

*Cóputas: un nuevo enfoque para la modelización de
campos aleatorios*

DANNA LESLEY CRUZ REYES

CÓDIGO: 01832419



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, D.C.
MAYO DE 2013

*Cóputas: un nuevo enfoque para la modelización de
campos aleatorios*

DANNA LESLEY CRUZ REYES

CÓDIGO: 01832419

TRABAJO DE TESIS PARA OPTAR AL TÍTULO DE
MAGISTER EN ESTADÍSTICA

DIRECTOR
EDILBERTO CEPEDA CUERVO, PH.D
PH.D EN MATEMÁTICAS - PROFESOR ASOCIADO DEPARTAMENTO DE
ESTADÍSTICA.



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, D.C.
MAYO DE 2013

Título en español

Cópulas: un nuevo enfoque para la modelización de campos aleatorios

Title in English

Cópulas, a new approach to the modeling of random fields

Resumen: El estudio de fenómenos espacio-temporales que fluctúan aleatoriamente se modelan por medio de métodos geoestadísticos, desarrollados originalmente para predecir la distribución de probabilidad para operaciones mineras, (Giraldo, 2010). Algunos de estos métodos son: el variograma, que describe la estructura de dependencia espacial bajo la media, y el Kriging, como la metodología de predicción espacial. Ambos métodos son sensibles a observaciones atípicas y están fuertemente influenciadas por la distribución marginal del campo aleatorio subyacente. En este trabajo, como una alternativa a la modelización tradicional de campos aleatorios, se propone el uso de funciones cópula. En la primera etapa, se realiza el análisis exploratorio de datos, buscando identificar la forma de la estructura de dependencia espacial. Por ello, se presenta la cópula empírica, que permite identificar la dependencia mediante el cálculo de rangos y la distribución empírica. Posteriormente, se define la distribución multivariada del campo en términos de sus marginales con ayuda del teorema de Sklar, formando una función de distribución alterna para el campo. Dicha distribución quedará expresada en términos de la distancia entre los lugares de observación. En la tercera etapa, se describen los procedimientos empleados para identificar la cópula que mejor se aproxima a la cópula empírica, la cual se denomina cópula óptima, seleccionada a partir de un conjunto de cópulas, llamadas cópulas elípticas, usando el estadístico desarrollado por (Genest, 2009) centrado en un blanket test. Una vez detectada la cópula que corresponderá al modelo de dependencia espacial, se realiza predicción, aplicando los procedimientos propuestos por (Kazianka, 2010) y (Kazianka, 2012), llamados kriging indicador, disyuntivo y trans-gaussiano, todos estos se aplican usando el modelo cópula.

Abstract: The study of spatiotemporal phenomena that fluctuate randomly modeled using geostatistical methods, originally developed to predict the probability distribution for mining operations (Giraldo, 2010). Some of these methods are: the variogram, which describes the spatial dependence structure under the average, and the Kriging, such as spatial prediction methodology. Both methods are sensitive to outliers and are strongly influenced by the marginal distribution of the underlying random field. In this work, as an alternative to traditional random field modeling, we propose the use of mating features. In the first stage, we performed exploratory data analysis, seeking to identify the form of the spatial dependence structure. Therefore, we present the empirical copula, which identifies the dependence by calculating ranges and the empirical distribution. Subsequently, it defines the multivariate distribution of the field in terms of its marginal theorem using the Sklar, forming alternating distribution function for the field. This distribution will be expressed in terms of the distance between the observation points. In the third stage, we describe the procedures used to identify the copula that best approximates the empirical copula, which is called optimal intercourse, selected from a set of copulation, copulation calls elliptical, using the statistic developed by (Genest, 2009) focused on a "blanket test". Once detected the copula which corresponds to the spatial dependence model, prediction is performed by applying the procedures proposed by (Kazianka, 2010) and

(Kazianka, 2012), called indicator kriging, disjunctive and trans-Gaussian, these are applied using the copula model.

Palabras clave: Cópulas, geoestadística, estadística espacial, tendencia, modelos Gaussianos.

Keywords: Copula, geostatistics, spatial statistics, trend, Gaussian model.

Nota de aceptación

Trabajo de tesis

Aprobado

“Mención Meritoria o Laureada”

Jurado

Jurado

Jurado

Director

Codirector

Bogotá, D.C.,

Agradecimientos

Doy gracias al profesor José Alberto Vargas, PH.D, quien fue el primero en confiar en mí para realizar mi maestría, al departamento de estadística de la Universidad Nacional de Colombia, a mi Alma máter, el proyecto de matemáticas de la Universidad Distrital Francisco José de Caldas. Y le agradezco mucho a mi directo de tesis Edilberto Cepeda Cuervo PH.D, por toda su paciencia y su gran colaboración.

Índice general

| | |
|---|------------|
| Índice general | I |
| Índice de tablas | III |
| Índice de figuras | IV |
| Introducción | VI |
| 1. Introducción a Cópulas | 1 |
| 1.1. Definición de Cópula | 1 |
| 1.2. Cotas de Fréchet Hoeffding | 3 |
| 1.3. Familia de Cópulas Arquimedianas | 3 |
| 1.3.1. Construcción de las cotas de Fréchet-Hoeffding | 5 |
| 1.4. Familia de cópulas elípticas | 5 |
| 1.4.1. Cópula Gaussiana | 5 |
| 1.4.2. t-cópula | 6 |
| 1.5. Medidas de dependencia y cópulas | 6 |
| 1.6. Estimación de los parámetros de las cópulas | 7 |
| 1.6.1. Estimación paramétrica | 7 |
| 1.6.2. Estimación semiparamétrica | 8 |
| 1.6.3. Estimación no paramétrica | 9 |
| 2. Geoestadística | 10 |
| 2.1. Campo aleatorio | 10 |
| 2.2. Estacionariedad | 11 |
| 2.3. Modelos Gaussianos | 12 |
| 2.3.1. Familia Matérn | 13 |

| | |
|---|-----------|
| 2.3.2. Anisotropía geométrica | 13 |
| 3. Geoestadística usando cópulas | 14 |
| 3.1. Descripción del campo aleatorio usando cópulas | 14 |
| 3.2. Cópulas empíricas bivariadas | 16 |
| 3.2.1. Algoritmo para la aplicación de las cópulas empíricas bivariadas | 16 |
| 3.3. Medida de asimetría | 19 |
| 3.4. Tendencia espacial | 21 |
| 4. Familia de cópulas no gaussianas | 22 |
| 4.1. Cópula χ^2 | 23 |
| 4.2. Cópula V - transformada normal | 25 |
| 5. Estimación de parámetros | 29 |
| 5.1. Inferencia para marginales continuas | 29 |
| 5.2. Selección de la mejor cópula | 32 |
| 6. Interpolación espacial | 34 |
| 6.1. Kriging Indicador y Kriging disyuntivo | 34 |
| 6.2. Plug-in de estimación | 35 |
| 7. Aplicación de cópulas: Catástrofe en Chernóbil | 37 |
| 7.1. Datos de Chernóbil | 37 |
| 7.2. Cópula Gaussiana | 38 |
| 7.3. Cópula no Gaussiana: V -transformada normal | 42 |
| 7.4. Cópula Gaussiana incluyendo anisotropía | 45 |
| 7.5. Cópula no Gaussiana incluyendo anisotropía | 46 |
| 7.6. Cópula Gaussiana con tendencia | 47 |
| 7.7. Resumen | 48 |
| 8. Cópulas Discretas | 50 |
| 8.1. Existencia y Unicidad | 50 |
| 8.2. Inferencia para cópulas con marginales discretas en Geoestadística | 52 |
| Conclusiones | 57 |
| Bibliografía | 58 |

Índice de tablas

| | |
|---|----|
| 1.1. Algunas cópulas Arquimedianas y sus generadores. | 5 |
| 3.1. Resumen estadístico de los datos de Gomel | 17 |
| 7.1. Valores de los estadísticos T_n y S_n | 39 |
| 7.2. Valor p los estadísticos T_n y S_n | 40 |
| 7.3. Valores de los estadísticos T_n y S_n | 42 |
| 7.4. Valor p de los estadísticos T_n y S_n | 42 |
| 7.5. Valores de MSE y MAE para modelos de dependencia | 49 |

Índice de figuras

| | | |
|-------|--|----|
| 3.1. | Observaciones de los datos de la ciudad Gomel | 18 |
| 3.2. | Gráfica de contorno con las localizaciones y los valores de las realizaciones | 18 |
| 3.3. | Densidades de las cópulas empíricas bivariadas | 19 |
| 3.4. | Rango de correlación | 20 |
| 4.1. | Transformaciones V -Normal | 26 |
| 4.2. | Transformaciones V -Normal con parámetros: $m = 0$, $k = 1$ y $\alpha = 1$ (Esquina superior izquierda), $m = 0$, $k = 3$ y $\alpha = 1$ (Esquina superior derecha), $m = 0$, $k = 2,5$ y $\alpha = 0,5$ (Esquina inferior izquierda) y $m = 1$, $k = 2,5$ y $\alpha = 0,5$ (Esquina inferior izquierda) | 27 |
| 4.3. | Cópula V -transformada Normal | 28 |
| 7.1. | Observaciones de los datos de Gomel | 38 |
| 7.2. | Observaciones y datos interpolados de los datos de Gomel | 39 |
| 7.3. | Gráfico de dispersión de los pares de datos de rango transformados, es decir, densidad de la cópula empírica bivariado (columna de la izquierda), cópula teórica bivariada Gaussiana (columna de la derecha) | 40 |
| 7.4. | Media de la predicción con la cópula Gaussiana | 41 |
| 7.5. | Desviación estándar de la predicción con la cópula Gaussiana | 41 |
| 7.6. | Ancho de los intervalos de confianza de validación cruzada para la cópula Gaussiana | 41 |
| 7.7. | Valores de validación cruzada PIT para la cópula Gaussiana | 42 |
| 7.8. | Gráfico de dispersión de los pares de datos de rango transformados, es decir, densidad de la cópula empírica bivariado (columna de la izquierda), cópula teórica bivariada χ^2 -cópula (columna de la derecha) | 43 |
| 7.9. | Media de la predicción con la cópula V | 44 |
| 7.10. | Desviación estándar de la predicción con la cópula V | 44 |
| 7.11. | Gomel datos: ancho de los intervalos de confianza de validación cruzada para la cópula χ^2 | 45 |

| | |
|---|----|
| 7.12. Valores de validación cruzada PIT para la cópula χ^2 | 45 |
| 7.13. Media de la predicción con la cópula Gaussiana con anisotropía. | 46 |
| 7.14. Desviación estándar de la predicción con la cópula Gaussiana con anisotropía. | 47 |
| 7.15. Media de la predicción con la cópula no Gaussiana con anisotropía. | 47 |
| 7.16. Desviación estándar de la predicción con la cópula no Gaussiana con anisotropía. | 48 |
| 8.1. Simulación proceso Gaussiano con marginales Poisson | 54 |
| 8.2. Cópula empírica para simulación proceso Gaussiano con marginales Poisson, $h = 0,15$ | 54 |
| 8.3. Cópula empírica para simulación proceso Gaussiano con marginales Poisson, $h = 0,19$ | 54 |
| 8.4. Cópula empírica para simulación proceso gaussiano con marginales Poisson, $h = 0,198$ | 55 |
| 8.5. Cópula empírica para simulación proceso gaussiano con marginales Poisson, $h = 0,21$ | 55 |
| 8.6. Cópula gaussiana con marginales Poisson | 56 |

Introducción

Las cópulas describen la estructura de dependencia entre variables aleatorias, no es extraño que la palabra cópula insinúa vínculo o unión, proviene del latín y su significado es conexión o lazo que une dos cosas distintas, fue utilizada por primera vez por Sklar en su célebre teorema en 1959 [Ayyad, 2008], para describir funciones de distribución multivariadas definidas sobre el cubo unidad $[0; 1]^n$ enlazando variables aleatorias con funciones de distribución de una sola dimensión.

Las cópulas han sido aplicadas en diferentes campos, por ejemplo, en actuaría en la cual diversos procesos en los seguro de accidentes involucran pares de variables correlacionadas; un ejemplo destacado es la pérdida y los gastos que se le asignan a una reclamación en una empresa de seguros, en estos casos los modelos de cópulas son ampliamente utilizados ya que puede ser computacionalmente fácil, genera medidas de asociación no paramétricas y el modelo es lo suficientemente flexible como para tener una posibilidad de ajustar los datos. [Klugman, 1999]. En el área financiera son utilizadas en el modelado de los activos y la gestión de riesgos,[Bouye, 2000] y [Wu, 2006]. En ingeniería, Sotirios P.Chatzis y Yiannis Demiris presentaron una nueva aplicación de la cópula, en su artículo “The copula echo state network”, en el cual se propone una nueva metodología para el modelado de datos secuenciales, basado en la postulación de las redes neuronales recurrentes y sistemas dinámicos, [Sotirios, 2011]. En particular, el desarrollo de esta tesis se enfocará al área de geoestadística con el desarrollo de una herramienta para modelizar campos aleatorios utilizando cópulas.

La geoestadística se concentra en el estudio de fenómenos que se desarrollan en el tiempo y el espacio mientras que pasan por fluctuaciones aleatorias, más formalmente, analiza las realizaciones de un campo aleatorio $\{Z(\mathbf{x}) : \mathbf{x} \in D\}$ donde $D \subset \mathbb{R}$, cuya realización, $z(\mathbf{x})$, representa el valor de interés registrado en la medición con respecto a cierto sistema de referencia \mathbf{x} .

Actualmente existen herramientas que para modelar la variabilidad espacial. La primera fue usada en principios de los cincuenta por Danie G. Krige, en Sudáfrica ampliando técnicas estadísticas para la estimación de las reservas de minerales. En los años sesenta el trabajo de Krige fue formalizado por el matemático Georges Matheron, desde entonces ha sido ampliamente utilizado en áreas como la minería, la industria petrolera, hidrología, meteorología, oceanografía, el control del medio ambiente, la ecología del paisaje y la agricultura. A pesar de estos desarrollos, el modelado espacial a menudo se basa en

hipótesis Gaussianas que muchas veces no se considera realista para los tipos de datos y/ó se reportan datos atípicos que causan problemas en las investigaciones.

Bárdossy en el año 2006 fue el pionero en proponer el uso de cópulas para describir variabilidad espacial, (Bárdossy et al., 2008) [Bárdossy, 2008] realiza modelación de campos aleatorios continuos, (Kazianka et al., 2010a) [Kazianka, 2010b], adopta la metodología de Bárdossy y realiza una extensión considerando modelos de tendencia y campos aleatorios discretos, (Kazianka et al., 2010b) [Kazianka, 2010a] muestran cómo se pueden incorporar en un marco bayesiano mediante la asignación de probabilidades apriori de todos los parámetros del modelo, en esta tesis se propone una extensión a estos modelos incluyendo las cópulas radialmente asimétricas.

Introducción a Cópulas

En la introducción se mencionó que la palabra cópula insinuaba un tipo de unión o lazo entre dos variables aleatorias y que dicha función podría describir una estructura de dependencia, estas afirmaciones se formaliza en lenguaje matemático en este capítulo. En primer lugar y siguiendo la definición de Nelsen [Nelsen, 1998], se presenta la definición de subcópulas para posteriormente definir la función cópula como una subcópula con dominio en el cubo unidad.

1.1. Definición de Cópula

Definición 1. (Adaptado de [Nelsen, 1998]). Una subcópula bivariada es una función C' tal que cumple las siguientes propiedades:

1. $DomC' = S_1 \times S_2$, donde S_1 y S_2 son subconjuntos de $[0, 1]$,
2. C' es decreciente,
3. para todo $(u, v) \in S_1 \times S_2$,
 - $C'(u, 1) = u$,
 - $C'(1, v) = v$,
 - $C'(1, 1) = 1$,
4. para todo $(u, v) \in S_1 \times S_2$, $0 \leq C'(u, v) \leq 1$. Así, el rango de C' es $[0, 1]$.

Definición 2. (Adaptado de [Nelsen, 1998]). Una cópula bivalente C es una subcópula cuyo dominio es $[0, 1] \times [0, 1]$ tal que cumple con las siguientes propiedades:

- Para cada u, v en $[0, 1]$, se cumple que:

$$C(u, 0) = 0 = C(0, v),$$

y,

$$C(u, 1) = u, \quad y, \quad C(1, v) = v.$$

- Para cada u_1, u_2, v_1, v_2 en $[0, 1]$ tal que $u_1 \leq u_2$ y $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0. \quad (1.1)$$

Por tanto, es posible establecer una relación entre las cópulas y las funciones de distribución. Considere una función distribución $F(y_1, y_2)$ con funciones de distribución marginales $F_1(y_1), F_2(y_2)$, entonces,

$$y_1 = F^{-1}(u_1) \sim F_1, y_2 = F^{-1}(u_2) \sim F_2,$$

donde, u_1 y u_2 son variables aleatorias con distribución uniforme. De esta forma, se puede escribir:

$$F(y_1, y_2) = F(F_1^{-1}(u_1), F_2^{-1}(u_2)) \quad (1.2)$$

$$= P[U_1 \leq u_1, U_2 \leq u_2] \quad (1.3)$$

$$= C(u_1, u_2) \quad (1.4)$$

Además, se puede concluir que para $F(y_1)$ y $F(y_2)$ continuas, la cópula $C(u_1, u_2)$ es única.

La definición original de la función cópula fue dada por Sklar en 1959, aquí la presentamos:

Nous appellerons copule (a n dimensions) tout fonction C continue et non-deroissante au sens employ pour une fonction de repartition a n dimensions d finie sur le produit Cartesien de n intervalles fermes $[0, 1]$ et satisfaisant aux conditions $C(0, \dots, 0) = 0$ et $C(1, \dots, 1, u, 1, \dots, 1) = u$. (Tomada de [Bouye, 2000]).

En base a esta definición, Sklar propuso su importante teorema, ya que proporciona una forma de analizar la estructura de dependencia de las distribuciones bivariadas sin conocer el comportamiento las distribuciones marginales.

Teorema 1. Teorema de Sklar

Sea $F(y_1, y_2)$ una función de distribución multivariada con marginales $F_1(y_1), F_2(y_2)$; entonces existe una cópula C tal que

$$F(y_1, y_2) = C(F_1(y_1), F_2(y_2); \theta)$$

donde θ es el parámetro de la cópula llamado parámetro de dependencia, el cual mide la dependencia entre las distribuciones marginales. [Trivedi, 2007].

1.2. Cotas de Fréchet Hoeffding

Teorema 2. (Tomado de Ayyad [Ayyad, 2008]). Sea $C' : S_1 \times S_2 \mapsto [0, 1]$ una subcópula; para todo $(u, v) \in S_1 \times S_2$ se cumple que

$$\max(u + v - 1, 0) \leq C'(u, v) \leq \min(u, v).$$

Denotando de la siguiente forma:

$$W(u, v) = \max(u + v - 1, 0), \quad (1.5)$$

y,

$$M(u, v) = \min(u, v),$$

se tiene que,

$$W(u, v) \leq C'(u, v) \leq M(u, v) \quad (1.6)$$

Dado que toda cópula es una subcópula, las desigualdades (1.6) se cumplen para las cópulas. Las cotas de Fréchet-Hoeffding representan los extremos entre los cuales toma valores toda cópula.

nota 1. La densidad c asociada a la cópula C esta dada por

$$c(u_1, u_2) = \frac{f(F_1^{-1}(u_1), F_2^{-1}(u_2))}{\prod_{i=1}^2 f_i(F_i^{-1}(u_i))},$$

donde $f(\cdot)$ denota la función de densidad de $F(y_1, y_2)$ y $f_i(\cdot)$ las densidades de cada una de las $F_i(y_i)$, para $i = 1, 2$. [Kazianka, 2010b].

1.3. Familia de Cópulas Arquimedianas

Las cópulas arquimedianas son una familia importante de cópulas debido a su fácil implementación, acotación y simulación. Para definir la familia de cópulas Arquimedianas se sigue a [Ayyad, 2008].

