

MODELADO DE DATOS DE TRANSICIÓN BINARIOS
A TRAVÉS DE UNA CÓPULA DISCRETA

T E S I S
QUE PARA OBTENER EL GRADO DE:
M A E S T R O E N C I E N C I A S
(MATEMÁTICAS APLICADAS E INDUSTRIALES)

P R E S E N T A:
LUIS CARLOS PÉREZ RUIZ

**UNIVERSIDAD AUTÓNOMA METROPOLITANA
UNIDAD IZTAPALAPA**

**MAESTRÍA EN CIENCIAS
(MATEMÁTICAS APLICADAS E INDUSTRIALES)**

**MODELADO DE DATOS DE TRANSICIÓN BINARIOS
A TRAVÉS DE UNA CÓPULA DISCRETA**

T E S I S
QUE PARA OBTENER EL GRADO DE:
M A E S T R O E N C I E N C I A S

P R E S E N T A:
LUIS CARLOS PÉREZ RUIZ

ASESOR: DR. GABRIEL ESCARELA PÉREZ

Diciembre de 2007.

Índice general

1. Introducción	1
1.1. Datos Dependientes y su Modelado	1
1.2. Organización de los Capítulos	2
1.3. Ejemplo Ilustrativo	4
2. Modelos Marginales y GEE	5
2.1. Ecuaciones Estimadoras Basadas en Verosimilitud	5
2.2. Ecuaciones Estimadoras Basadas en Cuasi-verosimilitud	9
2.3. Ecuaciones Estimadoras Generalizadas	11
2.4. Aplicación al Ejemplo Ilustrativo	15
2.5. Datos faltantes	18
3. Modelos de Transición y GLM	23
3.1. Cadenas de Markov de Tiempo Discreto	23
3.2. Modelos de Transición	27
3.3. Modelos de Transición de Respuestas Binarias	30
3.4. Aplicación al Ejemplo Ilustrativo	32
4. Modelos Marginales y de Transición con Cópulas	36
4.1. Cópulas Bivariadas	37

	II
4.1.1. Conceptos Básicos	37
4.1.2. Familias de Cópulas Paramétricas	42
4.1.3. Medidas de Dependencia	46
4.2. Modelo de Transición de Primer Orden	51
4.3. Diagnósticos	56
4.4. Aplicación al Ejemplo Ilustrativo	59
5. Consideraciones Finales	68
A. Programa en R	71
Bibliografía	77

Capítulo 1

Introducción

1.1. Datos Dependientes y su Modelado

En la recopilación de datos para ciertos tipos de estudios estadísticos se presentan comúnmente dos situaciones prácticas en las que la suposición de independencia entre observaciones específicas no es plausible. En una primera situación, los datos se toman de sujetos que de alguna manera están relacionados entre sí. Por ejemplo, para realizar un estudio de inferencia estadística sobre la estatura de los hombres de 30 años de edad de todo el país, se toma una muestra de diferentes regiones geográficas. Es claro que las estaturas de los hombres de una misma región presentarán cierta dependencia por cuestiones de herencia, medio ambiente y otros factores, que no se verá reflejada entre las estaturas de hombres de diferentes regiones. Para obtener inferencias válidas en este caso lo conducente es formar grupos de hombres por área geográfica de características similares. Una vez organizados de esta manera se les conoce como *datos agrupados*.

En la segunda situación, los datos se toman de un mismo sujeto pero en diferentes momentos a través del tiempo. Un ejemplo de esto es cuando se quiere establecer la eficacia de un nuevo tratamiento médico. Para ello, se aplica el tratamiento a una

muestra de diferentes personas sin parentesco entre sí, y después se les hace un seguimiento temporal durante el cual se van recolectando los resultados. Claramente las respuestas al tratamiento reportadas por una persona, presentarán alguna dependencia por factores propios de cada persona, mientras que las respuestas al tratamiento entre personas pueden considerarse como independientes. A datos recolectados de esta manera se les conoce como *datos longitudinales*.

Ambas situaciones deberán ser descritas por modelos que tomen en cuenta la correlación entre respuestas, ya que de otra manera se obtendrán inferencias estadísticas erróneas. Hay tres modelos usuales que son una extensión de los Modelos Lineales Generalizados para tratar con datos correlacionados:

- Modelos Marginales
- Modelos de Transición
- Modelos de Efectos Aleatorios

En cada uno de ellos se modela tanto la regresión entre la respuesta y las variables explicativas, como la correlación entre respuestas. Este trabajo se enfocará sólo en los modelos marginales y de transición para datos longitudinales, ya que son estos modelos sobre los que se propone la nueva metodología de cópulas.

1.2. Organización de los Capítulos

En el capítulo 2 se trabaja con el método de ecuaciones estimadoras generalizadas propuesto inicialmente por Liang y Zeger (1986) para ajustar modelos marginales de regresión. Primero se hace una introducción a la teoría de los modelos lineales generalizados para después hacer una extensión natural a las ecuaciones estimadoras

basadas en cuasi-verosimilitud. Hecho esto, se pasa a la obtención de las ecuaciones estimadoras generalizadas y a su uso práctico en un ejemplo ilustrativo con datos faltantes.

En el capítulo 3 se trabaja con el método de modelos lineales generalizados introducido por Nelder y Wedderburn (1972) para ajustar modelos de transición según la propuesta de Diggle *et al.* (1994). En la primera sección se introduce al tema de las cadenas de Markov de tiempo discreto, con el propósito de que en la siguiente sección se desarrolle un modelo de transición con una estructura de Markov. Después se aplica dicho modelo a respuestas binarias para concluir con su utilización en el ejemplo ilustrativo pero esta vez sólo con datos completos.

En el capítulo 4 se propone el método de cópulas presentadas originalmente por Sklar (1959) para ajustar ambos modelos: marginales y de transición. En primer lugar se hace una revisión general de las cópulas bivariadas en cuanto a conceptos básicos como definición, propiedades, representación gráfica, etc. También se revisan diversas familias de cópulas de un sólo parámetro como las elípticas y las Arquimedianas junto con sus propiedades. Después se estudian las medidas de dependencia usuales para respuestas normales. En una sección posterior se desarrolla el modelo de transición de primer orden por medio de cópulas discretizadas para respuestas binarias. Para concluir el capítulo, se aplica el método de cópulas con varias estructuras de dependencia al ejemplo ilustrativo con datos faltantes, y posteriormente se realizan los diagnósticos correspondientes.

Finalmente, en el capítulo 5 se dan las conclusiones y consideraciones finales con respecto a la comparación entre los diferentes métodos y modelos. También se propone una perspectiva de nuevas líneas de investigación que pueden ser desarrolladas para ampliar la teoría y el uso de las cópulas en el modelado multivariado.

1.3. Ejemplo Ilustrativo

Como ejemplo ilustrativo se utilizará la base de datos proporcionada por la Dra. Lori A. Crane del Centro de Ciencias de la Salud de la Universidad de Colorado. Esta base de datos se origina por una investigación que tiene por objetivo evaluar un tratamiento que reduzca la exposición al sol de los niños. Los infantes fueron incluidos en el estudio antes de los 6 meses de edad, y repartidos aleatoriamente en un grupo de control y en un grupo intervenido. Los padres de los niños del grupo intervenido recibieron paquetes de protección solar y consejos sobre salud infantil de niños de entre 2 y 36 meses de edad. Las inspecciones anuales a los padres a 1, 2 y 3 años de edad del niño fueron usadas para evaluar el impacto de la intervención. En este estudio se categorizó la variable **respuesta** “evitó la exposición al sol del mediodía” como *nunca* o *raramente* (codificada en 0) y como *siempre* o *frecuentemente* (codificada en 1). Las dos variables explicativas son el factor **Tratamiento** y el **tiempo** de la observación. Hubo en el estudio 675 niños, de los cuales 466 tienen completas sus tres respuestas. Los restantes 209 (31 %) tiene todo tipo de respuestas faltantes, por lo que hay niños con sólo una o dos observaciones registradas según se muestra a continuación.

Tabla 1.1: Respuestas faltantes por casos.

Respuestas faltantes	Número de casos	Porcentaje del total
Año 1	37	5
Año 2	27	4
Año 3	79	12
Años 1 y 2	13	2
Años 1 y 3	6	1
Años 2 y 3	47	7
Totales	209	31

Capítulo 2

Modelos Marginales y GEE

En un modelo marginal, la regresión entre la variable respuesta y las variables explicativas es el principal objetivo estadístico, y se modela separadamente de la correlación entre observaciones, cuya naturaleza queda como un objetivo menor. En la regresión se modela la distribución marginal, que es la respuesta promedio de la población que tiene valores comunes en sus variables explicativas. Por ejemplo, para respuestas binarias en un estudio longitudinal que busca medir la efectividad de cierto tratamiento, las distribuciones marginales son las probabilidades de éxito del tratamiento en los diferentes grupos de individuos y para cada unidad temporal.

2.1. Ecuaciones Estimadoras Basadas en Verosimilitud

La teoría de los Modelos Lineales Generalizados (GLM), introducida inicialmente por Nelder y Wedderburn (1972), se basa en la suposición primordial de independencia entre las observaciones, y por lo tanto, son modelos que pueden construirse a partir de funciones de verosimilitud. Además, esta teoría contiene un importante y fundamental concepto unificador que es la *familia exponencial* de distribuciones de probabilidad.

Miembros de esta familia son las distribuciones binomial, gamma, normal, normal inversa, binomial negativa, geométrica y Poisson. Lo que caracteriza a todo miembro de esta familia es que su función de densidad de probabilidad (pdf) puede expresarse de la forma siguiente

$$f(y; \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\} \quad (2.1)$$

donde $b(\cdot)$ está en función del parámetro natural o de posición θ , mientras que $a(\cdot)$ y $c(\cdot)$ son funciones del parámetro de dispersión o de escala ϕ .

El *componente lineal* del modelo se establece como $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$, donde \mathbf{X} es la matriz de variables explicativas y $\boldsymbol{\beta}$ es el vector de los coeficientes de regresión a estimar. Finalmente, el modelo de regresión se construye a través del uso de una *función de enlace* monótona y diferenciable g , que conecta el valor esperado de las respuestas $E(\mathbf{y}) = \boldsymbol{\mu}$ con el componente lineal del modelo, esto es, $\boldsymbol{\eta} = g(\boldsymbol{\mu})$, o tomando la función de enlace inversa, $\boldsymbol{\mu} = g^{-1}(\boldsymbol{\eta})$. Puede haber diferentes alternativas para la función de enlace de cierta distribución de la familia exponencial, pero si escogemos aquella tal que $g(\mu) = \theta$, entonces diremos que g es la función de enlace canónica.

Otra característica importante en la estructura de las distribuciones de la familia exponencial, es que la varianza de las respuestas puede ser escrita en función de la media como $\text{var}(\mathbf{y}) = V(\boldsymbol{\mu})$, donde V es llamada *función de varianza*. De hecho, se pueden obtener expresiones generales para $E(y)$ y $\text{var}(y)$ de la siguiente manera. De la definición de pdf

$$\int_{-\infty}^{\infty} f(y; \theta, \phi) dy = 1 \quad (2.2)$$

Obviando los límites de integración y derivando (2.2) con respecto a θ

$$\frac{d}{d\theta} \int f(y; \theta, \phi) dy = 0 \implies \int \frac{df(y; \theta, \phi)}{d\theta} dy = 0 \quad (2.3)$$

$$\frac{d^2}{d\theta^2} \int f(y; \theta, \phi) dy = 0 \implies \int \frac{d^2 f(y; \theta, \phi)}{d\theta^2} dy = 0 \quad (2.4)$$

Al derivar (2.1) con respecto a θ

$$\frac{df(y; \theta, \phi)}{d\theta} = \left[\frac{y - b'(\theta)}{a(\phi)} \right] f(y; \theta, \phi) \quad (2.5)$$

Integrando (2.5) y usando el resultado (2.3)

$$\int \frac{df(y; \theta, \phi)}{d\theta} dy = \frac{1}{a(\phi)} \int y f(y; \theta, \phi) dy - \frac{b'(\theta)}{a(\phi)} \int f(y; \theta, \phi) dy = 0 \quad (2.6)$$

Sustituyendo en (2.6) la definición de valor esperado $E(y) = \int y f(y) dy$ y usando el resultado (2.2), finalmente obtenemos

$$\frac{1}{a(\phi)} E(y) - \frac{b'(\theta)}{a(\phi)} = 0 \implies \boxed{E(y) = b'(\theta) = \mu} \quad (2.7)$$

Mediante un procedimiento análogo, se obtiene la expresión para $\text{var}(y)$. Al derivar (2.5) con respecto a θ

$$\frac{d^2 f(y; \theta, \phi)}{d\theta^2} = \left[\frac{y - b'(\theta)}{a(\phi)} \right]^2 f(y; \theta, \phi) - \frac{b''(\theta)}{a(\phi)} f(y; \theta, \phi) \quad (2.8)$$

Integrando (2.8) y usando los resultados (2.4) y (2.7)

$$\int \frac{d^2 f(y; \theta, \phi)}{d\theta^2} dy = \frac{1}{[a(\phi)]^2} \int [y - E(y)]^2 f(y; \theta, \phi) dy - \frac{b''(\theta)}{a(\phi)} \int f(y; \theta, \phi) dy = 0 \quad (2.9)$$

Sustituyendo en (2.9) la definición de varianza $\text{var}(y) = \int [y - E(y)]^2 f(y) dy$ y usando el resultado (2.2), finalmente obtenemos

$$\frac{1}{[a(\phi)]^2} \text{var}(y) - \frac{b''(\theta)}{a(\phi)} = 0 \implies \boxed{\text{var}(y) = b''(\theta) a(\phi) = \frac{d\mu}{d\theta} a(\phi)} \quad (2.10)$$

Ahora, se obtendrán las ecuaciones estimadoras basadas en la función de verosimilitud para la familia exponencial de los GLM. La pdf de una sola respuesta es

$$f(y_i; \theta, \phi) = \exp \left\{ \frac{y_i \theta - b(\theta)}{a(\phi)} + c(y_i, \phi) \right\} \quad (2.11)$$

y la pdf conjunta de n respuestas independientes es

$$f(y_1, \dots, y_n; \theta, \phi) = \prod_{i=1}^n \exp \left\{ \frac{y_i \theta - b(\theta)}{a(\phi)} + c(y_i, \phi) \right\} \quad (2.12)$$

La *función de verosimilitud* es simplemente la pdf conjunta donde ahora se consideran las respuestas como dadas y los parámetros desconocidos

$$L(\theta, \phi | y_1, \dots, y_n) = \prod_{i=1}^n \exp \left\{ \frac{y_i \theta - b(\theta)}{a(\phi)} + c(y_i, \phi) \right\} \quad (2.13)$$

La función log-verosimilitud está dada por

$$\mathcal{L}(\theta, \phi | y_1, \dots, y_n) = \sum_{i=1}^n \left\{ \frac{y_i \theta - b(\theta)}{a(\phi)} + c(y_i, \phi) \right\} \quad (2.14)$$

Al derivar (2.14) con respecto a θ y usando el resultado (2.7)

$$\frac{\partial \mathcal{L}}{\partial \theta} = \sum_{i=1}^n \frac{y_i - b'(\theta)}{a(\phi)} = \sum_{i=1}^n \frac{y_i - \mu_i}{a(\phi)} \quad (2.15)$$

Los estimadores de máxima verosimilitud (MLE) de los coeficientes de regresión β y del parámetro de dispersión ϕ , se obtienen al resolver por separado las *ecuaciones estimadoras* correspondientes

$$\frac{\partial \mathcal{L}}{\partial \beta} = \mathbf{0} \quad \text{y} \quad \frac{\partial \mathcal{L}}{\partial \phi} = 0$$

Como el principal objetivo es obtener los MLE de β , se puede ver a ϕ como un parámetro auxiliar y su estimación como un objetivo menor; es por ello, que sólo se desarrollarán las ecuaciones estimadoras para β . Usando la regla de la cadena y los resultados (2.10) y (2.15)

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \beta} &= \left(\frac{\partial \mathcal{L}}{\partial \theta} \right) \left(\frac{\partial \theta}{\partial \mu} \right) \left(\frac{\partial \mu}{\partial \eta} \right) \left(\frac{\partial \eta}{\partial \beta} \right) \\ &= \left(\sum_{i=1}^n \frac{y_i - \mu_i}{a(\phi)} \right) \left(\frac{a(\phi)}{\text{var}(y_i)} \right) \left(\frac{\partial \mu_i}{\partial \eta_i} \right) (\mathbf{x}_i) \end{aligned} \quad (2.16)$$

Sabiendo que $\text{var}(y_i) = V(\mu_i)$, finalmente las ecuaciones estimadoras para la familia exponencial de los GLM son

$$\sum_{i=1}^n \frac{y_i - \mu_i}{V(\mu_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \mathbf{x}_i = \mathbf{0} \quad (2.17)$$

Los coeficientes β no aparecen directamente en las ecuaciones estimadoras porque están incluidos a través de η y μ . En notación matricial las ecuaciones estimadoras quedan como

$$\mathbf{X}' \mathbf{D} \mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{0} \quad (2.18)$$

donde $\mathbf{D} = \text{diag}\{\partial \mu_i / \partial \eta_i\}$ y $\mathbf{V} = \text{diag}\{V(\mu_i)\}$.

2.2. Ecuaciones Estimadoras Basadas en Cuasi-verosimilitud

En las ecuaciones estimadoras anteriores la elección de la distribución de la respuesta determina la relación media-varianza. Sin embargo, al observar las ecuaciones se puede notar la siguiente propiedad interesante en ellas, dependen de la distribución de la respuesta sólo a través de su media y su varianza. Por lo que Wedderburn (1974) propuso un atinado enfoque alternativo que extiende los resultados ya obtenidos, ver la media y la función de varianza como partes integrales de las ecuaciones estimadoras, independientemente de la distribución que las origina. Bajo este nuevo enfoque, ahora se pueden elegir libremente las funciones de varianza y de enlace, que pueden pertenecer o no a la familia exponencial, y aplicarlas en las ecuaciones estimadoras obtenidas en la sección anterior sin necesidad de alterarlas.

Si se escogen funciones de varianza y enlace que pertenecen a distribuciones de la familia exponencial, implica que las ecuaciones estimadoras tienen asociada en forma

implícita la función de verosimilitud de la distribución. Este es el caso de ecuaciones estimadoras basadas en verosimilitud que nos proporcionan los MLE's ordinarios.

Por otro lado, si se escogen funciones de varianza y enlace que no pertenecen a distribuciones de la familia exponencial, implica que las ecuaciones estimadoras no tendrán asociada ninguna función de verosimilitud. En este caso a las ecuaciones estimadoras se les conoce como basadas en *cuasi-verosimilitud* y nos proporcionan estimadores de máxima cuasi-verosimilitud (MQE) propiamente dichos.

Visto así, se puede decir que la cuasi-verosimilitud es una generalización de la verosimilitud, ya que nos permite incluir otras funciones diferentes de las funciones-miembro de la familia exponencial. Se debe tener en mente que ésta metodología fue desarrollada para situaciones en las que era necesario obtener inferencias de experimentos en los cuales una función de verosimilitud no puede ser construida. Se puede tener un modelo razonable y una función de varianza conocida, pero tal vez no se tenga información suficiente que sugiera alguna distribución para la respuesta.

