente requiere transformar los puntos :  $T(\sum_{k=0}^n P_k B_k^n(t)) = \sum_{k=0}^n T(P_k) B_k^n(t)$ .

*Envolvente convexa (Convex Hull).* Las curvas de Bézier siempre se encuentran dentro de la envolvente convexa de los puntos de control. Esta propiedad garantiza que la curva esté ubicada en la proximidad de los puntos de control. Además también garantiza que si los puntos de control son visibles en una pantalla, entonces la totalidad de la curva también es visible (ver Figura A.4).

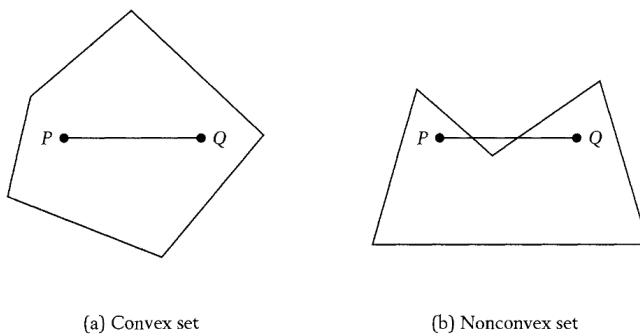


Figura A.4: Sean  $P$  y  $Q$  dos puntos en el conjunto  $S$ . (a) el segmento rectilíneo  $\overline{PQ}$  yace completamente dentro del conjunto convexo  $S$ ; (b) ejemplo que muestra un conjunto no convexo, en el cual parte del segmento rectilíneo yace fuera del conjunto.

*Simetría.* Invertir el orden de los puntos de control produce la misma curva de Bézier, pero con orientación contraria.

*Interpolación de los puntos extremos.* A diferencia de los polinomios de Lagrange, las curvas de Bézier generalmente no interpolan los puntos de control, pero siempre interpolan el primer y el último punto. Por ejemplo, un requisito de las funciones de distribución es que tome los valores 0 y 1 lo cual se da automáticamente con los polinomios de Bernstein-Bézier.

*No-degeneración.* Las curvas de Bézier son no-degeneradas, es decir, la curva nunca se colapsa un solo punto al menos que todos los puntos de control sean el mismo.

*Derivada.* Si  $F^{(k)}(x)$  definida en  $[0, 1]$  tiene un signo dado, entonces  $B$  tiene el mismo signo en el mismo rango. Por ejemplo, si  $F(x)$  existe y es no-negativa, de tal manera que es convexa, entonces su correspondiente representación de Bézier satisface las mismas características (Phillips, 2003, p. 254). Además, si  $F$  pertenece al espacio de las funciones con derivadas continuas hasta la  $k$ -ésima derivada ( $C_k[0, 1]$ ) para algún entero  $k \geq 0$ , entonces  $B^{(k)}$  converge uniformemente a  $F^{(k)}$  (Phillips, 2003, p. 258). Cuando  $k = 1$  se tiene que  $f(x) = F'(x)$ , la densidad de probabilidad, es aproximada mediante la derivada de la curva de Bézier.

Esta propiedad también explica que las curvas de Bézier preserven la forma de la curva. La derivada se puede escribir como combinación lineal de dos polinomios de grado menor:

$$B'(u) = n \sum_{i=0}^{n-1} (\Delta P_i) B_i^{n-1}(u) \quad (\text{A.6})$$

$$= n \sum_{i=0}^{n-1} (P_{i+1} - P_i) B_i^{n-1}(u) \quad (\text{A.7})$$

Reescribiéndola de otra manera:

$$B'(u|P_0, \dots, P_n) = n (B(u|P_0, \dots, P_n) - B(u|P_0, \dots, P_{n-1})) \quad (\text{A.8})$$

*Disminución de la variabilidad.* Intersecte una curva de Bézier en el plano con una línea. Cuente el número de intersecciones entre la línea y la curva  $N_{lc}$  y entre la línea y el polígono de control  $N_{lp}$ . Esta propiedad asegura que  $N_{lc} \leq N_{lp}$ . Por ejemplo, si el polígono de control oscila  $w$  veces, entonces la curva no oscilará más que eso (Mann, 2006).

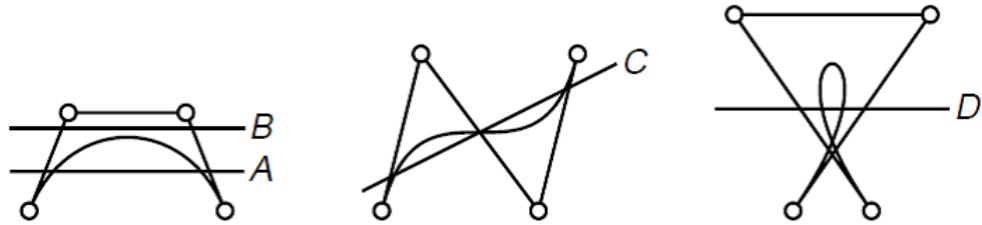


Figura A.5: Líneas que intersectan la curva de Bézier y sus puntos de control.

Otra alternativa no paramétrica, quizás más conocidas que las curvas de Bézier, son los splines, funciones polinomiales por tramos. Dentro de un contexto probabilístico, en <http://vita.had.co.nz/papers/density-estimation.pdf> se muestra una comparación de implementaciones con el enfoque de splines en el software estadístico R.