Sean X, Y variables aleatorias absolutamente continuas con función de distribución conjunta $H(x, y)$ y distribuciones marginales $F(x)$ y $G(y)$. La familia de las cópulas Arquimedianas cumplen la siguiente propiedad. Para alguna función $\lambda(\cdot)$, la cual debe ser positiva en el intervalo $[0, 1]$, será posible escribir:

$$\lambda[H(x, y)] = \lambda[F(x)]\lambda[G(y)],$$

y $H(x, y)$ como la suma de las marginales $F(x)$ y $G(Y)$ si se define $\phi(t) = -\ln[\lambda(t)]$:

$$\begin{aligned} \phi[C(u, v)] &= \phi[H(x, y)] \\ &= \phi(u) + \phi(v) \\ &= \phi[F(x)] + \phi[G(y)] \end{aligned}$$

donde $C(u, v)$ corresponde a la cópula Arquimediana y la función ϕ se denomina generadora de la cópula, de manera formal, se tiene la siguiente definición:

Definición 3. Se llama pseudo inversa de una función $\phi(\cdot)$ a otra función $\phi^{-1}(\cdot)$ con dominio $[0, \infty+]$ y rango $[0, 1]$ definida del siguiente modo:

$$\phi^{-1}(t) = \begin{cases} \phi^{-1}(t) & 0 \leq t \leq \phi(0), \\ 0 & \phi(0) < t < \infty+, \end{cases}$$

Puede demostrarse que $\phi^{-1}(\cdot)$ es también continua, no creciente en $[0, +\infty]$ y estrictamente decreciente en $[0, \phi(0)]$. Además, $\phi^{-1}(\phi(u)) = u$ en $[0, 1]$ y

$$\phi^{-1}(t) = \begin{cases} t & 0 \leq t \leq \phi(0) \\ \phi(0) & \phi(0) < t < \infty+ \end{cases} = \min(t, \phi(0))$$

La forma de la cópula Arquimediana queda definida por el siguiente teorema:

Teorema 3. Sean $\phi(\cdot)$ y $\phi^{-1}(\cdot)$ las funciones definidas en la definición (3) y sea $C : [0, 1]^2 \mapsto [0, 1]$ una función dada por:

$$C(u, v) = \phi^{-1}(\phi(u) + \phi(v)). \quad (1.7)$$

y ϕ es convexa, entonces C satisface las condiciones dadas en la definición [2], es decir, $C(\cdot, \cdot)$ es una cópula sí y solamente sí ϕ es convexa.

Por tanto, la construcción de las cópulas Arquimedianas queda dependiendo de la función generadora ϕ , siempre y cuando ϕ sea continua, decreciente y convexa, con dominio en $[0, 1]$ y $\phi(1) = 0$. Se puede ver este tipo de construcción de cópulas por medio del siguiente ejemplo.

Ejemplo 1. Cópula Producto (Tomado de [Ayyad, 2008]). Sea $\phi(y) = -\ln(t)$, para $t \in [0, 1]$, se tiene que:

$$\phi(0) = +\infty, \quad \mathbf{y} \quad \phi(1) = 0,$$

cuya función inversa queda determina por:

$$\phi^{-1}(t) = \exp(-t).$$

Entonces, se puede determinar la cópula de la siguiente forma:

$$\begin{aligned} C(u, v) &= \phi^{-1}(\phi(u) + \phi(v)) = \exp(-[-\ln(u) + (-\ln(v))]) \\ &= \exp(\ln(u) + \ln(v)) \\ &= \exp(\ln(u))\exp(\ln(v)) \\ &= uv. \end{aligned}$$

Se puede concluir que la cópula producto es una cópula Arquimediana.

1.3.1. Construcción de las cotas de Fréchet-Hoeffding

(Tomado de [Nelsen, 1998]). Sea $\phi(t) = 1 - t$ para $t \in [0, 1]$. Entonces, $\phi^{-1}(t) = 1 - t$ para $t \in [0, 1]$ y 0 para $t > 1$; es decir, $\phi^{-1}(t) = \max(1 - t, 0)$. La cópula queda definida de la siguiente forma:

$$C(u, v) = \max(u + v - 1, 0)$$

por lo que corresponde a la cota inferior $W(u, v)$ de Fréchet-Hoeffding. De esta forma se prueba que W es una cópula Arquimediana.

La cota superior M no es una cópula Arquimediana. (Ver [Nelsen, 1998]).

Numerosas cópulas han sido creadas bajo este paradigma de construcción y por tanto forman parte de esta familia. A continuación se presenta la siguiente tabla con algunas de las cópulas Arquimedianas, la construcción de cada una es análoga a la construcción de la cópula producto en el ejemplo (1).

| Nombre | $C(u_1, u_2; \theta)$ | $\phi(t)$ | Rango de θ |
|---------|--|--|---------------------|
| Clayton | $(u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}}$ | $\theta^{-1}(t^{-\theta} - 1)$ | $(0, \infty)$ |
| Frank | $-\frac{1}{\theta} \log \left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1} \right)$ | $-\log \left(\frac{e^{-\theta t} - 1}{e^{-\theta} - 1} \right)$ | $(-\infty, \infty)$ |
| Gumbel | $\exp(-(\tilde{u}_1^\theta + \tilde{u}_2^\theta)^{1/\theta})$, donde $\tilde{u}_j = -\log(u_j)$ | $(-\log(t))^\theta$ | $[1, \infty)$ |

TABLA 1.1. Algunas cópulas Arquimedianas y sus generadores.

1.4. Familia de cópulas elípticas

La familia de cópulas elípticas son las cópulas generadas por distribuciones cuyos contornos de sus densidades tiene la misma forma elíptica de la Normal incluyendo las distribuciones con colas más y menos pesadas, como son la distribución Normal y la t de Student. La principal desventaja de esta familia de cópulas es que carecen de una expresión cerrada de la distribución lo que puede ocasionar problemas computacionales. A continuación se presenta la cópula Normal y t de Student. La familia de cópulas elípticas son tomadas de [Trivedi, 2007].

1.4.1. Cópula Gaussiana

La cópula gaussiana se deriva de la distribución gaussiana multivariada. La forma de los contornos de la densidad de la cópula que muestran simetría y la facilidad en su implementación, la convierten en una de las cópulas mas conocidas. La cópula gaussiana toma la forma:

$$\begin{aligned}
C(u_1, u_2; \theta) &= \Phi_G(\Phi^{-1}(u_1), \Phi^{-1}(u_2)), \\
&= \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi(1-\theta^2)^{1/2}} \left(\frac{-(s^2 - 2\theta st + t^2)}{2(1-\theta^2)} \right) ds dt
\end{aligned}$$

donde Φ es la función de distribución normal estándar y $\Phi_G(u_1, u_2)$ es la distribución normal bivariada con parámetro de correlación θ restringido al intervalo $(-1, 1)$.

A medida que el parámetro de dependencia θ se aproxima a -1 y 1 , la cópula normal alcanza las cotas de Fréchet-Hoeffding inferior y superior, respectivamente. La cota inferior de Fréchet-Hoeffding muestra que la cópula gaussiana en su cota inferior posee un valor de correlación, $\theta = -1$, esto quiere decir, dependencia negativa:

$$W^n(x) = C_1(u_1, u_n),$$

Mientras que la cota superior de Fréchet Hoeffding muestra un valor de correlación $\theta = 1$ mostrando dependencia positiva:

$$M^n(x) = C_1(u_1, u_n).$$

la cópula normal permite por igual, grados de dependencia positiva o negativa, es por esto que es una de las más utilizadas.

1.4.2. t-cópula

La t-cópula es una cópula con dos parámetros de dependencia, se construye usando la distribución bivariada t de Student con v grados de libertad y correlación ρ , de la siguiente forma:

$$C^t(u_1, u_2; \theta_1, \theta_2) = t_{2,v,\rho}(t_{1,v}^{-1}(u_1), t_{1,v}^{-1}(u_2)), \quad (1.8)$$

donde $t_{1,v}$ y $t_{2,v,r}$ son distribuciones t de Student univariadas y bivariadas con v grados de libertad, respectivamente.

Los dos parámetros de dependencia se pueden denotar como (θ_1, θ_2) . El parámetro θ_1 controla el peso de las colas, es decir, si $\theta_1 < 3$, no existe varianza y para $\theta_1 < 5$, el cuarto momento no existe. Si $\theta_1 \mapsto \infty$, entonces, $C^t(u_1, u_2, \theta_1, \theta_2) \mapsto \Phi_G(u_1, u_2; \theta_2)$. [Denuit, 2005].

la t-cópula es usada cuando se necesitan capturar resultados de dependencia en valores extremos.

1.5. Medidas de dependencia y cópulas

Debido a que la cópula es una caracterización de dependencia de un par de variables aleatorias X e Y , tiene sentido una relación entre los coeficientes no paramétricos de dependencia, τ de Kendall y Rho de Spearman, para esto se definen estas dos medidas:

- τ de Kendall: Sean (X_1, Y_1) y (X_2, Y_2) dos pares de variables aleatorias independientes con función de distribución bivariada H . El coeficiente de correlación de rangos τ de Kendall está dado por:

$$\tau_n = \frac{P_n - Q_n}{\binom{n}{2}} \quad (1.9)$$

donde, P_n con el número de pares concordantes y Q_n discordantes. [Ayyad, 2008]

- Rho de Spearman (ρ_s): Sean X e Y dos variables aleatorias con funciones de distribución marginales F y G y función de distribución conjunta H . El coeficiente de correlación de rangos de Spearman es:

$$\rho_s(X, Y) = \rho(F(X), G(Y)),$$

por tanto, ρ_s es la correlación lineal de las variables aleatorias X e Y transformadas a través de su función de distribución.

Siguiendo a [Denuit, 2005], para funciones marginales continuas, las medidas de correlación de rangos se puede escribir en términos de la cópula C asociada a F , G y H :

$$\tau(X, Y) = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1,$$

y,

$$\rho_s(X, Y) = 12 \int_0^1 \int_0^1 [C(u, v) - uv] dudv.$$

1.6. Estimación de los parámetros de las cópulas

En esta sección se presenta los procedimientos utilizados para realizar la estimación de parámetros de la cópula. Se realiza una extensión a las cópulas multivariadas, la cual su definición es análoga al caso bivariado, debido a que la bibliografía encontrada para esta sección considera cópulas multivariadas.

1.6.1. Estimación paramétrica

Suponga que se tiene n realizaciones independientes de una distribución multivariada, $\{(X_{i1}, \dots, X_{ip})^T\}$ y que la distribución marginal esta denotada como $F_i(x)$ y función de densidad $f_i(x)$ para cada $i = 1, \dots, p$, y sea c la función de densidad de la cópula C . Sea β el vector de los parámetros de las marginales y α el vector de parámetros de la cópula. Por tanto, el vector de parámetros que se debe estimar es $\theta = (\beta^T, \alpha^T)^T$. La función logverosimilitud es:

$$l(\theta) = \sum_{i=1}^n \log c\{F_1(X_{i1}; \beta), \dots, X_{ip}; \alpha; \beta\} + \sum_{i=1}^n \sum_{j=1}^p \log f_i(X_{ij}; \beta).$$

El estimador ML de θ es

$$\hat{\theta}_{ML} = \operatorname{argmax}_{\theta \in \Theta} l(\theta), \quad (1.10)$$

donde Θ es el espacio de parámetros. [Yan, 2007].

1.6.2. Estimación semiparamétrica

(Adaptado de [Jaworski et al., 2010].) La estimación semiparamétrica de la cópula se basa en las representaciones de densidad y descomposiciones para el logaritmo de la verosimilitud de los modelos de dependencia, es decir, la densidad f de una función n -dimensional F con marginales F_1, \dots, F_n y correspondientes densidades univariadas f_1, \dots, f_n se puede representar como:

$$f(x_1, \dots, x_n) = c(F_1(x_1), \dots, F_n(x_n)) \prod_{i=1}^n f_i(x_i) \quad (1.11)$$

donde $c(u_1, \dots, u_n; \theta) = \frac{\partial^n C(u_1, \dots, u_n)}{\partial u_1 \dots \partial u_n}$ es la densidad de la cópula multivariada $C(u_1, \dots, u_n; \theta)$.

La descomposición para la función verosimilitud $L = \sum_{j=1}^n \log f(x_1^{(j)}, \dots, x_n^{(j)})$ de una muestra aleatoria se puede escribir como:

$$L = L_C + \sum_{i=1}^n L_i, \quad (1.12)$$

donde,

$$L_C = \sum_{j=1}^n \log c(F_1(x_1^j), \dots, F_n(x_n^j)) \quad (1.13)$$

que corresponde a la log-verosimilitud en la estructura de dependencia en los datos, representados por la cópula C , es decir, los parámetros de la cópula, y

$$L_i = \sum_{j=1}^n \log f_i(x_i^j), \quad i = 1, \dots, n, \quad (1.14)$$

que corresponde a la log-verosimilitud de cada marginal.

La estimación semiparamétrica se realiza en dos etapas: la primera etapa consiste en estimar las marginales univariadas F_i , no paramétricamente, por ejemplo, si F_i es la función de distribución empírica o sus versiones escaladas. En la segunda etapa, los parámetros de la cópula se estiman utilizando máxima verosimilitud.

Por tanto, dada una estimación no paramétrica \hat{F}_i de las marginales univariadas F_i , es natural la estimación del parámetro θ de la cópula de la siguiente forma:

$$\hat{\theta} = \operatorname{argmax}_{\theta} L_C(\theta) = \operatorname{argmax}_{\theta} \sum_{j=1}^n \log c(\hat{F}_1(x_1^j), \dots, \hat{F}_n(x_n^j); \theta)$$

El estimador resultante semi-paramétrico del parámetro de dependencia θ es consistente y asintóticamente normal, bajo condiciones de regularidad adecuados.

1.6.3. Estimación no paramétrica

(Adaptado de [Jaworski et al., 2010].) El procedimiento de estimación no paramétrica para cópulas está inspirado por las distribuciones empíricas, es decir, un estimador de la cópula $C(u_1, \dots, u_n)$ esta dado por:

$$\hat{C}(u_1, \dots, u_n) = \hat{F}(\hat{F}_1^{-1}(u_1), \dots, \hat{F}_n^{-1}(u_n)) \quad (1.15)$$

donde \hat{F} es la estimación no paramétrica de la función de distribución n -dimensional F y $\hat{F}_1^{-1}, \dots, \hat{F}_n^{-1}$ las estimaciones de las pseudo inversas $\hat{F}_1^{-1} = \{t | F_1(t) \geq s\}$ de las marginales univariadas F_1, \dots, F_n . Clásicamente, \hat{F} se estima de la función de distribución empírica n -dimensional:

$$\hat{F}(x_1, \dots, x_n) = \frac{1}{T} \sum_{t=1}^T I(X_1 \leq x_1, \dots, X_n \leq x_n)$$

y F_i^{-1} se estima de la distribución empírica univariada:

$$\hat{F}_i = \frac{1}{T} \sum_{t=1}^T I(X_i \leq x_i),$$

donde $I(\cdot)$ denota la función indicadora.

La experiencia refleja que para las diferentes familias de cópulas los estimadores suelen comportarse diferente, mas adelante se tratará este tema con mayor profundidad.

Geoestadística

Si se asocia una variable aleatoria a cada punto en un lattice con coordenadas $x = (x_1, \dots, x_n)$ en un espacio n -dimensional, el conjunto de estas variables se conoce como un proceso estocástico, que no puede ser visto como un proceso determinístico ya que son sucesos que ocurren a través del tiempo con fluctuaciones aleatorias, la aplicación de estas variables a la modelización de procesos o fenómenos geológicos en el espacio se conoce como geoestadística.

2.1. Campo aleatorio

Para definir campo aleatorio, es necesario recordar que es un proceso estocástico:

Definición 4. Sea $T \subset \mathbb{R}^n$, un proceso estocástico es una colección de variables aleatorias

$$\{X_t\}_{t \in T}$$

definida en un espacio de probabilidad (Ω, σ, P) asume valores reales.

Definición 5. Sea $T \subset \mathbb{R}^n$ un conjunto, un campo aleatorio se define como una función

$$(\omega, t) : \Omega \times T \mapsto \mathbb{R}^n$$

de tal forma que $Z(\omega, t)$ es un vector aleatorio para cada $t \in T$.

La función $Z(\omega, t)$, en adelante se notará:

$$\{Z(x) : x \in D\} \tag{2.1}$$

donde D , llamado dominio, es un subconjunto fijo de \mathbb{R}^n .

El campo aleatorio (2.1) está definido a través de una función de distribución finito dimensional:

$$F_{x_1, \dots, x_n}(z_1, \dots, z_n) = P[Z(x_1) \leq z_1, \dots, Z(x_n) \leq z_n] \tag{2.2}$$

En algunos textos, $z(\cdot)$ es llamada variable regionalizada. [Cressie, 1991]

Definición 6. Momentos de un campo aleatorio

(Tomado de [Giraldo, 2010]) Sea $\{Z(x) : x \in D\}$ con $D \subset \mathbb{R}^n$ el campo aleatorio, con función de distribución conjunta dada por (2.2), se tiene que:

1. $E(Z(x_i)) = m(x_i)$
2. $V(Z(x_i)) = E[Z(x_i) - m(x_i)]^2 = \sigma_i^2$
3. $C(Z(x_i), Z(x_j)) = E[Z(x_i) - m(x_i)][Z(x_j) - m(x_j)]$
4. $\gamma(Z(x_i), Z(x_j)) = \frac{1}{2}E[Z(x_i) - Z(x_j)]^2$

El momento (3) es la función de autocovarianza y el (4) función de semivarianza.

2.2. Estacionariedad

La estacionariedad es una de las propiedades más importantes de los campos aleatorios, dependiendo de nuestros propósitos, se estudiarán tres tipos de estacionariedad: Estricta, de segundo orden e intrínseca. En [Lloyd, 2006] se enfatiza que la estacionariedad no es una propiedad de los datos si no que es una propiedad del campo aleatorio, ya que en muchos casos se relaciona erróneamente la homogeneidad de los datos con la estacionariedad.

Definición 7. Estacionariedad (Adaptado de [Giraldo, 2010]). El campo aleatorio es estacionario si su función de distribución conjunta es invariante respecto a cualquier translación del vector $\mathbf{h} \in \mathbf{R}^d$, es decir,

$$F_{x_1, \dots, x_n}(z_1, \dots, z_n) = F_{x_1 + \mathbf{h}, \dots, x_n + \mathbf{h}}(z_1, \dots, z_n) \quad (2.3)$$

Definición 8. Estacionariedad de segundo orden

(Adaptado de [Cressie, 1991]) Si para un campo aleatorio del tipo (2.1), se tiene que:

$$E(Z(x)) = \mu; \quad \forall x \in D, \quad (2.4)$$

implica que, $F_x(z) = P(Z(x) \leq z)$ no depende de la localización x , además si

$$\text{cov}(Z(x_1), Z(x_2)) = C(x_1 - x_2), \quad \forall x_1, x_2 \in D, \quad (2.5)$$

es decir, la covarianza para $(Z(x), Z(x + \mathbf{h}))$, $\forall x_1, x_2 \in D$ siempre existe, entonces el campo aleatorio se define estacionario de segundo orden. Además si $C(x_1 - x_2)$ es una función que depende de \mathbf{h} , entonces $C(\cdot)$ es llamada isotrópico.

La estacionaridad o estacionaridad fuerte esta definida bajo la condición:
 $F_{x_1, \dots, x_n}(z_1, \dots, z_n) = F_{x_1 + \mathbf{h}, \dots, x_n + \mathbf{h}}(z_1, \dots, z_n)$.

Por tanto, la estacionariedad fuerte implica estacionariedad de segundo orden.

Definición 9. Estacionariedad intrínseca (Adaptado de [Lloyd, 2006]). Si para un campo aleatorio se tiene que la esperanza de la variable no debe depender de x y la varianza de los incrementos es finita. Es decir, el variograma $\gamma(\mathbf{h})$ exista y depende de únicamente \mathbf{h} , entonces se puede escribir:

$$\gamma(\mathbf{h}) = \frac{1}{2}E[(Z(x) - Z(x - \mathbf{h}))^2], \quad \text{para todo } \mathbf{h}, \quad (2.6)$$

Definición 10. Suponga que $\mu(x) = E(Z(x))$, existe para todo $x \in D$; se llama $\mu(x)$ a la tendencia espacial.

Definición 11. Dada una función de covarianza $C(\cdot)$ en \mathbb{R}^n , se dice que es una función de covarianza isotrópica si dados $x, x' \in \mathbb{R}^n$ se cumple:

$$C(x) = C(x') \quad \text{cuando} \quad \|x\| = \|x'\|$$

La isotropía no implica estacionariedad, existen campos aleatorios isotrópicos que no son estacionarios, un ejemplo puede ser el movimiento Browniano multiparamétrico de Lévy.

2.3. Modelos Gaussianos

Un campo aleatorio $Z(x)$ esta determinado por alguna distribución multivariada,

$$P(Z(x_1) \leq z_1, \dots, Z(x_n) \leq z_n) \quad (2.7)$$

para algún $n \in \mathbb{N}$ aleatorio, y $x_1, \dots, x_n \in D \subset \mathbb{R}^n$.