La solución de las ecuaciones estimadoras se obtiene por medio de técnicas de optimización. Un método común emplea la expansión en series de Taylor de las ecuaciones estimadoras $\mathbf{S}(\boldsymbol{\beta}) = \mathbf{0}$, esto es

$$\mathbf{0} = \mathbf{S}(\boldsymbol{\beta}^{(0)}) + (\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}) \mathbf{S}'(\boldsymbol{\beta}^{(0)}) + \frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)})^2 \mathbf{S}''(\boldsymbol{\beta}^{(0)}) + \dots$$

Tomando sólo los dos primeros términos, tenemos la aproximación lineal

$$\mathbf{0} \approx \mathbf{S}(\boldsymbol{\beta}^{(0)}) + (\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}) \mathbf{S}'(\boldsymbol{\beta}^{(0)})$$

$$\boldsymbol{\beta} \approx \boldsymbol{\beta}^{(0)} - \frac{\mathbf{S}(\boldsymbol{\beta}^{(0)})}{\mathbf{S}'(\boldsymbol{\beta}^{(0)})}$$

Escribiendo ésta relación en notación matricial, iteramos la solución por medio de

$$\boldsymbol{\beta}^{(k)} = \boldsymbol{\beta}^{(k-1)} + \left[-\frac{\partial}{\partial \boldsymbol{\beta}} \mathbf{S}(\boldsymbol{\beta}^{(k-1)}) \right]^{-1} \mathbf{S}(\boldsymbol{\beta}^{(k-1)}) \quad (2.19)$$

Esto es, dado un estimador inicial $\boldsymbol{\beta}^{(0)}$, éste se actualiza usando la ecuación (2.19). Aquí una técnica de optimización usual es por mínimos cuadrados re-ponderados iterativos (IRLS). Cuando hay parámetros auxiliares con sus ecuaciones estimadoras respectivas, el método de optimización actualiza el valor de $\boldsymbol{\beta}$, después actualiza el valor del parámetro auxiliar, y así sucesivamente, alternando entre sus ecuaciones estimadoras por separado.

2.3. Ecuaciones Estimadoras Generalizadas

Los datos agrupados o datos longitudinales siempre generan respuestas multivariadas $(y_1, y_2, \dots, y_{n_i})$ generalmente dependientes. En el caso univariado donde se puede asumir independencia, se especifica un modelo para μ_i y una función de varianza $V(\mu_i)$. Mientras que en el caso multivariado, el modelo se aplica a la distribución marginal de cada y_t .

Las ecuaciones estimadoras basadas en cuasi-verosimilitud para GLM's en este caso multivariado son

$$\sum_{i=1}^n \sum_{t=1}^{n_i} \frac{y_{it} - \mu_{it}}{V(\mu_{it})} \left(\frac{\partial \mu_{it}}{\partial \eta_{it}} \right) \mathbf{x}_{it} = \mathbf{0} \quad (2.20)$$

que reescritas en forma matricial con respecto a los grupos o sujetos quedan como

$$\sum_{i=1}^n \mathbf{X}_i' \mathbf{D}_i \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0} \quad (2.21)$$

donde $\mathbf{D}_i = \text{diag}\{\partial \mu_{it} / \partial \eta_{it}\}$ y $\mathbf{V}_i = \text{diag}\{V(\mu_{it})\}$. Es claro que la matriz \mathbf{V}_i puede descomponerse en

$$\mathbf{V}_i = \mathbf{V}_i^{1/2} \mathbf{I} \mathbf{V}_i^{1/2} \quad (2.22)$$

Si nos enfocamos a la distribución marginal de las respuestas, para las cuales el valor esperado y la función de varianza están promediados a través de los grupos o

sujetos, entonces la matriz identidad en (2.22) es realmente la matriz de correlación de las observaciones dentro de un grupo o sujeto. Por lo que esta descomposición hace ver que las ecuaciones estimadoras tratan a las observaciones dentro de los grupos o sujetos como independientes, y en consecuencia su modelo asociado es llamado *modelo independiente*.

Las ecuaciones estimadoras *generalizadas* (GEE) propuestas por Liang y Zeger (1986) son una extensión de las ecuaciones estimadoras anteriores donde simplemente reemplazan la matriz identidad con una matriz de correlación más general a la que llaman *matriz de correlación de trabajo* (working correlation matrix). Esto es

$$\mathbf{V}_i = \mathbf{V}_i^{1/2} \mathbf{R}(\boldsymbol{\alpha}) \mathbf{V}_i^{1/2} \quad (2.23)$$

donde \mathbf{V}_i se convierte en una matriz de varianzas-covarianzas o matriz de dispersión. \mathbf{R} se estima a través del vector de parámetros $\boldsymbol{\alpha}$, lo que permite describir con relativa facilidad varias estructuras usuales para la matriz de correlación. Finalmente, las GEE para datos dependientes en modelos marginales quedan como

$$\sum_{i=1}^n \mathbf{X}_i' \mathbf{D}_i \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0} \quad (2.24)$$

Las estructuras más comunes para especificar la matriz de correlación de trabajo son las siguientes.

1. Intercambiable (exchangeable)

$$\mathbf{R}(\boldsymbol{\alpha}) = \begin{bmatrix} 1 & \alpha & \alpha & \cdots & \alpha \\ \alpha & 1 & \alpha & \cdots & \alpha \\ \alpha & \alpha & 1 & \cdots & \alpha \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \alpha & \cdots & 1 \end{bmatrix} \quad \text{ó} \quad R_{uv} = \begin{cases} 1 & \text{si } u = v \\ \alpha & \text{si } u \neq v \end{cases}$$

Esta estructura es apropiada para datos agrupados en los cuales es plausible que no halla dependencia temporal y cualquier permutación entre las observaciones es válida. Lo primero que se estima son los residuales de Pearson, y con ellos el parámetro de escala.

$$\hat{r}_{it} = \frac{y_{it} - \hat{\mu}_{it}}{\sqrt{V(\hat{\mu}_{it})}}, \quad \hat{\phi} = \frac{1}{\sum_{i=1}^n n_i} \sum_{i=1}^n \sum_{t=1}^{n_i} \hat{r}_{it}^2 \quad (2.25)$$

Después se estima α , que en este caso es un escalar.

$$\hat{\alpha} = \frac{1}{\hat{\phi}} \sum_{i=1}^n \left\{ \frac{\sum_{u=1}^{n_i} \sum_{v=1}^{n_i} \hat{r}_{iu} \hat{r}_{iv} - \sum_{u=1}^{n_i} \hat{r}_{iu}^2}{n_i(n_i - 1)} \right\} \quad (2.26)$$

2. Autorregresivo de k-ésimo orden (AR(k))

$$R_{uv} = \begin{cases} 1 & \text{si } u = v \\ \alpha^{|u-v|} & \text{si } u \neq v \end{cases}$$

Esta estructura es apropiada para datos longitudinales en los cuales es plausible que halla dependencia temporal y un orden natural entre las observaciones. Si la respuesta se distribuye normalmente, esto es análogo a un proceso autorregresivo de tiempo continuo (AR). Los residuales de Pearson y el parámetro de escala se estiman por medio de (2.25). Después se estima α , que en este caso es un vector.

$$\hat{\alpha} = \frac{1}{\hat{\phi}} \left[\sum_{i=1}^n \left(\frac{\sum_{t=1}^{n_i-0} \hat{r}_{i,t} \hat{r}_{i,t+0}}{n_i}, \dots, \frac{\sum_{t=1}^{n_i-k} \hat{r}_{i,t} \hat{r}_{i,t+k}}{n_i} \right) \right] \quad (2.27)$$

3. Estacionario de k-ésimo orden (stationary(k))

$$R_{uv} = \begin{cases} 1 & \text{si } u = v \\ \alpha^{|u-v|} & \text{si } |u - v| \leq k \\ 0 & \text{si } |u - v| > k \end{cases}$$

Esta estructura es una alternativa del modelo autorregresivo donde se supone que la correlación existe sólo entre las k+1 unidades temporales adyacentes. Los residuales de Pearson y el parámetro de escala se estiman por medio de (2.25). Después se estima α por medio de (2.27).

4. No estacionario de g-ésimo orden (nonstationary(g))

$$R_{uv} = \begin{cases} 1 & \text{si } u = v \\ \alpha_{uv} & \text{si } |u - v| \leq g \\ 0 & \text{si } |u - v| > g \end{cases}$$

A diferencia del modelo estacionario, en esta estructura se asume que las correlaciones alrededor de la diagonal principal no son constantes. Una desventaja en la estimación de esta estructura es que no se garantiza su invertibilidad y problemas numéricos pueden ser encontrados, especialmente para bases de datos desbalanceadas, o cuando el tamaño de la matriz es relativamente grande con respecto al número de grupos o sujetos en estudio. Los residuales de Pearson y el parámetro de escala se estiman por medio de (2.25). Después se estima α por medio de las siguientes ecuaciones.

$$\begin{aligned} I(i, u, v) &= \begin{cases} 1 & \text{si el grupo } i \text{ tiene observaciones indexadas } u \text{ y } v \\ 0 & \text{de otro modo} \end{cases} \\ g_{u,v} &= \left(\sum_{i=1}^n I(i, u, v) \right)^{-1} \\ \mathbf{G} &= \begin{pmatrix} g_{1,1}\widehat{r}_{i,1}^2 & g_{1,2}\widehat{r}_{i,1}\widehat{r}_{i,2} & \cdots & g_{1,n_i}\widehat{r}_{i,1}\widehat{r}_{i,n_i} \\ g_{2,1}\widehat{r}_{i,2}\widehat{r}_{i,1} & g_{2,2}\widehat{r}_{i,2}^2 & \cdots & g_{2,n_i}\widehat{r}_{i,2}\widehat{r}_{i,n_i} \\ \vdots & \vdots & \ddots & \vdots \\ g_{n_i,1}\widehat{r}_{i,n_i}\widehat{r}_{i,1} & g_{n_i,2}\widehat{r}_{i,n_i}\widehat{r}_{i,2} & \cdots & g_{n_i,n_i}\widehat{r}_{i,n_i}^2 \end{pmatrix} \\ \widehat{\alpha} &= \frac{\sum_{i=1}^n n_i}{\sum_{i=1}^n \sum_{t=1}^{n_i} \widehat{r}_{i,t}^2 / n_i} \mathbf{G} \end{aligned} \quad (2.28)$$

5. Sin estructura (unstructured)

$$R_{uv} = \begin{cases} 1 & \text{si } u = v \\ \alpha_{uv} & \text{si } u \neq v \end{cases}$$

Este caso es el más general de todas las estructuras y es igual al modelo no estacionario pero con el orden g máximo. Al igual que en el caso anterior, la

estimación de esta estructura no garantiza su invertibilidad y problemas numéricos pueden ser encontrados, especialmente para bases de datos desbalanceadas, o cuando el tamaño de la matriz es relativamente grande con respecto al número de grupos o sujetos en estudio. Los residuales de Pearson y el parámetro de escala se estiman por medio de (2.25). Después se estima α por medio de (2.28).

6. Fija (fixed)

Este modelo puede ser impuesto si se tiene conocimiento previo de la estructura de correlación por medio de una fuente externa. Con este modelo ya no se estima la matriz de correlación en cada paso de la iteración, ya que se toma la matriz como dada. Otro uso de este modelo es posibilitar la utilización de algún tipo de estructura que no esté directamente soportada como opción en un paquete de software específico.

7. Libre especificación (free specification)

Esta puede ser cualquier estructura definida por el usuario a partir de suposiciones hechas de acuerdo al tipo de experimento realizado, y que no comprenda ninguna de las formas estructurales anteriores.

2.4. Aplicación al Ejemplo Ilustrativo

Como un análisis preliminar del tratamiento que intenta reducir la exposición al sol de los niños, primero se modelará sólo la parte de la base de datos cuyos sujetos tiene sus tres respuestas completas clasificadas como se muestra en la Tabla 2.1.

Tabla 2.1: Clasificación de las respuestas completas por grupo.

Grupo	Respuestas en tres años								Total
	000	001	010	011	100	101	110	111	
Control	13	3	4	5	32	20	26	129	232
Tratamiento	9	4	7	3	18	18	32	143	234

En la Tabla 2.2 se reportan las proporciones muestrales de respuestas favorables (codificadas como 1) para las 6 distribuciones marginales que se tienen. Por ejemplo, de la Tabla 2.1, la proporción de respuestas favorables para el grupo de control al primer año de seguimiento es de $(32 + 20 + 26 + 129)/232 = 0.892$.

Tabla 2.2: Proporciones marginales muestrales de respuestas favorables.

Grupo	Proporciones muestrales		
	Año 1	Año 2	Año 3
Control	0.892	0.707	0.677
Tratamiento	0.902	0.791	0.718

De la Tabla 2.2 se puede observar que la proporción de respuestas favorables aparentemente (1) decrece con el tiempo para ambos grupos; (2) decrece a una tasa mayor en el grupo de control; y (3) es ligeramente mayor en el grupo intervenido durante los tres años. Para averiguar que tan ciertas son estas aseveraciones y así determinar si las diferencias presentes entre grupos y entre años son estadísticamente significativas, se ajustará un modelo de regresión marginal con función de enlace probit. Para ajustar el modelo con GEE se utilizará la función `geese` implementada en la biblioteca `geepack` del software estadístico R.

Tabla 2.3: Estimadores GEE con estructura de correlación AR(1).

Parámetro	Estimador	Error Est.	Valor P
Intercepto	1.48006	0.13053	0.00000
tiempo	-0.36014	0.05304	0.00000
Tratamiento	0.10777	0.19435	0.57924
tiempo:Trat	0.01564	0.07763	0.84031
α	0.38108	0.05534	0.00000

El valor del parámetro de correlación α refleja una dependencia entre observaciones relativamente buena. Además, la interacción del tratamiento con el tiempo no es significativa, por lo que no hay evidencia suficiente de que la proporción de respuestas favorables decrezca a una tasa mayor en el grupo de control. Por ello, se ajustará de nuevo el modelo sin dicha interacción.

Tabla 2.4: Estimadores GEE con estructura de correlación AR(1).

Parámetro	Estimador	Error Est.	Valor P
Intercepto	1.46375	0.10596	0.00000
tiempo	-0.35266	0.03878	0.00000
Tratamiento	0.14224	0.09828	0.14784
α	0.38111	0.05537	0.00000

Tampoco hay evidencia de que las proporciones de respuestas favorables sean mayores en el grupo intervenido que en el grupo de control; de hecho, con el tratamiento sólo se obtienen $\exp(0.14224) \approx 1.153$ veces más respuestas favorables que sin él. Sin embargo, lo que sí se puede afirmar, es que dichas proporciones decrecen con el tiempo para ambos grupos. Por lo tanto, se ajustará el modelo final sin la variable **Tratamiento**.

Tabla 2.5: Estimadores GEE con estructura de correlación AR(1).

Parámetro	Estimador	Error Est.	Valor P
Intercepto	1.53132	0.09660	0.00000
tiempo	-0.35170	0.03860	0.00000
α	0.38446	0.05459	0.00000

Entonces, el modelo marginal probit resultante es

$$\Phi^{-1}[P(Y_t = 1)] = 1.53132 - 0.35170 \text{ tiempo}$$

donde Φ es la función de distribución acumulada (CDF) normal estándar.

Finalmente, las probabilidades marginales estimadas de respuestas favorables se muestran a continuación.

Tabla 2.6: Probabilidades marginales estimadas de respuestas favorables.

Probabilidades estimadas		
Año 1	Año 2	Año 3
0.881	0.796	0.683

2.5. Datos faltantes

Cuando en una base de datos hay valores faltantes y ésta se analiza sólo con los valores observados como si los valores faltantes no existieran, pueden obtenerse estimadores sesgados como veremos fue el caso en la sección anterior.

Los sujetos con valores faltantes se pueden clasificar en dos grandes grupos:

- Aquellos en los que una vez que hay un valor faltante, el resto de los valores también son faltantes (dropouts).
- Aquellos que tienen uno o varios valores faltantes, pero cuyo último valor no es faltante (intermittent).

En general, el análisis con valores faltantes intermitentes se dificulta más que con dropouts, debido a la gran variedad de patrones de valores faltantes que tiene que ser tomados en cuenta.

Para investigar el proceso que genera a los valores faltantes, primero se dividen las respuestas \mathbf{Y} en respuestas observadas \mathbf{Y}_o y respuestas faltantes \mathbf{Y}_m , donde $\mathbf{Y} = (\mathbf{Y}_o, \mathbf{Y}_m)$. Después se construye una matriz indicadora \mathbf{M} de valores faltantes, cuyos elementos son

$$M_{it} = \begin{cases} 1 & \text{si } Y_{it} \text{ es faltante} \\ 0 & \text{si } Y_{it} \text{ es observada} \end{cases}$$

Si $P(\mathbf{M}|\mathbf{Y}) = P(\mathbf{M})$ para toda \mathbf{Y} , entonces \mathbf{M} es independiente de \mathbf{Y}_o y de \mathbf{Y}_m . En este caso, el proceso que genera a los valores faltantes es llamado *completamente aleatorio* (MCAR). Bajo este supuesto, las inferencias estadísticas obtenidas con modelos basados en verosimilitud y con GEE son válidas; es decir, un análisis usando sólo los valores observados no es sistemáticamente sesgado.

Si $P(\mathbf{M}|\mathbf{Y}) = P(\mathbf{M}|\mathbf{Y}_o)$ para toda \mathbf{Y}_m , entonces \mathbf{M} es dependiente de \mathbf{Y}_o e independiente de \mathbf{Y}_m . Para este caso, el proceso que genera a los valores faltantes es llamado *aleatorio* (MAR). Bajo este supuesto, las inferencias estadísticas obtenidas con modelos basados en verosimilitud y con GEE ponderadas son válidas. La dificultad de las GEE ponderadas es que aún requieren ciertas suposiciones sobre el

mecanismo de datos faltantes, y además, todavía no se encuentran implementadas en la mayoría del software estadístico disponible.

Si $P(\mathbf{M}|\mathbf{Y}) = P(\mathbf{M}|\mathbf{Y}_m)$ para toda \mathbf{Y}_o , entonces \mathbf{M} es dependiente de \mathbf{Y}_m e independiente de \mathbf{Y}_o . Para este caso, el proceso que genera a los valores faltantes es llamado *no aleatorio* o *no ignorable* (NMAR). Aquí es necesario un análisis más complejo para modelar $P(\mathbf{M}|\mathbf{Y})$. Bajo este supuesto, las inferencias estadísticas obtenidas con modelos basados en verosimilitud y con GEE suelen ser sesgadas y deben tomarse con cautela.

Un análisis en presencia de muchos datos faltantes debe hacerse con precaución. Generalmente, poco se sabe del proceso que genera los valores faltantes, y las suposiciones acerca de éste no pueden ser verificadas. En ausencia de un modelo del mecanismo de datos faltantes, al menos se pueden comparar los resultados de analizar la base de datos completa con todos los valores disponibles, con los resultados del análisis usando sólo la parte de la base de datos con grupos o sujetos sin valores faltantes. Si dichos resultados difieren sustancialmente, las conclusiones deberán ser tentativas hasta que el mecanismo de valores faltantes pueda ser modelado.