# Apéndice **B**

## Software

### B.1. Introducción

Para poder llevar a cabo el objetivo de este informe se pretende primero reproducir los resultados del artículo de Carnicero (2013). Esto con el fin de probar los programas generados por el estudiante del doctorado y así, posteriormente aplicarlos en el objetivo de este trabajo de investigación. Procediendo de esta manera, se descargaron de internet (<http://infomet.am.ub.es/clima/gijon/>) los datos circulares-lineales del artículo de Carnicero los cuales corresponden a datos de dirección del viento y precipitación del 06/Noviembre/1994–31/Enero/2009 de una estación meteorológica en Somió (España). Posteriormente, para verificar la validez de los datos se realizó el mismo análisis exploratorio que se muestra en el artículo con el fin de confirmar que los datos son los mismos. Las direcciones del viento están dadas en grados (0°-359°).

### B.2. Validación del software

A continuación se muestran los gráficos análogos a los presentados en dicho artículo.

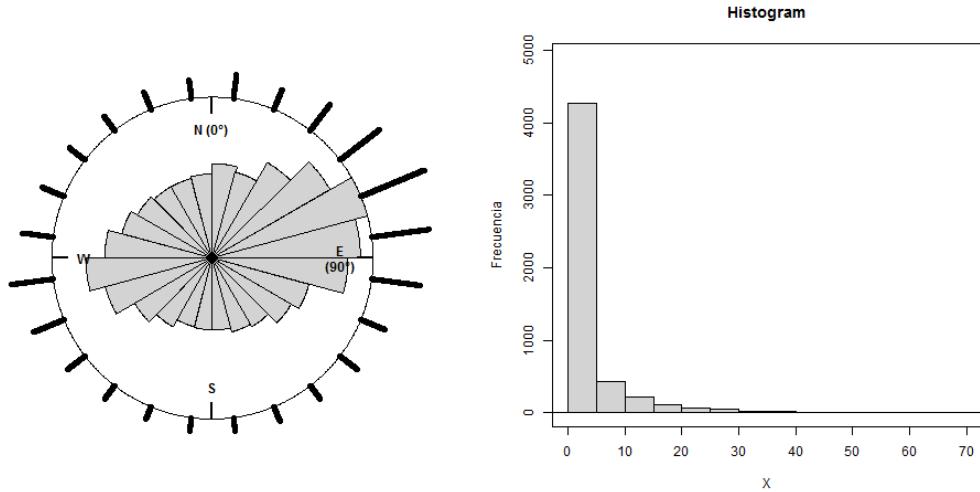


Figura B.1: (izquierda) Roseta de orientaciones, (derecha) histograma de precipitación.

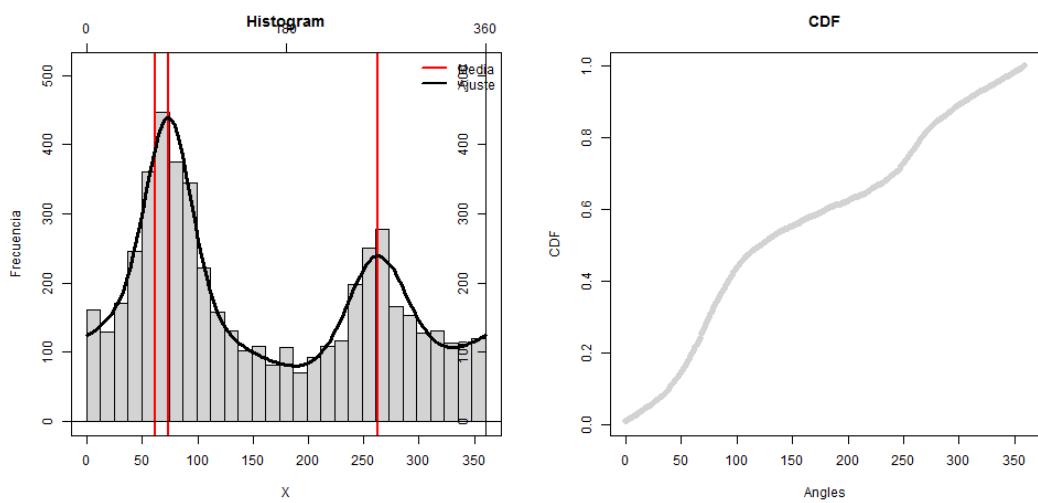


Figura B.2: Figuras equivalentes a la Fig. 2 del artículo. (recuadros superiores) Densidad marginal estimada, (recuadros inferiores) funciones de distribución acumulada, para la dirección del viento (izquierda) y para la precipitación pluvial (derecha)