Debido a que la caracterización completa en muchos casos suele ser imposible, la media $m(x) = E(Z(x))$ y la función de covarianza $K(x_i, x_j) = Cov(Z(x_i), Z(x_j))$ son indispensables para el manejo adecuado de estos modelos.

Por tanto, $m(x)$ representa la tendencia del modelo y $K(x_i, x_j)$ representa la estructura de dependencia del campo aleatorio.

Si se supone isotropía, es decir, que la media se considera constante y la función de covarianza $K(x_1, x_2)$ dependa de x_1 y x_2 solo a través de $\mathbf{h} = \|x_1 - x_2\|_2$, donde $\|\cdot\|_2$ denota la distancia Euclidiana, entonces K es llamada función de autocovarianza isotrópica.

El modelo Gaussiano, es quizá el mas importante modelo geoestadístico, se define como:

$$Y_i = \mathbf{f}(\mathbf{x}_k)^T \boldsymbol{\beta} + Z(\mathbf{x}_i), \quad i = 1, \dots, n, \quad (2.8)$$

donde, $Z(\mathbf{x})$ es un campo aleatorio isotrópico, con función de correlación $K_{\sigma^2, \boldsymbol{\theta}}$, \mathbf{f} es el vector de localizaciones dependientes de las variables explicativas y $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ el vector de parámetros de regresión.

2.3.1. Familia Matérn

Una de las familias que se consideran en esta tesis, es la familia de funciones de autocovarianza isotrópicas Matérn:

$$K_{\sigma^2, \boldsymbol{\theta}}(h) = \sigma^2 \left((1 - \nu_2) + \frac{\nu_2}{2^{\kappa-1} \Gamma(\kappa)} \left(\frac{2\kappa^{\frac{1}{2}} h}{\vartheta_1} \right) \right), \quad (2.9)$$

donde, \mathcal{K}_κ denota la función modificada de Bessel de orden $\kappa > 0$, $\vartheta_1 > 0$ es el “rango”, $\vartheta_2 \in [0, 1]$ es el llamado “efecto nugget” siendo la proporción de la varianza atribuida a errores de medición, y $\boldsymbol{\theta} = (\vartheta_1, \vartheta_2, \kappa)$ denota el vector de correlación de parámetro. El parámetro κ controla la uniformidad del proceso correspondiente.

2.3.2. Anisotropía geométrica

La isotropía débil es una suposición bastante fuerte ya que los procesos no son, por lo general, independientes de la dirección, y pueden mostrar un comportamiento anisotrópico.

Una extensión popular de campos aleatorios isotrópicos es considerar la conversión en campos aleatorios isotrópicos después de una transformación lineal sobre las coordenadas. Esta variante especial de la anisotropía se denomina anisotropía geométrica.

Sea $Z_1(x)$ un campo aleatorio isotrópico en \mathbb{R}^d con función de covarianza K_1 y media μ . Para un campo aleatorio $Z(x) = Z_1(Tx)$, donde $T \in \mathbb{R}^{d \times d}$, se obtiene que:

$$E(Z(x)) = \mu, \quad (2.10)$$

y la función de covarianza corresponde a

$$Cov(Z(x_1), Z(x_2)) = K_1(\|T(x_1 - x_2)\|_2), \quad (2.11)$$

Al corregir la anisotropía geométrica se debe revertir la transformación de coordenadas, ya que $Z(T^{-1}x)$ tiene la misma media que $Z(x)$ pero función de covarianza isotrópica K_1 .

Para la corrección de coordenadas, se tiene:

$$T^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix}, \quad (2.12)$$

donde, λ y φ se denomina relación de anisotropía y el ángulo de anisotropía, respectivamente.

Geoestadística usando cópulas

Con el fin de describir la estructura de dependencia en un campo aleatorio en el área de geoestadística, la literatura clásica se enfoca en métodos usuales, como son las propiedades de suavizamiento de los campos aleatorios, la función de autocorrelación, variogramas y técnicas para interpolación espacial, kriging simple, kriging universal, cokriging, kriging disyuntivo, kriging Bayesiano entre otros. En este trabajo se presenta el análisis de estructura de dependencia a través de cópulas proponiendo una familia de cópulas radialmente asimétricas.

En este capítulo se describe la forma en que se aplicarán las funciones cópulas en el campo de geoestadística.

3.1. Descripción del campo aleatorio usando cópulas

Bardossy [Bárdossy, 2008] presentó un método diferente de los métodos usuales en geoestadística, para el modelado de dependencia espacial por medio de cópulas que aquí se adopta, se pretende describir algunas de las distribuciones bivariadas necesarias del campo aleatorio por medio de cópulas.

Se asume un campo aleatorio estacionario $\{Z(x)|x \in D\}$ donde $D \in \mathbb{R}^2$ es el área de interés. Se nota \mathbf{h} el vector de separación entre dos puntos.

Sea F_Z la distribución univariante del proceso espacial, debido a que el campo aleatorio es estacionario F_Z es la misma para cada localización $x \in D$.

Con el teorema de Sklar, se puede establecer un modelo multivariante del campo tomando $F_Z = F_1 = F_2 = \dots = F_n$ tal que,

$$H(x_1, \dots, x_n) = C(F_1, F_2, \dots, F_n), \quad (3.1)$$

Entonces, la relación entre dos localizaciones separadas por el vector \mathbf{h} esta caracterizada por la distribución bivalente:

$$P(Z(x) \leq z_1, Z(x + \mathbf{h}) \leq z_2) = C_{\mathbf{h}}(F_Z(z_1), F_Z(z_2)), \quad (3.2)$$

por tanto, la estructura de dependencia quedaría descrita por la función copula $C_{\mathbf{h}}$ en función del vector \mathbf{h} , esto implica que la cópula podría describir la estructura completa de dependencia a diferencia de los variogramas que solo describen con respecto a la media. Por otro lado, la elección de la cópula C será determinada aplicando varias familias de cópulas al modelo de la ecuación (3.2) y comparando los diferentes resultados de cada familia.

Es de esperarse que no todas las familias de cópulas continuas sean apropiadas para este modelo, se puede suponer una cópula asimétrica, es decir, para dos localizaciones x_1 y x_2 la cópula asimétrica en x_1 y x_2 es la misma que en x_2 y x_1 , de manera general se tiene que:

$$C_{\mathbf{h}}(u_1, \dots, u_n) = C_{\mathbf{h}}(u_{\pi(1)}, \dots, u_{\pi(N)}), \quad (3.3)$$

para una permutación arbitraria π .

Además, tenemos que añadir las siguientes restricciones:

- Cuando $\|\mathbf{h}\| \rightarrow \infty$ entonces, $C_{\mathbf{h}}(u) \rightarrow \Pi^n(x)$, ya que se quiere independencia sobre localizaciones muy alejadas entre si,
- Cuando $\|\mathbf{h}\| \rightarrow 0$ entonces, $C_{\mathbf{h}}(u) \rightarrow M^n(x)$ o equivalentemente, en localizaciones muy próximas entre si, queremos dependencia muy fuerte.

Estas condiciones son fundamentales para la construcción de la cópula, permite realizar un filtro sobre las cópulas que se podrían proponer, por ejemplo, se puede descartar la familia Farlie-Gumbel-Morgenstern debido a que no cumplen estas propiedades, o permite la construcción de cópulas.

La metodología de construcción de cópulas para geoestadística se basa en transformaciones de las variables, esto puede verse como una relación entre cópulas y los variogramas indicadores. Normalmente, para un valor de corte β , el variograma indicador es:

$$\gamma = F_s^{-1}(\beta) - C_s(\mathbf{h}, F_s^{-1}(\beta), F_s^{-1}(\beta)) \quad (3.4)$$

y los indicadores de cross covarianza,

$$\gamma_{\beta_1, \beta_2}(\mathbf{h}) = \frac{1}{2} \min(F_s^{-1}(\beta_1), F_s^{-1}(\beta_2)) - C_{\mathbf{h}}(F_s^{-1}(\beta_1), F_s^{-1}(\beta_2)) \quad (3.5)$$

En este sentido, las funciones cópulas no contienen información diferente o excesiva en comparación con el variograma indicador y el cross variograma, pero permiten un manejo conjunto y una presentación diferente, con una parametrización más parsimoniosa

[Bárdossy, 2006].

A continuación se presenta las cópulas que van a ser utilizadas en el desarrollo de esta tesis, se debe suponer que debido a que los campos aleatorios más importantes son los gaussianos es de esperarse que las cópulas también lo sean, la cópula gaussiana definida en la siguiente ecuación (3.6) con marginal $F_Z = \Phi_{\mu, \sigma^2}$, donde μ y σ denota la media y varianza respectivamente, es ampliamente usada en el marco geoestadístico.

$$C_{\Sigma}^G = \Phi_{0, \Sigma}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n)). \quad (3.6)$$

La cópula gaussiana se convierte en una función de \mathbf{h} , suponiendo que la función de correlación sigue uno de los modelos paramétricos conocidos, por ejemplo, el modelo Matern.

3.2. Cópulas empíricas bivariadas

Las cópulas empíricas son usadas por primera vez en el área de geoestadística, por Haussler [Haslauer, 2010], quien implementó las cópulas bivariadas empíricas, describiendo la estructura de dependencia entre variables aleatorias. En esta tesis se consideran este tipo de cópulas como una metodología de exploración de datos, de tal forma que se pueda considerar la forma de la distribución que puedan tener, para esto el campo aleatorio debe cumplir la condición de estacionariedad fuerte.

El uso de cópulas tiene como consecuencia las siguientes ventajas, [Haslauer, 2010]:

- La distribución marginal, que podría distorsionar la estructura de dependencia, se filtra usando cópulas. Así, quedaría definida únicamente por los datos.
- La cópula permite una mejora en la cuantificación de la incertidumbre en la interpolación.

3.2.1. Algoritmo para la aplicación de las cópulas empíricas bivariadas

Las cópulas son usadas para explorar la estructura de dependencia entre dos variables aleatorias sin considerar las distribuciones marginales de cada variable. Las cópulas empíricas es el caso más simple de construcción pero no es computacionalmente óptimo, aun así, estas cópulas se pueden evaluar en diferentes direcciones y ángulos para cada par de puntos y generar una idea de la forma de la estructura de dependencia del campo aleatorio.

Según Haslauer, [Haslauer, 2010], se debe considerar el campo aleatorio estacionario y la construcción de la cópula correspondiente se puede realizar con el siguiente algoritmo:

1. Se calcula la distribución empírica marginal $F_Z(z)$ de las observaciones.

2. Para algún vector \mathbf{h} dado, se calcula el conjunto $S(\mathbf{h})$ definido como:

$$S(\mathbf{h}) = \{(F_Z(z(s_i)), F_Z(z(s_j))) \mid |s_i - s_j| \approx \mathbf{h}\} \quad (3.7)$$

3. Debido a que $S(h)$ es un conjunto de pares de puntos definidos en el cuadrado unidad, es posible dividirlo en una grilla de dimensiones $k \times k$ y calcula la función de densidad de la cópula empírica dado un vector \mathbf{h} usando la siguiente ecuación:

$$\begin{aligned} g_{i,j} &= c^* \left(\frac{2i-1}{2k}, \frac{2j-1}{2k} \right) \\ &= \frac{k^2}{|S(h)|}; (u, v) \in S(h) \mid \frac{i-1}{k} < u < \frac{i}{k} \quad \frac{j-1}{k} < v < \frac{j}{k} \end{aligned} \quad (3.8)$$

(3.10)

con $|S(\mathbf{h})|$ el número de parejas que tienen como vector de separación \mathbf{h} .

Ejemplo 2. En este ejemplo, se aplicará la cópula empírica a un conjunto de datos, los datos utilizados son de la región de Gomel (Estado de Bielorrusia) diez años después de la catástrofe de Chernobyl en abril de 1986, el peor accidente nuclear ocurrido en la ex UdSSR (ahora Ucrania). El gobierno ruso construyó una red donde se estudia la concentración de la cantidad de radiactividad. El conjunto de datos que utilizamos son mediciones de Cs137, un isótopo radiactivo.

La primera parte, consiste en realizar la exploración de datos, el conjunto posee 148 localizaciones del nivel de concentración del isótopo radiactivo, en la Figura (3.1) se muestra esta distribución, en la Figura (3.2) se muestra una interpolación de cuadrícula biarmónica con el uso del software matemático MATLAB.

Se puede ver que la mayoría de los datos se encuentran hacia la izquierda y no corresponde a una distribución Gaussiana. En la tabla (2) se muestran el resumen estadístico de los datos:

| min | 1st Q | median | 3rd Q | max |
|--------|--------|--------|--------|---------|
| 0.0400 | 0.6800 | 1.6500 | 5.3700 | 32.4100 |

TABLA 3.1. Resumen estadístico de los datos de Gomel

De la misma forma que el variograma, la cópula se calcula para distancias definidas, \mathbf{h} la cual representa la distancia de separación entre los puntos, usualmente llamado lag. La cópula presenta la dependencia fuerte a lo largo de los diferentes cuantiles, no solo bajo la media a diferencia del variograma. De esta forma obtenemos los datos transformados y las densidades de las cópulas empíricas en la Figura (3.3) de cuatro diferentes lags $\mathbf{h} = 10km, 20km, 40km$ y $60km$.

Se puede inferir, de la Figura (3.3) que es muy probable encontrar un punto con un valor de función de distribución de 0,8 el cual está separado de otro punto con el mismo valor de distribución a 10 km de distancia. Por otro lado, es poco probable, encontrar una

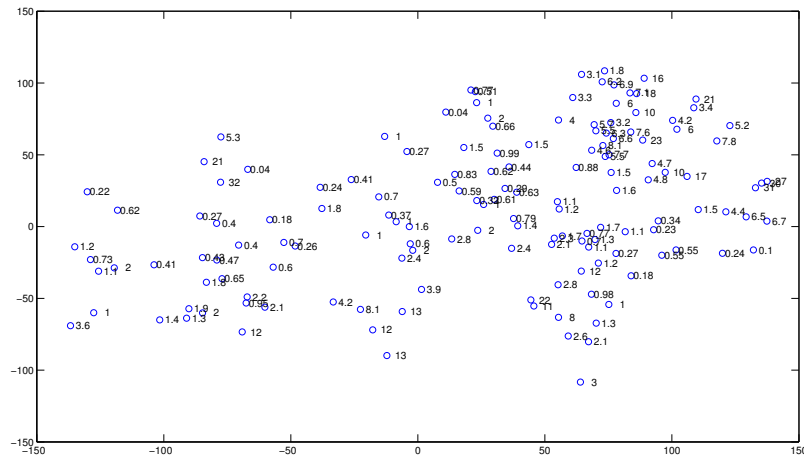


FIGURA 3.1. Observaciones de los datos de la ciudad Gómel

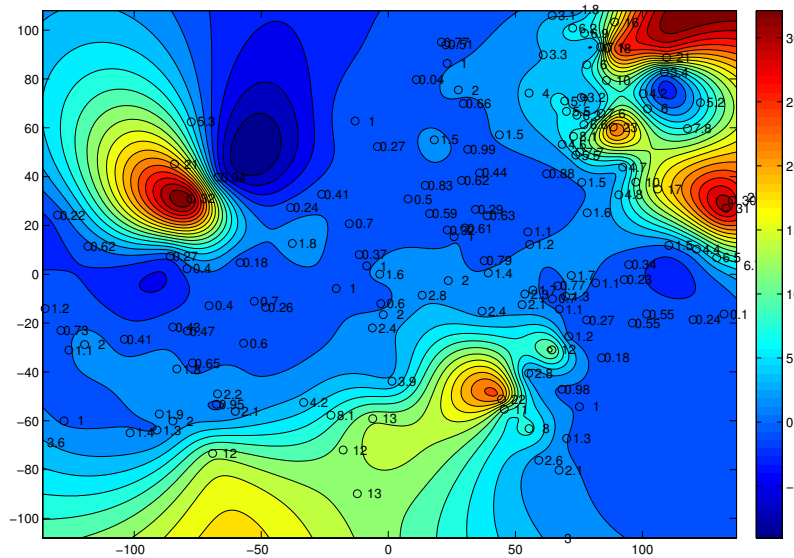


FIGURA 3.2. Gráfica de contorno con las localizaciones y los valores de las realizaciones

punto con un valor de función de distribución de 0,9 a 10 km de otro punto con un valor de función de distribución de 0,3. En general, si la distancia crece, la forma de la cópula se vuelve menos específica lo que indicaría una independencia a distancias grandes.

La cópula empírica permite visualizar la estructura de dependencia y muestra como la dependencia varía de acuerdo al valor de lag tomado, como se puede ver en la Figura (3.3). La forma de la cópula resulta ser menos específica cada que incrementa el valor de lag, lo cual indica un decaimiento de la dependencia entre puntos lejanos. Sin embargo es necesaria la medida de dependencia de Rango:

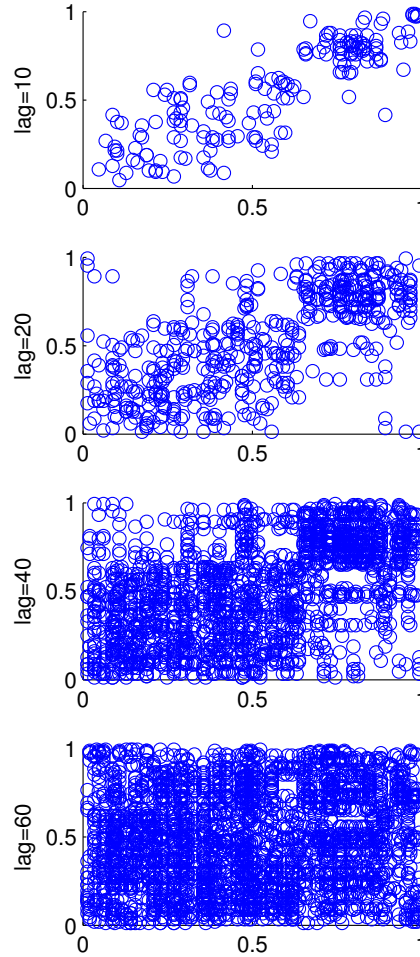


FIGURA 3.3. Densidades de las cópulas empíricas bivariadas

Definición 12. El rango, representa el grado de dependencia espacial, es calculado por:

$$Rank(h) = \frac{1}{12\mathbf{n}(\mathbf{h})} \sum_{x_i - x_j \approx h} (F_n(z(x_i)) - \frac{1}{2})(F_n(z(x_j)) - \frac{1}{2})$$

donde $\mathbf{n}(\mathbf{h})$ representa el número de parejas de puntos que cumplen que la distancia entre ellos es aproximadamente \mathbf{h} .

En la Figura (3.4) se muestra la gráfica de los valores del rango de correlación aumenta a medida que aumenta la distancia indicando dependencia espacial.

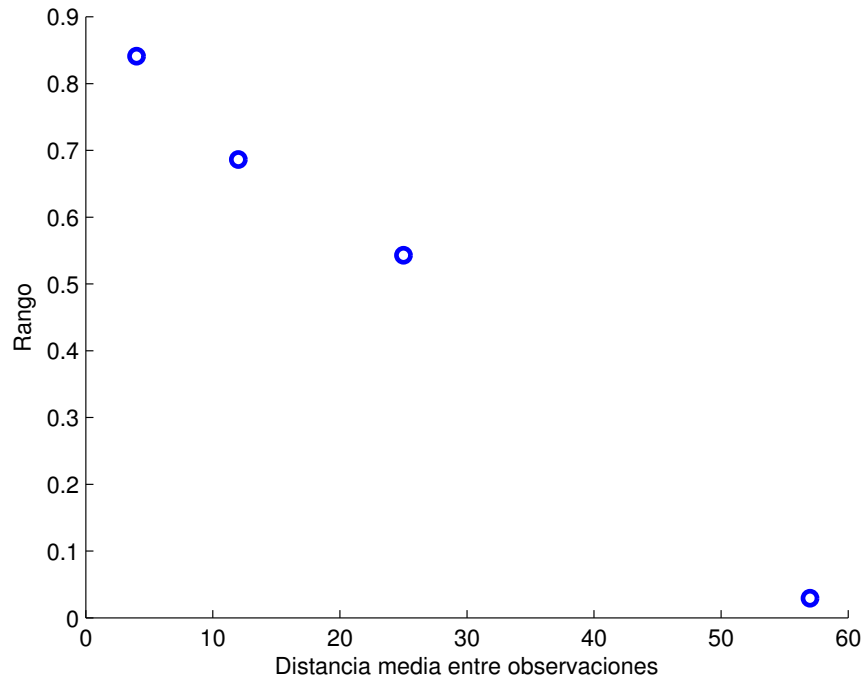


FIGURA 3.4. Rango de correlación

3.3. Medida de asimetría

Muchas de las metodologías en el análisis geoestadístico requieren del supuesto de campos aleatorios gaussianos pero este supuesto es falso en datos reales, se puede inferir, por ejemplo, en la Figura (3.3) en la cual se estudian las cópulas empíricas, que los datos no poseen simetría exhibiendo una clara dependencia no-gaussiana. Para solucionar este problema, se debe definir la medida de asimetría y la asimetría radial, en Bardossy [Bárdossy, 2008] define la medida de asimetría como :

Definición 13. Sea X e Y dos variables aleatorias, con función de cópula definida $C(u, v)$, la medida de asimetría se define como:

$$A(u) = \frac{2u - 1 + C(1 - u, 1 - u)}{C(u, u)}, \quad (3.11)$$

indicando que si en valores altos la dependencia es mayor que en valores bajos, $A(u) > 1$, o lo opuesto, $A(u) < 1$, si la dependencia es la misma $A(u) \approx 1$.