Retomando el ejemplo del tratamiento que reduce la exposición al sol de los niños, ahora se analizará la base de datos completa que incluye a los sujetos con todas sus respuestas y a los sujetos con respuestas faltantes. Como en este caso se tienen valores faltantes tanto intermitentes como dropouts, investigar el proceso específico que los genera puede resultar bastante complejo. En su lugar, al menos comparemos los resultados que nos arroje este nuevo análisis con los resultados del análisis previo de la sección anterior. A continuación se muestra el reporte del ajuste del modelo marginal probit.

Tabla 2.7: Estimadores GEE con estructura de correlación AR(1).

Parámetro	Estimador	Error Est.	Valor P
Intercepto	1.40400	0.11366	0.00000
tiempo	-0.34689	0.04783	0.00000
Tratamiento	0.12814	0.17200	0.45627
tiempo:Trat	0.02874	0.07141	0.68736
α	0.37794	0.04631	0.00000

El valor del parámetro de correlación α se mantiene prácticamente igual que en el análisis previo. Por otro lado, la interacción del tratamiento con el tiempo de nuevo no es significativa, por lo que no hay evidencia suficiente de que la proporción de respuestas favorables decrezca a una tasa mayor en el grupo de control. Se ajustará de nuevo el modelo sin esta interacción.

Tabla 2.8: Estimadores GEE con estructura de correlación AR(1).

Parámetro	Estimador	Error Est.	Valor P
Intercepto	1.37560	0.09279	0.00000
tiempo	-0.33336	0.03562	0.00000
Tratamiento	0.18948	0.08526	0.02625
α	0.37803	0.04655	0.00000

Contrario al análisis previo, con este nuevo ajuste y a un nivel de significancia estándar del 0.05, resulta que ahora el tratamiento sí es estadísticamente significativo. Es decir, las proporciones de respuestas favorables son mayores en el grupo intervenido que en el grupo de control; aunque no por mucho, pues con el tratamiento sólo se obtienen $\exp(0.18948) \approx 1.21$ veces más respuestas favorables que sin él. Esto también puede verse del hecho de que el tratamiento no sea significativo al 0.01. Sin embargo,

lo que sí se puede afirmar categóricamente, es que dichas proporciones decrecen con el tiempo para ambos grupos.

Entonces, el modelo marginal probit resultante es

$$\Phi^{-1}[P(Y_t = 1)] = 1.37560 - 0.33336 \text{ tiempo} + 0.18948 \text{ Tratamiento}$$

Finalmente, las probabilidades marginales estimadas de respuestas favorables se muestran a continuación.

Tabla 2.9: Probabilidades marginales estimadas de respuestas favorables.

Grupo	Probabilidades estimadas		
	Año 1	Año 2	Año 3
Control	0.851	0.761	0.646
Tratamiento	0.891	0.816	0.714

Las diferencias obtenidas entre los resultados de ambos ajustes con GEE, muestran la importancia de incluir en el análisis a los sujetos con respuestas faltantes, ya que sus respuestas observadas siguen aportando suficiente información al modelo. Sin embargo, con GEE las inferencias serán válidas si el proceso de datos faltantes es MCAR. Desafortunadamente, en la práctica, la distinción entre MCAR y MAR usualmente no puede hacerse con mucha certeza. Esta es la gran desventaja del método de GEE ante valores faltantes por asumir una suposición tan restrictiva como lo es MCAR.

Capítulo 3

Modelos de Transición y GLM

En un modelo de transición, la correlación entre las observaciones y_0, y_1, y_2, \dots , se identifica como la influencia explícita que ejercen las observaciones previas y_{t-1}, y_{t-2}, \dots , sobre la observación actual y_t . Si estas observaciones previas son tratadas como variables explicativas adicionales, es claro que la correlación entre observaciones se puede modelar junto con el proceso de regresión estándar entre la respuesta actual y demás variables explicativas. En la regresión se modelan distribuciones condicionales de transición, que son las probabilidades de obtener cierta respuesta, dadas respuestas anteriores y valores específicos de posibles variables explicativas.

3.1. Cadenas de Markov de Tiempo Discreto

Sea un *espacio de estados* finito $S = \{s_1, s_2, \dots, s_N\}$ donde cada s_i representa un *estado*, y sea $\{Y_t\} = Y_0, Y_1, Y_2, \dots$, una sucesión tal que la *variable aleatoria* Y toma valores de S en el tiempo $t = 0, 1, 2, \dots$, respectivamente. Se define lo siguiente.

- Un *vector de probabilidades* en el tiempo t denotado por $\mathbf{p}^{(t)} = [p_1^{(t)} \ p_2^{(t)} \ \dots \ p_N^{(t)}]$, donde $p_i^{(t)} = \Pr(Y_t = s_i)$ es la *probabilidad marginal*. Si se cumple lo siguiente

$\sum_{i=1}^N p_i^{(t)} = 1 \forall t$, entonces el vector de probabilidades es una *distribución*. A $\mathbf{p}^{(0)}$ se le llama *distribución inicial*.

- Una *matriz de transición* entre los tiempos $t - 1$ y t con $t > 0$ denotada por

$$\mathbf{P}^{(t)} = \begin{bmatrix} p_{11}^{(t)} & p_{12}^{(t)} & \cdots & p_{1N}^{(t)} \\ p_{21}^{(t)} & p_{22}^{(t)} & \cdots & p_{2N}^{(t)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1}^{(t)} & p_{N2}^{(t)} & \cdots & p_{NN}^{(t)} \end{bmatrix},$$

donde $p_{ij}^{(t)} = \Pr(Y_t = s_j | Y_{t-1} = s_i)$ es la *probabilidad de transición*. Si se cumple que $\sum_{j=1}^N p_{ij}^{(t)} = 1 \forall i$, entonces la matriz de transición es *estocástica*. $\mathbf{P}^{(0)}$ se define como la matriz identidad, esto es, $\mathbf{P}^{(0)} = \mathbf{I}$.

- La sucesión $\{Y_t\}$ es una *cadena de Markov* (CM) con distribución inicial $\mathbf{p}^{(0)}$ y con matrices de transición estocásticas $\mathbf{P}^{(t)}$ si cumple con la siguiente *propiedad de Markov*.

$$\Pr(Y_t = s_j | Y_{t-1} = s_i, \dots, Y_0 = s_0) = \Pr(Y_t = s_j | Y_{t-1} = s_i) = p_{ij}^{(t)} \quad \forall t,$$

donde $\Pr(Y_{t-1} = s_i, \dots, Y_0 = s_0) > 0$. Es decir, la probabilidad de que Y tome cierto valor de S en un tiempo dado, sólo depende de éste valor y del valor precedente. Si $\mathbf{P}^{(t)}$ es invariante en el tiempo, $\mathbf{P}^{(t)} = \mathbf{P}^{(t+1)} = \mathbf{P} \forall t > 0$, entonces se dice que la CM es *homogénea*. En caso contrario, tenemos una CM *no homogénea*. La propiedad de Markov define la “memoria” u “orden” de la cadena, que en este caso es igual a uno.

- Una CM de *orden* k es aquella que cumple con la siguiente propiedad de Markov generalizada para toda t .

$$\Pr(Y_t = s_t | Y_{t-1} = s_{t-1}, \dots, Y_0 = s_0) = \Pr(Y_t = s_t | Y_{t-1} = s_{t-1}, \dots, Y_{t-k} = s_{t-k}).$$

Algunas veces es posible descomponer el espacio de estados de una CM en subconjuntos más pequeños llamados *clases*, cada una de las cuales es relativamente más fácil de tratar, y una vez juntas dan un mejor entendimiento de la cadena completa.

El hecho de que de un estado s_i se puede pasar a un estado s_j con $p_{ij} > 0$, se denota como $s_i \rightarrow s_j$. Se dice que un estado s_i se *comunica* con un estado s_j si se cumple que $s_i \rightarrow s_j$ y $s_j \rightarrow s_i$. Es claro que $s_i \rightarrow s_j$ y $s_j \rightarrow s_k$ implica que $s_i \rightarrow s_k$. Entonces, una clase es el conjunto de estados que se comunican entre sí en forma directa o a través de terceros estados.

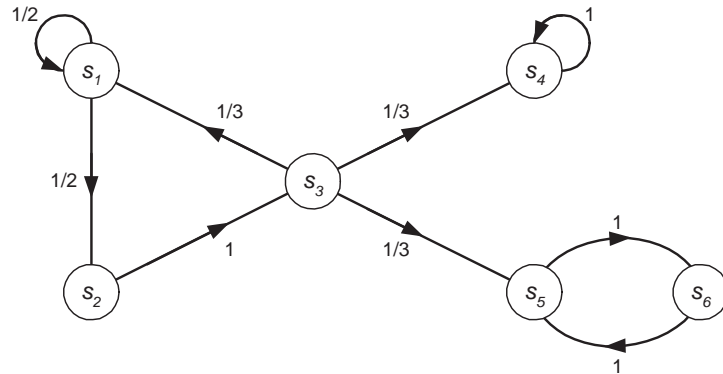
Se dice que una clase es *cerrada* si una vez entrando en ella ya no es posible salir de la misma. Una clase es *recurrente* (o *persistente*) sii es una clase cerrada, en caso contrario es una clase *transitoria*. Si una clase recurrente se compone de un sólo estado s_i , éste es llamado estado *absorbente*, donde $p_{ii} = 1$. Los estados no absorbentes son llamados *transitorios*, donde $0 \leq p_{ii} < 1$. Una CM es *absorbente* si tiene al menos un estado absorbente.

Una CM es *irreductible* si su espacio de estados no puede ser descompuesto en clases más pequeñas, por lo que el espacio de estados forma una sola clase, en la que siempre se puede pasar de cualquier estado a cualquier otro estado con probabilidad de transición positiva. Un resultado importante es que una CM irreductible con un espacio de estados finito, es siempre recurrente.

Por ejemplo, en la siguiente CM absorbente

$$\mathbf{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

cuyo diagrama correspondiente es



las clases son $\{s_1, s_2, s_3\}$, $\{s_4\}$ y $\{s_5, s_6\}$, donde $\{s_1, s_2, s_3\}$ es una clase transitoria, s_4 es un estado absorbente y $\{s_5, s_6\}$ es una clase recurrente.

Uno de los principales objetivos en el análisis de una CM es el cálculo de distribuciones futuras dadas las matrices de transición y distribuciones precedentes. Para una CM no homogénea las distribuciones subsecuentes se obtienen de la siguiente manera.

$$\mathbf{p}^{(1)} = \mathbf{p}^{(0)} \mathbf{P}^{(1)}, \quad \mathbf{p}^{(2)} = \mathbf{p}^{(1)} \mathbf{P}^{(2)} = \mathbf{p}^{(0)} \mathbf{P}^{(1)} \mathbf{P}^{(2)}, \dots$$

En general,

$$\mathbf{p}^{(n)} = \mathbf{p}^{(n-1)} \mathbf{P}^{(n)} = \mathbf{p}^{(0)} \mathbf{P}^{(1)} \dots \mathbf{P}^{(n)}.$$

En caso de que la CM sea homogénea se tiene que

$$\mathbf{p}^{(1)} = \mathbf{p}^{(0)} \mathbf{P}, \quad \mathbf{p}^{(2)} = \mathbf{p}^{(1)} \mathbf{P} = \mathbf{p}^{(0)} \mathbf{P}^2, \dots$$

En general,

$$\mathbf{p}^{(n)} = \mathbf{p}^{(n-1)} \mathbf{P} = \mathbf{p}^{(0)} \mathbf{P}^n,$$

donde la matriz \mathbf{P}^n es llamada matriz de transición de n pasos. Si \mathbf{P} es positiva (sin elementos cero) y conforme n se incrementa, $\mathbf{p}^{(n)}$ se aproxima a un *estado estable* o

distribución de equilibrio que satisface $\mathbf{p} = \mathbf{pP}$. También se pueden usar los términos *estacionaria* o *invariante* con el mismo significado. Otro resultado importante, es que toda CM recurrente tiene asociada una única distribución de equilibrio positiva.

3.2. Modelos de Transición

Los modelos de transición son una extensión de los GLMs que describen la distribución condicional de cada respuesta y_{it} como una función explícita de las respuestas previas $y_{it-1}, y_{it-2}, \dots, y_{i0}$ y de las variables explicativas x_{it} . Se supondrá el caso en el que los tiempos de observación están igualmente espaciados. Para simplificar la notación, se denota la historia de las respuestas previas y el valor pasado y presente de las variables explicativas para el sujeto i al tiempo t como

$$H_{it} = \{y_{it-1}, y_{it-2}, \dots, y_{i0}; x_{it}, x_{it-1}, \dots, x_{i0}\}.$$

Los modelos de transición más usuales son los que siguen una estructura tipo CM de orden k , en los cuales la distribución condicional de y_{it} dada H_{it} sólo depende de las k respuestas previas

$$H_{it} = \{y_{it-1}, y_{it-2}, \dots, y_{it-k}; x_{it}, x_{it-1}, \dots, x_{it-k}\}.$$

Al número natural k se le llama *orden* del modelo de transición.

Como las distribuciones condicionales pertenecen a la familia exponencial, la pdf condicional tiene la forma

$$f(y_{it}|H_{it}) = \exp \left\{ \frac{y_{it}\theta_{it} - b(\theta_{it})}{a(\phi)} + c(y_{it}, \phi) \right\}, \quad (3.1)$$

donde $a(\cdot)$, $b(\cdot)$ y $c(\cdot)$ son funciones conocidas. La media y varianzas condicionales son

$$\mu_{it}^c = E(Y_{it}|H_{it}) = b'(\theta_{it}) \quad \text{y} \quad v_{it}^c = \text{var}(Y_{it}|H_{it}) = b''(\theta_{it})a(\phi).$$

Las funciones de enlace g y varianza V satisfacen

$$g(\mu_{it}^c) = \mathbf{x}_{it}^T \boldsymbol{\beta} + \sum_{r=1}^s f_r(H_{it}; \boldsymbol{\alpha}) \quad \text{y} \quad v_{it}^c = V(\mu_{it}^c),$$

donde $f_r(\cdot)$ es una función adecuada del vector de los coeficientes de regresión de las respuestas previas $\boldsymbol{\alpha}$.

Es decir, el modelo de transición expresa la media condicional como una combinación lineal de las variables explicativas y las respuestas previas, que simplemente son tratadas como variables explicativas adicionales. Se asume que el pasado afecta al presente a través de la suma de s términos, cada uno de los cuales puede depender de los k valores previos.

Los modelos de transición también pueden ser ajustados usando máxima verosimilitud. La distribución conjunta de las respuestas Y_{i0}, \dots, Y_{in_i} , se factoriza de la forma

$$f(y_{i0}, \dots, y_{in_i}) = f(y_{i0}) f(y_{i1}|y_{i0}) f(y_{i2}|y_{i1}, y_{i0}) \cdots f(y_{in_i}|y_{in_i-1}, \dots, y_{i0}),$$

mientras que en una CM de orden k se tiene

$$f(y_{it}|y_{it-1}, \dots, y_{i0}) = f(y_{it}|y_{it-1}, \dots, y_{it-k}),$$

por lo que la contribución del sujeto i a la función de verosimilitud de un modelo de orden k es

$$f(y_{i0}, \dots, y_{in_i}) = f(y_{i0}, \dots, y_{ik-1}) \prod_{t=k}^{n_i} f(y_{it}|y_{it-1}, \dots, y_{it-k}),$$

que al introducir explícitamente las variables explicativas queda como

$$L_i(\boldsymbol{\beta}, \boldsymbol{\alpha}|y_{i0}, \dots, y_{in_i}; \mathbf{x}) = f(y_{i0}, \dots, y_{ik-1}; \mathbf{x}) \prod_{t=k}^{n_i} f(y_{it}|H_{it}).$$

En la función de verosimilitud anterior, la distribución condicional $f(y_{it}|H_{it})$ está especificada por la ecuación (3.1), mientras que la verosimilitud de la distribución marginal $f(y_{i0}, \dots, y_{ik-1}; \mathbf{x})$ no está especificada directamente. Para un modelo de respuestas continuas, en el que Y_{i0}, \dots, Y_{ik-1} es normal multivariada y la estructura de covarianza para Y_{it} es débilmente estacionaria, la distribución marginal puede ser determinada en su totalidad de la distribución condicional sin introducir parámetros desconocidos adicionales. Por lo tanto, una función de verosimilitud completa puede ser usada para ajustar modelos autorregresivos Gaussianos.

Por otro lado, para modelos de respuestas categóricas, la distribución marginal no puede ser especificada de la distribución condicional sin suposiciones adicionales, por lo que una función de verosimilitud completa no está disponible. Sin embargo, una alternativa para estimar $\boldsymbol{\beta}$ y $\boldsymbol{\alpha}$ es maximizar la función de *verosimilitud condicional*

$$L(\boldsymbol{\beta}, \boldsymbol{\alpha} | y_1, \dots, y_n; \mathbf{x}) = \prod_{i=1}^n \prod_{t=k}^{n_i} f(y_{it}|H_{it}). \quad (3.2)$$

Cuando se maximiza (3.2) se presentan dos casos distintos a considerar. En el primer caso, $f_r(H_{it}; \boldsymbol{\alpha}) = \alpha_r f_r(H_{it})$, por lo que la función de enlace es una combinación lineal de $\boldsymbol{\beta}$ y $\boldsymbol{\alpha}$ y el procedimiento para obtener sus MLEs condicionales es el estándar para GLM's de datos independientes. Simplemente se hace una regresión entre Y_{it} y el vector de variables explicativas extendidas $(\mathbf{x}_{it}, f_1(H_{it}), \dots, f_s(H_{it}))$ de dimensión $p + s$. Este caso es el que considera al modelar respuestas binarias.

El segundo caso ocurre cuando $f_r(H_{it}; \boldsymbol{\alpha}, \boldsymbol{\beta}) = \alpha_r f_r(H_{it}, \boldsymbol{\beta})$, y se considera al modelar, por ejemplo, respuestas contables. Un algoritmo de estimación para este caso proporciona las siguientes ecuaciones estimadoras condicionales.

$$\sum_{i=1}^n \sum_{t=k}^{n_i} \frac{y_{it} - \mu_{it}^c}{V(\mu_{it}^c)} \left(\frac{\partial \mu_{it}^c}{\partial \boldsymbol{\delta}} \right) = \mathbf{0},$$

donde $\boldsymbol{\delta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$. Estas ecuaciones estimadoras son el análogo condicional de las ecuaciones (2.20). Se puede aún formular el procedimiento de estimación en base a IRLS.

3.3. Modelos de Transición de Respuestas Binarias

Un modelo de transición de primer orden para respuestas binarias comprende una CM con un espacio de estados $\{0, 1\}$ y matrices de transición entre los tiempos $t - 1$ y t con $t > 0$ dadas por

$$\mathbf{P}^{(t)} = \begin{bmatrix} p_{00}^{(t)} & p_{01}^{(t)} \\ p_{10}^{(t)} & p_{11}^{(t)} \end{bmatrix},$$

donde $p_{ab}^{(t)} = \Pr(Y_{it} = b | Y_{it-1} = a)$, $a, b = 0, 1$ son las probabilidades de transición.