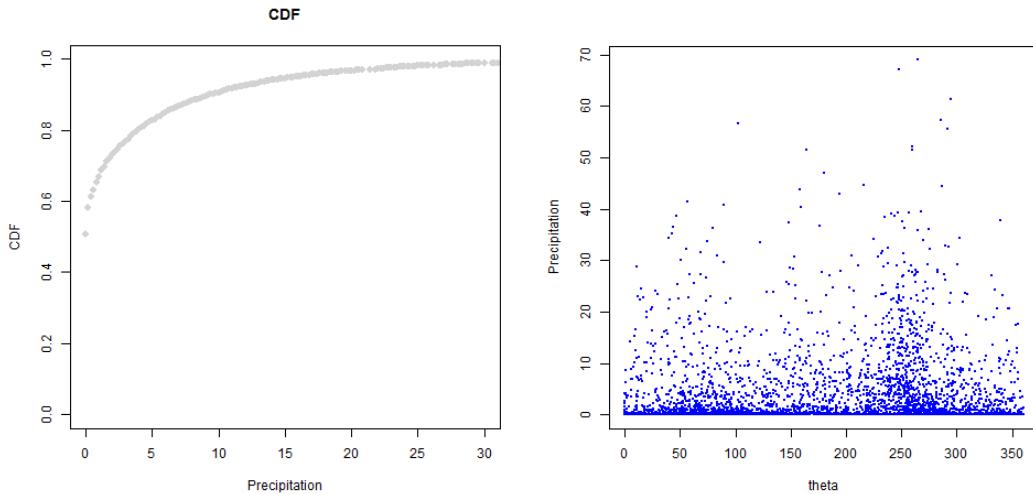


Figura B.3: Figura equivalentes a la Fig. 4 del artículo. Diagrama de dispersión de precipitación y dirección del viento (theta).

A continuación se procede a implementar los modelos matemáticos (Carnicero, 2013) de cópulas no paramétricas para datos circulares-lineales. Estos modelos fueron implementados en el paquete estadístico R.

La densidad de cópula obtenida se muestra en la siguiente figura:

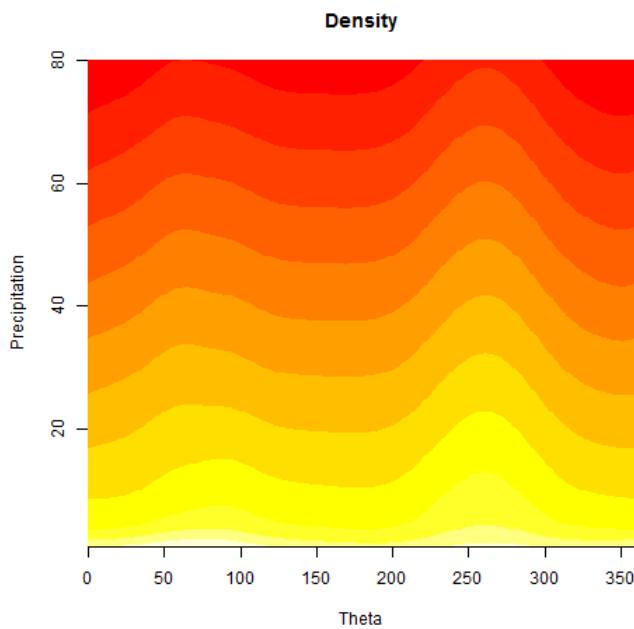


Figura B.4: Densidad de cópula.

Una vez reproducidos los datos del trabajo de Carnicero, se sabe que los programas computacionales funcionan correctamente. El siguiente paso sería implementar la simulación de dicho modelo de cópula.

Dado un modelo de dependencia (cópula), uno de los objetivos perseguidos al modelar la función de densidad de probabilidad conjunta es simular valores de las variables aleatorias, lo cual se llevará a cabo utilizando la metodología (Erdely-Ruiz & Díaz-Viera, 2012) de la sección denominada *Simulación conjunta de variables aleatorias* de la *Unidad Teórica B*. Atendiendo al punto 1 de dicha metodología es necesario simular valores con una distribución uniforme en el intervalo  $[0, 1]$ , ya que estos son requeridos por la cópula. Esto se puede conseguir con la función *runif* del paquete *base*. Recuérdese que la cópula bivariada tiene dominio  $[0, 1]^2$ . Para el segundo paso se requieren a) la Cúpula; b) obtener una función inversa. Al terminar este segundo paso se cuenta con dos conjuntos de valores  $\{u_i\}$  y  $\{v_i\}$  que están distribuidos uniformemente pero que conservan la misma dependencia que los datos originales. Por último, para el tercer paso, los datos simulados,  $\{(x_i, y_i)\}$ , se obtienen utilizando funciones cuantiles no paramétricas de  $X$  y  $Y$  respectivamente. Estos funciones  $\tilde{Q}$  y  $\tilde{R}$  son estimadas mediante polinomios de Bernstein-Kantorovich (Muñoz-Pérez y Fernández-Palacín, 1987). La Figura 2 muestra un ejemplo de un conjunto de datos a los cuales se les modeló su función de distribución con los polinomios de Bernstein-Kantorovich. Por lo tanto  $(x_i, y_i) = (\tilde{Q}(u_i), \tilde{R}(v_i))$ . La implementación computacional de estas funciones cuantiles se encuentra en el paquete *lmomco* bajo el nombre de *dat2bernqua*.