Definición 14. (Adaptado de [Nelsen, 1998].) Sean X y Y variables aleatorias continuas con distribución conjunta C y marginales F y G , respectivamente. Sea (a, b) un punto en \mathbb{R}^2 . Entonces (X, Y) son radialmente simétricas sobre (a, b) si y solo si,

$$C(a + x, b + y) = \bar{C}(a - x, b - y), \quad \text{para todo } (x, y) \in \mathbb{R}^2, \quad (3.12)$$

donde $\bar{C} = P(U > u, V > v)$.

La cópula gaussiana es simétrica, por tanto $A(u) = 1$, además también es radialmente simétrica, es decir, satisface:

$$c(u_1, u_2) = c(1 - u_1, 1 - u_2).$$

por tanto, no parece una cópula apropiada para el conjunto de datos en la Figura (3.3) ya que estos poseen asimetría positiva.

3.4. Tendencia espacial

Uno de los supuestos que se consideran en el campo geoestadístico es definir la distribución marginal univariada igual para todos los datos, implicando la ausencia de tendencia espacial, en la práctica, la forma de modelar $\mu(x)$ se realiza a través de un modelo de regresión. Los modelos de este tipo son llamados modelos de superficie de tendencia.

Por tanto, con el propósito de parametrizar las distribuciones marginales univariadas F_Z , siempre que la media μ aparezca explícitamente en la expresión analítica de la función de densidad marginal univariada f_Z , se tiene,

$$g(\mu_i) = \eta(y_i) = y_i^T \beta, \quad \text{para todo } i = 1, \dots, n,$$

donde, g es la función de enlace, $\eta(y_i)$ es la superficie de respuesta, y_i es el vector de covarianzas correspondiente a la localización x_i y β es el vector de los parámetros de regresión.

Familia de cópulas no gaussianas

En este capítulo se presenta una familia de cópulas asimétricas multivariadas no gaussianas, en respuesta a la necesidad que surge en geoestadística para ajustar una cópula asimétrica debido a la naturaleza de este tipo de datos. Hasta ahora solo se ha considerado la cópula gaussiana, pero esta no sólo expresa simetría, sino también dependencia de simetría radial, tal que los cuantiles altos y bajos de la distribución tienen propiedades iguales de dependencia. Este supuesto pocas veces se cumple con datos reales.

Una cópula bivariada expresa dependencia simétrica si:

$$C(u_1, u_2) = C(1 - u_1, 1 - u_2) - 1 + u_1 + u_2, \quad (4.1)$$

y para la densidad:

$$c(u_1, u_2) = c(1 - u_1, 1 - u_2), \quad (4.2)$$

la simetría se entiende con respecto al eje menor $u_2 = 1 - u_1$ del cuadrado unidad.

La cópula empírica describe la estructura de dependencia de los datos observados para cualquier par de puntos, de la misma forma que actúan los variogramas empíricos, las cópulas empíricas deberán disponer de cópulas teóricas que cumplan la condición de varianza no negativa. Para Bárdossy, [Bárdossy, 2006] la cópula multivariante teórica debe cumplir con las siguientes propiedades:

1. Debe poseer amplia gama de dependencia, un conjunto geográficamente cerca de puntos debería tener una dependencia arbitrariamente fuerte, mientras que los puntos distantes debe ser independientes,
2. la cópula multivariante debe tener una parametrización de tal manera que la estructura de dependencia refleje la posición geométrica de la correspondiente serie de puntos, esto es debido a que la propia cópula no permite una fácil visualización de la dependencia, sin embargo, su densidad revela las características de la dependencia.

La construcción de una familia de cópulas no es trivial, actualmente existen numerosas cópulas pero no cumplen las condiciones necesarias para ser utilizadas en geoestadística y en otros casos la construcción es imposible ya que se debe tener estabilidad con las marginales. Liebscher [Liebscher, 2008] propone la construcción de una familia de cópulas asimétricas multivariadas con el siguiente lema (1):

Lema 1. Sea $(U_1^{(1)}, \dots, U_d^{(1)})^T, \dots, (U_k^{(1)}, \dots, U_d^{(k)})^T$ donde k son vectores independientes con funciones de distribución $H_1, \dots, H_k \in [0, 1]^d$ respectivamente. Además, sea $g_{ji} : [0, 1] \rightarrow [0, 1]$ para $j = 1, \dots, k$ e $i = 1, \dots, d$ funciones con la propiedad de que cada una de ellas es continua y estrictamente creciente o idénticamente igual a 1. Supongamos que $g_{ji}(1) = 1$ para $j = 1, \dots, k, i = 1, \dots, d$. Se define $g_{ji}^{-1} := 0$ para $v \leq g_{ji}(0)$. Entonces $u = (u_1, \dots, u_d^T) \mapsto H(u) = \prod_{j=1}^k H_j(g_{j1}(u_1), \dots, g_{jd}(u_d))$ es la función de distribución de $(\max_{j \in J_i} g_{ji}^{-1}(U_i^j))_{i=1, \dots, d}$, $J_i = j \in 1, \dots, k : g_{ji} \neq 1$

A pesar de que los resultados generan cópulas asimétricas, debido a la naturaleza de su construcción, con cópulas que no pertenezcan a las familia arquimediana resulta casi imposible obtener resultados analíticos y ecuaciones explícitas para su correcta aplicación en geoestadística, entonces se propone la siguiente construccin:

4.1. Cópula χ^2

Una de las ventajas de considerar las funciones cópulas, es que son invariantes bajo transformaciones monótonas aplicadas a las marginales. Este hecho es ampliamente utilizado ya que transformaciones como Box-Cox no afecta el resultado, debido a que la cópula no esta influenciada por la marginal de las variables aleatorias. Esto representa una ventaja comparado con el uso de variogramas y covarianzas, ya que éstas dependen de las marginales ya mencionadas. Haciendo uso de esta propiedad y con el fin de encontrar cópulas apropiadas, seguiremos la metodología de construcción de Bárdossy [Bárdossy, 2006].

La cópula puede estar combinada con las distribuciones marginales univariadas de la variable espacial que definen el modelo y según el teorema de Sklar, para una función de distribución multivariada con marginales absolutamente continuas, existe una única cópula C tal que:

$$C(u_1, \dots, u_n) = F(F_n^{-1}(u_1), \dots, F_n^{-1}(u_n)), \quad (4.3)$$

Esto implica, que por el contrario, cualquier cópula C puede ser utilizada para combinar cualquier colección de distribuciones univariantes con el fin de crear una distribución multivariada. Este procedimiento frecuentemente aplicado se puede ver en cópulas como la cópula normal multivariante o la cópula t .

Las cópulas serán construidas según la ecuación (4.3). Se supone $g(t)$, una función no monótona y sea $Y \sim N(m, \Gamma)$, una variable aleatoria n -dimensional con media $m^T =$

(m, \dots, m) y matriz de covarianza Γ . Por simplicidad, se asume varianzas unitarias para las marginales. Sea J definida para cada coordenada $j = 1, \dots, n$ tal que:

$$J_i = g(Y_i) \quad (4.4)$$

la ecuación (4.4) define una distribución multivariada la cual no necesariamente es Gaussiana, en particular, si $g(t) = t^2$, la cópula que genera es la cópula no central χ^2 con un grado de libertad, ya que,

$$J_i = Y_i^2,$$

si $m = 0$, entonces todas las marginales se distribuyen χ^2 con un grado de libertad. Para el caso multivariado, se tiene:

$$\begin{aligned} J_1(t) &= P(Y^2 < t) \\ &= F(\sqrt{t}) - F(-\sqrt{t}) \\ &= \Phi(\sqrt{t} - m) - \Phi(-\sqrt{t} - m) \end{aligned}$$

derivando, la función de densidad tiene la siguiente forma:

$$j_1(t) = \frac{\phi(\sqrt{t} - m) - \phi(-\sqrt{t} - m)}{2\sqrt{t}}$$

donde Φ y ϕ son las funciones de distribución y densidad de la distribución normal estándar respectivamente. La correspondiente función de distribución queda definida con:

$$\begin{aligned} J_n(t_1, \dots, t_n) &= P(-\sqrt{t_1} < Y_1 < \sqrt{t_1}, \dots, -\sqrt{t_n} < Y_n < \sqrt{t_n}) \\ &= \sum_{i=0}^{2^n-1} (-1)^i \Phi(\epsilon_i - m), \\ &= \sum_{i=0}^{2^n-1} (-1)^{\sum_{j=1}^n i_j} \Phi_{m, \Sigma}(\epsilon_i), \end{aligned}$$

donde,

$$\epsilon_i^T = ((-1)^{i_1} \sqrt{t_1}, \dots, (-1)^{i_n} \sqrt{t_n}),$$

con $i_j = 0$ o 1 , y,

$$i = \sum_{j=0}^{n-1} i_j 2^j, \quad (4.5)$$

y la función de densidad j_n tiene la forma:

$$j_n = \frac{\sum_{i=0}^{2^n-1} \phi_{m,\Sigma}(\epsilon_i)}{2^n \sqrt{\prod_{i=1}^n z_i}}. \quad (4.6)$$

Para obtener las funciones de densidad y distribución de la cópula, se puede calcular a través de las distribuciones marginales y la distribución conjunta usando la ecuación (4.3) para calcular la función de distribución y la siguiente ecuación para la función de densidad:

$$c_n(u_1, \dots, u_n) = \frac{j_n(x_1, \dots, x_n)}{j_1(x_1) \dots j_1(x_n)}, \quad (4.7)$$

donde, $u_i = J_1(x_i)$.

4.2. Cópula V – transformada normal

Bárdossy y Li [Bárdossy, 2008] proponen una familia de distribuciones que se obtienen a través de una transformación no-monotónica de la cópula gaussiana multivariante, llamada cópula V –transformada. En este trabajo se propone dos extensiones de esta metodología: La primera es la inclusión de tendencia y la segunda es el método de indicador kriging e interpolación usando cópulas que se presentará mas adelante.

Definición 15. Sea $\mathbf{Y} \sim N(0, \Gamma)$ una variable aleatoria n – dimensional con media $0^T = (0, \dots, 0)$ y matriz de covarianza Γ . Todas las marginales se suponen con varianza unitaria. Sea \mathbf{X} definida para cada coordenada $j = 1, \dots, n$ de tal forma:

$$\mathbf{X}_j = \begin{cases} k(Y_j - m)^\alpha & Y_j \geq m, \\ m - Y_j & Y_j < m, \end{cases}$$

donde k es una constante positiva y α y m números reales arbitrarios, a manera de ejemplo, en la Figura (4.1), se presenta algunas transformaciones, se puede observar que la cópula recibe este nombre debido a la forma de V en la gráfica. La consecuencia más importante es que para valores menores que m , los cuales son los valores menores en el espacio original, se convierten, en el nuevo espacio en los valores mayores; tal que el valor m se transforma en el valor más pequeño en el nuevo espacio, en cuanto a los valores mayores que m en el espacio original, la transformación produce un concentramiento o una división dependiendo de la configuración de k y α . Este efecto produce que la dependencia asociada con los valores cercanos o iguales a m influya en la dependencia de los valores bajos en el nuevo espacio. [Jing, 2010].

El efecto de la transformación también puede explicarse por el cambio en la distribución marginal. Figura (4.2) muestra cómo la densidad en la distribución cambia después de la transformación. Se puede notar las siguientes características:

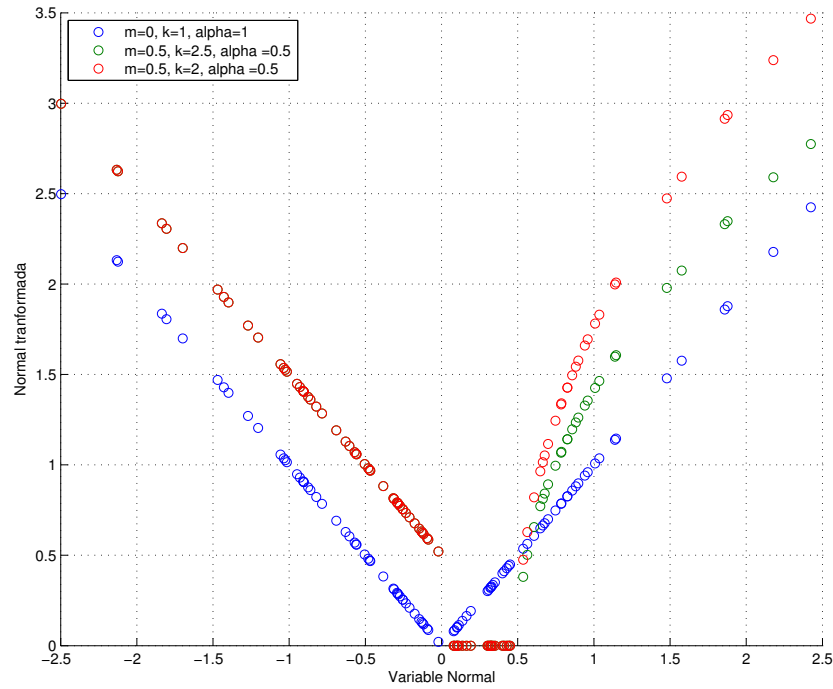


FIGURA 4.1. Transformaciones V-Normal

- Por lo general, los datos se dispersan después de la transformación,
- Si los valores de k y α son invariables, pero el valor de m incrementa, la densidad está más concentrada a la mediana y la distribución es más simétrica.
- Si $k = 1$, $\alpha = 1$ la transformada V-normal se aproxima a la distribución χ^2 con un grado de libertad.

La función de distribución marginal de X es:

$$\begin{aligned}
 H(x) &= P(X \leq x) \\
 &= P\left(Y < \left(\frac{x}{k}\right)^{\frac{1}{\alpha}} + m\right) + P(Y > x - m) \\
 &= \Phi\left(\left(\frac{x}{k}\right)^{\frac{1}{\alpha}} + m\right) - \Phi(-x + m)
 \end{aligned}$$

se obtiene la función de densidad:

$$h(x) = \frac{1}{k\alpha} \left(\frac{x}{k}\right)^{\frac{1}{\alpha}-1} \phi\left(\left(\frac{x}{k}\right)^{\frac{1}{\alpha}} + m\right) + \phi(-x + m)$$

tal que $\Phi(\cdot)$ y $\phi(\cdot)$ son las funciones de distribución y densidad de la normal estándar, respectivamente.

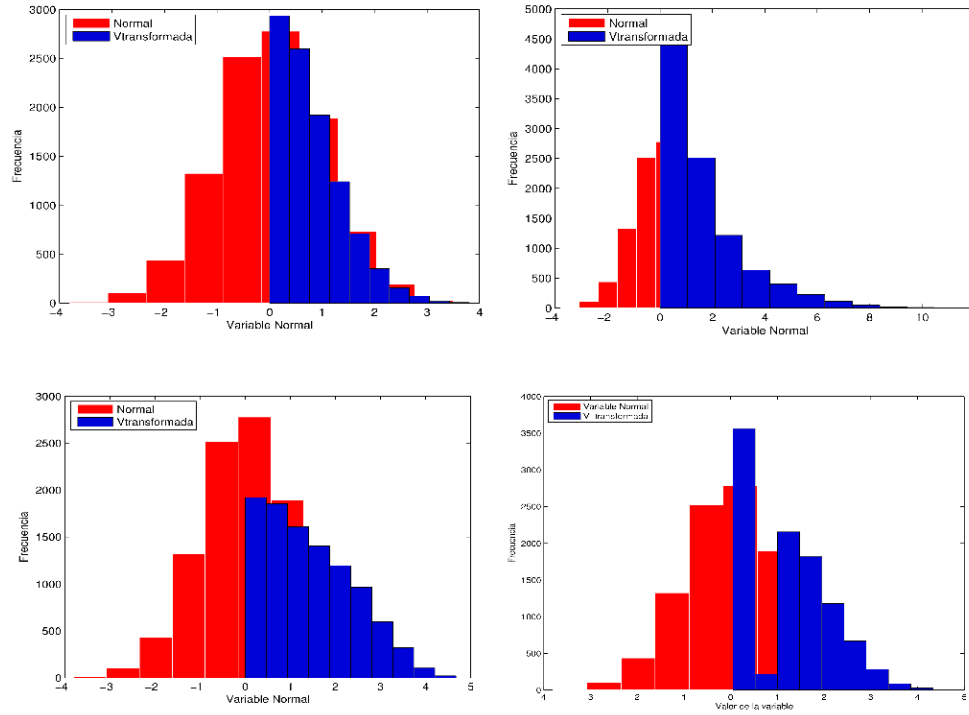


FIGURA 4.2. Transformaciones V -Normal con parámetros: $m = 0$, $k = 1$ y $\alpha = 1$ (Esquina superior izquierda), $m = 0$, $k = 3$ y $\alpha = 1$ (Esquina superior derecha), $m = 0$, $k = 2,5$ y $\alpha = 0,5$ (Esquina inferior izquierda) y $m = 1$, $k = 2,5$ y $\alpha = 0,5$ (Esquina inferior izquierda)

Para el caso particular bivariado, se tiene la función de distribución:

$$\begin{aligned}
 V(x_1, x_2) &= P(X_1 < x_1, X_2 < x_2) \\
 &= P\left(Y_1 < \left(\frac{x_1}{k}\right)^{\frac{1}{\alpha}} + m, Y_2 < \left(\frac{x_2}{k}\right)^{\frac{1}{\alpha}} + m\right) \\
 &\quad - P\left(Y_1 < -x_1 + m, Y_2 < \left(\frac{x_2}{k}\right)^{\frac{1}{\alpha}} + m\right) - P\left(Y_1 < \left(\frac{x_1}{k}\right)^{\frac{1}{\alpha}} + m, Y_2 < -x_2 + m\right) \\
 &\quad + P(Y_1 < x_1 + m, Y_2 > -x_2 - m) \\
 &= \Phi\left(\left(\frac{x_1}{k}\right)^{\frac{1}{\alpha}} + m, \left(\frac{x_2}{k}\right)^{\frac{1}{\alpha}} + m\right) \\
 &\quad - \Phi\left(\left(\frac{x_1}{k}\right)^{\frac{1}{\alpha}} + m, -x_2 + m\right) - \Phi\left(-x_1 + m, \left(\frac{x_2}{k}\right)^{\frac{1}{\alpha}} + m\right) + \Phi(-x_1 + m, -x_2 + m)
 \end{aligned}$$

y la función de densidad:

$$\begin{aligned}
 v(x_1, x_2) &= \frac{1}{k^2 \alpha^2} \left(\frac{x_1 x_2}{k^2} \right)^{\frac{1}{\alpha} - 1} \phi \left(\left(\frac{x_1}{k} \right)^{\frac{1}{\alpha}} + m, \left(\frac{x_2}{k} \right)^{\frac{1}{\alpha}} + m \right) \\
 &+ \frac{1}{k \alpha} \left(\frac{x_1}{k} \right)^{\frac{1}{\alpha} - 1} \phi \left(\left(\frac{x_1}{k} \right)^{\frac{1}{\alpha}} + m, -x_2 + m \right) \\
 &+ \frac{1}{k \alpha} \left(\frac{x_2}{k} \right)^{\frac{1}{\alpha} - 1} \phi \left(-x_1 + m, \left(\frac{x_2}{k} \right)^{\frac{1}{\alpha}} + m \right) \\
 &+ \phi(-x_1 + m, -x_2 + m)
 \end{aligned}$$

La Figura (4.3) se muestra la densidad de algunas cópulas bivariadas, en todos los casos $\rho = 0,8$ pero m va aumentando, tal que $m = 0, 1, 3, 15$ y 50 . Se puede notar que las cópulas son asimétricas y similares a los resultados de la distribución empírica obtenidas en la (3.3). Además, cada vez que incrementa m la distribución es mas simétrica, así que se puede concluir que si $m \rightarrow \infty$ la cópula converge a la cópula gaussiana, [Jing, 2010]. Para $m = 0, k = 1$ y $\alpha = 1$ la cópula que se genera es la cópula χ^2 , en la figura se ubica en la esquina superior derecha.