En la parte de regresión, se modelan las probabilidades de transición en función de las variables explicativas y las respuestas previas. Por ejemplo, en un modelo de regresión probit tenemos que

$$\begin{aligned} \mu_{it}^c &= \mathbb{E}(Y_{it} | Y_{it-1}) = \Pr(Y_{it} = 1 | Y_{it-1} = a), \quad a = 0, 1 \quad \text{y} \\ v_{it}^c &= \mathbb{V}(Y_{it} | Y_{it-1}) = \mu_{it}^c(1 - \mu_{it}^c), \end{aligned}$$

con función de enlace

$$g(\mu_{it}^c) = \text{probit}(\mu_{it}^c) = \Phi^{-1}(\mu_{it}^c) = \mathbf{x}_{it}^T \boldsymbol{\beta} + y_{it-1} \mathbf{x}_{it}^T \boldsymbol{\alpha},$$

donde se incluye la interacción de las variables explicativas con la respuesta previa. Para un modelo sin interacción, el término $y_{it-1} \mathbf{x}_{it}^T \boldsymbol{\alpha} = \alpha y_{it-1}$; es decir, las variables explicativas tienen el mismo efecto en la probabilidad de la respuesta, aún si la respuesta previa es igual a cero o igual a uno.

Bajo este modelo, las matrices de transición quedan especificadas como

$$\mathbf{P}^{(t)} = \begin{bmatrix} \Phi(-\mathbf{x}_{it}^T \boldsymbol{\beta}) & \Phi(\mathbf{x}_{it}^T \boldsymbol{\beta}) \\ \Phi(-\mathbf{x}_{it}^T \boldsymbol{\beta} - \mathbf{x}_{it}^T \boldsymbol{\alpha}) & \Phi(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{x}_{it}^T \boldsymbol{\alpha}) \end{bmatrix}.$$

Un modelo de transición de segundo orden está caracterizado por la matriz

		Y_{it}	
Y_{it-2}	Y_{it-1}	0	1
0	0	p_{000}	p_{001}
0	1	p_{010}	p_{011}
1	0	p_{100}	p_{101}
1	1	p_{110}	p_{111}

donde $p_{abc} = \Pr(Y_{it} = c | Y_{it-1} = b, Y_{it-2} = a)$, $a, b, c = 0, 1$ son las probabilidades de transición. Análogamente al modelo de primer orden, en un modelo de regresión probit tenemos que

$$\begin{aligned} \mu_{it}^c &= E(Y_{it} | Y_{it-1}, Y_{it-2}) = \Pr(Y_{it} = 1 | Y_{it-1} = b, Y_{it-2} = a), \quad a, b = 0, 1 \quad \text{y} \\ v_{it}^c &= V(Y_{it} | Y_{it-1}, Y_{it-2}) = \mu_{it}^c (1 - \mu_{it}^c), \end{aligned}$$

con función de enlace

$$g(\mu_{it}^c) = \text{probit}(\mu_{it}^c) = \Phi^{-1}(\mu_{it}^c) = \mathbf{x}_{it}^T \boldsymbol{\beta} + y_{it-1} \mathbf{x}_{it}^T \boldsymbol{\alpha}_1 + y_{it-2} \mathbf{x}_{it}^T \boldsymbol{\alpha}_2 + y_{it-1} y_{it-2} \mathbf{x}_{it}^T \boldsymbol{\alpha}_3,$$

donde se incluyen todas las interacciones de las variables explicativas con las respuestas previas.

Un caso especial importante del modelo anterior ocurre cuando no hay interacciones entre las respuestas pasadas y las variables explicativas, por lo que la función de enlace queda como

$$g(\mu_{it}^c) = \text{probit}(\mu_{it}^c) = \Phi^{-1}(\mu_{it}^c) = \mathbf{x}_{it}^T \boldsymbol{\beta} + \alpha_1 y_{it-1} + \alpha_2 y_{it-2} + \alpha_3 y_{it-1} y_{it-2}.$$

En este caso, las respuestas previas afectan la probabilidad de una respuesta positiva, pero los efectos de las variables explicativas son los mismos sin hacer caso de la historia. Incluso en esta situación, aún se debe escoger entre modelos de transición de diferente orden. Por ejemplo, se podría empezar con un modelo de tercer orden que puede ser escrito de la forma

$$\begin{aligned}\Phi^{-1}(\mu_{it}^c) &= \mathbf{x}_{it}^T \boldsymbol{\beta} + \alpha_1 y_{it-1} + \alpha_2 y_{it-2} + \alpha_3 y_{it-3} \\ &\quad + \alpha_4 y_{it-1} y_{it-2} + \alpha_5 y_{it-1} y_{it-3} + \alpha_6 y_{it-2} y_{it-3} + \alpha_7 y_{it-1} y_{it-2} y_{it-3}.\end{aligned}$$

Un modelo de segundo orden puede ser utilizado si los datos son consistentes con $\alpha_3 = \alpha_5 = \alpha_6 = \alpha_7 = 0$; un modelo de primer orden estará implicado si se cumple que $\alpha_r = 0$ para $r = 2, \dots, 7$. Como todo coeficiente de regresión, la interpretación y valor de $\boldsymbol{\beta}$ depende de las otras variables explicativas en el modelo, en particular en aquellos en los que las respuestas previas han sido incluidas. Cuando las inferencias acerca de $\boldsymbol{\beta}$ son el objetivo estadístico, es primordial revisar su sensibilidad respecto al orden asumido en el modelo de transición.

3.4. Aplicación al Ejemplo Ilustrativo

Retomando el análisis del tratamiento que intenta reducir la exposición al sol de los niños, sólo se modelará la parte de la base de datos cuyos sujetos tienen sus tres respuestas completas clasificadas como se muestra en la Tabla 3.1.

En la tabla 3.2 se reportan las proporciones muestrales para las cuatro matrices de transición de primer orden que se tienen. Por ejemplo, de la Tabla 3.1, la proporción de respuestas favorables (codificadas como 1) para el grupo de control al segundo año de seguimiento, dado que la respuesta previa fue no favorable (codificada como 0), es de $(4 + 5)/(13 + 3 + 4 + 5) = 0.36$.

Tabla 3.1: Clasificación de las respuestas completas por grupo.

Grupo	Respuestas en tres años								Total
	000	001	010	011	100	101	110	111	
Control	13	3	4	5	32	20	26	129	232
Tratamiento	9	4	7	3	18	18	32	143	234

Tabla 3.2: Proporciones de transición muestrales.

Grupo	Año previo	Año 2		Año 3	
		0	1	0	1
Control	0	0.6400	0.3600	0.6618	0.3382
	1	0.2512	0.7488	0.1829	0.8171
Tratamiento	0	0.5652	0.4348	0.5510	0.4490
	1	0.1706	0.8294	0.2108	0.7892

De la tabla 3.2 se puede observar que la proporción de respuestas favorables aparentemente (1) es similar con el tiempo para ambos grupos; (2) es mayor si la respuesta previa es también favorable; y (3) es ligeramente mayor en el grupo intervenido que en el grupo de control. Para averiguar que tan ciertas son las aseveraciones de este análisis exploratorio y así determinar si las diferencias presentes entre grupos y entre años son estadísticamente significativas, se hará uso de un modelo de transición con función de enlace probit. Para ajustar el modelo con GLM se utilizará la función `glm` implementada en el software estadístico R.

Primero se ajustaron modelos con todo tipo de interacciones, y al verificar que ninguna de ellas era estadísticamente significativa se eliminaron del modelo. En la Tabla 3.3 se muestra el reporte del ajuste del modelo probit sin interacciones.

Tabla 3.3: Estimadores GLM con función de enlace probit.

Parámetro	Estimador	Error Est.	Valor P
Intercepto	-0.42373	0.27442	0.123
tiempo	0.02800	0.09337	0.764
Tratamiento	0.14217	0.09136	0.120
Previo	1.10852	0.11435	0.000

Como podemos ver, no hay evidencia suficiente para suponer que las proporciones de respuestas favorables sean diferentes con el tiempo o mayores en el grupo intervenido. Sin embargo, lo que sí se puede afirmar, es que dichas proporciones aumentan fuertemente si la respuesta previa es también favorable. Entonces, el modelo marginal probit es

$$\begin{aligned} \Phi^{-1}[\Pr(Y_{it} = 1 | Y_{it-1} = \text{Previo})] = & - 0.4237 + 0.028 \text{ tiempo} \\ & + 0.1422 \text{ Tratamiento} + 1.1085 \text{ Previo} \end{aligned}$$

donde Φ es la CDF normal estándar.

Finalmente, las probabilidades de transición estimadas se muestran a continuación.

Tabla 3.4: Probabilidades de transición estimadas.

Grupo	Año previo	Año 2		Año 3	
		0	1	0	1
Control	0	0.6435	0.3565	0.6330	0.3670
	1	0.2294	0.7706	0.2210	0.7790
Tratamiento	0	0.5892	0.4108	0.5783	0.4217
	1	0.1886	0.8114	0.1812	0.8188

La desventaja de los modelos de transición bajo esta propuesta de Diggle *et al.* (1994), es que no pueden incluirse sujetos con datos faltantes, lo cual provoca pérdida de información relevante. Esto se debe a que el ajuste del modelo se hace sobre proporciones obtenidas sólo de frecuencias observadas. Por ejemplo, en este caso, el ajuste del modelo de transición se hizo sobre las proporciones de la Tabla 3.2 obtenidas de las frecuencias de la Tabla 3.1.

Capítulo 4

Modelos Marginales y de Transición con Cópulas

Las cópulas son funciones C que relacionan, juntan -o bien “copulan”- funciones de distribución conjuntas multivariadas con sus funciones de distribución marginales univariadas. Esto es,

$$C : (0, 1)^m \rightarrow (0, 1) \mid C [F_1(y_1), F_2(y_2), \dots, F_m(y_m)] = F(y_1, y_2, \dots, y_m).$$

Las cópulas fueron presentadas originalmente por Sklar (1959) quien resolvió algunos problemas formulados por M. Fréchet sobre la relación entre una función de distribución conjunta y sus funciones de distribución marginales. En este sentido las cópulas proporcionan una estructura general unificadora para el modelado de distribuciones multivariadas. En la actualidad, las cópulas se han convertido en una poderosa y popular herramienta de modelado multivariado en muchos campos de la investigación, sobre todo donde la dependencia entre varias variables aleatorias es de gran interés y para las cuales la suposición de normalidad multivariada puede ser cuestionable.

4.1. Cópulas Bivariadas

4.1.1. Conceptos Básicos

El siguiente teorema de Sklar (1959) es central en la teoría de cópulas y es el fundamento para muchas, si no es que la mayoría, de las aplicaciones de dicha teoría a la estadística. Este teorema aclara el rol que juegan las cópulas en la relación entre funciones de distribución multivariadas y sus marginales univariadas.

Teorema 4.1 (Sklar) *Sean Y_1 y Y_2 variables aleatorias con función de distribución conjunta F y funciones de distribución marginales F_1 y F_2 , respectivamente. Entonces existe una cópula C tal que satisface*

$$F(y_1, y_2) = C[F_1(y_1), F_2(y_2)], \quad (4.1)$$

para toda $y_1, y_2 \in \mathbb{R}$. Si F_1 y F_2 son continuas, entonces C es única sobre rango $F_1 \times$ rango F_2 . Inversamente, si C es una cópula y F_1, F_2 son funciones de distribución, entonces F definida en la ecuación (4.1) es una función de distribución conjunta con marginales F_1 y F_2 .

La ecuación (4.1) proporciona una expresión para funciones de distribución conjuntas en términos de una cópula y dos funciones de distribución marginales. Pero (4.1) puede ser invertida para expresar las cópulas en términos de una función de distribución conjunta y las inversas de las dos funciones de distribución marginales. Sin embargo, si una marginal no es estrictamente creciente entonces no tiene una inversa en el sentido usual. Esto proporciona un método de construcción de cópulas a partir de funciones de distribución conjuntas. Si $v_1 = F_1(y_1)$ y $v_2 = F_2(y_2)$, entonces $y_1 = F_1^{-1}(v_1)$ y $y_2 = F_2^{-1}(v_2)$. Por lo tanto,

$$C(v_1, v_2) = F(F_1^{-1}(v_1), F_2^{-1}(v_2)). \quad (4.2)$$

Una *cópula bivariada* es una función $C : (0, 1)^2 \rightarrow (0, 1)$ que satisface las siguientes propiedades:

- Para todo v_1, v_2 en $(0, 1)$,

$$\begin{aligned}\lim_{v_j \rightarrow 1} C(v_1, v_2) &= v_{3-j}, \\ \lim_{v_j \rightarrow 0} C(v_1, v_2) &= 0;\end{aligned}$$

donde $j = 1, 2$.

- Para todo u_1, u_2, w_1, w_2 en $(0, 1)$ tales que $u_1 \leq u_2$ y $w_1 \leq w_2$,

$$C(u_2, w_2) - C(u_2, w_1) - C(u_1, w_2) + C(u_1, w_1) \geq 0.$$

Debido a esta última propiedad, algunos autores se refieren a la cópula C como *bi-creciente* o *cuasi-monótona*.

Aquí un hecho importante es que una cópula es también creciente en cada uno de sus argumentos.

Por otro lado, es posible demostrar que si C es una cópula, entonces para todo (v_1, v_2) en $(0, 1)^2$,

$$W(v_1, v_2) = \max\{v_1 + v_2 - 1, 0\} \leq C(v_1, v_2) \leq \min\{v_1, v_2\} = M(v_1, v_2).$$

donde W y M son también cópulas. Como una consecuencia del teorema de Sklar, si Y_1 y Y_2 con variables aleatorias con una función de distribución conjunta F y marginales F_1 y F_2 , respectivamente, entonces para todo $(y_1, y_2) \in \mathbb{R}^2$,

$$\max\{F_1(y_1) + F_2(y_2) - 1, 0\} \leq F(y_1, y_2) \leq \min\{F_1(y_1), F_2(y_2)\}.$$

Esta desigualdad es conocida como la *desigualdad de las cotas de Fréchet-Hoeffding*, y las funciones W y M como las cotas inferior y superior, respectivamente.

También se puede establecer la continuidad uniforme de las cópulas en su dominio vía una condición de Lipschitz en $(0, 1)^2$ de la forma siguiente. Para todo $(u_1, u_2), (w_1, w_2)$ en $(0, 1)^2$,

$$|C(u_2, w_2) - C(u_1, w_1)| \leq |u_2 - u_1| + |w_2 - w_1|.$$

Esto implica que la representación gráfica de una cópula sea una superficie continua $z = C(u, v)$ dentro del cubo unitario $(0, 1)^3$; y situada entre las gráficas de las cotas de Frèchet $z = W(u, v)$ y $z = M(u, v)$, como se muestra en la Figura 4.1. Como ejemplo, se incluyó la gráfica de una cópula importante llamada *cópula producto* o *cópula independiente* $z = \Pi(u, v) = uv$.

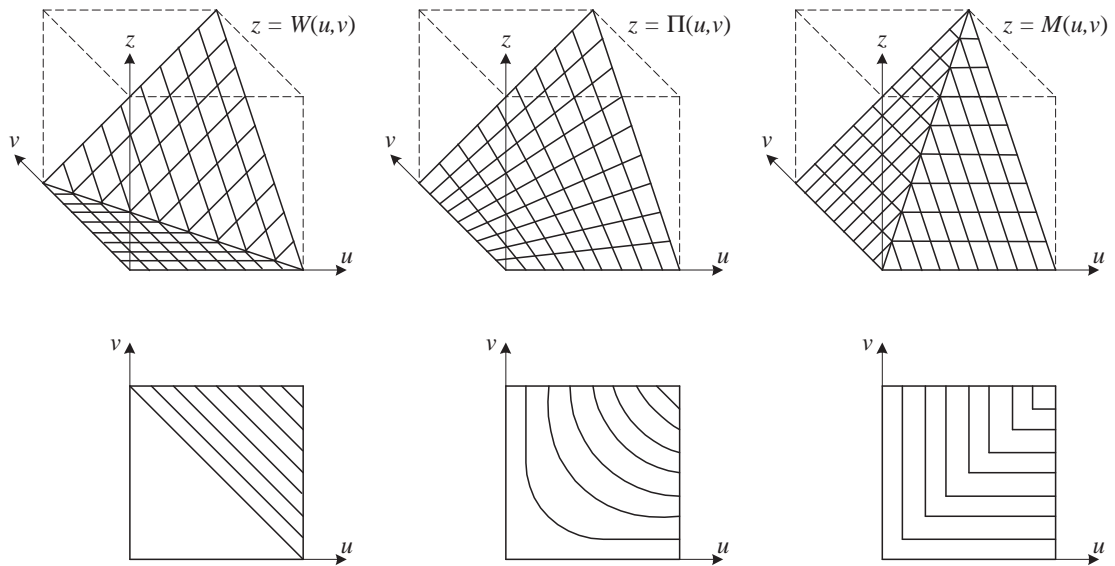


Figura 4.1: Gráficas y diagramas de contornos de las cópulas W , Π y M .

Se puede hablar de un orden parcial en el conjunto de cópulas llamado *orden concordante*, en el que W y M son las copulas mínima y máxima, respectivamente. Es un orden parcial y no un orden total porque no todo par de cópulas es comparable.

Con una extensión apropiada de su dominio a \mathbb{R}^2 , toda cópula bivariada es una función de distribución conjunta con funciones de distribución marginales uniformes en $(0, 1)$. Es decir, si C es una cópula en $(0, 1)^2$, se extiende a \mathbb{R}^2 por medio de F_C definida como

$$F_C(x, y) = \begin{cases} 0 & : x \leq 0 \text{ o } y \leq 0 \\ C(x, y) & : (x, y) \in (0, 1)^2 \\ x & : y \geq 1, x \in (0, 1) \\ y & : x \geq 1, y \in (0, 1) \\ 1 & : x \geq 1 \text{ y } y \geq 1 \end{cases}$$

Entonces, F_C es una función de distribución conjunta con funciones de distribución marginales uniformes $U(0,1)$. De hecho, con bastante frecuencia es útil pensar en las cópulas como la restricción a $(0, 1)^2$ de funciones de distribución conjuntas cuyas marginales son $U(0,1)$. En la siguiente figura se muestra la gráfica de la superficie $z = F_C(x, y)$, donde $C(x, y) = \Pi(x, y) = xy$ como caso especial.

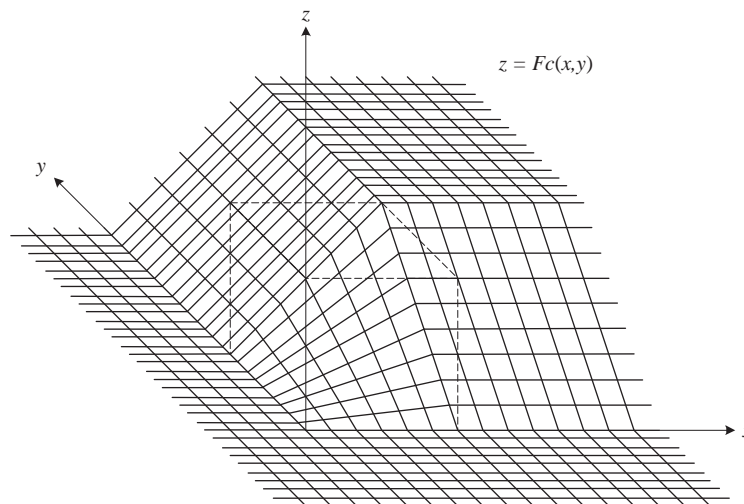


Figura 4.2: Gráfica de $z = F_C(x, y)$.