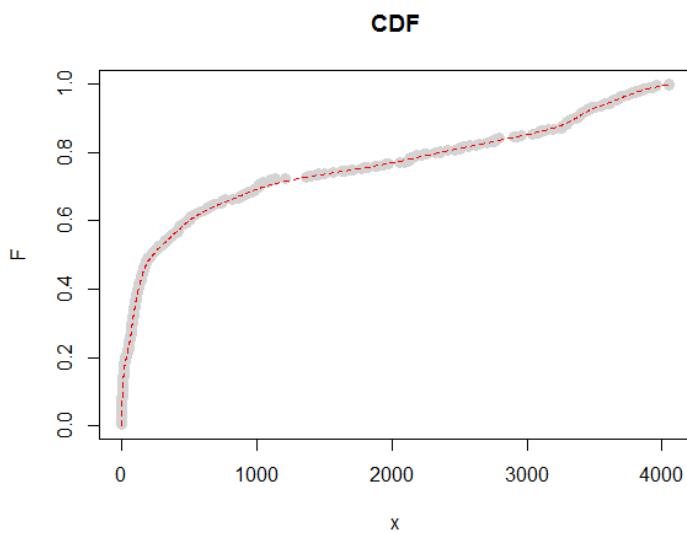


Figura B.5: Función de densidad de probabilidad empírica y ajustada con el enfoque no paramétrico utilizando polinomios de Bernstein-Kantorovich. En gris se muestran los datos observados y la línea es el ajuste.

A continuación se muestran histogramas de valores simulados con el enfoque de polinomios de Bernstein-Kantorovich.

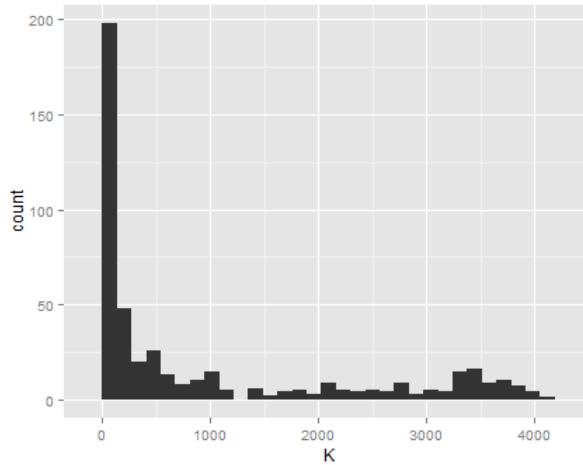


Figura B.6: Histogramas de ejemplo de simulación no paramétrica utilizando polinomios de Bernstein-Kantorovich para modelar la función de distribución de probabilidad. 473 datos (izquierda) y 100 datos simulados (derecha).

No solamente se muestra la versatilidad del modelo si no también se observó una tendencia en el tiempo de cómputo.

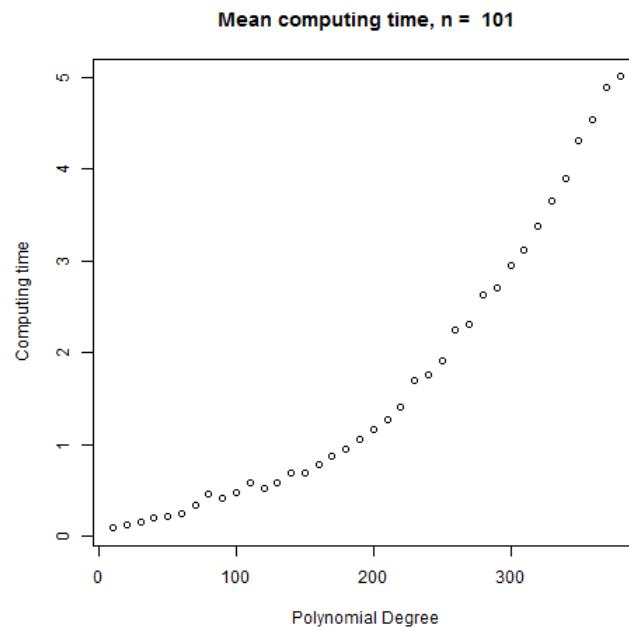


Figura B.7: Tiempo de cómputo de simulaciones de la cópula de Bernstein en función del grado del polinomio.

Además, se detectó casi nula dependencia de las simulaciones con respecto al grado del polinomio:

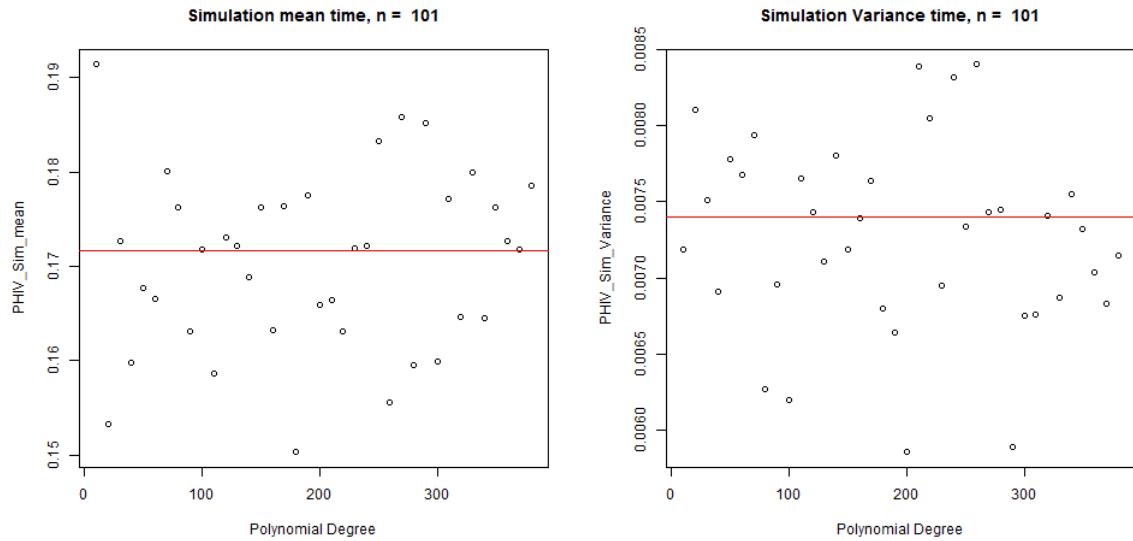


Figura B.8: Media (recuadro izquierdo) y varianza (recuadro derecho) en función del grado del polinomio de Bernstein de datos simulados.

En todos los casos se utilizó el mismo paquete desarrollado por los autores de este trabajo. La computadora utilizada es una laptop Lenovo ideapad z475, con 4 núcleos y 4 Gb de memoria RAM.

## Bibliografía

- AAS, K., CZADO, C., FRIGESSI, A., Y BAKKEN, H. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics* **44**(2):182–198 (2009) 64
- ADLER, P.M. Y THOVERT, J.F. *Fractures and Fracture Networks, Theory and Applications of Transport in Porous Media*, tomo 15. Springer Netherlands, Dordrecht (1999) 1, 20
- AGOSTINELLI, C. Y LUND, U. R package circular: Circular Statistics (version 0.4-7) (2013) 79
- ASQUITH, W. lmomco: L-Moments, Censored L-Moments, Trimmed L-Moments, L-Comoments, and Many Distributions (2017) 40
- BALANKIN, A., MORALES-MATAMOROS, D., CASTREJON, F., PACHECO, C., ZALDIVAR, J., Y VELASQUILLO-MARTÍNEZ, L.G. Distribution of Fracture Apertures in Rocks (2001) 2
- BÁRDOSSY, A. Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research* **42**(11):W11416 (2006) 2
- BERGHAUS, B., BUCHER, A., Y VOLGUSHEV, S. Weak convergence of the empirical copula process with respect to weighted metrics. *Bernoulli* **23**(1):743–772 (2017) 2, 47
- BONNET, E., BOUR, O., ODLING, N.E., DAVY, P., MAIN, I., COWIE, P., Y BERKOWITZ, B. Scaling of fracture systems in geological media. *Reviews of Geophysics* **39**(3):347 (2001) 40
- BOUR, O. Y DAVY, P. Connectivity of random fault networks following a power law fault length distribution. *Water Resources Research* **33**(7):1567–1583 (1997) 40
- BOURBIAUX, B., BASQUET, R., CACAS, M.C., DANIEL, J.M., Y SARDA, S. An Integrated Workflow to Account for Multi-Scale Fractures in Reservoir Simulation Models: Implementation and Benefits. En *SPE-78489-MS*, págs. 1–16. Society of Petroleum Engineers, SPE (2002) 1
- BOX, M.J. Bias in Nonlinear Estimation. *Journal of the Royal Statistical Society. Series B (Methodological)* **33**(2):171–201 (1971) 2

- BUCHER, A. Y VOLGUSHEV, S. Empirical and sequential empirical copula processes under serial dependence. *arXiv:1111.2778 [math, stat]* (2011). ArXiv: 1111.2778 2, 47
- CACAS, M.C., DANIEL, J.M., Y LETOUZEY, J. Nested geological modelling of naturally fractured reservoirs. *Petroleum Geoscience* 7(S):S43–S52 (2001). <http://pg.lyellcollection.org/content/7/S/S43.full.pdf+html> 1
- CARNICERO, J.A., AUSÍN, M.C., Y WIPER, M.P. Non-parametric copulas for circular-linear and circular-circular data: an application to wind directions. *Stochastic Environmental Research and Risk Assessment* 27(8):1991–2002 (2013) x, 2, 47, 55, 57, 58, 59, 60, 61, 95
- CASELLA, G. Y BERGER, R.L. *Statistical inference*. 2<sup>a</sup> edición. DUXBURY ADVANCED SERIES (2002) 37
- CHERNIH, A., MAJ, M., Y VANDUFFEL, S. Beyond Correlations: The Use and Abuse of Copulas in Economic Capital Calculations. *Belgian Actuarial Bulletin* 7(1) (2007) 2
- CHILÈS, J.P. Y DELFINER, P. *Geostatistics: modeling spatial uncertainty*, tomo 497. John Wiley & Sons (2012) 1
- DAHLQUIST, G. Y BJÖRCK, A. *Numerical Methods in Scientific Computing: Volume 1*. Society for Industrial and Applied Mathematics, Philadelphia (2008) 40
- DAVIS, G.H., REYNOLDS, S.J., Y KLUTH, C.F. *Structural Geology of Rocks and Regions*. 3<sup>a</sup> edición. Wiley, Hoboken (2011) 78
- DE VILLIERS, J. *Mathematics of Approximation*. Atlantis Press, Paris (2012) 100
- DEHEUVELS, P. La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. *Acad. Roy. Belg. Bull. Cl. Sci.* (5), 65(6), págs. 274–292 (1979) 2, 47, 65
- DÍAZ-VIERA, A.E.Â.M. A vine and gluing copula model for permeability stochastic simulation. En *16th ASMDA Conference Proceedings* (2015) 63
- DÍAZ-VIERA, M.A. SPATIAL STOCHASTIC MODELS AS A POWERFUL TOOL FOR UNCERTAINTY MODELING IN HETEROGENEOUS RESERVOIR FORMATIONS. *Geological Society of America Abstracts with Programs*, 44(3):20 (2012) 45
- DÍAZ-VIERA, M.A. Y ANGUIANO-ROJAS, P. Stochastic modeling of permeability in naturally fractured carbonate reservoirs applying a Monte-Carlo simulation method with t-copulas. *SPWLA 47th Annual Logging Symposium* (2006) 45
- DÍAZ-VIERA, M.A. Y CASAR-GONZÁLEZ, R. Stochastic simulation of complex dependency patterns of petrophysical properties using t-copulas. En *Proceedings of IAMG05: GIS and Spatial Analysis*, tomo 2, págs. 749–755 (2005a) 2, 45