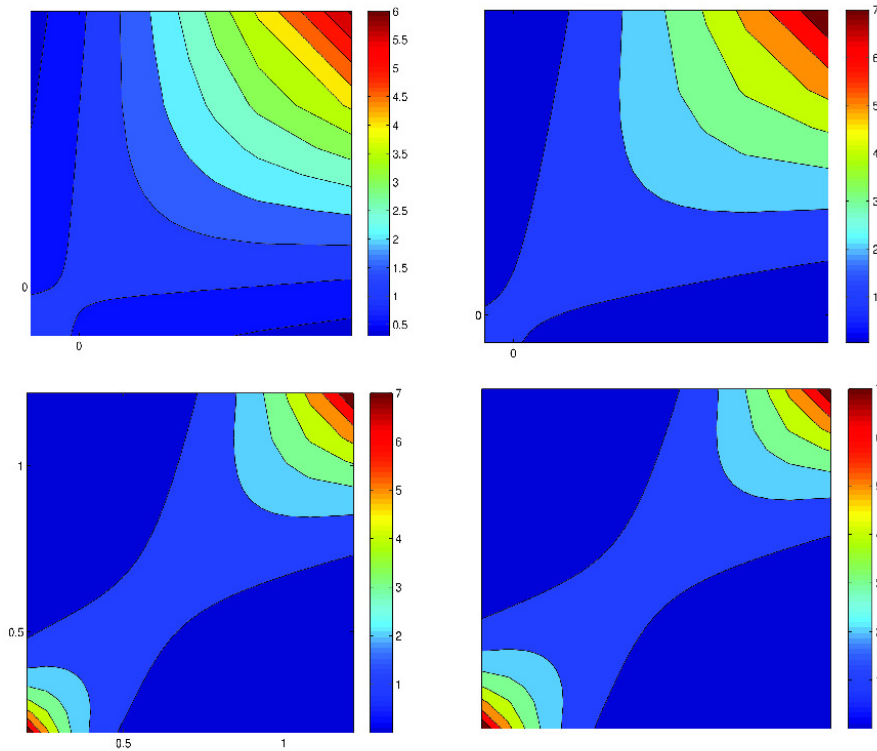


FIGURA 4.3. Cópula V-transformada Normal

Estimación de parámetros

Para realizar una estimación adecuada se deben considerar los tres tipos de parámetros que puede tener un modelo de copula espacial, según Kazianka, [Kazianka, 2010b], estos tipos son:

- θ que define el parámetro de correlación de la cópula,
- λ que define los parámetros de la cópula y,
- η correspondiente a los parámetros de tendencia espacial incluyendo los valores β de la regresión para la distribución marginal,

Por tanto, se puede notar, $C_{\mathbf{h}} = C_{\theta, \lambda}$ como la cópula y la marginal $F_Z = F_{\eta}$. Kazianka [Kazianka, 2010b] propone una estimación conjunta de todos los parámetros del modelo utilizando máxima verosimilitud tanto para funciones marginales discretas y continuas.

5.1. Inferencia para marginales continuas

Para la estimación de parámetros de las cópulas se debe considerar el vector de parámetros como:

$$\Theta = (\mu, \sigma, D_i^1, L_i^1, D_j^2, L_j^2), \quad (5.1)$$

realizando una partición en subconjuntos de tamaños aleatorios, tal que cada subconjunto S_k contiene $n(k) \leq 2$ observaciones, esto es

$$S_k = \{x_{k,1}, \dots, x_{k,n(k)}\} \quad \text{tal que } S_k \cap S_j = \emptyset \text{ si } k \neq j,$$

tal que,

$$S = \cup_{k=1}^K S_k = S,$$

para conjuntos S_k , la verosimilitud del vector de parámetros se obtiene calculando la densidad de la cópula:

$$c(S_k, \theta) = c(F_Z(Z(x_{k,1})), \dots, F_Z(Z(x_{k,n(k)}))), \theta$$

y,

$$L(\theta|Z(x_1), \dots, Z(x_n)) = \prod_{k=1}^K c(S_k, \theta).$$

Debido a que los parámetros de la cópula Gaussiana, V - transformada normal y χ^2 dependen de la estructura de correlación, se pueden estimar de la siguiente forma:

$$\Gamma = (\rho_{i,j})_{l,l}^{n,n}$$

donde, $\rho_{i,j}$ corresponde al coeficiente de correlación de Pearson que depende del vector $lags$, $h = x_i - x_j$:

$$\rho_{i,j} = R(h) = \sum_{j=0}^J D_j r_j(h, L_j),$$

donde, $r_j(h, L)$ denota la función de correlación correspondiente al modelo (Matérn, exponencial, entre otras), L representa el rango, D debe cumplir $\sum_{j=0}^J D_j = 1$ y representa la máxima contribución para cada modelo (sill).

Entonces, para el caso de funciones marginales continuas, se tiene que la verosimilitud puede ser evaluada de,

$$l(\Theta; D) = c_{\theta,\lambda}(F_\eta(z(x_1)), \dots, F_\eta(z(x_n))) \times \prod_{i=1}^n f_\eta(z(x_i)), \quad (5.2)$$

donde, $\Theta = (\theta, \lambda, \eta)$ el vector de parámetros, $c_{\theta,\lambda}$ es la función de densidad de la cópula y f_η la función de densidad de la marginal, y $D = \{z(x_1), \dots, z(x_n)\}$ un conjunto de realizaciones, por tanto, el estimador de máxima verosimilitud es:

$$\hat{\Theta} = \operatorname{argmax}_{\Theta} l(\Theta; D) \quad (5.3)$$

En general, este método no es muy eficaz computacionalmente, cuando se trata de un número grande de dimensiones es casi imposible ya que cópulas como la cópula χ^2 o la normal no poseen una forma definida para la función de distribución, además, se sabe por experiencia, que la estimación de parámetros para las funciones cópulas son óptimas dependiendo del tipo de cópula, por esta razón, se presenta diferentes formas para estimar los parámetros.

Kazianka, [Kazianka, 2010b] propone una metodología usando la probabilidad marginal y sus densidades modificando la probabilidad de la ecuación (5.2) para la estimación de parámetros para las cópulas χ^2 y Gaussiana. Bajo el supuesto de que los diferentes pares

de observaciones son independientes, se deberá maximizar la probabilidad de pares, de tal forma que:

$$l(\Theta; D) = \prod_{i,j \in \{1, \dots, n\}, i \neq j} c_{\theta, \lambda}(F_{\eta}(z(x_1)), \dots, F_{\eta}(z(x_n))) \times \prod_{k \in \{i, j\}} f_{\eta}(z_k) \quad (5.4)$$

tal que puede ser simplificada considerando las parejas preseleccionadas con una distancia de separación específica, volviendo la estimación más rápida computacionalmente.

Es claro que uno de los mayores problemas que se puede presentar con este tipo de metodología, es que suele no ser eficiente, por tanto, Song [Song, 2003] aplica una forma de lograr mayor eficacia realizando particiones donde se desarrolla el siguiente algoritmo, (adaptado de [Kazianka, 2010b]):

- Se realiza la siguiente partición:

$$l(\Theta; D) = l_w(\eta; D) + l_e(\Theta; D),$$

tal que

$$l_w(\eta; D) = \sum_{i=1}^n \log f_{\eta}(z(x_i)),$$

y,

$$l_e(\Theta; D) = \frac{1}{2} \mu_{\eta}^T (I^n - \Sigma^{-1}) \mu_{\eta} - \frac{1}{2} \log \det \Sigma,$$

con, $\mu_{\eta} = (\Phi^{-1}(F_{\eta}(z(x_1))), \dots, \Phi^{-1}(F_{\eta}(z(x_n))))$ y $\eta = (\beta, \sigma^2)$.

- Paso 1. Se obtiene η^1 resolviendo $\frac{\partial l_w(\eta; D)}{\partial \eta} = 0$,
- Paso 2. Se obtiene θ^1 resolviendo $\frac{\partial l_e(\eta; D, \eta^1)}{\partial \theta} = 0$,
- Paso k=2, 3, ... Se resuelve $\frac{\partial l_e(\eta; D, \eta^{k-1})}{\partial \theta} = 0$ para θ^k y $\frac{\partial l_w(\eta; D)}{\partial \eta} = -\frac{\partial l_e(\eta; D, \eta^{k-1})}{\partial \eta} \Big|_{\eta = \eta^{k-1}}$ para η^k , de esta forma se obtiene el vector de parámetros Θ^k .

Para el cálculo de las derivadas se debe implementar un algoritmo computacional, tal que cada derivada resulta:

$$\frac{\partial l_e(\Theta; D)}{\partial \beta} = \frac{\partial \mu_{\eta}^T}{\partial \beta} (I^n - \Sigma^{-1}) \mu_{\eta},$$

$$\frac{\partial l_e(\Theta; D)}{\partial \sigma^2} = \frac{\partial \mu_{\eta}^T}{\partial \sigma^2} (I^n - \Sigma^{-1}) \mu_{\eta},$$

$$\frac{\partial l_e(\Theta; D)}{\partial \theta_j} = \frac{1}{2} \left[\frac{\partial \Sigma}{\partial \theta_j} (\Sigma^{-1} \mu_{\eta} \mu_{\eta}^T \Sigma^{-1} - \Sigma^{-1}) \right],$$

donde el θ_j por lo general representan el parámetro de nugget, rango y la parámetro de suavizamiento del modelo de correlación de la función específica.

5.2. Selección de la mejor cópula

La selección de una copula apropiada para un conjunto de datos es un estudio que ha generado interés en muchas investigaciones. En [Genest, 2008b] considera que a este punto, las investigaciones se centran en procedimientos desarrollados para probar la dependencia de estructuras específicas como la cópula normal o la familia Clayton igualmente popular, usando parámetros arbitrarios, estadísticos de rango, kernels y demás, Genest [Genest, 2008a] usando un **blanket test** validó el método de selección para un conjunto de datos de la cópula apropiada. Lo siguiente es una adaptación del trabajo de Genest [Genest, 2008b].

El objetivo es probar la siguiente de hipótesis:

$$H_0 : C \in C_0 \quad (5.5)$$

tal que la estructura de dependencia esta representada por la familia paramétrica de cópulas C_0 . Dado que la cópula C de un vector aleatorio es invariante para transformaciones continuas, estrictamente crecientes, parece que la única opción razonable consiste en basar la inferencia en las estadísticas máximamente invariables con respecto a esta serie de transformaciones, es decir, rangos.

Una colección de funciones $U_1 = (U_{11}, \dots, U_{1d}), \dots, U_n = (U_{n1}, \dots, U_{nd})$ de pseudo-observaciones deducidas de rangos,

$$U_{ij} = \frac{R_{ij}}{n+1} = n \frac{\hat{F}_j(X_{ij})}{n+1},$$

donde el factor $\frac{n}{n+1}$ de escala se introduce para evitar problemas de valores límite en las cópulas. Las pseudo-observaciones U_1, \dots, U_n pueden ser interpretadas como una muestra de la cópula C .

Por tanto, este test realiza un procedimiento paramétrico de bootstrapping y el estadístico de Kolmogorov-Smirnov, T_n , ó el estadístico de Cramer-von Mises, S_n , tal que:

$$S_n = \int_{[0,1]^2} C'_n(u)^2 dC_n(u), \quad (5.6)$$

y,

$$T_n = \sup_{u \in [0,1]^2} |C'_n(u)|, \quad (5.7)$$

con $C'_n = \sqrt{n}(C_n - C_{\hat{\Theta}})$, donde C_n es la cópula empírica y $C_{\hat{\Theta}}$ es la estimación bajo la hipótesis nula.

En el marco geoestadístico, para realizar la selección de la cópula, se presenta el algoritmo propuesto por Kazianka, [Kazianka, 2010b]:

1. Para cada h_1, \dots, h_r se calcula la cópula empírica $C_n^{h_1}, \dots, C_n^{h_r}$,
2. Se estima los parámetros de la cópula teórica, para cada cópula del item anterior,
3. Se calcula los estadísticos Cramer-von Mises o Kolmogorov-Smirnov para cada *lag*, notados como $T_n^{h_1}, \dots, T_n^{h_r}$ ó $S_n^{h_1}, \dots, S_n^{h_r}$,
4. Para un número $N \gg$ se repiten los siguientes pasos:
 - a. Simular un campo aleatorio tal que la cópula es la cópula estimada en el paso 2,
 - b. Calcular la copula empírica $C_{n,k}^{h_1}, \dots, C_{n,k}^{h_r}$,
 - c. Se estima los parámetros de la cópula teórica de los campos simulados, se denotarán como $\hat{\Theta}_k$. Tal que para cada *lag* existirá una correspondiente cópula teórica $C_{\hat{\Theta}_k}^{h_1}, \dots, C_{\hat{\Theta}_k}^{h_r}$,
5. Para clases *lag* h_1, \dots, h_r , la aproximación del p-valor esta dado por:

$$p_{h_j} = \frac{1}{N} \sum_{k=1}^N I(T_{n,k}^{h_j} > T_n^{h_j}),$$

y,

$$p_{h_j} = \frac{1}{N} \sum_{k=1}^N I(S_{n,k}^{h_j} > S_n^{h_j}),$$

donde, I es una función indicadora y $j = 1, \dots, r$.

La necesidad de crear la cópula empírica y luego calcular sus estimaciones hace que el proceso sea computacionalmente difícil debido a que se debe asignar *lags* que sean $\mathbf{h} \gg$, no obstante, es un método eficaz ya que proporciona una mayor confiabilidad para el procesos de selección de cópula.

Interpolación espacial

Uno de los propósitos de geoestadística, es lograr realizar predicción, los métodos clásicos, dependen de la elección del modelo de variograma y los mas usuales son los métodos de krigeado. En este trabajo, no se considera variogramas, Bárdossy [Bárdossy, 2008] y Kazianka [Kazianka, 2010b] proponen diferentes tipos métodos para interpolación espacial usando cópulas que aquí adoptaremos.

6.1. Kriging Indicador y Kriging disyuntivo

El indicador Kriging es un estimador de la probabilidad acumulada hasta el limite z , definido en la función indicadora ([Giraldo, 2010]), esto se hace por cokriging de variables indicadoras $I(Z(x) \leq z_j)$ donde los valores de z_j son cuantiles.

Los cross-/auto-covariances se puede derivar de una cópula espacial a través de:

$$\gamma_{z_j}(x_1, x_2) = C_{\hat{\theta}, \hat{\lambda}}(F_{\hat{\eta}}(z_j; x_1), F_{\hat{\eta}}(z_j; x_2)) - F_{\hat{\eta}}(z_j; x_1)F_{\hat{\eta}}(z_j; x_2), \quad (6.1)$$

$$\gamma_{z_j, z_k}(x_1, x_2) = C_{\hat{\theta}, \hat{\lambda}}(F_{\hat{\eta}}(z_j; x_1), F_{\hat{\eta}}(z_k; x_2)) - F_{\hat{\eta}}(z_j; x_1)F_{\hat{\eta}}(z_k; x_2), \quad (6.2)$$

tal que $F_{\hat{\eta}}(\cdot : x)$ puede depender de la localización x . En este caso $\gamma_{z_j}(x_1, x_2)$ y $\gamma_{z_j, z_k}(x_1, x_2)$ son no estacionarios. Si se asume tendencia constante, entonces las ecuaciones anteriores quedarían simplificadas de la siguiente forma:

$$\gamma_{z_j}(h) = C_{\hat{\theta}, \hat{\lambda}}(F_{\hat{\eta}}(z_j), F_{\hat{\eta}}(z_j)) - F_{\hat{\eta}}(z_j)^2, \quad (6.3)$$

$$\gamma_{z_j, z_k}(h) = C_{\hat{\theta}, \hat{\lambda}}(F_{\hat{\eta}}(z_j), F_{\hat{\eta}}(z_k)) - F_{\hat{\eta}}(z_j)F_{\hat{\eta}}(z_k), \quad (6.4)$$

Al unir las ecuaciones (6.1) y (6.2) en el procedimiento de cokriging, se llega al kriging indicador, basado en el modelo de cópula espacial. Por otro lado, una de las evidentes ventajas del uso de cópulas es que las ecuaciones correspondientes (6.1) y (6.2) son válidas

tanto para marginales discretas como continuas.

Para el caso discreto, tomando en un número finito de valores m , cada función $f(Z(x))$, se tiene:

$$f(Z(x)) = \sum_{i=1}^m f_i I(Z(x) \leq i), \quad (6.5)$$

El estimador del kriging disyuntivo (DK) y el de cokriging (CK) se obtienen de la siguiente forma:

$$[f(Z(x))]^{DK} = \sum_{i=1}^m f_i [I(Z(x)) \leq i]^{CK}, \quad (6.6)$$

tal que el kriging disyuntivo esta definido por la cópula bivariada.

6.2. Plug-in de estimación

Este tipo de estimación es quizá, el mas completo que se ha visto hasta ahora, propuesto por Kazianka [Kazianka, 2010b] en el marco bayesiano se puede tener una distribución completa de predicción para cada variable de $V(x_0)$ en la localización desconocida x_0 , la cual, tomaría la forma de:

$$p(v(x_0)|D) = \int p(v(x_0)|\Theta, D)p(\Theta|D)d\Theta. \quad (6.7)$$

Además, se puede tomar $p(v(x_0)|D) = p(v(x_0)|\hat{\Theta}, D)$.

En el modelo de cópula espacial la $p(v(x_0)|D)$ es la densidad de la cópula condicional de $V(x_0)$, tal que

$$c_{\hat{\theta}, \hat{\lambda}}(v(x_0)|D) = \frac{c_{\hat{\theta}, \hat{\lambda}}(v(x_0), v(x_1), \dots, v(x_n))}{c_{\hat{\theta}, \hat{\lambda}}(v(x_1), \dots, v(x_n))}, \quad (6.8)$$

donde $v(x_i) = F_{\hat{\eta}}(z(x_i))$, con $i = 1, \dots, n$ y debido a que la función $V(s)$ esta definida en $[0, 1]$, entonces, no puede ocurrir que el rango no pertenezca al conjunto $[0, 1]$.

La cópula condicional se puede escribir de la siguiente forma:

$$c_{\hat{\theta}, \hat{\lambda}}(v(x_0)|D) = \frac{d(F^{-1}(v(x_0))|\hat{\Theta}, D)}{f(F^{-1}(v(x_0)))}, \quad (6.9)$$

tal que la cópula es construida por una marginal F con densidad f y densidad condicional d por ejemplo, la cópula Gaussiana esta construida por $F = \Phi$, $f = \phi$ y la densidad Gaussiana $d = \phi_{\mu, \sigma^2}$ con media $\mu = \Sigma_{12}\Sigma_{22}^{-1}a$ y varianza $\sigma^2 = 1 - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ donde, Σ_{22} es la matriz de correlación de las localizaciones conocidas, $\Sigma_{12} = \Sigma_{21}^T$ es el vector

de correlaciones entre las localizaciones conocidas y donde se tomará la predicción, $a = (\Phi^{-1}(v(x_1)), \dots, \Phi^{-1}(v(x_n)))^T$ (Tomado de [Kazianka, 2010b]).

La densidad de $Z(x_0)$ puede ser calculada usando la transformación Jacobiana. El determinante del Jacobiano es la densidad f_z ,

$$p(z(x_0)|\hat{\Theta}, D) = c_h(F_Z(z(x_0))|\hat{\Theta}, D)f_Z(z(x_0)). \quad (6.10)$$

El estimador Bayesiano para $Z(x_0)$ bajo la pérdida cuadrática es la media de la distribución predictiva,

$$\hat{Z}(x_0) = E(Z(x_0)|\hat{\Theta}, D). \quad (6.11)$$

Kazianka [Kazianka, 2010b] deriva un estimador similar al rank order kriging, de tal forma que resulta:

$$\hat{Z}(x_0) = \int_0^1 F_Z^{-1}(v(x_0))c_h(v(x_0)|\hat{\Theta}, D)dv(x_0). \quad (6.12)$$

La predicción de la varianza $\hat{\sigma}^2(x_0)$ puede ser calculada como:

$$\hat{\sigma}^2(x_0) = \int_0^1 (F_Z^{-1}(v(x_0)) - \hat{Z}(x_0))^2 c_h(v(x_0)|\hat{\Theta}, D)dv(x_0). \quad (6.13)$$

Aplicación de cópulas: Catástrofe en Chernóbil

En abril de 1986, ocurrió el peor accidente nuclear en Chernóbil en la ex USSR (ahora Ucrania). La central nuclear de Chernóbil, situada a 100 kilómetros al norte de Kiev, tenía 4 reactores. En un día de abril a las 1:23 am la reacción en cadena en un reactor perdió el control, la creación de explosiones y una bola de fuego, que voló el acero pesado del reactor y la tapa de concreto. El desastre destruyó el reactor Chernóbil-4 y mató a 30 personas, entre ellas 28 por exposición a la radiación, 209 más fueron tratados por envenenamiento agudo por radiación y entre estos, 134 casos fueron confirmados. Grandes áreas de Bielorrusia, Ucrania, Rusia y más allá estaban contaminados en diversos grados. El desastre de Chernóbil fue un evento único y el único accidente en la historia de la energía nuclear comercial. Ahora 26 años después del accidente todavía más de 3 millones de niños sufren en de estos efectos, la zona alrededor del reactor todavía está muy contaminada, la naturaleza está muerta y no hay vida silvestre.