Finalmente, veamos ciertos resultados importantes con respecto a las derivadas parciales de las cópulas. Si $F_1(y_1)$, $F_2(y_2)$ y la cópula $C(v_1, v_2)$ son diferenciables, entonces la función de densidad conjunta de (Y_1, Y_2) puede expresarse como:

$$\begin{aligned}
f(y_1, y_2) &= \frac{\partial^2 F(y_1, y_2)}{\partial y_1 \partial y_2} \\
&= \frac{\partial}{\partial y_1} \left\{ \frac{\partial}{\partial y_2} C [F_1(y_1), F_2(y_2)] \right\} \\
&= \frac{\partial}{\partial y_1} \left\{ \frac{\partial}{\partial F_2(y_2)} C [F_1(y_1), F_2(y_2)] \cdot \frac{dF_2(y_2)}{dy_2} \right\} \\
&= f_2(y_2) \frac{\partial}{\partial y_1} \left\{ \frac{\partial}{\partial F_2(y_2)} C [F_1(y_1), F_2(y_2)] \right\} \\
&= f_2(y_2) \frac{\partial}{\partial F_1(y_1)} \left\{ \frac{\partial}{\partial F_2(y_2)} C [F_1(y_1), F_2(y_2)] \right\} \cdot \frac{dF_1(y_1)}{dy_1} \\
&= f_1(y_1) f_2(y_2) \frac{\partial^2 C [F_1(y_1), F_2(y_2)]}{\partial F_1(y_1) \partial F_2(y_2)} \\
&= f_1(y_1) f_2(y_2) c [F_1(y_1), F_2(y_2)].
\end{aligned}$$

donde $f_1(y_1)$ y $f_2(y_2)$ son las funciones de densidad marginales de Y_1 y Y_2 , respectivamente; y c es la *función de densidad de la cópula*. Como consecuencia, se tiene que la función de densidad condicional de Y_2 dada Y_1 puede convenientemente expresarse de la siguiente forma:

$$f(y_2 | y_1) = \frac{f(y_1, y_2)}{f_1(y_1)} = f_2(y_2) c [F_1(y_1), F_2(y_2)]. \quad (4.3)$$

4.1.2. Familias de Cópulas Paramétricas

Si se tiene una colección de cópulas, entonces, como una consecuencia del teorema de Sklar, automáticamente se tiene una colección de distribuciones bivariadas o multivariadas con cualesquiera distribuciones marginales que se desee. Claramente esto puede ser útil en modelado y simulación. Además, la naturaleza no paramétrica de la dependencia entre dos variables aleatorias es expresada por la cópula. Por esto, es conveniente tener una variedad de cópulas a nuestra disposición.

Las dos clases de cópulas de un parámetro más frecuentemente usados son las cópulas *elípticas* y las cópulas *Arquimedianas*. Una cópula elíptica es la correspondiente a una distribución elíptica; es decir, aquellas cuya función de densidad tiene contornos que son elipses concéntricos con excentricidad constante. Estas cópulas se construyen por medio del *método de inversión* vía (4.2).

Dentro de las cópulas elípticas, una familia importante es la *Gaussiana* o *normal* bivariada, que tiene la forma

$$C_\rho(v_1, v_2) = \Phi_2[\Phi^{-1}(v_1), \Phi^{-1}(v_2)], \quad (v_1, v_2)^T \in (0, 1)^2, \quad (4.4)$$

donde Φ_2 es la función de distribución normal bivariada con media $(0, 0)^T$ y varianza común 1, por lo que la matriz de varianza-covarianza es igual a la matriz de correlación \mathbf{R} , que es una matriz no singular de 2×2 cuyos elementos fuera de la diagonal principal son los coeficientes de correlación de Pearson $\rho \in (-1, 1)$, y los elementos en la diagonal principal son iguales a uno. Φ^{-1} es la función inversa de la función de distribución normal estándar Φ . La función de densidad de la cópula está dada por

$$c_\rho(v_1, v_2) = \frac{\phi_2[\Phi^{-1}(v_1), \Phi^{-1}(v_2)]}{\phi[\Phi^{-1}(v_1)] \phi[\Phi^{-1}(v_2)]},$$

donde ϕ es la función de densidad normal estándar y ϕ_2 es la función de densidad

normal bivariada definida como

$$\phi_2(\mathbf{z}) = (2\pi)^{-1} |\mathbf{R}|^{-1/2} \exp \left\{ -\frac{1}{2} \mathbf{z}^T \mathbf{R}^{-1} \mathbf{z} \right\}, \quad \mathbf{z}^T \in \mathbb{R}^2.$$

Nótese que si Y_1 y Y_2 están normalmente distribuidas, la función de densidad conjunta resultante generada con la cópula Gaussiana se reduce a la función de densidad normal bivariada usual.

Conforme a la ecuación (4.3) se tiene que la función de densidad condicional de Y_2 dada Y_1 correspondiente a la cópula Gaussiana está dada por

$$f_{2|1}(y_2 | y_1) = f_2(y_2) \frac{\phi_2 [\Phi^{-1}(F_1(y_1)), \Phi^{-1}(F_2(y_2))]}{\phi[\Phi^{-1}(F_1(y_1))] \phi[\Phi^{-1}(F_2(y_2))]}.$$

Después de algo de álgebra laboriosa se puede demostrar que

$$f_{2|1}(y_2 | y_1) = \frac{f_2(y_2)}{\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2} \left[\frac{(s_2 - \rho s_1)^2}{1-\rho^2} - s_2^2 \right] \right\},$$

donde $s_i = \Phi^{-1}[(F_i(y_i))]$ para $i = 1, 2$.

El uso de la cópula Gaussiana bivariada es atractivo ya que codifica la dependencia en la misma forma en que la distribución normal bivariada lo hace usando el parámetro de dependencia ρ , con la diferencia de que lo hace para variables aleatorias con cualesquiera marginales arbitrarias. Esta cópula tiene la capacidad de capturar el rango completo de dependencia ya que $C_{-1} = W$, $C_0 = \Pi$ y $C_1 = M$.

Por otro lado, las cópulas Arquimedianas tienen un amplio rango de aplicaciones en el modelado estadístico por varias razones: (1) la facilidad con la cual pueden ser construidas; (2) la gran variedad de estructuras de dependencia presentes en las familias de cópulas que pertenecen a esta clase; y (3) las varias propiedades deseables que poseen los miembros de esta clase. El término ‘‘Arquimediano’’ fue introducido por Ling (1965) ya que esta clase de cópulas satisfacen el axioma Arquimediano para los números reales positivos en su propia versión.

Así como la cópula $\Pi(u, v) = uv$ puede expresarse como el producto de sus argumentos y *funciones de cópula* pueden expresarse como el producto de *funciones de sus argumentos* $\phi[C(u, v)] = \phi(u)\phi(v)$ para alguna función $\phi(t)$, también una función de cópula puede expresarse como la suma de funciones de sus argumentos $\varphi[C(u, v)] = \varphi(u) + \varphi(v)$ para alguna función $\varphi(t)$. Como el interés primordial es tener una expresión que pueda ser usada para la construcción de cópulas, se requiere resolver la relación $\varphi[C(u, v)] = \varphi(u) + \varphi(v)$ para $C(u, v)$, esto es, $C(u, v) = \varphi^{-1}[\varphi(u) + \varphi(v)]$, donde φ^{-1} es la función inversa de φ . Dicho esto, se obtiene (y se puede probar) el siguiente resultado.

Una cópula Arquimediana tiene la forma

$$C_\alpha(v_1, v_2) = \varphi^{-1}[\varphi(v_1) + \varphi(v_2)], \quad (v_1, v_2)^T \in (0, 1)^2, \quad (4.5)$$

donde α es el *parámetro de dependencia* y φ es una función continua, decreciente y convexa $\varphi : (0, 1] \rightarrow [0, \infty) \mid \varphi(1) = 0$, llamada *generador* de la cópula C . Un generador determina únicamente una cópula Arquimediana. Por lo tanto, esta representación ayuda a identificar la forma de la cópula.

Algunas propiedades algebraicas de una cópula Arquimediana C con generador φ son:

- C es *simétrica*; $C(v_1, v_2) = C(v_2, v_1) \quad \forall v_1, v_2 \in (0, 1)$.
- C es *asociativa*; $C[C(v_1, v_2), w] = C[v_1, C(v_2, w)] \quad \forall v_1, v_2, w \in (0, 1)$.
- Si $k > 0$ es una constante, entonces $k\varphi$ es también un generador de C .

En la Tabla 4.1 se muestran diferentes generadores junto con el rango de su parámetro de dependencia, que determinan familias importantes de cópulas Arquimedianas de un parámetro.

Tabla 4.1: Familias de cópulas Arquimedianas de un parámetro.

	Generador $\varphi(t)$	Parámetro α	Cópula $C_\alpha(v_1, v_2)$	Casos especiales
1	$-\log t$	No aplica	$v_1 v_2$	Π
2	$1-t$	No aplica	$\max\{v_1+v_2-1, 0\}$	W
3	$(-\log t)^\alpha$	$\alpha \geq 1$	$\exp\left\{-\left[(-\log v_1)^\alpha + (-\log v_2)^\alpha\right]^{1/\alpha}\right\}$	$C_1=\Pi, C_\infty=M$
4	$t^{-\alpha}-1$	$\alpha > 0$	$(v_1^{-\alpha}+v_2^{-\alpha}-1)^{-1/\alpha}$	$C_0=\Pi, C_\infty=M$
5	$-\log[1-(1-t)^\alpha]$	$\alpha \geq 1$	$1-\left[(1-v_1)^\alpha+(1-v_2)^\alpha-(1-v_1)^\alpha(1-v_2)^\alpha\right]^{1/\alpha}$	$C_1=\Pi, C_\infty=M$
6	$-\log\frac{\exp(\alpha t)-1}{\exp(\alpha)-1}$	$\alpha \neq 0$	$-\frac{1}{\alpha}\log\left\{1+\frac{[\exp(\alpha v_1)-1][\exp(\alpha v_2)-1]}{\exp(\alpha)-1}\right\}$	$C_{-\infty}=W, C_0=\Pi$ $C_\infty=M$

Nota: La cópula M no es Arquimediana.

La familia 3 fue introducida por Gumbel (1960), por lo que muchos autores se refieren a ella como la familia *Gumbel*. Sin embargo, dado que el nombre de Gumbel está asociado a otra familia Arquimediana que también aparece en Hougaard (1986), Hutchinson y Lai (1990) se refieren a ella como la familia *Gumbel-Hougaard*.

La familia 4 fue presentada por Clayton (1978), Oakes (1982, 1986), Cox y Oakes (1984), y Cook y Johnson (1981, 1986). Genest y MacKay (1986) le llamaron la familia *Cook-Johnson* generalizada; Hutchinson y Lai (1990) le llamaron la familia *Pareto*; mientras que Genest y River (1993) le llamaron la familia *Clayton*.

La familia 5 es presentada en Joe (1993, 1997), y aparece en Frank (1981).

La familia 6 es la familia *Frank*, la cual es introducida en Frank (1979) en un contexto no estadístico. Algunas de las propiedades estadísticas de esta familia fueron presentadas en Nelsen (1986) y Genest (1987).

4.1.3. Medidas de Dependencia

La estructura de dependencia entre variables aleatorias representada por las cópulas permite un camino fácil en el estudio de las medidas de dependencia o asociación entre dichas variables aleatorias. Existen varias formas de describir y medir esta dependencia. En lo que respecta al término de “coeficiente de correlación”, generalmente en la literatura es aplicado para referirse a una medida de dependencia lineal entre variables aleatorias como es el caso del coeficiente de correlación de Pearson, mientras que el término de “medida de dependencia” es usado para las medidas tales como la tau de Kendall y la rho de Spearman, las cuales “midan” una forma de dependencia conocida como concordancia.

De manera informal, un par de variables aleatorias son concordantes si valores grandes (pequeños) de una de ellas tienden a estar asociados con valores grandes (pequeños) de la otra. Es decir, sean (x_i, y_i) y (x_j, y_j) dos observaciones de un vector (X, Y) de variables aleatorias continuas. Se dice que (x_i, y_i) y (x_j, y_j) son *concordantes* si $x_i < x_j$ y $y_i < y_j$, o $x_i > x_j$ y $y_i > y_j$. Similarmente, se dice que (x_i, y_i) y (x_j, y_j) son *discordantes* si $x_i < x_j$ y $y_i > y_j$, o $x_i > x_j$ y $y_i < y_j$. Alternativamente, (x_i, y_i) y (x_j, y_j) son concordantes si $(x_i - x_j)(y_i - y_j) > 0$ y discordantes si $(x_i - x_j)(y_i - y_j) < 0$.

1. Coeficiente de correlación de Pearson

Esta medida de dependencia lineal es la forma más tradicional para evaluar la asociación entre dos variables aleatorias continuas Y_1 y Y_2 en una distribución bivariada. Se define como

$$\frac{\text{Cov}(Y_1, Y_2)}{[\text{Var}(Y_1) \text{Var}(Y_2)]^{1/2}} \quad (4.6)$$

El uso del coeficiente de correlación de Pearson es conveniente pues es muy fácil de calcular y es una medida natural de dependencia en distribuciones elípticas;

de hecho, a menudo es usado en la familia Gaussiana o normal bivariada. Sin embargo, para algunos casos de parejas aleatorias continuas cuyas distribuciones no son elípticas, la utilidad de este coeficiente es dudosa pues su valor depende no solo de la cópula sino también de las distribuciones marginales, por lo que es afectado por cambios de escala no lineales.

2. *La tau de Kendall*

La versión muestral de esta medida de dependencia se define en términos de la concordancia como sigue: Sea $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ una muestra aleatoria de n pares de observaciones de un vector (X, Y) de variables aleatorias continuas; y sean c y d el número de pares concordantes y discordantes, respectivamente. Entonces, la tau de Kendall para la muestra se define como

$$\tau = \frac{c - d}{c + d} = \frac{c}{c + d} - \frac{d}{c + d}.$$

Equivalentemente, τ es la probabilidad de concordancia menos la probabilidad de discordancia para un par de observaciones (x_i, y_i) y (x_j, y_j) que son elegidas aleatoriamente de la muestra. Por lo que la versión poblacional de la tau de Kendall para un vector (X, Y) de variables aleatorias continuas con función de distribución conjunta F se define en base a dicha diferencia. Sean (X_1, Y_1) y (X_2, Y_2) vectores aleatorios independientes e idénticamente distribuidos, cada uno con función de distribución conjunta F . Entonces, la tau de Kendall para la población se define como

$$\begin{aligned} \tau &= \Pr[(X_1 - X_2)(Y_1 - Y_2) > 0] - \Pr[(X_1 - X_2)(Y_1 - Y_2) < 0] \\ &= 2\Pr[(X_1 - X_2)(Y_1 - Y_2) > 0] - 1. \end{aligned} \tag{4.7}$$

Se puede demostrar que si X y Y son variables aleatorias continuas cuya cópula de un parámetro es C_α , entonces en términos de la cópula la tau de Kendall

poblacional para X y Y se expresa como

$$\tau_\alpha = 4 \iint_{\mathbf{I}^2} C_\alpha(v_1, v_2) dC_\alpha(v_1, v_2) - 1, \quad (4.8)$$

donde $dC_\alpha(v_1, v_2) = c_\alpha(v_1, v_2)dv_1dv_2$.

En caso de que C_α sea una cópula Arquimediana generada por φ , entonces la tau de Kendall está dada por

$$\tau_\alpha = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt. \quad (4.9)$$

La tau de Kendall $\tau \in [-1, 1]$ con $\tau = 0$ si X y Y son variables aleatorias independientes.

3. *La rho de Spearman*

Al igual que la tau de Kendall, la versión poblacional de esta medida de dependencia está basada en concordancia y discordancia. Para obtenerla, sean ahora (X_1, Y_1) , (X_2, Y_2) , y (X_3, Y_3) tres vectores aleatorios independientes con una función de distribución conjunta común F . Entonces la rho de Spearman poblacional se define como proporcional a la probabilidad de concordancia menos la probabilidad de discordancia para los dos vectores (X_1, Y_1) y (X_2, Y_3) ; es decir, un par de vectores con las mismas marginales, pero un vector tiene función de distribución F , mientras las componentes de los otros son independientes:

$$\rho = 3\{\Pr[(X_1 - X_2)(Y_1 - Y_3) > 0] - \Pr[(X_1 - X_2)(Y_1 - Y_3) < 0]\}, \quad (4.10)$$

donde 3 es un coeficiente de normalización. (Se pudo haber usado el par (X_3, Y_2) igual de bien). Se puede demostrar que si X y Y son variables aleatorias continuas cuya cópula de un parámetro es C_α , entonces en términos de la cópula la rho de Spearman poblacional para X y Y se expresa como

$$\rho_\alpha = 12 \iint_{\mathbf{I}^2} v_1 v_2 dC_\alpha(v_1, v_2) - 3 = 12 \iint_{\mathbf{I}^2} C_\alpha(v_1, v_2) dv_1 dv_2 - 3. \quad (4.11)$$

En la Tabla 4.2 se muestran algunas familias de cópulas junto con las medidas de dependencia tau de Kendall y rho de Spearman correspondientes. En el caso de la familia Frank se usan las funciones llamadas “Debye”, definidas como

$$D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{e^t - 1} dt \quad \text{y} \quad D_k(-x) = D_k(x) + \frac{kx}{k+1}, \quad \text{para } k = 1, 2.$$

Tabla 4.2: Familias de cópulas y sus medidas de dependencia.

Familia	Parámetro	tau de Kendall τ	rho de Spearman ρ
Normal	$-1 \leq \rho \leq 1$	$(2/\pi) \operatorname{sen}^{-1}(\rho)$	$(6/\pi) \operatorname{sen}^{-1}(\rho/2)$
Gumbel	$\alpha \geq 1$	$(\alpha-1)/\alpha$	Sin forma cerrada
Clayton	$\alpha > 0$	$\alpha/(\alpha+2)$	Forma complicada
Frank	$\alpha \neq 0$	$1 - \frac{4}{\alpha} [D_1(-\alpha) - 1]$	$1 - \frac{12}{\alpha} [D_2(-\alpha) - D_1(-\alpha)]$

Un punto a destacar de la tabla anterior es que hay una relación uno-a-uno entre el parámetro de dependencia y las correspondientes medidas de dependencia. Además, podemos ver que las familias Clayton y Gumbel sólo permiten modelar dependencia positiva, mientras que las familias Normal y Frank permiten modelar dependencia tanto positiva como negativa.