- DÍAZ-VIERA, M.A. Y CASAR-GONZÁLEZ, R. Stochastic simulation of complex dependency patterns of petrophysical properties using t-copulas. *Proceedings of IAMG05: GIS and Spatial Analysis* **2**:749–755 (2005b) 45
- DÍAZ-VIERA, M.A., ANGUIANO-ROJAS, P., MOUSATOV, A., KAZATCHENKO, E., Y MARKOV, M. Stochastic modeling of permeability in double porosity carbonates applying a monte-carlo simulation method with t-copulas. En *SPWLA 47th Annual Logging Symposium*, págs. 1–15 (2006) 45
- DIETRICH, P., HELMIG, R., SAUTER, M., HÖTZL, H., KÖNGETER, J., Y TEUTSCH, G. *Flow and Transport in Fractured Porous Media*. Springer-Verlag, Berlin, Heidelberg (2005) 20
- DISSMANN, J.F. *Statistical Inference for Regular Vines and Application*. Diplomarbeit, Technische Universität München (2010) 63, 64
- DOWD, P.A., XU, C., MARDIA, K.V., Y FOWELL, R.J. A Comparison of Methods for the Stochastic Simulation of Rock Fractures. *Mathematical Geology* **39**(7):697–714 (2007) 16
- ELMO, D., ROGERS, S., STEAD, D., Y EBERHARDT, E. Discrete Fracture Network approach to characterise rock mass fragmentation and implications for geomechanical upscaling. *Mining Technology* **123**(3):149–161 (2014) 1
- EMBRECHTS, P. Y HOFERT, M. A note on generalized inverses. *Mathematical Methods of Operations Research* **77**(3):423–432 (2013) 40
- EMBRECHTS, P., MCNEIL, A., Y STRAUMANN, D. Correlation: Pitfalls and Alternatives. *Risk Magazine* **12**(5):69–71 (1999) 2
- ERDELY, A. Y DÍAZ-VIERA, M.A. Nonparametric bivariate modeling of petrophysical porosity-permeability well log data. *Workshop on Copula Theory and Its Applications Preliminary Book of Abstracts Posters August 22, 2009* pág. 1 (2009) 2, 45, 62
- ERDELY, A. Y DÍAZ-VIERA, M.A. Modeling porosity-permeability data by gluing copulas. *AGU Fall Meeting Abstracts* **1**:1038 (2011) 45
- ERDELY, A. Y DÍAZ-VIERA, M.A. Joint Porosity-permeability Stochastic Simulation and Spatial Median Regression by Nonparametric Copula Modeling. En *GI\_Forum 2012: Geovizualisation, Society and Learning*, págs. 346 – 354. Herbert Wichmann Verlag, VDE VERLAG GMBH, Berlin/Offenbach, Wichman, Berlin (2012) x, 45, 51, 62
- ERDELY, A., DÍAZ-VIERA, M.A., Y HERNÁNDEZ-MALDONADO, V. Trivariate nonparametric dependence modeling of petrophysical properties. *Mathematical and Numerical Modeling in Porous Media: Applications in Geosciences* pág. 195 (2012) 45