7.1. Datos de Chernóbil

En la región de Gomel, el gobierno ruso construyó una red donde se estudia la concentración de la cantidad de radiactividad. El conjunto de datos que utilizamos son mediciones de Cs137, un isótopo radiactivo.

Se analiza el conjunto de datos que corresponde a 148 localizaciones $x_i = (x_{1i}, x_{2i})^T$, $i = 1, \dots, 148$ en la región de Gomel. Los datos son observados diez años después del accidente de Chernóbil. En la Figura (7.1) se muestra las realizaciones donde se encontró el isótopo radiactivo,

Se puede ver que la mayoría de los valores son pequeños, sin embargo, en la parte noreste, noroeste y sur de la región algunos valores relativamente grandes se producen.

El análisis de datos en copulas bivariadas permite realizar predicción, en la Figura (7.2) se muestra las localizaciones observadas marcadas con una x roja, y los puntos azules la grilla de interpolación, donde se realizarán las respectivas predicciones.

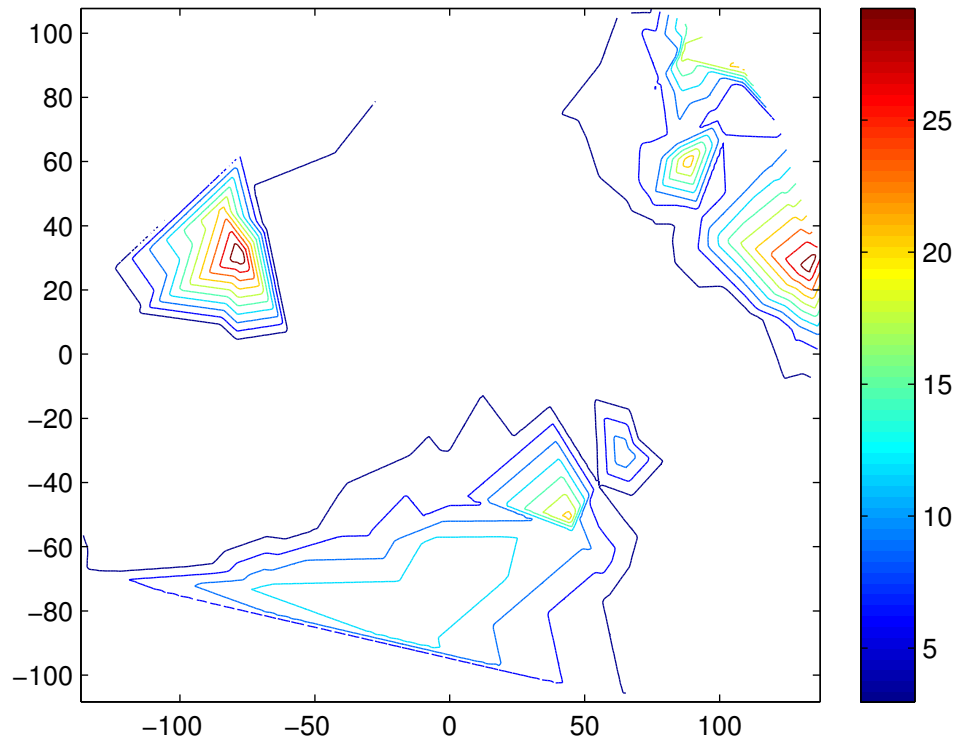


FIGURA 7.1. Observaciones de los datos de Gomel

7.2. Cópula Gaussiana

Siguiendo a [Kazianka, 2010b], para encontrar un marginal univariada apropiada, se prueba las distribuciones univariadas normal, gamma, el valor generalizado extremo (GEV) y la distribución de log-normal, para todas ellas se calcula las estimaciones de máxima verosimilitud, asumiendo que las observaciones son independientes. La familia que corresponde al máximo de todos los valores de probabilidad maximizadas se selecciona. Los valores obtenidos son, Log-Normal = $-285,38$, Gamma = $-288,85$, Normal = $-432,30$, Box-cox = $-285,38$.

Por tanto, la elección que se toma será de la distribución marginal Box-cox con parámetro $\gamma = 0,0032$.

La selección de una cópula Gaussiana se puede justificar por medio del ajuste de bondad propuesto en el capítulo 5, es recomendable realizar un número grande de simulaciones, pero en este caso es casi imposible ya que la complejidad computacional es muy alta, por tanto, según Genest [Genest, 2008a], se pueden realizar 150 simulaciones, los resultados de la prueba muestran que (Ver Tabla (7.2)):

Los valores de la prueba de Kolmogorov-Smirnov parecen ser menores que los valores de p para el de Cramer-von Mises. Esto es debido a que la prueba de Kolmogorov-Smirnov

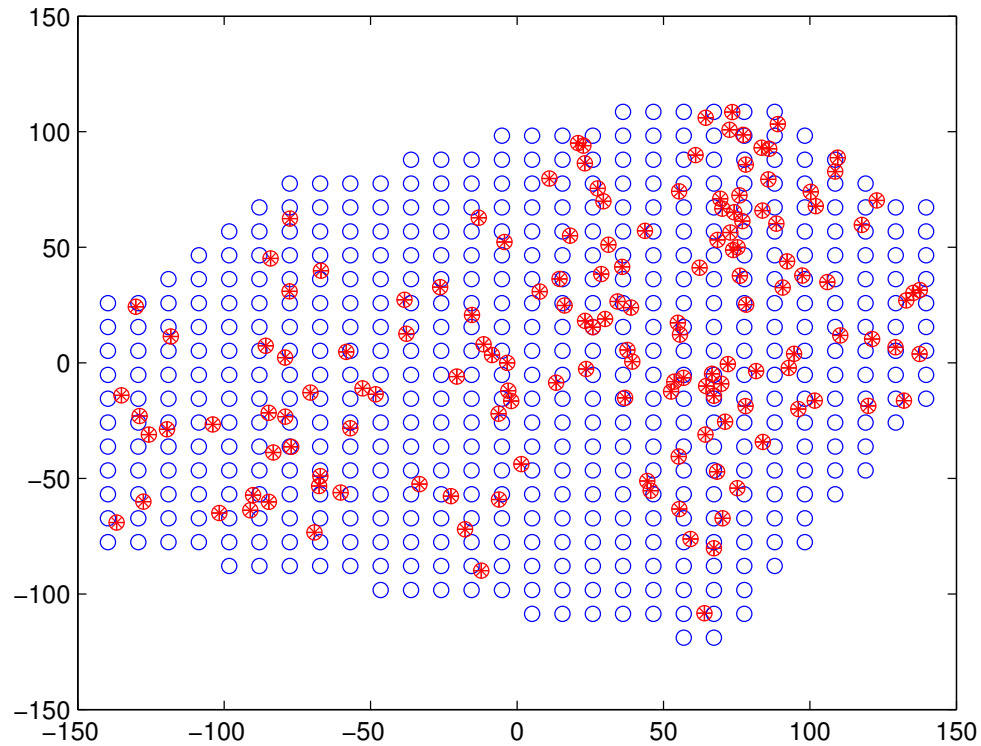


FIGURA 7.2. Observaciones y datos interpolados de los datos de Gomel

| Valor del estadístico | | | |
|-----------------------|--------|---------|----------|
| h | 0 – 10 | 10 – 20 | 90 – 100 |
| T_n | 0,99 | 0,99 | 0,9867 |
| S_n | 21,66 | 35,1352 | 79,07 |

TABLA 7.1. Valores de los estadísticos T_n y S_n

es insensible a valores extremos. Se puede asumir que el modelo propuesto se ajusta a los datos. A pesar de esto, se puede comparar en la Figura (7.3) la cópula empírica con la teórica, mostrando de manera gráfica que el modelo no puede ser simétrico.

Aun así, se realiza la estimación de los parámetros, se utiliza el modelo de correlación Matérn incluyendo un término efecto de pepita,

$$\rho(h) = \nu_2 I(h=0) + \frac{1 - \nu_2}{2^{\kappa-1} \Gamma(\kappa)} \left(\frac{2\kappa^{1/2} h}{\nu_1} \right)^{\kappa} K_{\kappa} \left(\frac{2\kappa^{1/2} h}{\nu_1} \right) \quad (7.1)$$

tal que $\nu_1 \in [0, \infty]$ corresponde al parámetro de rango, $\nu_2 \in [0, 1]$ el parámetro del efecto pepita y κ es el parámetro de suavizamiento. Entonces se debe calcular cinco parámetros, los parámetros correspondientes a la transformación Box–cox, γ , y los parámetros de la función de correlación $\theta = (\nu_1, \nu_2, \kappa)$. Los resultados de la estimación son: $\hat{\gamma} = 0,090$,

| Resultado p -valor con 95 % | | | |
|-------------------------------|--------|---------|----------|
| \mathbf{h} | 0 – 10 | 10 – 20 | 90 – 100 |
| T_n | 0,33 | 0,999 | 0,99 |
| S_n | 0,99 | 0,99 | 0,999 |

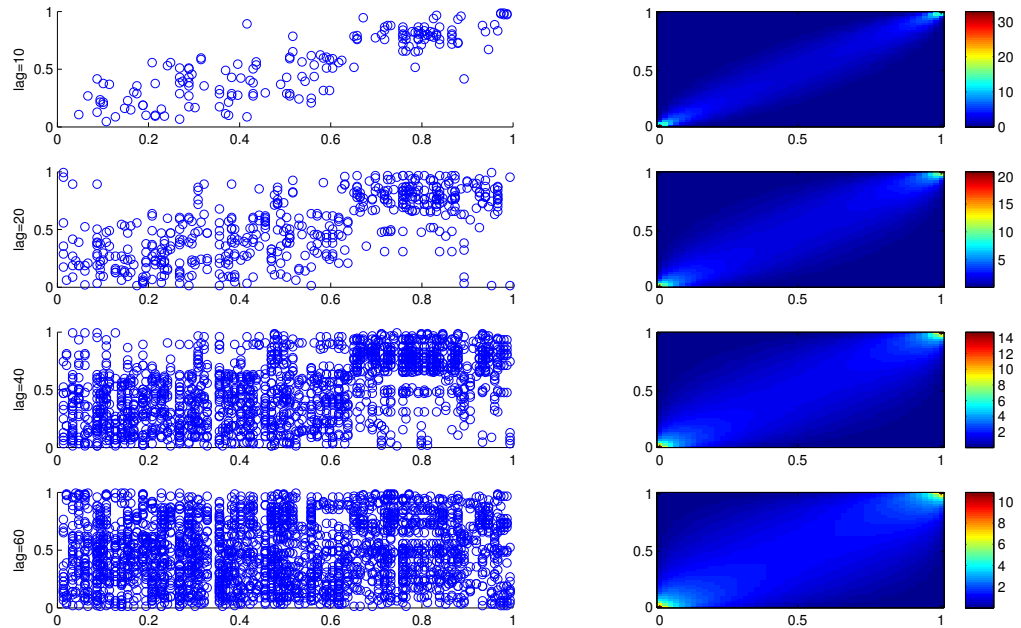
TABLA 7.2. Valor p los estadísticos T_n y S_n 

FIGURA 7.3. Gráfico de dispersión de los pares de datos de rango transformados, es decir, densidad de la cópula empírica bivariado (columna de la izquierda), cópula teórica bivariada Gaussiana (columna de la derecha)

$$\hat{\nu}_1 = 61,92, \hat{\nu}_2 = 0,0539 \text{ y } \hat{\kappa} = 0,8650.$$

Se considera el campo isotrópico y se realiza la correspondiente interpolación espacial Plug-in: Las Figuras (7.4) y (7.5) muestran la media y la desviación estándar de los valores predichos para el modelo de cópula Gaussiano,

Las Figuras (7.4) y (7.5) revelan que la media del modelo espacial de la cópula Gaussiana en la parte sur y noroeste aumentan y las desviaciones estándar se concentran.

El MSE es de 17,9995. Este valor podrá ser usado para comparar con las cópulas no Gaussianas el ajuste del modelo.

Otra forma de verificar los resultados de la predicción es la validación cruzada de los intervalos de confianza, puede servir como una herramienta para la comparación cualitativa de las actuaciones de predicción. En la Figura (7.6) se muestra la esa validación mostrando que posee un porcentaje de cobertura que está siempre muy cerca del nivel nominal (línea azul).

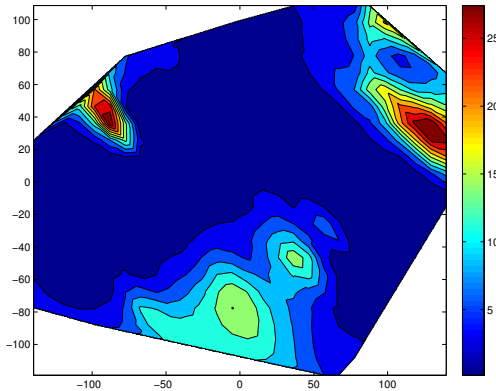


FIGURA 7.4. Media de la predicción con la cópula Gaussiana

FIGURA 7.5. Desviación estándar de la predicción con la cópula Gaussiana

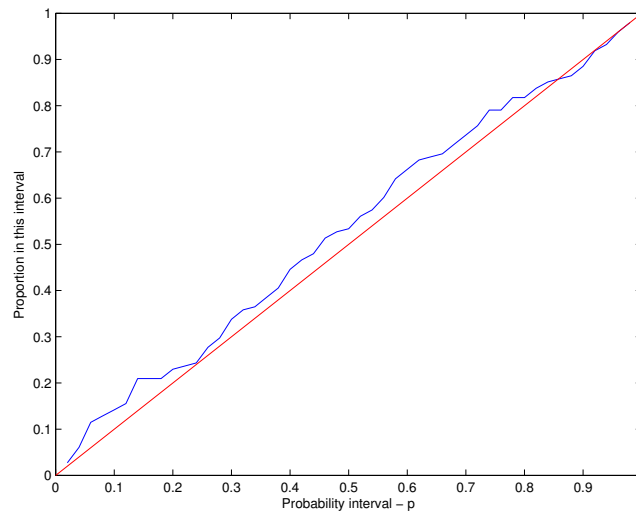


FIGURA 7.6. Ancho de los intervalos de confianza de validación cruzada para la cópula Gaussiana

Otro método para comparar la calibración de los diferentes modelos es el valor PIT. Este es el valor de la supresión basado en la distribución predictiva, la ventaja de utilizar el valor PIT es la capacidad de relacionarse con el diagnóstico de la distribución espacial, la trama de los valores PIT pueden indicar un rendimiento deficiente del modelo en una subregión de la región de interés (Tomado de [Kazianka, 2010b]). Los valores de PIT que quedan fuera del intervalo de confianza del 95 % se muestran resaltados. Para la cópula Gaussiana la Figura que representa el valor de PIT se muestra en la Figura (7.7). Los valores en rojo indican subpredicción mientras que los valores en azul indican valores de sobreestimación:

Se puede ver que no existe mayor predicción en la región norte–occidental como se esperaba, ya que existen pocas observaciones en esta área. De la misma forma sucede con

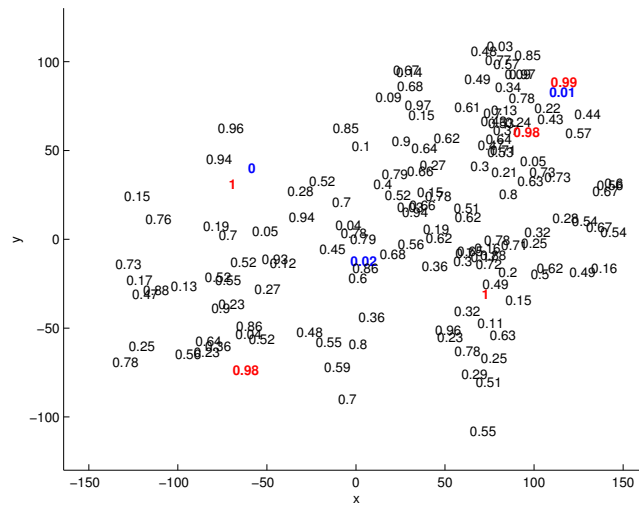


FIGURA 7.7. Valores de validación cruzada PIT para la cópula Gaussiana

la región sur–este.

7.3. Cópula no Gaussiana: V –transformada normal

Para la cópula no Gaussiana, se estudiará la cópula V –transformada, se selecciona la distribución marginal log Normal, con esto se pretende tomar en cuenta la propiedad fundamental de las cópula que permite a la estructura de dependencia no ser influenciada con las distribuciones marginales. De la misma forma que en la anterior sección, la selección de la cópula V se puede justificar por medio del ajuste de bondad propuesto en el capítulo 5, los resultados de la prueba muestran que (Ver Tabla (7.3)):

| Valor del estadístico: | | | |
|------------------------|--------|---------|----------|
| \mathbf{h} | 0 – 10 | 10 – 20 | 90 – 100 |
| T_n | 0,99 | 0,99 | 0,98 |
| S_n | 22,37 | 44,91 | 88,07 |

TABLA 7.3. Valores de los estadísticos T_n y S_n

| Resultado p –valor con 95 % | | | |
|-------------------------------|--------|---------|----------|
| \mathbf{h} | 0 – 10 | 10 – 20 | 90 – 100 |
| T_n | 0,99 | 0,999 | 0,99 |
| S_n | 0,99 | 0,99 | 0,99 |

TABLA 7.4. Valor p de los estadísticos T_n y S_n

Realizando una comparación con los p -valores de la cópula Gaussiana, se puede notar que son mas bajos que los de la cópula V -transformada. Por tanto, se puede asumir que el modelo propuesto se ajusta a los datos. Se puede comparar en la Figura (7.8) la cópula empírica con la teórica, mostrando un mejor ajuste que la cópula Gaussiana de la (7.3).

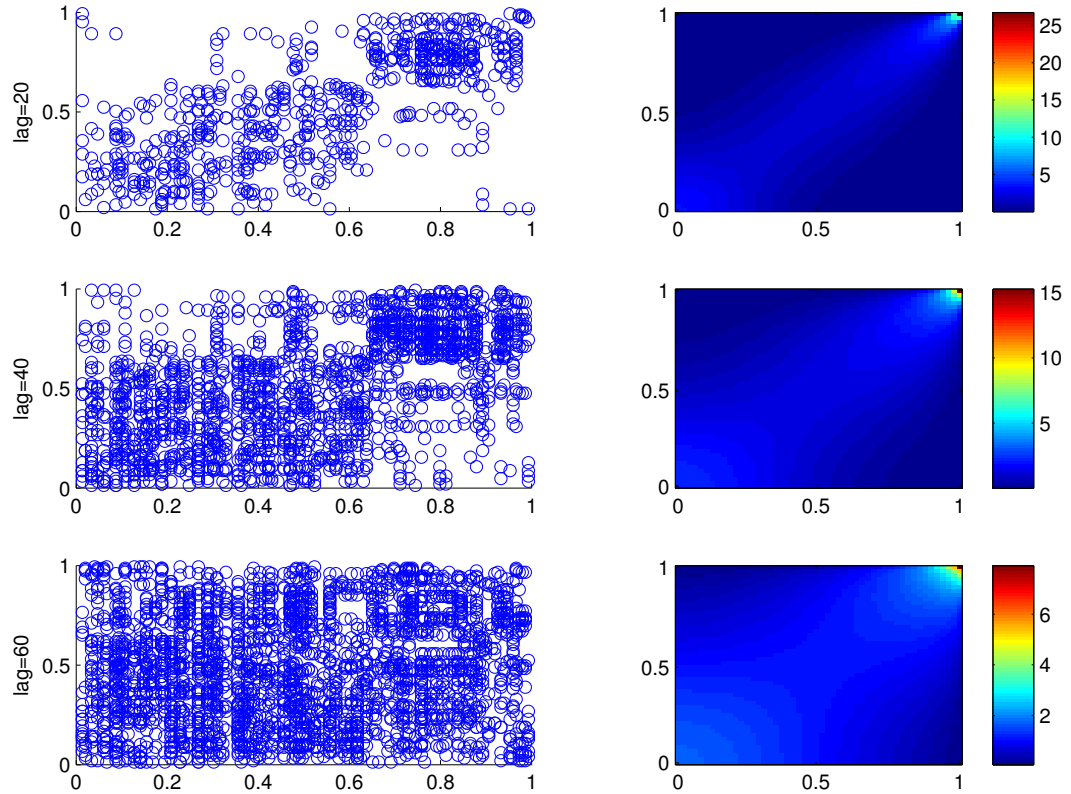


FIGURA 7.8. Gráfico de dispersión de los pares de datos de rango transformados, es decir, densidad de la cópula empírica bivariado (columna de la izquierda), cópula teórica bivariada χ^2 -cópula (columna de la derecha)

En este caso, se debe calcular 6 parámetros, los parámetros correspondientes a la marginal Log Normal μ y σ^2 , a la función de correlación, que en este caso se usara Matern, (7.1) $\theta = (\nu_1, \nu_2, \kappa)$ y el parámetro correspondiente a la cópula V , $m = 1, 27$. Se considera el caso especial para $k = 1$ y $\alpha = 1$, es decir la cópula χ^2 debido a la facilidad computacional, Los resultados de las estimaciones son: $\hat{\mu} = 0,595$, $\hat{\sigma} = 1,37$, $\hat{\nu}_1 = 100,023$, $\hat{\nu}_2 = 0,0576$ y $\hat{\kappa} = 10$.