En la Tabla 4.3 se dan los valores de los parámetros de dependencia de algunas familias de cópulas que corresponden a valores dados de la tau de Kendall y de la rho de Spearman de 0 a 1 en incrementos de 0.1. Algunos valores se obtienen por integración numérica simple o doble, o por medio de simulación por el método de Monte Carlo. De los valores de estas tablas se sugiere que la rho de Spearman es mayor que la tau de Kendall para estas familias.

Tabla 4.3: Valores de los parámetros de dependencia de algunas familias de cópulas.

τ	Normal	Frank	Clayton	Joe	Gumbel
0.0	0	0	0	1	1
0.1	0.156	0.91	0.22	1.19	1.11
0.2	0.309	1.86	0.50	1.44	1.25
0.3	0.454	2.92	0.86	1.77	1.43
0.4	0.588	4.16	1.33	2.21	1.67
0.5	0.707	5.74	2.00	2.86	2.00
0.6	0.809	7.93	3.00	3.83	2.50
0.7	0.891	11.4	4.67	5.46	3.33
0.8	0.951	18.2	8.00	8.77	5.00
0.9	0.988	20.9	18.0	14.4	10.0
1.0	1	∞	∞	∞	∞

ρ	Normal	Frank	Clayton	Joe	Gumbel
0.0	0	0	0	1	1
0.1	0.105	0.60	0.14	1.12	1.07
0.2	0.209	1.22	0.31	1.27	1.16
0.3	0.313	1.88	0.51	1.46	1.26
0.4	0.416	2.61	0.76	1.69	1.38
0.5	0.518	3.45	1.06	1.99	1.54
0.6	0.618	4.47	1.51	2.39	1.75
0.7	0.717	5.82	2.14	3.00	2.07
0.8	0.813	7.90	3.19	4.03	2.58
0.9	0.908	12.2	5.56	6.37	3.73
1.0	1	∞	∞	∞	∞

4.2. Modelo de Transición de Primer Orden

Para series de tiempo con variables respuesta normales, un modelo estándar es el autorregresivo (AR). Para series de tiempo con variables respuesta no-normales, un modelo estándar es el de transición cuya estructura es la de una cadena de Markov. Un modelo de transición de primer orden con cualesquiera distribuciones marginales univariadas dadas se puede construir por medio de una cópula bivariada. De hecho, esta es una de las aplicaciones más importantes de las cópulas. Esto se puede ver como una generalización del modelo AR(1) que admite cualquier marginal posible, ya que AR(1) aparecería como un caso especial de una cópula normal bivariada que permite sólo marginales normales.

Más aún, para un análisis de regresión de datos multivariados, la media marginal es de primordial interés y siempre se espera que sea modelada como una función explícita de las variables explicativas. Esto es lo que motiva el desarrollo de distribuciones multivariadas que puedan disponer de marginales dadas. Para ello la cópula es el método idóneo ya que permite la construcción de distribuciones multivariadas a través de marginales de forma cerrada, sobre todo para distribuciones discretas.

Para construir un modelo de transición de primer orden para respuestas discretas tales como Y_1 y Y_2 , primero se aplica la derivada de Radon-Nikodym a la ecuación (4.1) para obtener la función de densidad conjunta de (Y_1, Y_2) expresada como:

$$f(y_1, y_2) = \sum_{j=1}^2 \sum_{k=1}^2 (-1)^{j+k} C(v_{1j}, v_{2k}), \quad (4.12)$$

donde $v_{i1} = F_i(y_i)$ y $v_{i2} = F_i(y_i - 1)$. Al desarrollar las sumatorias de la ecuación anterior, tenemos que

$$\begin{aligned}
f(y_1, y_2) &= C(v_{11}, v_{21}) - C(v_{11}, v_{22}) - C(v_{12}, v_{21}) + C(v_{12}, v_{22}) \\
&= C[F_1(y_1), F_2(y_2)] - C[F_1(y_1), F_2(y_2 - 1)] - \\
&\quad C[F_1(y_1 - 1), F_2(y_2)] + C[F_1(y_1 - 1), F_2(y_2 - 1)]. \quad (4.13)
\end{aligned}$$

Como consecuencia, se tiene que la función de densidad condicional de Y_2 dada Y_1 puede convenientemente expresarse de la siguiente forma:

$$\begin{aligned}
f(y_2|y_1) &= \frac{f(y_1, y_2)}{f_1(y_1)} \\
&= \{C[F_1(y_1), F_2(y_2)] - C[F_1(y_1), F_2(y_2 - 1)] - \\
&\quad C[F_1(y_1 - 1), F_2(y_2)] + C[F_1(y_1 - 1), F_2(y_2 - 1)]\} / f_1(y_1). \quad (4.14)
\end{aligned}$$

En caso de que ambas respuestas Y_i ($i = 1, 2$) sean variables aleatorias binarias, sus funciones de densidad marginales son

$$f_i(y_i) = p_i^{y_i}(1 - p_i)^{1-y_i} = \begin{cases} 1 - p_i & : y_i = 0 \\ p_i & : y_i = 1 \end{cases}$$

mientras que sus funciones de distribución marginales son

$$F_i(y_i) = \begin{cases} 0 & : y_i < 0 \\ 1 - p_i & : 0 \leq y_i < 1 \\ 1 & : y_i \geq 1 \end{cases}$$

Entonces, de la ecuación (4.13) y tomando en cuenta las propiedades de una cópula bivariada, las funciones de densidad conjunta de (Y_1, Y_2) respecto a las cuatro posibles combinaciones de valores que pueden tomar ambas variables son

$$\begin{aligned}
f(0, 0) &= C(1 - p_1, 1 - p_2) - C(1 - p_1, 0) - C(0, 1 - p_2) + C(0, 0) \\
&= C(1 - p_1, 1 - p_2),
\end{aligned}$$

$$\begin{aligned} f(0, 1) &= C(1 - p_1, 1) - C(1 - p_1, 1 - p_2) - C(0, 1) + C(0, 1 - p_2) \\ &= 1 - p_1 - C(1 - p_1, 1 - p_2), \end{aligned}$$

$$\begin{aligned} f(1, 0) &= C(1, 1 - p_2) - C(1, 0) - C(1 - p_1, 1 - p_2) + C(1 - p_1, 0) \\ &= 1 - p_2 - C(1 - p_1, 1 - p_2), \end{aligned}$$

$$\begin{aligned} f(1, 1) &= C(1, 1) - C(1, 1 - p_2) - C(1 - p_1, 1) + C(1 - p_1, 1 - p_2) \\ &= 1 - (1 - p_2) - (1 - p_1) + C(1 - p_1, 1 - p_2) \\ &= p_1 + p_2 - 1 + C(1 - p_1, 1 - p_2). \end{aligned}$$

Para hacer uso de este modelo en un análisis de regresión, una manera conveniente de incluir variables explicativas \mathbf{x}_i en el modelo marginal es por medio de la función de enlace probit, lo cual implica que $\text{probit}(\mu_i) = \Phi^{-1}(p_i) = \boldsymbol{\beta}^T \mathbf{x}_i$ o $p_i = \Phi(\boldsymbol{\beta}^T \mathbf{x}_i)$. Usando la propiedad de simetría de la función de distribución normal estándar se tiene que $1 - p_i = \Phi(-\boldsymbol{\beta}^T \mathbf{x}_i)$. Por otro lado, si se elige a la cópula Gaussiana con parámetro de dependencia ρ como cópula del modelo, entonces obtenemos

$$\begin{aligned} C_\rho(1 - p_1, 1 - p_2) &= \Phi_2[\Phi^{-1}(1 - p_1), \Phi^{-1}(1 - p_2)] \\ &= \Phi_2\{\Phi^{-1}[\Phi(-\boldsymbol{\beta}^T \mathbf{x}_1)], \Phi^{-1}[\Phi(-\boldsymbol{\beta}^T \mathbf{x}_2)]\} \\ &= \Phi_2(-\boldsymbol{\beta}^T \mathbf{x}_1, -\boldsymbol{\beta}^T \mathbf{x}_2). \end{aligned}$$

En general, la función de densidad conjunta para dos respuestas (Y_{t-1}, Y_t) queda como

$$f(y_{t-1}, y_t) = \begin{cases} \Phi_2(-\boldsymbol{\beta}^T \mathbf{x}_{t-1}, -\boldsymbol{\beta}^T \mathbf{x}_t) & : y_{t-1} = 0, y_t = 0 \\ \Phi(-\boldsymbol{\beta}^T \mathbf{x}_{t-1}) - \Phi_2(-\boldsymbol{\beta}^T \mathbf{x}_{t-1}, -\boldsymbol{\beta}^T \mathbf{x}_t) & : y_{t-1} = 0, y_t = 1 \\ \Phi(-\boldsymbol{\beta}^T \mathbf{x}_t) - \Phi_2(-\boldsymbol{\beta}^T \mathbf{x}_{t-1}, -\boldsymbol{\beta}^T \mathbf{x}_t) & : y_{t-1} = 1, y_t = 0 \\ \Phi(\boldsymbol{\beta}^T \mathbf{x}_{t-1}) + \Phi(\boldsymbol{\beta}^T \mathbf{x}_t) - 1 + \Phi_2(-\boldsymbol{\beta}^T \mathbf{x}_{t-1}, -\boldsymbol{\beta}^T \mathbf{x}_t) & : y_{t-1} = 1, y_t = 1 \end{cases}$$

Como consecuencia, se tiene que las probabilidades de transición de y_{t-1} a y_t están dadas por sus correspondientes funciones de densidad condicional de Y_t dada Y_{t-1} de la siguiente forma:

$$p_{y_{t-1} y_t}^{(t)} = f(y_t|y_{t-1}) = \begin{cases} f(y_{t-1}, y_t)/\Phi(-\boldsymbol{\beta}^T \mathbf{x}_{t-1}) & : y_{t-1} = 0, y_t = 0 \\ f(y_{t-1}, y_t)/\Phi(-\boldsymbol{\beta}^T \mathbf{x}_{t-1}) & : y_{t-1} = 0, y_t = 1 \\ f(y_{t-1}, y_t)/\Phi(\boldsymbol{\beta}^T \mathbf{x}_{t-1}) & : y_{t-1} = 1, y_t = 0 \\ f(y_{t-1}, y_t)/\Phi(\boldsymbol{\beta}^T \mathbf{x}_{t-1}) & : y_{t-1} = 1, y_t = 1 \end{cases}$$

Entonces, matriz de transición que está representada por una cadena de Markov binaria entre los tiempos $t - 1$ y t con $t > 1$ es de la forma

$$\mathbf{P}^{(t)} = \begin{bmatrix} p_{00}^{(t)} & p_{01}^{(t)} \\ p_{10}^{(t)} & p_{11}^{(t)} \end{bmatrix},$$

mientras que el vector de probabilidades marginales queda como

$$\mathbf{p}^{(t)} = \begin{bmatrix} p_0^{(t)} & p_1^{(t)} \end{bmatrix}, \quad \text{donde} \quad p_{y_t}^{(t)} = \begin{cases} \Phi(-\boldsymbol{\beta}^T \mathbf{x}_t) & : y_t = 0 \\ \Phi(\boldsymbol{\beta}^T \mathbf{x}_t) & : y_t = 1 \end{cases}$$

Para estimar tanto los coeficientes de regresión $\boldsymbol{\beta}$ como el parámetro de dependencia ρ por máxima verosimilitud, primero deberá establecerse cuál es la contribución de cada sujeto a la función de verosimilitud. Si se retoma como ejemplo ilustrativo y objetivo de este trabajo a la base de datos de exposición al sol de los niños, vemos que hay respuestas faltantes que se tienen que tomar en cuenta para determinar la contribución de estos sujetos a la función de verosimilitud. Dicha base de datos consiste de n sujetos con la posibilidad de tres respuestas cada uno igualmente espaciadas en el tiempo y_{i1}, y_{i2}, y_{i3} para el i -ésimo sujeto, además de las variables explicativas correspondientes.

Sea $L = \prod_{i=1}^n L_i$ la función de verosimilitud, L_i es la contribución a la misma del i -ésimo sujeto que dependiendo de sus respuestas (“na” indica una respuesta faltante) toma las formas siguientes:

Respuestas	Contribución L_i
$y_{i1} \ y_{i2} \ y_{i3}$	$f(y_{i1})f(y_{i2} y_{i1})f(y_{i3} y_{i2}) = f(y_{i1}, y_{i2})f(y_{i2}, y_{i3})/f(y_{i2})$
$y_{i1} \ y_{i2} \ \text{na}$	$f(y_{i1})f(y_{i2} y_{i1}) = f(y_{i1}, y_{i2})$
$\text{na} \ y_{i2} \ y_{i3}$	$f(y_{i2})f(y_{i3} y_{i2}) = f(y_{i2}, y_{i3})$
$y_{i1} \ \text{na} \ y_{i3}$	$f(y_{i1})f(y_{i3} y_{i1})$
$y_{i1} \ \text{na} \ \text{na}$	$f(y_{i1})$
$\text{na} \ y_{i2} \ \text{na}$	$f(y_{i2})$
$\text{na} \ \text{na} \ y_{i3}$	$f(y_{i3})$

Aunque no se indicó explícitamente, en cada función de densidad de la tabla anterior se incluyen las variables explicativas presentes y pasadas. En el caso especial de la secuencia de respuestas y_{i1} , na, y_{i3} , las funciones de densidad condicional entre respuestas no adyacentes $f(y_{i3}|y_{i1})$ son las probabilidades de transición $p_{y_{i1}y_{i3}}$ que a su vez son los elementos de la matriz de transición correspondiente $\mathbf{P}_{13} \equiv \mathbf{P}^{(2)}\mathbf{P}^{(3)}$. En cuanto a la estimación del parámetro de dependencia ρ , por conveniencia se estimó el parámetro $\text{arctanh}(\rho)$ que toma valores en \mathbb{R} para asegurar que $-1 < \rho < 1$.

Dado que este modelo de transición basado en cópulas es no lineal, es necesario usar métodos numéricos para obtener los estimadores de máxima verosimilitud MLE's. Para el cálculo de la cópula Gaussiana por medio de la función de distribución normal bivariada se utilizó la función `pnorm` implementada en la biblioteca

`normt` del paquete estadístico R. Además, se usó la función `nlm` del mismo paquete para minimizar la función $-\log L = -\sum_{i=1}^n \log L_i$.

Para identificar las variables explicativas apropiadas a ser incluidas en el modelo, se considera el *criterio de información bayesiano* (BIC) como el criterio principal para la elección del modelo y se compara con el *criterio de información de Akaike* (AIC). Ambos criterios se definen como

$$\text{AIC} = -2 \log L + 2 n_p,$$

$$\text{BIC} = -2 \log L + n_p \log(n_t),$$

donde n_p denota el número de parámetros en el modelo y n_t el número total de observaciones no faltantes. El modelo elegido es aquel para el cual alguno de ambos criterios es el menor.

4.3. Diagnósticos

Las suposiciones hechas acerca de un modelo de transición en particular pueden ser revisadas por medio del cálculo de los residuales de cuantiles aleatorizados propuestos por Dunn y Smyth (1996). Bajo el modelo asumido, tales residuales se distribuyen normalmente y se encuentran invirtiendo la función de distribución condicional ajustada para cada valor de la variable respuesta. Después, sólo algunas gráficas simples son necesarias para revisar que dichos residuales son valores observados de variables aleatorias independientes y con una distribución normal estándar, además de indicar la calidad del ajuste.

Si el modelo de transición es para respuestas discretas, la función de distribución condicional de Y_t dada Y_{t-1} está dada por

$$F(y_t|y_{t-1}) = \{C[F(y_{t-1}), F(y_t)] - C[F(y_{t-1} - 1), F(y_t)]\} / f(y_{t-1}). \quad (4.15)$$

En caso de que las respuestas Y_{t-1} y Y_t sean binarias, las funciones de distribución condicional respecto a las cuatro posibles combinaciones de valores son

$$\begin{aligned} F(0|0) &= [C(1 - p_{t-1}, 1 - p_t) - C(0, 1 - p_t)] / (1 - p_{t-1}) \\ &= C(1 - p_{t-1}, 1 - p_t) / (1 - p_{t-1}), \end{aligned}$$

$$\begin{aligned} F(0|1) &= [C(1, 1 - p_t) - C(1 - p_{t-1}, 1 - p_t)] / p_{t-1} \\ &= [1 - p_t - C(1 - p_{t-1}, 1 - p_t)] / p_{t-1}, \end{aligned}$$

$$\begin{aligned} F(1|0) &= [C(1 - p_{t-1}, 1) - C(0, 1)] / (1 - p_{t-1}) \\ &= (1 - p_{t-1}) / (1 - p_{t-1}) = 1, \end{aligned}$$

$$\begin{aligned} F(1|1) &= [C(1, 1) - C(1 - p_{t-1}, 1)] / p_{t-1} \\ &= p_{t-1} / p_{t-1} = 1. \end{aligned}$$

Usando la función de enlace probit en el modelo marginal y la cópula Gaussiana

$$F(y_t|y_{t-1}) = \begin{cases} \Phi_2(-\boldsymbol{\beta}^T \mathbf{x}_{t-1}, -\boldsymbol{\beta}^T \mathbf{x}_t) / \Phi(-\boldsymbol{\beta}^T \mathbf{x}_{t-1}) & : y_{t-1} = 0, y_t = 0 \\ [\Phi(-\boldsymbol{\beta}^T \mathbf{x}_t) - \Phi_2(-\boldsymbol{\beta}^T \mathbf{x}_{t-1}, -\boldsymbol{\beta}^T \mathbf{x}_t)] / \Phi(\boldsymbol{\beta}^T \mathbf{x}_{t-1}) & : y_{t-1} = 1, y_t = 0 \\ 1 & : y_{t-1} = 0, y_t = 1 \\ 1 & : y_{t-1} = 1, y_t = 1 \end{cases}$$

Si el modelo de transición es para respuestas discretas, los residuales de cuantiles aleatorizados están definidos como $r_t = \Phi^{-1}(u_t)$, donde u_t es un valor aleatorio de una distribución uniforme en el intervalo $(a_t, b_t]$ con $a_t = F(y_t - 1|y_{t-1})$ y $b_t = F(y_t|y_{t-1})$. Para la primera observación de cada sujeto $a_1 = F(y_1 - 1)$ y $b_1 = F(y_1)$, y en consecuencia

$$y_1 = 0 \implies a_1 = 0 \quad y \quad b_1 = 1 - p_1 = \Phi(-\boldsymbol{\beta}^T \mathbf{x}_1),$$

$$y_1 = 1 \implies a_1 = 1 - p_1 = \Phi(-\boldsymbol{\beta}^T \mathbf{x}_1) \quad y \quad b_1 = 1.$$

Para las siguientes observaciones de cada sujeto

$$y_{t-1} = 0, y_t = 0 \implies a_t = 0 \quad y \quad b_t = \Phi_2(-\boldsymbol{\beta}^T \mathbf{x}_{t-1}, -\boldsymbol{\beta}^T \mathbf{x}_t) / \Phi(-\boldsymbol{\beta}^T \mathbf{x}_{t-1}),$$

$$y_{t-1} = 0, y_t = 1 \implies a_t = \Phi_2(-\boldsymbol{\beta}^T \mathbf{x}_{t-1}, -\boldsymbol{\beta}^T \mathbf{x}_t) / \Phi(-\boldsymbol{\beta}^T \mathbf{x}_{t-1}) \quad y \quad b_t = 1,$$

$$y_{t-1} = 1, y_t = 0 \implies a_t = 0 \quad y$$

$$b_t = [\Phi(-\boldsymbol{\beta}^T \mathbf{x}_t) - \Phi_2(-\boldsymbol{\beta}^T \mathbf{x}_{t-1}, -\boldsymbol{\beta}^T \mathbf{x}_t)] / \Phi(\boldsymbol{\beta}^T \mathbf{x}_{t-1}),$$

$$y_{t-1} = 1, y_t = 1 \implies a_t = [\Phi(-\boldsymbol{\beta}^T \mathbf{x}_t) - \Phi_2(-\boldsymbol{\beta}^T \mathbf{x}_{t-1}, -\boldsymbol{\beta}^T \mathbf{x}_t)] / \Phi(\boldsymbol{\beta}^T \mathbf{x}_{t-1}) \quad y$$

$$b_t = 1.$$

4.4. Aplicación al Ejemplo Ilustrativo

Continuando con el análisis de la base de datos de exposición al sol de los niños, en primer lugar se ajustarán diversos modelos con cópulas de estructura independiente ($\rho = 0$).