- EXCHANGE, C.S.S. solver - Best method to find the zero of a decreasing function numerically (2017) 41
- FARRELL, J. *Programming Logic and Design, Comprehensive.* 8<sup>a</sup> edición. Course Technology, Stamford, CT (2014) 69
- FERMANIAN, J.D., RADULOVIC, D., Y WEGKAMP, M. Weak convergence of empirical copula processes. *Bernoulli* **10**(5):847–860 (2004) 2, 47
- FERNÁNDEZ-DURÁN, J.J. Modelling Ground-Level Ozone Concentration Using Copulas. tomo 707, págs. 406–416. AIP Publishing (2004) 57
- FISHER, N.I. *Statistical Analysis of Circular Data.* Cambridge University Press, Cambridge England; New York (1995) 89
- FISHER, N.I., LEWIS, T., Y EMBLETON, B.J.J. *Statistical Analysis of Spherical Data.* Cambridge University Press (1993) XIII, 41, 42, 43
- FOSSEN, H. *Structural Geology.* 1<sup>a</sup> edición. Cambridge University Press, Cambridge ; New York (2010) 77
- GIJBELS, I., VERAVERBEKE, N., Y OMELKA, M. Conditional copulas, association measures and their applications. *Comput. Stat. Data Anal.* **55**(5):1919–1932 (2011) 64
- GOLDMAN, R. *Pyramid algorithms: A dynamic programming approach to curves and surfaces for geometric modeling.* Morgan Kaufmann (2002) 51, 100
- GRINGARTEN, E. *Geometric modeling of fracture networks.* Tesis Doctoral, Stanford University (1997) 1
- GUDMUNDSSON, A., MOHAJERI, N., Y BARNETT, Z. Power-law size distributions in volcano-tectonics. *Geophysical Research Abstracts* **13** (2011) 40
- HASLAUER, C.P., LI, J., Y BÁRDOSSY, A. Application of Copulas in Geostatistics. En P.M. Atkinson y C.D. Lloyd (editores), *geoENV VII – Geostatistics for Environmental Applications*, número 16 en Quantitative Geology and Geostatistics, págs. 395–404. Springer Netherlands (2010). DOI: 10.1007/978-90-481-2322-3\_34 2
- HERNÁNDEZ-MALDONADO, V., DÍAZ-VIERA, M.A., Y ERDELY, A. Joint porosity-permeability stochastic simulation by non-parametric copulas. *CRC/Press Balkema Taylor & Francis Group*:Ch. 11, 209 (2012a) 45
- HERNÁNDEZ-MALDONADO, V., DÍAZ-VIERA, M.A., Y ERDELY, A. A joint stochastic simulation method using the Bernstein copula as a flexible tool for modeling nonlinear dependence structures between petrophysical properties. *Journal of Petroleum Science and Technology* **90**:112–123 (2012b) 3, 45

- HERNÁNDEZ-MALDONADO, V., DÍAZ-VIERA, M.A., Y ERDELY, A. A trivariate Bernstein copula for joint stochastic simulation of petrophysical properties. *Petroleum Science and Technology* (2012c) 3, 45
- HERNÁNDEZ-MALDONADO, V.M. Y DÍAZ-VIERA, M.A. *Simulación estocástica espacial de propiedades petrofísicas usando cópulas de Bernstein*. Ph. D., Instituto Mexicano del Petróleo, Mexico City (2014) 45
- JAMMALAMADAKA, S.R. Y SENGUPTA, A. *Topics in circular statistics, Series on Multivariate Analysis*, tomo 5. World Scientific, River Edge, N.J. (2001) XIII, 41, 43, 89
- JOE, H. *Dependence Modeling with Copulas*. 1<sup>a</sup> edición. Chapman and Hall/CRC, Boca Raton (2014) 46, 64
- JOHNSON, R.A. Y WEHRLY, T.E. Some Angular-Linear Distributions and Related Regression Models. *Journal of the American Statistical Association* **73**(363):602–606 (1978) 56
- KARNIADAKIS, G. Y KIRBY, R.M. *Parallel scientific computing in C++ and MPI: a seamless approach to parallel algorithms and their implementation*. Cambridge University Press, New York (2003) 25, 41
- KAT, H.M. The Dangers of Using Correlation to Measure Dependence. *The Journal of Alternative Investments* **6**(2):54–58 (2003) 2
- KAUERMANN, G. Y SCHELLHASE, C. Flexible pair-copula estimation in D-vines using bivariate penalized splines. *Statistics and Computing* **24**(6):1–20 (2013) 56
- KAZIANKA, H. spatialCopula: A Matlab toolbox for copula-based spatial analysis. *Stochastic Environmental Research and Risk Assessment* **27**(1):121–135 (2012) 2
- KAZIANKA, H. Y PILZ, J. Spatial Interpolation Using Copula-Based Geostatistical Models. En P.M. Atkinson y C.D. Lloyd (editores), *geoENV VII – Geostatistics for Environmental Applications*, número 16 en Quantitative Geology and Geostatistics, págs. 307–319. Springer Netherlands (2010). DOI: 10.1007/978-90-481-2322-3\_27 2
- KAZIANKA, H. Y PILZ, J. Bayesian spatial modeling and interpolation using copulas. *Geoinformatics for Environmental Surveillance* **37**(3):310–319 (2011) 2
- KING, G. How not to lie with statistics: Avoiding common mistakes in quantitative political science. *American Journal of Political Science* págs. 666–687 (1986) 2
- LANTUÉJOUL, C. *Geostatistical simulation : models and algorithms*. Springer, Berlin ; New York (2002) 23, 24