Después de estimar los valores de los parámetros, se puede realizar la correspondiente predicción en los puntos de la Figura (7.2), se muestra la media y la desviación estándar correspondientes en las Figuras (7.9) y (7.10), donde se puede observar que el modelo espacial de la cópula Gaussiana y la cópula χ^2 muestran resultados diferentes en la parte sur y noroeste de la región, la desviación estándar en el modelo χ^2 lleva a los medios de predicción similares o incluso más altos que el modelo gaussiano espacial en el sur y en

el noreste, las desviaciones estándar son significativamente más pequeños. La desviación estándar en la cópula χ^2 es sólo mayor en la zona noroeste debido a los muy altos valores pronosticados y el pequeño número de observaciones.

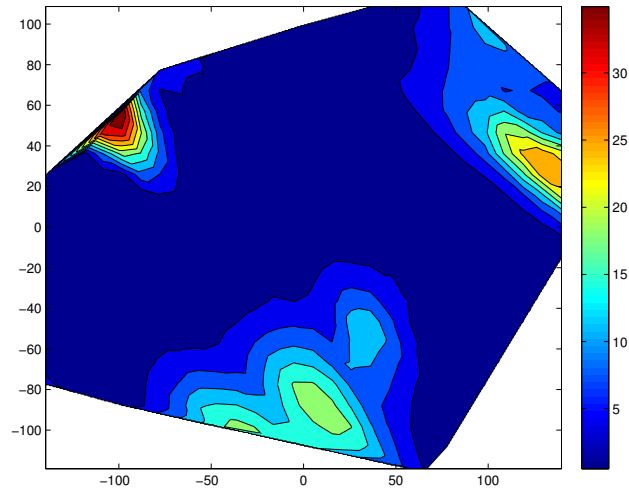


FIGURA 7.9. Media de la predicción con la cópula V

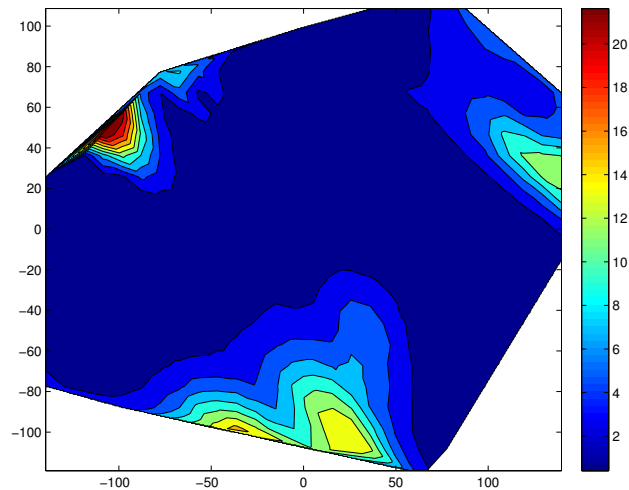


FIGURA 7.10. Desviación estándar de la predicción con la cópula V

En la Figura (7.11) se muestra la validación cruzada, se puede notar que posee un porcentaje de cobertura mayor que el de la cópula Gaussiana ya que está siempre cerca del nivel nominal (línea azul). También, se puede encontrar mejores resultados en la Figura (7.12) de valores de PIT.

Las probabilidades predichas son diferentes para el modelo de cópula χ^2 y la cópula Gaussiana sobre todo en la parte sur y noroeste de la región, donde el modelo de cópula

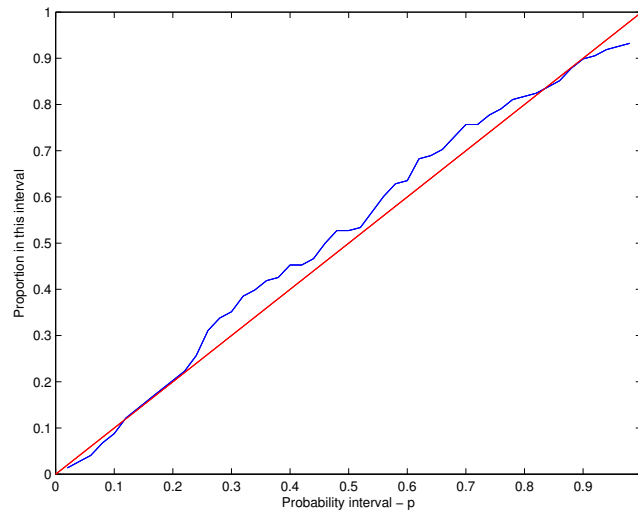


FIGURA 7.11. Gomel datos: ancho de los intervalos de confianza de validación cruzada para la cópula χ^2

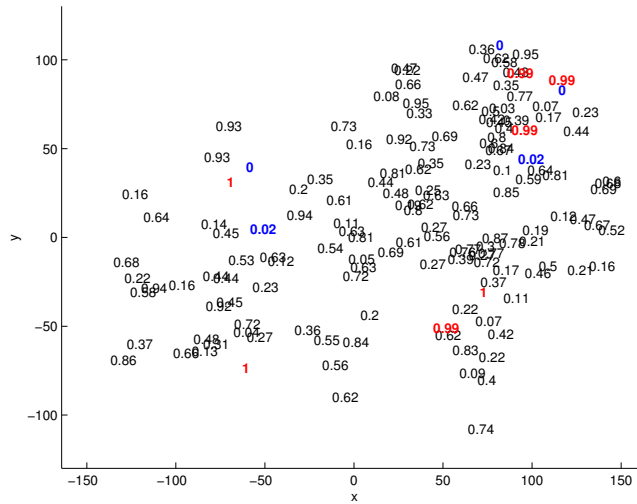


FIGURA 7.12. Valores de validación cruzada PIT para la cópula χ^2

χ^2 tiene una menor probabilidad de tomar los valores por debajo de los umbrales.

El valor de: $MSE = 16,835$, que resulta menor que el valor para la cópula Gaussiana. Por otro lado, se observa que los intervalos de confianza son mucho más cortos para los modelos basados en la cópula V - transformada, un hecho que también se refleja en las predicciones de las desviaciones estándar.

7.4. Cópula Gaussiana incluyendo anisotropía

Hasta ahora, solo se han considerado campos isotrópicos y los modelos de cópulas se han ajustado bien a los datos, pero se debe recordar que considerar modelos isotrópicos no es una premisa realista, ya que el campo puede carecer de esta característica, por tanto, en esta sección se considera anisotropía, entonces se deben estimar los parámetros a y ϕ correspondientes a la transformación:

$$(x_{i1}^{iso}, x_{i2}^{iso})^T = \begin{pmatrix} \cos \varphi & \sin \varphi \\ -a \sin \varphi & a \cos \varphi \end{pmatrix}, \quad (7.2)$$

Como consecuencia, se deben estimar siete parámetros, lo cual puede generar problemas computacionales, se recomienda máxima verosimilitud para el caso de la copula Gaussiana. En este caso, se realiza una transformación Box-cox con parámetros $\hat{\mu}$ y $\hat{\sigma}$, la función de correlación, que en este caso se usara Matern, (7.1) $\theta = (\nu_1, \nu_2, \kappa)$, los parámetros correspondientes a la transformación de anisotropía, a y ϕ , por tanto: $\hat{\mu} = 0,08551$, $\hat{\sigma} = 1,6314$, $\hat{\nu}_1 = 58,227$, $\hat{\nu}_2 = 0,343$, $\hat{\kappa} = 0,05$, $\hat{\phi} = 2,37$ y $\hat{a} = 1,34$.

Después de realizar la estimación de los parámetros, se realiza la predicción de la media y la desviación estándar, en las Figuras (7.13) y (7.14) y con base en los modelo cópula y la grilla de predicción definida en la Figura (7.2) se presentan los resultados de predicción.

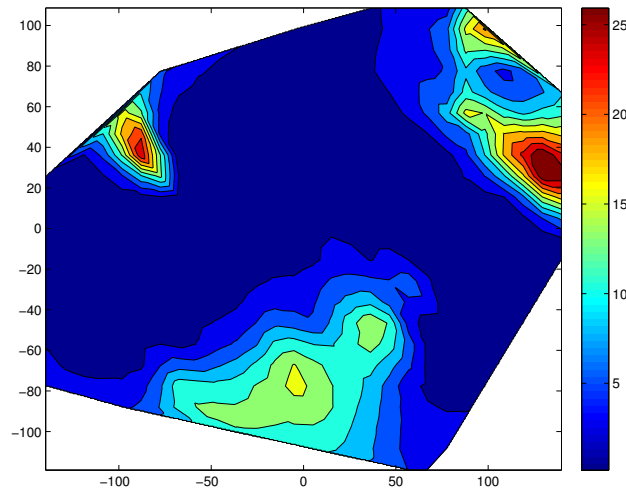


FIGURA 7.13. Media de la predicción con la cópula Gaussiana con anisotropía.

7.5. Cópula no Gaussiana incluyendo anisotropía

De manera análoga a la cópula gaussiana, se modela con la cópula V incluyendo anisotropía, se estiman los parámetros correspondientes al modelo, esta cópula es considerablemente mas pesada computacionalmente, se deben estimar los siete parámetros usuales y el parámetro correspondiente a la cópula V , $\hat{m} = 1,27$. Los resultados son:

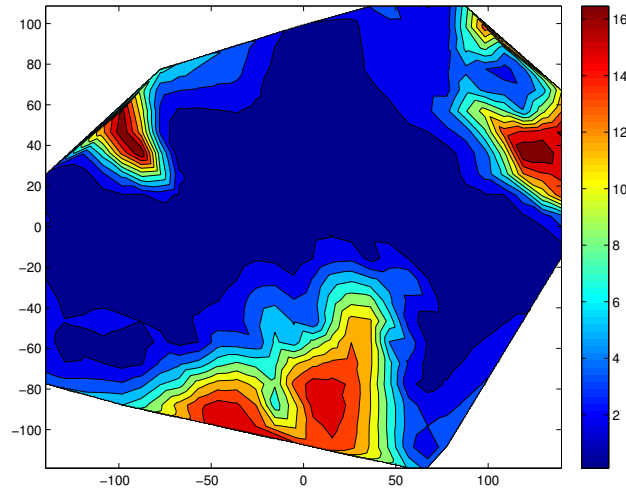


FIGURA 7.14. Desviación estándar de la predicción con la cópula Gaussiana con anisotropía.

$$\hat{\mu} = 0,001, \hat{\sigma} = 1,37, \hat{\nu}_1 = 69,223, \hat{\nu}_2 = 0,05, \hat{\kappa} = 10, \hat{\phi} = 2,37, \hat{a} = 1,34 \text{ y } \hat{\lambda} = 1.$$

Se realiza la predicción de la media y la desviación estándar. Con base en los modelo cópula y la grilla de predicción definida en la Figura (7.2), se presentan estos resultados en las Figuras (7.15) y (7.16):

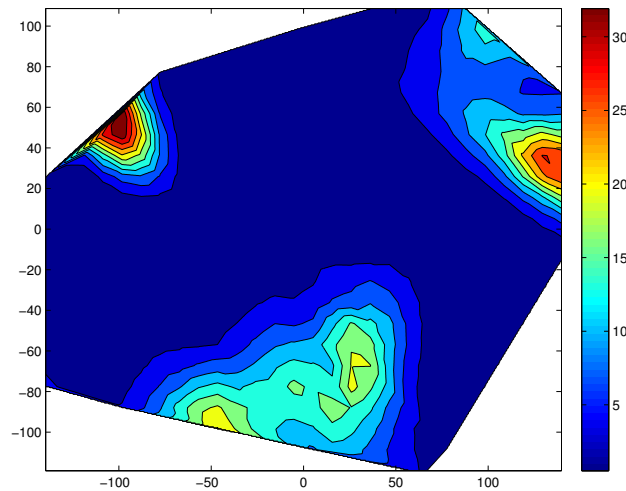


FIGURA 7.15. Media de la predicción con la cópula no Gaussiana con anisotropía.

7.6. Cópula Gaussiana con tendencia

En esta sección se consideran modelos con tendencia espacial. Se realizará una parametrización de μ de la distribución log normal, que a sido escogida para ser la

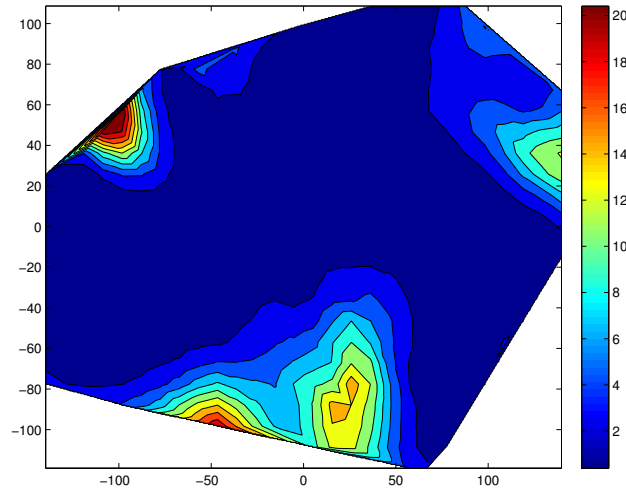


FIGURA 7.16. Desviación estándar de la predicción con la cópula no Gaussiana con anisotropía.

distribución marginal del modelo. Cabe recordar, que es necesario que la distribución marginal tenga el parámetro μ explícito.

EL análisis exploratorio de datos, muestra una posible tendencia cuadrática, tal que se μ queda definida para la i -ésima observación:

$$\mu_i(x_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i}^3 + \beta_4 x_{4i}^4 + \beta_5 x_{1i} x_{2i}, \quad (7.3)$$

donde $\beta = (\beta_0, \dots, \beta_5)^T$ es el vector de regresión.

Para realizar esta estimación no es necesario considerarla en el modelo cópula, ya que se tendrían cinco variables mas y seria pesado computacionalmente, por tanto, se estiman por mínimos cuadrados en el software Rcran.

Los resultados que generan son $\hat{\beta}_0 = 0,9260$, $\hat{\beta}_1 = 0,092$, $\hat{\beta}_2 = -0,0034$, $\hat{\beta}_3 = 0,0003$, $\hat{\beta}_4 = 0,0004$ y $\hat{\beta}_5 = 0,0001$.

Después de realizar la correspondiente predicción con la copula gaussiana y correlación exponencial, se nota que el valor de $MSE = 17,62$, la cual resulta no ser buena en comparación con la cópula V .

En conclusión, realizar un modelo de tendencia no es un proceso óptimo para este conjunto de datos, ya que no genera mejores resultados que la cópula V y el proceso de estimación e interpolación no es óptimo debido a que para la cópula V se debe estimar doce parámetros.

7.7. Resumen

En ausencia de datos de prueba que se utiliza para realizar una adecuada validación cruzada como un método cuantitativo para evaluar el desempeño del modelo, utilizamos el valor de MSE y los resultados obtenidos del Kriging ordinario por [Kazianka, 2010b]:

$$MSE = \frac{1}{2} \sum_{i=1}^n (\hat{Z}_{-i}(x_i) - z(x_i))^2 \quad (7.4)$$

| MSE /MAE | V -transformada | Gaussiana | Kriging Ordinario |
|-------------|-------------------|--------------|-------------------|
| Isotropía | 16,835/2,019 | 17,9995/2,01 | 2,23 |
| Anisotropía | 16,9404/1,90 | 19,0244/2,03 | 2,75 |

TABLA 7.5. Valores de MSE y MAE para modelos de dependencia

Es evidente que el modelo de la copula V transformada posee el valor mas bajo de MSE independiente de si se considera anisotropía geométrica. Las observaciones que toman valores entre 3 y 20 se predicen con mucha más precisión en el caso de la copula V transformada.

Por tanto, el kriging ordinario es claramente el peor método. Y se puede decir, que los modelos de cópula se ajustan bien a los datos, pero la cópula gaussiana es mas fácil de utilizar que la cópula V en cuanto a nivel de exigencia computacional, ya que la cópula V presento graves problemas que en muchos casos hicieron imposible el calculo. A pesar de esto, es la que mejor se ajusta, la razón es debido a su comportamiento radialmente asimétrico y que los datos cumplen con esta condicon. La inclusión de la anisotropía geométrica y modelos de tendencia conduce a un mejor rendimiento predictivo. La metodología clásica kriging ordinario no es apropiado para este conjunto de datos, ya que no tiene en cuenta que en el lado derecho de la distribución esta sesgado.

Cóputas Discretas

En los capítulos anteriores se a descrito la forma en la que se utilizan las cóputas para construir estructuras de dependencia, pero únicamente se consideran variables aleatorias con distribuciones marginales continuas y en algunos casos se puede presentar la necesidad de implementar otro tipo de cóputas.

En el artículo de Genest [Genest, 2007] se proponen las bases teóricas para cóputas discretas, el cual expone las ventajas y un método de estimación de parámetros, un trabajo similar se encuentra en Berkes [Berkes, 2009] el cual realiza un análisis estadístico de dependencia a un conjunto de datos de respuestas neuronales con cóputas y marginales que se distribuyen Poisson. Panagiotelis [Panagiotelis, 2011], realiza un estudio similar con marginales discretas y extiende su construcción hacia un tipo de cóputas llamadas “vine cóputas”.

En algunos casos geoestadísticos, se puede tener variables aleatorias cuyas realizaciones pertenezcan al conjunto de los números naturales de tal forma, que los mas conveniente es utilizar distribuciones discretas. En el marco de esta tesis, donde el principal objetivo es utilizar cóputas para datos geoestadísticos y siguiendo el teorema de Sklar, [1] el cual garantiza la existencia de la cóputa para una función de distribución conjunta $H(X, Y)$ de las variables aleatorias X e Y , se propone una extensión a este teorema considerando X y Y variables aleatorias discretas. Kazianka, [Kazianka, 2010b] introduce cóputas en geoestadística para marginales discretas, en base en este último artículo realizamos este capítulo, sin embargo, optaremos por una metodología diferente para realizar la inferencia y estimación de parámetros, ya que la complejidad computacional de Kazianka es alta.

8.1. Existencia y Unicidad

Sea X e Y las variables aleatorias con distribución conjunta H con marginales F y G , es decir,

$$H(x, y) = P(X \leq x, Y \leq y), \quad F(x) = P(X \leq x), \quad G(y) = P(Y \leq y) \quad (8.1)$$

Siguiendo a [Genest, 2007], se define:

Definición 16. Sea (X, Y) una pareja de variables aleatorias tal que $P(X \leq x, Y \leq y) = H(x, y)$, con marginales $F(x) = P(X \leq x)$, $G(y) = P(Y \leq y)$, $\forall x, y \in \mathbb{R}$, se tiene:

1. A es el conjunto de las subcópulas asociadas a H , es decir, la colección de funciones $A : [0, 1]^2 \mapsto [0, 1]$ tal que $\forall x, y \in \mathbb{R}$ y,

$$H(x, y) = A(F(x), G(y)) \quad (8.2)$$

2. $B : [0, 1]^2 \mapsto [0, 1]$ es la función definida $\forall u, v \in [0, 1]$ por:

$$B(u, v) = H(F^{-1}(u), G^{-1}(v)), \quad (8.3)$$

3. $C : [0, 1]^2 \mapsto [0, 1]$ es la función definida $\forall u, v \in [0, 1]$ por:

$$C(u, v) = H(F^{-1}(u_{\leftarrow}), G^{-1}(v_{\leftarrow})), \quad (8.4)$$

4. D es la función de distribución de la pareja $(F(X), G(Y))$,
5. E es la función de distribución de la pareja $(F(X_{\rightarrow}), G(Y_{\rightarrow}))$.