Modelo: tiempo * Tratamiento

Parámetro	Estimador	Error Est.	Valor P	— I. C. 95% —
Intercepto	1.42289	0.13062	0.00000	1.16687 1.67891
tiempo	-0.36216	0.05876	0.00000	-0.47732 -0.24699
Tratamiento	0.12868	0.19176	0.50219	-0.24717 0.50454
tiempo:Trat	0.02947	0.08549	0.73033	-0.13810 0.19704

Modelo: tiempo + Tratamiento

Parámetro	Estimador	Error Est.	Valor P	— I. C. 95% —
Intercepto	1.39409	0.10005	0.00000	1.19799 1.59018
tiempo	-0.34827	0.04269	0.00000	-0.43194 -0.26460
Tratamiento	0.19045	0.06839	0.00536	0.05641 0.32450

Modelo: tiempo

Parámetro	Estimador	Error Est.	Valor P	— I. C. 95% —
Intercepto	1.47965	0.09542	0	1.29263 1.66668
tiempo	-0.34577	0.04258	0	-0.42923 -0.26232

Modelo: Tratamiento

Parámetro	Estimador	Error Est.	Valor P	— I. C. 95% —
Intercepto	0.68697	0.04592	0.00000	0.59696 0.77699
Tratamiento	0.17887	0.06715	0.00773	0.04726 0.31048

Modelo: *nulo*

Parámetro	Estimador	Error Est.	Valor P	— I. C. 95% —
Intercepto	0.77219	0.03344	0	0.70664 0.83774

Tomando los niveles de significancia del 1 %, 5 % y 10 % como referencias estándares, y según los reportes anteriores de ajuste de los modelos, ambas variables explicativas **tiempo** y **Tratamiento** son estadísticamente significativas aún al 1 %, mientras que la interacción entre ellas no lo es ni siquiera al 10 %. Ahora se se ajustarán los modelos con cópulas de estructura dependiente AR(1).

Modelo: tiempo * Tratamiento

Parámetro	Estimador	Error Est.	Valor P	—	I. C. 95 %	—
Intercepto	1.39596	0.12779	0.00000	1.14549	1.64643	
tiempo	-0.34386	0.05507	0.00000	-0.45181	-0.23592	
Tratamiento	0.14142	0.18761	0.45097	-0.22629	0.50913	
tiempo:Trat	0.02034	0.08000	0.79935	-0.13647	0.17714	
arctanh(ρ)	0.64221	0.06573	0.00000	0.51338	0.77103	
ρ	0.56640	—	—	0.47257	0.64753	

Modelo: tiempo + Tratamiento

Parámetro	Estimador	Error Est.	Valor P	—	I. C. 95 %	—
Intercepto	1.37604	0.10066	0.00000	1.17876	1.57333	
tiempo	-0.33428	0.04007	0.00000	-0.41281	-0.25574	
Tratamiento	0.18435	0.08183	0.02426	0.02397	0.34472	
arctanh(ρ)	0.64220	0.06571	0.00000	0.51341	0.77100	
ρ	0.56640	—	—	0.47259	0.64751	

Modelo: tiempo

Parámetro	Estimador	Error Est.	Valor P	—	I. C. 95 %	—
Intercepto	1.45861	0.09371	0	1.27493	1.64228	
tiempo	-0.33148	0.03990	0	-0.40968	-0.25328	
arctanh(ρ)	0.64896	0.06566	0	0.52026	0.77766	
ρ	0.57097	—	—	0.47790	0.65136	

Modelo: Tratamiento

Parámetro	Estimador	Error Est.	Valor P	—	I. C. 95 %	—
Intercepto	0.70104	0.05525	0.00000	0.59275	0.80932	
Tratamiento	0.16853	0.08010	0.03538	0.01153	0.32553	
arctanh(ρ)	0.60937	0.06318	0.00000	0.48553	0.73320	
ρ	0.54368	—	—	0.45066	0.62502	

Modelo: *nulo*

Parámetro	Estimador	Error Est.	Valor P	—	I. C. 95 %	—
Intercepto	0.78183	0.04009	0	0.70326	0.86041	
$\operatorname{arctanh}(\rho)$	0.61574	0.06311	0	0.49204	0.73944	
ρ	0.54816	—	—	0.45584	0.62881	

Según los reportes anteriores de ajuste de los modelos, la variable explicativa **tiempo** sigue siendo estadísticamente significativa aún al 1%, mientras que **Tratamiento** ahora sólo lo es al 5% y la interacción entre ellas no lo es ni siquiera al 10%. Comparando estos resultados con los resultados proporcionados por el ajuste de los modelos con cópulas de estructura independiente, podemos ver que el hecho de no haber tomado en cuenta la dependencia entre respuestas provocó que se sobreestimara la significancia estadística de la variables explicativa **Tratamiento**.

Con respecto a los valores estimados del parámetro de dependencia ρ , es interesante observar como permanecen en el mismo rango sin ser afectados por el conjunto de variables explicativas incluidas en el modelo. Para probar la significancia de ρ es posible realizar una prueba de hipótesis con la hipótesis nula $H_0 : \rho = 0$, que corresponde al modelo independiente, contra la hipótesis alternativa $H_1 : \rho \neq 0$, que corresponde al modelo AR(1); por las propiedades de las muestras grandes como en este caso, la hipótesis nula es rechazada dado que ninguno de los intervalos de confianza de ρ incluyen al cero. Por lo tanto, la dependencia entre respuestas adyacentes es significativa y el parámetro ρ debe ser tomado en cuenta.

En la Tabla 4.4 se muestran los valores de las deviancias, AIC's y BIC's reportados durante el ajuste de los modelos anteriores con ambas estructuras de dependencia.

Tabla 4.4: Deviancias, AIC's y BIC's de diferentes modelos.

Modelo	Independiente			AR(1)		
	Deviancia	AIC	BIC	Deviancia	AIC	BIC
tiempo * Trat	1768.57	1776.57	1798.44	1660.45	1670.45	1697.79
tiempo + Trat	1768.69	1774.69	1791.09	1660.52	1668.52	1690.39
tiempo	1776.47	1780.47	1791.40	1665.59	1671.59	1688.00
Tratamiento	1837.06	1841.06	1852.00	1732.30	1738.30	1754.70
nulo	1844.18	1846.18	1851.65	1736.72	1740.72	1751.66

De la tabla anterior se puede observar que de acuerdo al criterio AIC el modelo de mejor ajuste para ambas estructuras de dependencia es el que incluye tanto al tiempo como al Tratamiento sin su interacción. Lo mismo sucede si se usa el criterio BIC con la estructura independiente. Sin embargo, con la estructura de dependencia AR(1) el criterio BIC elige como el modelo de mejor ajuste aquel que sólo incluye al tiempo como variable explicativa con efectos realmente significativos.

Si se elige al modelo tiempo + Trat con estructura de dependencia AR(1), entonces las probabilidades marginales y de transición estimadas y su intervalo de confianza del 95% para el grupo de control están dadas por los siguientes vectores marginales y matrices de transición.

$$\hat{\mathbf{p}}^{(1)} = \begin{bmatrix} 0.149 & 0.851 \\ (0.094, 0.222) & (0.778, 0.906) \end{bmatrix}, \quad \hat{\mathbf{p}}^{(2)} = \begin{bmatrix} 0.240 & 0.760 \\ (0.144, 0.362) & (0.638, 0.856) \end{bmatrix},$$

$$\hat{\mathbf{p}}^{(3)} = \begin{bmatrix} 0.354 & 0.646 \\ (0.210, 0.524) & (0.476, 0.790) \end{bmatrix};$$

$$\hat{\mathbf{P}}^{(2)} = \begin{bmatrix} 0.579 & 0.421 \\ (0.475, 0.679) & (0.321, 0.525) \\ 0.180 & 0.820 \\ (0.110, 0.272) & (0.728, 0.890) \end{bmatrix}, \quad \hat{\mathbf{P}}^{(3)} = \begin{bmatrix} 0.659 & 0.341 \\ (0.538, 0.768) & (0.232, 0.462) \\ 0.258 & 0.742 \\ (0.155, 0.385) & (0.615, 0.845) \end{bmatrix},$$

$$\hat{\mathbf{P}}_{13} = \begin{bmatrix} 0.490 & 0.510 \\ (0.337, 0.645) & (0.355, 0.663) \\ 0.331 & 0.669 \\ (0.197, 0.489) & (0.511, 0.803) \end{bmatrix}.$$

Mientras que para el grupo intervenido, las probabilidades marginales y de transición estimadas y su intervalo de confianza del 95% están dadas por los siguientes vectores marginales y matrices de transición

$$\hat{\mathbf{p}}^{(1)} = \begin{bmatrix} 0.110 & 0.890 \\ (0.048, 0.215) & (0.785, 0.952) \end{bmatrix}, \quad \hat{\mathbf{p}}^{(2)} = \begin{bmatrix} 0.186 & 0.814 \\ (0.080, 0.353) & (0.647, 0.920) \end{bmatrix},$$

$$\hat{\mathbf{p}}^{(3)} = \begin{bmatrix} 0.289 & 0.711 \\ (0.125, 0.514) & (0.486, 0.875) \end{bmatrix};$$

$$\hat{\mathbf{P}}^{(2)} = \begin{bmatrix} 0.534 & 0.466 \\ (0.392, 0.673) & (0.327, 0.608) \\ 0.143 & 0.857 \\ (0.064, 0.265) & (0.735, 0.936) \end{bmatrix}, \quad \hat{\mathbf{P}}^{(3)} = \begin{bmatrix} 0.615 & 0.385 \\ (0.453, 0.763) & (0.237, 0.547) \\ 0.214 & 0.786 \\ (0.096, 0.378) & (0.622, 0.904) \end{bmatrix},$$

$$\hat{\mathbf{P}}_{13} = \begin{bmatrix} 0.428 & 0.572 \\ (0.236, 0.637) & (0.363, 0.764) \\ 0.271 & 0.729 \\ (0.119, 0.481) & (0.519, 0.881) \end{bmatrix}.$$

Al comparar las probabilidades anteriores entre el grupo de control y el grupo intervenido, se ve que hay muy poca diferencia entre ellas. Además, el hecho de que la variable **Tratamiento** sea significativa al 5 % pero no al 1 %, hace pensar que para efectos prácticos el tratamiento pudo no haber tenido realmente efectos significativos. En consecuencia, y tomando en cuenta el criterio BIC, con cierta seguridad se puede elegir como modelo de mejor ajuste aquel que sólo incluye a la variable **tiempo** con estructura de dependencia AR(1). Es decir, al menos si hay evidencia suficiente de que la proporción de respuestas favorables durante el experimento va decreciendo con el tiempo. Entonces, bajo esta especificación, las probabilidades marginales y de transición estimadas y su intervalo de confianza del 95 % están dadas por los siguientes vectores marginales y matrices de transición

$$\hat{\mathbf{p}}^{(1)} = \begin{bmatrix} 0.130 & 0.870 \\ (0.082, 0.193) & (0.807, 0.918) \end{bmatrix}, \quad \hat{\mathbf{p}}^{(2)} = \begin{bmatrix} 0.213 & 0.787 \\ (0.128, 0.324) & (0.676, 0.872) \end{bmatrix},$$

$$\hat{\mathbf{p}}^{(3)} = \begin{bmatrix} 0.321 & 0.679 \\ (0.189, 0.482) & (0.518, 0.811) \end{bmatrix};$$

$$\hat{\mathbf{P}}^{(2)} = \begin{bmatrix} 0.560 & 0.440 \\ (0.460, 0.658) & (0.342, 0.540) \\ 0.161 & 0.839 \\ (0.098, 0.244) & (0.756, 0.902) \end{bmatrix}, \quad \hat{\mathbf{P}}^{(3)} = \begin{bmatrix} 0.640 & 0.360 \\ (0.522, 0.748) & (0.252, 0.478) \\ 0.235 & 0.765 \\ (0.140, 0.354) & (0.646, 0.860) \end{bmatrix},$$

$$\hat{\mathbf{P}}_{13} = \begin{bmatrix} 0.462 & 0.538 \\ (0.315, 0.613) & (0.387, 0.685) \\ 0.300 & 0.700 \\ (0.177, 0.450) & (0.550, 0.823) \end{bmatrix}.$$

Recordando que una respuesta favorable de los padres es aquella en la que sí evitaron la exposición al sol del niño, mientras que una respuesta desfavorable es aquella en la que no lo evitaron; de las probabilidades de transición anteriores se puede concluir que: (1) los padres con respuestas desfavorables durante cierto año, casi se dividieron

por igual con respuestas desfavorables y favorables al año siguiente, por lo que su contribución a la dependencia entre respuestas es relativamente pequeña, aunque aún en sentido positivo; y (2) en promedio el 80 % de los padres con respuestas favorables durante cierto año, mantuvieron sus respuestas en el mismo sentido al año siguiente, por lo que la dependencia positiva entre respuestas se debe a su contribución en buena medida.

Por otro lado, cuando la suposición de que un modelo de transición sigue una estructura de dependencia específica es cuestionable, es posible realizar un análisis gráfico para determinar que tal estructura de dependencia está siendo capturada por los datos. Para ello se grafica el *perfil* de la función log-verosimilitud con respecto al parámetro de dependencia ρ . Cuando se centra la atención a la forma local de dicho perfil en la vecindad de $\rho = \hat{\rho}$, y si la curva es muy plana, entonces el punto máximo podría depender sensiblemente de cualquier marginal, lo cual indica que la estructura de dependencia describe a los datos pobremente, Copas (1997).

En la Figura 4.3 se presenta el perfil de la función log-verosimilitud con respecto al parámetro de dependencia ρ para el modelo de transición con variable explicativa **tiempo** y cópula de estructura dependiente AR(1). Se puede observar que la curva es “bien comportada” en el sentido de que su forma no es nada plana para valores cercanos al máximo, por el contrario, su forma se asemeja más a una cuadrática. Por lo tanto, se puede concluir que la estructura elegida AR(1) captura bien la dependencia.

Respecto a los diagnósticos de las suposiciones hechas acerca del modelo de transición con la variable explicativa **tiempo** y cópula de estructura dependiente AR(1), en la Figura 4.4 se presentan las gráficas de los residuales de cuantiles aleatorizados en cuanto a su distribución. Las formas de estas gráficas sugieren que el modelo propuesto ajusta los datos razonablemente bien.

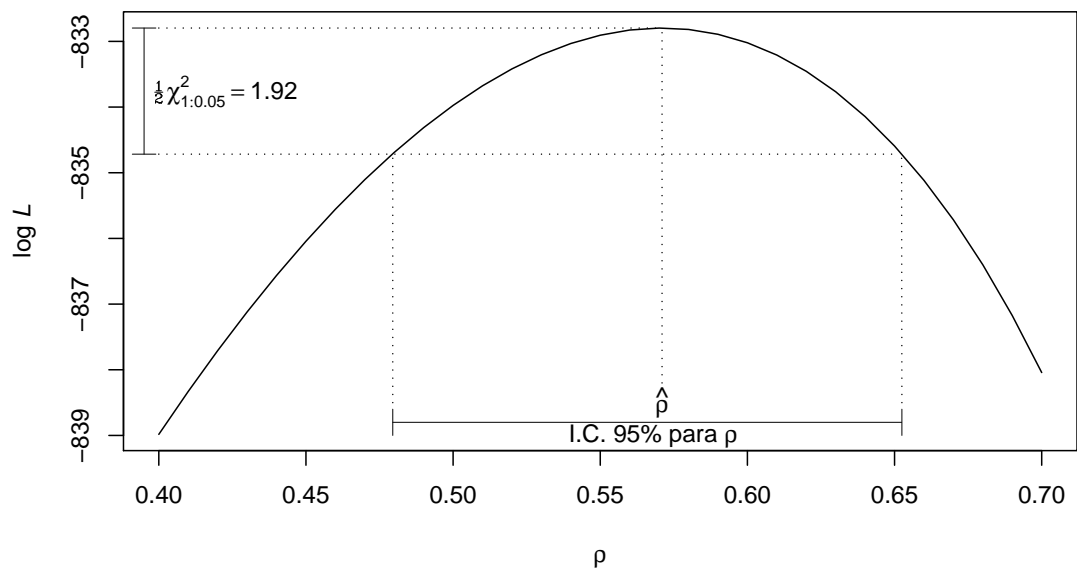


Figura 4.3: Perfil de la función log-verosimilitud con respecto al parámetro ρ .