- MAI, J.F. Y SCHERER, M. *Simulating copulas: stochastic models, sampling algorithms and applications*, tomo 4. World Scientific (2012) 63, 64
- MANN, S. A Blossoming development of splines. *Synthesis Lectures on Computer Graphics and Animation* **1**(1):1–108 (2006) 51, 100, 103
- MARDIA, K.V. Y JUPP, P.E. *Directional statistics*. Wiley, Chichester (2000) XIII, 42, 43, 89
- MATHOVERFLOW. na.numerical analysis - Fast root finding for strictly decreasing function (2017) 41
- MATLOFF, N.S. *Parallel computing for data science: with examples in R, C++ and CUDA*. Número 28 en Chapman & Hall/CRC: The R Series. CRC Press, Boca Raton (2016). OCLC: ocn913072956 25
- MENDOZA-TORRES, F., DÍAZ-VIERA, M.A., Y ERDELY, A. Bernstein copula modeling for 2D discrete fracture network simulations. *Journal of Petroleum Science and Engineering* **156**:710–720 (2017) 1, 77
- MILLER, D.M. Reducing Transformation Bias in Curve Fitting. *The American Statistician* **38**(2):124–126 (1984) 2
- MIRANKER, W.L. A Survey of Parallelism in Numerical Analysis. *SIAM Review* **13**(4):524–547 (1971) 41
- MUÑOZ-PÉREZ, J. Y FERNÁNDEZ-PALACÍN, A. Estimating the quantile function by Bernstein polynomials. *Computational Statistics & Data Analysis* **5**(4):391–397 (1987) 40, 62, 101, 108
- NELSEN, R.B. *An Introduction to Copulas*. Springer Series in Statistics. Springer New York, New York, NY (2006) 62
- NELSON, R.A. *Geologic analysis of naturally fractured reservoirs*. Gulf Professional Pub., Boston (2001) 1
- NIJMEIJER, M.J.P. A parallel root-finding algorithm. *LMS Journal of Computation and Mathematics* **18**(1):713–729 (2015) 41
- OLSON, J.E. Fracture aperture, length and pattern geometry development under biaxial loading: a numerical study with applications to natural, cross-jointed systems. *Geological Society, London, Special Publications* **289**(1):123–142 (2007) 2
- OVERFLOW, S. algorithm - Best parallel method for finding roots of a numerical function (2017) 41
- PHILLIPS, G.M. *Interpolation and Approximation by Polynomials*. CMS Books in Mathematics. Springer New York, New York, NY (2003) 54, 56, 93, 100, 101, 102

- QUARTERONI, A., SACCO, R., Y SALERI, F. *Numerical Mathematics*. 2<sup>a</sup> edición. Springer, Berlin ; New York (2006) 40
- RADULOVIC, D., WEGKAMP, M., Y ZHAO, Y. Weak convergence of empirical copula processes indexed by functions. *Bernoulli* **23**(4B):3346–3384 (2017) 2, 47
- RENTELN, P. *Manifolds, tensors, and forms: an introduction for mathematicians and physicists*. Cambridge University Press, Cambridge, UK ; New York (2014) 51
- ROBERT, C.P. Y CASELLA, G. *Introducing Monte Carlo Methods with R*. 2010<sup>a</sup> edición. Springer Verlag, New York (2009) 63
- ROMANO-PEREZ, C.A. *Modelos de Fracturas Discretas para la Simulación de Flujo Monofásico en Medios Porosos Fracturados*. Proyecto Fin de Carrera, Universidad Nacional Autónoma de México (2016) 16
- SALVADORI, G., DE MICHELE, C., KOTTEGODA, N.T., Y ROSSO, R. *Extremes In Nature: An Approach Using Copulas*. Springer (2007) 2, 39
- SANCETTA, A. Y SATCHELL, S. The Bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric Theory* **20**(3):535–562 (2004) 2, 51, 54, 56, 57
- SCHELLHASE, C. Y KAUERMANN, G. Density estimation and comparison with a penalized mixture approach. *Computational Statistics* **27**(4):757–777 (2012) 60, 97
- SCOTT, D.W. *Multivariate Density Estimation: Theory, Practice, and Visualization*. 1<sup>a</sup> edición. John Wiley & Sons, New York (1992) 55
- SEBER, G.A.F. Y WILD, C.J. *Nonlinear regression*. Wiley series in probability and statistics. Wiley-Interscience, Hoboken, N.J (2003) 2
- SINGHAL, B. Y GUPTA, R. *Applied Hydrogeology of Fractured Rocks*. 2<sup>a</sup> edición. Springer Netherlands, Dordrecht (2010) 77
- SKLAR, A. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris* **8**:229–231 (1959) 2, 45, 46, 57
- STOYAN, D., KENDALL, W.S., Y MECKE, J. *Stochastic geometry and its applications*. Wiley series in probability and mathematical statistics. Wiley, Chichester, W. Sussex, New York (1987). Rev. translation of: Stochastische Geometrie 1
- TU, R. A Study of the Parametric and Nonparametric Linear-Circular Correlation Coefficient. *Digital Commons Statistics* (2015) 82

- VENIT, S. Y DRAKE, E. *Prelude to Programming*. 6<sup>a</sup> edición. Pearson, Boston (2014) 69
- WEHRLY, T.E. Y JOHNSON, R.A. Bivariate models for dependence of angular observations and a related Markov process. *Biometrika* **67**(1):255–256 (1980) 56
- WEI, G. Y SCHEFFER, M. Smooth Nonparametric Bernstein Vine Copulas. *arXiv:1210.2043 [q-fin]* (2012). ArXiv: 1210.2043 93
- WIKIPEDIA ROOTFINDING2016. Root-finding algorithm (2016). Page Version ID: 752433885 41
- ZELLOU, A., HARTLEY, L., HOOGERDUIJN-STRATING, E., AL DHAHAB, S., BOOM, W., Y HADRAMI, F. Integrated Workflow Applied to the Characterization of a Carbonate Fractured Reservoir: Qarn Alam Field. En *SPE-81579-MS*, págs. 1 – 8. Society of Petroleum Engineers, SPE (2003) 1