Si F y G son continuas, entonces se tiene unicidad en la cópula y se cumple que:

$$A = \{B\} = \{C\} = \{D\} = \{E\} \quad (8.5)$$

En el caso discreto, esto no necesariamente se cumple, ya que se tiene saltos en las inversas de la distribución, solo se asegura la existencia de la cópula para H en A , pero no es única.

Genest,[Genest, 2007] propone el siguiente ejemplo para ilustrar la complejidad del problema:

Ejemplo 3. Sea X y Y una variable aleatoria Bernoulli con $P(X = 0) = p$ y $P(Y = 0) = 1$, además, $P(X = 0, Y = 0) = r \leq \min(p, q)$.

Por tanto, $A : [0, 1]^2 \mapsto [0, 1]$, tal que:

$$\begin{aligned} A(0, 0) &= 0, & A(0, q) &= 0, & A(0, 1) &= 0, \\ A(p, 0) &= 0, & A(p, q) &= r, & A(p, 1) &= p, \\ A(1, 0) &= 0, & A(1, q) &= q, & A(1, 1) &= 1, \end{aligned}$$

$$B(u, v) = \begin{cases} 0 & \text{si } u = 0 \text{ o } v = 0, \\ r & \text{si } (u, v) \in (0, p] \times (0, q], \\ q & \text{si } (u, v) \in (p, 1] \times (0, q], \\ p & \text{si } (u, v) \in (0, p] \times (q, 1], \\ 1 & \text{si } (u, v) \in (p, 1] \times (q, 1], \end{cases}$$

$$C(u, v) = \begin{cases} r & \text{si } (u, v) \in (0, p] \times (0, q], \\ q & \text{si } (u, v) \in (p, 1] \times (0, q], \\ p & \text{si } (u, v) \in (0, p] \times (q, 1], \\ 1 & \text{si } (u, v) \in (p, 1] \times (q, 1], \end{cases}$$

Se puede ver claras diferencias que no pueden ser ignoradas, por ejemplo, $B \in A$ mientras que $C \notin A$. Además, se puede verificar que $C = E$ que corresponde a la función de distribución de $(F(X_{\rightarrow}), G(Y_{\rightarrow}))$, la cual es diferente a $(F(X), G(Y))$.

Estas observaciones son utilizadas por Genest, [Genest, 2007] para introducir el siguiente teorema:

Teorema 4. *Sea A, B, C, D y E las funciones definidas en la definición 16. Si la función conjunta de (X, Y) es discreta, entonces:*

- $B \in A$ no es una función de distribución,
- $D \in A$ es una función de distribución pero no una cópula,
- $C = E \notin A$ es una función de distribución pero no una cópula,

La demostración de este teorema se encuentra en Genest [Genest, 2007].

Por tanto, para el caso discreto, la probabilidad de un resultado particular está determinado por una integral sobre la región de $[0, 1]^2$ correspondiente a ese resultado; cualesquiera dos cópulas que integran a los mismos valores en todas las regiones pueden producir la misma distribución conjunta.

8.2. Inferencia para cópulas con marginales discretas en Geoestadística

Las cópulas con marginales discretas no difieren demasiado para el caso continuo, se deben tener en cuenta los anteriores resultados y se aplican de la misma forma.

La estructura de dependencia esta caracterizada por las familias de cópulas paramétricas, por ejemplo, la cópula gaussiana definida como:

$$C_{\mu, \Sigma}(u, v) = \Phi_{\Sigma}(\phi^{-1}(u), \phi^{-1}(v)). \quad (8.6)$$

Entonces, es posible realizar una estimación de máxima verosimilitud considerando un modelo generativo donde los marginales uniformes se generan a partir de la densidad de la cópula, y a su vez, se utilizan para generar las variables discretas con el uso de la distribución inversa de la marginal de las funciones de distribución. Esta marginal puede ser de cualquier familia parametrizada de distribuciones univariantes discretas.

Recordando, se tiene que los parámetros a estimar son:

- θ que define el parámetro de correlación de la cópula,
- λ que define los parámetros de la cópula y,
- η correspondiente a los parámetros de tendencia espacial incluyendo los valores β de la regresión para la distribución marginal,

se puede notar, $C_h = C_{\theta,\lambda}$ como la cópula y la marginal $F_Z = F_\eta$. Por tanto, la función de verosimilitud para este caso, sería de la forma:

$$l(\Theta; D) = C_h(u_1, u_2) + C_h(u_1^-, u_2^-) - C_-(u_1^-, u_2) - C_h(u_1, u_2^-) \quad (8.7)$$

donde, $u_i = F_i(y_i)$ y $u_i^- = F_i(y_i - 1)$.

La estimación máxima verosimilitud, para el caso bivariado, se encuentra optimizado en el programa MATLAB y son rutinas relativamente eficientes.

El problema surge si se considera una tendencia no constante, en este caso, se puede tener en cuenta que la cópula gaussiana esta definida por la matriz de correlación Σ , que a su vez, depende la familia de las funciones de correlación correspondiente al modelo (Matérn, exponencial, entre otras) y se puede estimar fuera de la máxima verosimilitud. Para el caso de la marginal, si la media μ aparece explícitamente en la expresión analítica de la función de densidad marginal univariada f_Z , entonces,

$$g(\mu_i) = \eta(y_i) = y_i^T \beta, \quad \text{para todo } i = 1, \dots, n,$$

donde, g es la función de enlace, $\eta(y_i)$ es la superficie de respuesta, y_i es el vector de covarianzas correspondiente a la localización x_i y β es el vector de los parámetros de regresión, se puede estimar por mínimos cuadrados.

Ejemplo 4. A manera de ejemplo, se utiliza la simulación realizada por Diggle, [Diggle, 2007] con respuestas Poisson y un proceso gaussiano en el programa R. Se consideró un proceso gaussiano para simular las localizaciones con $\mu = 0,5$ y $\sigma^2 = 3$ y función de correlación Matérn con $\kappa = 1,5$ y $\phi = 0,2$. Las realizaciones son exponenciales con media Poisson, $\mu_i = \exp(0,5 + z(x_i))$. EL resultado y el cómputo de la cópula se realiza en MATLAB. En la Figura (8.1) se muestra la simulación realizada.

De la misma forma que procedemos anteriormente, se realiza una exploración de datos de la estructura de dependencia utilizando la cópula empírica, en las Figuras (8.2), (8.3), (8.4) y (8.5) se muestra la estructura de dependencia para $lags = 0,15, 0,19, 0,198, 0,21$, se puede notar una clara simetría en los puntos lo que puede indicar una cópula gaussiana. La cópula teórica se presenta en la Figura [8.6].

Los resultados muestran que es posible construir una estructura de dependencia para cópulas con marginales discretas, para el caso bivariado y que además resulta flexible, sin embargo, matemáticamente, la generalización de este método no es una tarea trivial debido

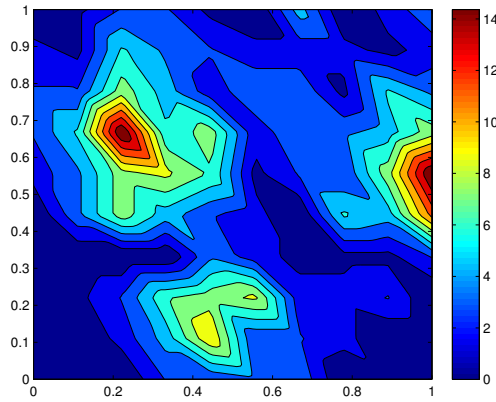
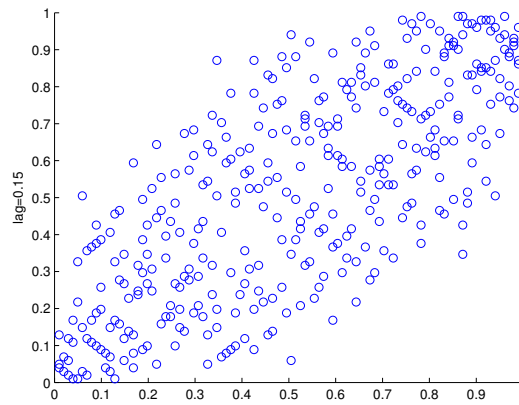
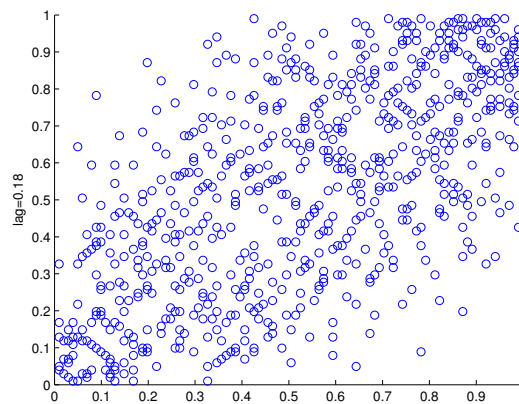


FIGURA 8.1. Simulación proceso Gaussiano con marginales Poisson

FIGURA 8.2. Cópula empírica para simulación proceso Gaussiano con marginales Poisson, $h = 0,15$ FIGURA 8.3. Cópula empírica para simulación proceso Gaussiano con marginales Poisson, $h = 0,19$

a la complejidad computacional que se puede presentar para un caso multivariado ($n > 2$).

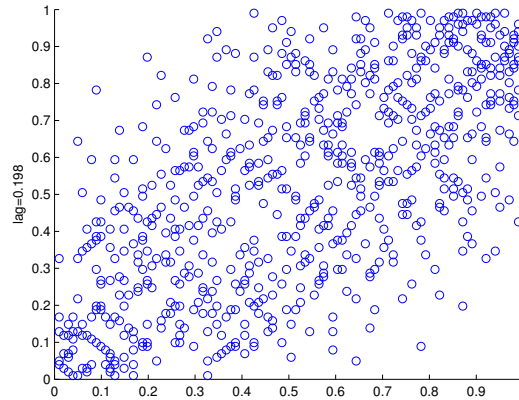


FIGURA 8.4. Cópula empírica para simulación proceso gaussiano con marginales Poisson, $h = 0,198$

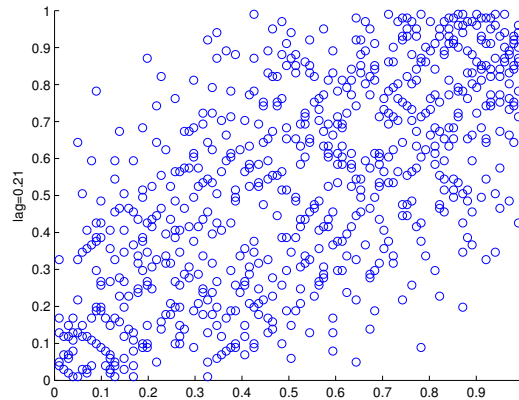


FIGURA 8.5. Cópula empírica para simulación proceso gaussiano con marginales Poisson, $h = 0,21$

Si la estructura de dependencia entre una población es relativamente homogénea, entonces, estas copulas pueden ser útiles, en el sentido de que se puede estimar a partir de una muestra mucho menor que la necesaria, por ejemplo, para una matriz de covarianza completa. Por otra parte, si las dependencias dentro de una población varían notablemente para diferentes pares de datos, la cópula gaussiana carece de la flexibilidad para capturar las dependencias extremas. En tales casos, se puede aplicar otra cópula de la familia elíptica ya que está parametrizada por la misma matriz de covarianza de la cópula gaussiana. Sin embargo, la cópula gaussiana se prohíbe para dimensiones altas, ya que la evaluación de la probabilidad requiere un número exponencial de evaluaciones de la gaussiana multivariada, que se debe calcular numéricamente convirtiendo el análisis en una labor imposible.

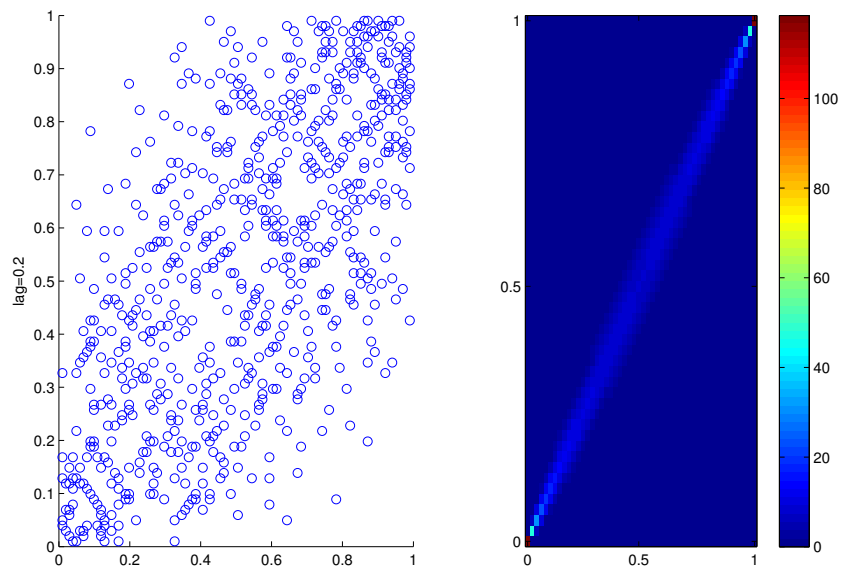


FIGURA 8.6. Cópula gaussiana con marginales Poisson

Conclusiones

En esta tesis, se realiza una revisión de literatura de la funciones cópulas en el marco geoestadístico. El concepto de función cópulas para marginales discretas y continuas se aplica como una nueva herramienta para describir y analizar la estructura de dependencia espacial en un marco geoestadístico. Se desarrollan estrategias específicas para utilizar las funciones cópulas y mejorar los métodos estadísticos en el análisis espacial, interpolación espacial y los procedimientos de toma de decisiones.

Por medio de las cópulas empíricas, se genera una metodología que permite producir una idea intuitiva de la forma en que la dependencia espacial parece corresponder a los modelos teóricos de copulas, ya sean con marginales discretas o continuas.

Se desarrolla un estadístico de prueba para la elección de las copulas que se ajusten a los datos geoestadísticos, en este caso, las copulas gaussiana y V probaron ser adecuadas para este tipo de datos mostrando que son adecuadas para el modelado espacial.

Una vez que la cópula es seleccionada, se realizó la estimación de los parámetros del modelo, a cada cópula se le asigno un proceso diferente de estimación ya que las pruebas realizadas mostraron la necesidad de algoritmos que mejoren el desempeño de la estimación, debido a que cada copula posee un nivel computacional diferente.

Finalmente, se realizó predicción usando cópulas y el indicador Kriging. La validación cruzada mostró que las copulas generan una completa estructura de dependencia, implicando que los resultados de predicción son mas fiables que los métodos tradicionales, además de mostrar que las copulas asimétricas se comportan mejor para datos que cumplen estas características. En el caso de la cópula discreta, se presentó un mejor ajuste con la cópula gaussiana, ya que los datos simulados presentaban una dependencia simétrica.

Bibliografía

- [Ayyad, 2008] Ayyad, C. Mateu, J. P. E. (2008). *Inferencia y modelización mediante cópulas*. Universidad Jaume.
- [Bárdossy, 2006] Bárdossy, A. (2006). Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research*, 42(11).
- [Bárdossy, 2008] Bárdossy, A. Li, J. (2008). Geostatistical interpolation using copulas. *Water Resources Research*, (7):W07412+.
- [Berkes, 2009] Berkes, P. Wood, F. P. J. (2009). Characterizing neural dependencies with copula models. pages 129–136.
- [Bickel and Doksum, 2000] Bickel, P. and Doksum, K. (2000). *Mathematical statistics: basic ideas and selected topics*. Number v. 1 in Mathematical Statistics: Basic Ideas and Selected Topics. Prentice Hall.
- [Bouye, 2000] Bouye, E. Durrlema, V. G. R. T. (2000). *Copulas for Finance A Reading Guide and Some Applications*. London.
- [Christian, 2007] Christian, G. Favre, A. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, pages 347–368.
- [Cressie, 1991] Cressie, N. (1991). *Statistics for spatial data*. Wiley series in probability and mathematical statistics: Applied probability and statistics. J. Wiley.
- [Cuervo, 2008] Cuervo, E. (2008). *Estadística matemática*.
- [Denuit, 2005] Denuit, M. (2005). *Actuarial theory for dependent risks: measures, orders and models*. Number v. 10 in Actuarial theory for dependent risks: measures, orders and models. Wiley.
- [Diggle, 2007] Diggle, P. Ribeiro, P. (2007). *Model-based Geostatistics*. Springer Series in Statistics. Springer.
- [Gaël, 2011] Gaël, B. Edzer, P. (2011). The pair-copula construction for spatial data: a new approach to model spatial dependency.
- [Genest, 2008a] Genest, C. Rémillard, B. (2008a). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Ann Inst Henri Poincare*, 44(6):1096–1127.

- [Genest, 2008b] Genest, C. Rémillard, B. B. D. (2008b). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44(2):199–213.
- [Genest, 2007] Genest, C. Neslehova, J. (2007). A primer on copulas for count data. *Astin Bulletin*, 37(2):475.
- [Giraldo, 2010] Giraldo, R. (2010). *Introducción a la Geoestadística: Teoría y Aplicación*. Evento: simposio nacional de estadística Ponencia.
- [Haslauer, 2010] Haslauer, C. Li, J. B. A. (2010). Application of Copulas in Geostatistics. 16:395–404.
- [Hengl, 2009] Hengl, T. (2009). *A Practical Guide to Geostatistical Mapping*. University of Amsterdam, Amsterdam, 2nd edition.
- [Jaworski et al., 2010] Jaworski, P., Durante, F., Härdle, W., and Rychlik, T. (2010). *Copula Theory and Its Applications: Proceedings of the Workshop Held in Warsaw, 25-26 September 2009*. Lecture Notes in Statistics. Springer.
- [Jing, 2010] Jing, L. (2010). *Application of copulas as a new geostatistical tool*. PhD thesis, Universität Stuttgart, Holzgartenstr. 16, 70174 Stuttgart.
- [Kazianka, 2010a] Kazianka, H. Pilz, J. (2010a). Bayesian spatial modeling and interpolation using copulas. *Computers & Geosciences*.
- [Kazianka, 2010b] Kazianka, H. Pilz, J. (2010b). Copula based geostatistical modeling of continuous and discrete data including covariates. *Stochastic environmental research and risk assessment*, 24(5):661–673.
- [Klugman, 1999] Klugman, A. Parsa, R. (1999). Fitting bivariate loss distributions with copulas. *Insurance: Mathematics and Economics*, 24(1-2):139–148.
- [Liebscher, 2008] Liebscher, E. (2008). Construction of asymmetric multivariate copulas. *J. Multivar. Anal.*, 99(10):2234–2250.
- [Lloyd, 2006] Lloyd, C. (2006). *Local models for spatial analysis*. CRC Press.
- [McCullagh and Nelder, 1989] McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models (Second edition)*. London: Chapman & Hall.
- [Nelder and Wedderburn, 1972] Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A, General*, 135:370–384.
- [Nelsen, 1998] Nelsen, R. (1998). *An Introduction to Copulas (Lecture Notes in Statistics)*. Springer, 1 edition.
- [Panagiotelis, 2011] Panagiotelis, A. Czado, C. J. H. (2011). Pair copula constructions for discrete data.
- [Song, 2003] Song, P. Fan, Y. K. J. (2003). Maximization by parts in likelihood inference. Working Papers 0319, Department of Economics, Vanderbilt University.
- [Sotirios, 2011] Sotirios, P. Yiannis, D. (2011). The Copula Echo State Network. *Pattern Recognition Journal*.

-
- [Späc̃k,] Späc̃k, G. Kazianka, H. P. J. Modeling and interpolation of non-gaussian spatial data: A comparative study.
- [Trivedi, 2007] Trivedi, K. Zimmer, M. (2007). *Copula modeling: an introduction for practitioners*. Foundations and Trends in Econometrics.
- [Wu, 2006] Wu, F. Sherris, M. (2006). Risk measures and capital for dependent risks. *Actuarial Foundation of the institute of actuaries of Australia*.
- [Yan, 2007] Yan, J. (2007). Enjoy the joy of copulas: With a package copula. *Journal of Statistical Software*, 21(4):34–78.