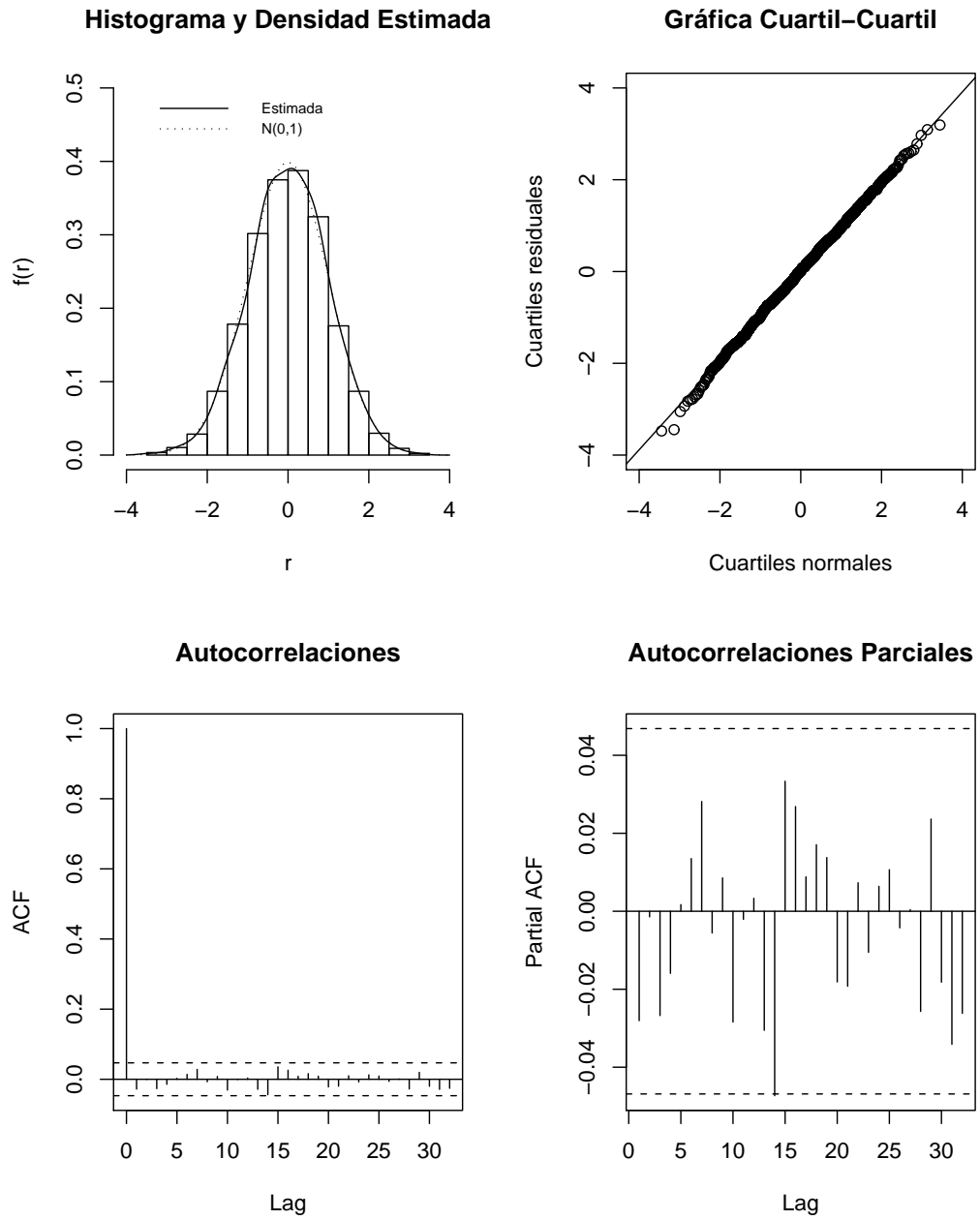


Figura 4.4: Gráficas de los residuales de cuantiles aleatorizados.

Capítulo 5

Consideraciones Finales

Como pudimos constatar en el capítulo anterior, la gran ventaja del uso de las cópulas en el modelado multivariado es que es un método muy general y flexible que nos permite obtener simultáneamente durante el mismo proceso de ajuste tanto los modelos marginales como los modelos de transición, mientras que el método de GEE propuesto en el capítulo 2 sólo es útil para obtener modelos marginales y el método de GLM propuesto en el capítulo 3 sólo lo es para obtener modelos de transición.

Además, y con respecto a los modelos marginales, la primera ventaja del método de cópulas sobre el método de GEE es que es menos dependiente de suposiciones; por ejemplo, el manejo de datos faltantes con cópulas es menos restrictivo ya que basta con que el proceso que los genera sólo sea aleatorio, mientras que con GEE tiene que ser completamente aleatorio. De hecho, cuando se ajustó el modelo con cópulas se supuso que el proceso que genera los datos faltantes es aleatorio, ya que la razón de que los datos falten en ciertos momentos es independiente del tipo de respuesta en estos momentos. Con esto se valida o justifica el ajuste del modelo con GEE, ya que si se comparan los reportes de ambos ajustes podemos ver que los resultados son muy parecidos como se reproduce a continuación.

Parámetro	GEE			Cópulas		
	Estimador	E. E.	Valor P	Estimador	E. E.	Valor P
Intercepto	1.376	0.093	0.000	1.376	0.101	0.000
tiempo	-0.333	0.036	0.000	-0.334	0.040	0.000
Tratamiento	0.189	0.085	0.026	0.184	0.082	0.024

Una segunda ventaja del método de cópulas sobre el método de GEE es que como el primero está basado en verosimilitud nos permite hacer una serie de inferencias como probar suposiciones, realizar diagnósticos, etc., mientras que con GEE no es posible, es limitado, o poco claro. Por poner un ejemplo, para la elección entre modelos que incluyen diferentes variables explicativas, se cuenta con criterios bien establecidos y relativamente fáciles de calcular que dependen de la deviancia, como lo son el AIC y el BIC, mientras que GEE cuenta con un criterio reciente que es una generalización del AIC basado en cuasi-verosimilitud identificado como QIC, pero que es computacionalmente costoso, a tal grado de que la mayoría del software estadístico comercial no lo tiene implementado internamente.

Sin embargo, hay que reconocer que la principal ventaja de GEE sobre métodos basados en verosimilitud como el que aquí se propone, es que computacionalmente es más simple de implementar, y por lo tanto, más rápido de ejecutar. Aunque algunos estadísticos son críticos con el método GEE por su falta de una función de verosimilitud, otros no ven esto como algo problemático, ya que ellos consideran a GEE más como un método de estimación que como un modelo propiamente dicho.

Por otro lado, y con respecto a los modelos de transición, la ventaja del método de cópulas sobre el método de GLM propuesto en el capítulo 3, es que este último no permite el manejo de datos faltantes.

Diferentes familias de cópulas, además de la Gaussiana, pueden también ser usadas en el modelado multivariado. Sin embargo, puede haber diferencias considerables en

las formas de las distribuciones condicionales generadas por el modelo de cópula discretizado. Por ello es importante elegir correctamente una familia de cópulas y una distribución marginal por cada distribución condicional en el modelo. Cómo hacer esto de la mejor manera permanece aún como una pregunta abierta. Por lo que, como siempre, es muy importante probar la validez del modelo propuesto antes de hacer cualquier inferencia.

Aunque el método presentado aquí está enfocado a datos igualmente espaciados, los modelos con cópulas Gaussianas discretizados se prestan para analizar datos longitudinales cuando cada sujeto es observado en diferentes momentos no igualmente espaciados. Si la suposición de dependencia tipo Markov permanece plausible, la metodología puede ser modificada para permitir un espaciado temporal general. Este aspecto requiere aún investigación adicional.

En principio, es posible extender el modelo de transición basado en cópulas descrito anteriormente a representaciones de mayor orden. Sin embargo, en la práctica esto puede ser bastante laborioso y computacionalmente costoso. Un método más parsimonioso que una cadena de Markov totalmente paramétrica puede ser el modelo de mezclas de distribuciones de transición (MTD) de orden p introducido por Raftery (1985), el cual está caracterizado por la siguiente función de densidad condicional

$$f(y_t|y_{t-1}, \dots, y_{t-p}) = \sum_{k=1}^p \omega_k f_k(y_t|y_{t-k}),$$

donde $\sum_{k=1}^p \omega_k = 1$ y $\omega_k \geq 0$ para $k = 1, \dots, p$. El desarrollo y uso práctico de este modelo de mezclas de Raftery se propone para un proyecto de investigación posterior.

Apéndice A

Programa en R

A continuación se presenta el programa desarrollado en el paquete estadístico R para ajustar los modelos marginales y de transición de primer orden por medio de la cópula bivariada Gaussiana.

```
library(mnormt)

shade <- read.table("shade.txt",header=F)
names(shade) <- c("id","time","treat","response")

mlogL <- function(p,Y,X){

  B <- p[1:dim(X)[2]]
  r <- tanh(p[dim(X)[2]+1])
  R <- matrix(c(1,r,r,1),2,2)

  Y1 <- numeric(); P1 <- Y1; f1 <- Y1; f12 <- Y1
  Y2 <- Y1; P2 <- Y1; f2 <- Y1; f23 <- Y1
  Y3 <- Y1; P3 <- Y1; f3 <- Y1; f31 <- Y1

  Li <- Y1

  M <- is.na(Y)
  j <- 0
```

```

for(h in 1:(length(Y)/3)){
  c <- 0; k <- 3 * h - 2

  if(M[k]==F & M[k+1]==F & M[k+2]==T){

    j <- j + 1

    if(Y[k]==0 & Y[k+1]==0)
      f12[j] <- pmnorm(c(-X[k,]%*%B,-X[k+1,]%*%B),varcov=R)

    else if(Y[k]==0 & Y[k+1]==1)
      f12[j] <- pnorm(-X[k,]%*%B) -
        pmnorm(c(-X[k,]%*%B,-X[k+1,]%*%B),varcov=R)

    else if(Y[k]==1 & Y[k+1]==0)
      f12[j] <- pnorm(-X[k+1,]%*%B) -
        pmnorm(c(-X[k,]%*%B,-X[k+1,]%*%B),varcov=R)

    else if(Y[k]==1 & Y[k+1]==1)
      f12[j] <- pnorm(X[k,]%*%B) + pnorm(X[k+1,]%*%B) - 1 +
        pmnorm(c(-X[k,]%*%B,-X[k+1,]%*%B),varcov=R)

    Li[j] <- f12[j]
  }

  else if(M[k]==F & M[k+1]==T & M[k+2]==F){

    j <- j + 1

    Y1[j] <- Y[k]
    P1[j] <- pnorm(X[k,]%*%B)
    f1[j] <- P1[j]^Y1[j]*(1-P1[j])^(1-Y1[j])

    p1 <- pnorm(X[k,]%*%B); q1 <- 1 - p1
    p2 <- pnorm(X[k+1,]%*%B); q2 <- 1 - p2
    p3 <- pnorm(X[k+2,]%*%B); q3 <- 1 - p3

    C12 <- pmnorm(c(-X[k,]%*%B,-X[k+1,]%*%B),varcov=R)
    C23 <- pmnorm(c(-X[k+1,]%*%B,-X[k+2,]%*%B),varcov=R)

    p00.12 <- C12 / q1
    p10.12 <- (q2 - C12) / p1
    p01.12 <- (q1 - C12) / q1
    p11.12 <- (1 - q1 - q2 + C12) / p1
  }
}

```

```

p00.23 <- C23 / q2
p10.23 <- (q3 - C23) / p2
p01.23 <- (q2 - C23) / q2
p11.23 <- (1 - q2 - q3 + C23) / p2

p00.13 <- p00.12*p00.23 + p01.12*p10.23
p10.13 <- p10.12*p00.23 + p11.12*p10.23
p01.13 <- p00.12*p01.23 + p01.12*p11.23
p11.13 <- p10.12*p01.23 + p11.12*p11.23

if(Y[k]==0 & Y[k+2]==0)
  f31[j] <- p00.13

else if(Y[k]==0 & Y[k+2]==1)
  f31[j] <- p01.13

else if(Y[k]==1 & Y[k+2]==0)
  f31[j] <- p10.13

else if(Y[k]==1 & Y[k+2]==1)
  f31[j] <- p11.13

Li[j] <- f1[j] * f31[j]
}

else if(M[k]==T & M[k+1]==F & M[k+2]==F){

  j <- j + 1

  if(Y[k+1]==0 & Y[k+2]==0)
    f23[j] <- pmnorm(c(-X[k+1, ]%%B, -X[k+2, ]%%B), varcov=R)

  else if(Y[k+1]==0 & Y[k+2]==1)
    f23[j] <- pnorm(-X[k+1, ]%%B) -
      pmnorm(c(-X[k+1, ]%%B, -X[k+2, ]%%B), varcov=R)

  else if(Y[k+1]==1 & Y[k+2]==0)
    f23[j] <- pnorm(-X[k+2, ]%%B) -
      pmnorm(c(-X[k+1, ]%%B, -X[k+2, ]%%B), varcov=R)

  else if(Y[k+1]==1 & Y[k+2]==1)
    f23[j] <- pnorm(X[k+1, ]%%B) + pnorm(X[k+2, ]%%B) - 1 +
      pmnorm(c(-X[k+1, ]%%B, -X[k+2, ]%%B), varcov=R)

  Li[j] <- f23[j]
}

```

```

else if(M[k]==F & M[k+1]==T & M[k+2]==T){

  j <- j + 1

  Y1[j] <- Y[k]
  P1[j] <- pnorm(X[k, ]%%B)
  f1[j] <- P1[j]^Y1[j]*(1-P1[j])^(1-Y1[j])

  Li[j] <- f1[j]
}

else if(M[k]==T & M[k+1]==F & M[k+2]==T){

  j <- j + 1

  Y2[j] <- Y[k+1]
  P2[j] <- pnorm(X[k+1, ]%%B)
  f2[j] <- P2[j]^Y2[j]*(1-P2[j])^(1-Y2[j])

  Li[j] <- f2[j]
}

else if(M[k]==T & M[k+1]==T & M[k+2]==F){

  j <- j + 1

  Y3[j] <- Y[k+2]
  P3[j] <- pnorm(X[k+2, ]%%B)
  f3[j] <- P3[j]^Y3[j]*(1-P3[j])^(1-Y3[j])

  Li[j] <- f3[j]
}

else{

  j <- j + 1

  Y2[j] <- Y[k+1]
  P2[j] <- pnorm(X[k+1, ]%%B)
  f2[j] <- P2[j]^Y2[j]*(1-P2[j])^(1-Y2[j])

  for(i in k:(k+1)){

    c <- c + 1

```

```

if(Y[i]==0 & Y[i+1]==0)
  if(c==1)
    f12[j] <- pmnorm(c(-X[i,]%%B,-X[i+1,]%%B),varcov=R)
  else
    f23[j] <- pmnorm(c(-X[i,]%%B,-X[i+1,]%%B),varcov=R)
else if(Y[i]==0 & Y[i+1]==1)
  if(c==1)
    f12[j] <- pnorm(-X[i,]%%B) -
      pmnorm(c(-X[i,]%%B,-X[i+1,]%%B),varcov=R)
  else
    f23[j] <- pnorm(-X[i,]%%B) -
      pmnorm(c(-X[i,]%%B,-X[i+1,]%%B),varcov=R)
else if(Y[i]==1 & Y[i+1]==0)
  if(c==1)
    f12[j] <- pnorm(-X[i+1,]%%B) -
      pmnorm(c(-X[i,]%%B,-X[i+1,]%%B),varcov=R)
  else
    f23[j] <- pnorm(-X[i+1,]%%B) -
      pmnorm(c(-X[i,]%%B,-X[i+1,]%%B),varcov=R)
else if(Y[i]==1 & Y[i+1]==1)
  if(c==1)
    f12[j] <- pnorm(X[i,]%%B) + pnorm(X[i+1,]%%B) - 1 +
      pmnorm(c(-X[i,]%%B,-X[i+1,]%%B),varcov=R)
  else
    f23[j] <- pnorm(X[i,]%%B) + pnorm(X[i+1,]%%B) - 1 +
      pmnorm(c(-X[i,]%%B,-X[i+1,]%%B),varcov=R)
}
Li[j] <- f12[j] * f23[j] / f2[j]
}
}
- sum(log(Li))
}

md <- model.matrix(~ time, data=shade)
np <- dim(md)[2]+1
mle <- nlm(f=mlogL, p=rep(0,np), Y=shade$response, X=md, hessian=T)

Est <- mle$estimate;      E.E <- sqrt(diag(solve(mle$hessian)))
V.z <- Est/E.E;          V.P <- 2*(1-pnorm(abs(V.z)))
L.B <- Est-1.96*E.E;     U.B <- Est+1.96*E.E

```

```
rho <- tanh(Est[dim(md)[2]+1])
LBr <- tanh(L.B[dim(md)[2]+1])
UBr <- tanh(U.B[dim(md)[2]+1])

N <- length(shade$response[!is.na(shade$response)])

mLL <- mle$minimum
Dev <- 2*mLL
aic <- Dev + 2*np
bic <- Dev + np*log(N)

print(data.frame(Parameter = c("Intercept", "time", "arctanh(rho)"),
                  Estimate = round(Est,5), Std.Err = round(E.E,5),
                  z.value = round(V.z,5), P.value = round(V.P,5),
                  LB.95 = round(L.B,5), UB.95 = round(U.B,5)))

print(data.frame(Parameter = c("rho"),
                  Estimate = round(rho,5),
                  LB.95 = round(LBr,5), UB.95 = round(UBr,5)))

cat("\nmLL:", mLL, " Deviance:", Dev, " AIC:", aic, " BIC:", bic, "\n")

cat("\nNumber of iterations:", mle$iterations)
cat("\nOptimization process code:", mle$code, "\n\n")
```


Bibliografía

- [1] Agresti, A. (2002). *Categorical Data Analysis*, Second Edition, Wiley Interscience, New Jersey.
- [2] Albert, P.S. y Waclawiw, M.A. (1998). A Two-State Markov Chain for Heterogeneous Transitional Data: A Quasi-Likelihood Approach. *Statistics in Medicine*, **17**, 1481-1493.
- [3] Albert, P.S. (1999). Longitudinal Data Analysis (Repeated Measures) in Clinical Trials. *Statistics in Medicine*, **18**, 1707-1732.
- [4] Albert, P.S. (2000). A Transitional Model for Longitudinal Binary Data Subject to Nonignorable Missing Data. *Biometrics*, **56**, 602-608.
- [5] Azzalini, A. (1994). Logistic Regression for Autocorrelated Data with Application to Repeated Measures. *Biometrika*, **81**, 767-775.
- [6] Benjamin, M.A.; Rigby, R.A. y Stasinopoulos D.M. (1998). Generalized Autoregressive Moving Average Models. *Journal of the American Statistical Association*, **98**, 214-223.
- [7] Berchtold, A. y Raftery, A.E. (2002). The Mixture Transition Distribution Model for High-Order Markov Chains and Non-Gaussian Time Series. *Statistical Science*, **17**, 328-356.

- [8] Copas, J.B. y Li, H.G. (1997). Inference for Non-random Samples. *Journal of the Royal Statistical Society B*, **59**, 55-95.
- [9] Diggle, P.J.; Liang, K.Y. y Zeger, S.L. (1994). *Analysis of Longitudinal Data*, Oxford University Press, New York.
- [10] Diggle, P.J.; Heagerty, P.J.; Liang, K.Y. y Zeger, S.L. (2002). *Analysis of Longitudinal Data*, Second Edition, Oxford University Press, Oxford.
- [11] Dunn, P.K. y Smyth, G.K. (1996). Randomized Quantile Residuals. *Journal of Computational and Graphical Statistics*, **5**, 236-244.
- [12] Fitzmaurice, G.M.; Laird, N.M. y Rotnitzky, A.G. (1993). Regression Models for Discrete Longitudinal Responses. *Statistical Science*, **8**, 284-299.
- [13] Frees, E.W. y Valdez, E.A. (1998). Understanding Relationships Using Copulas. *North American Actuarial Journal*, **2**, 1-25.
- [14] Halekoh, U.; Hojsgaard, S. y Yan, J. (2006). The R Package geePack for Generalized Estimating Equations. *Journal of Statistical Software*, **15**, 2.
- [15] Hardin, J.W. y Hilbe, J.M. (2003). *Generalized Estimating Equations*. Chapman & Hall, New York.
- [16] Heagerty, P.J. (1999). Marginally Specified Logistic-Normal Models for Longitudinal Binary Data. *Biometrics*, **55**, 688-698.
- [17] Heagerty, P.J. (2002). Marginalized Transition Models and Likelihood Inference for Longitudinal Categorical Data. *Biometrics*, **58**, 342-351.
- [18] Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.

- [19] Kopka, H. y Daly, P.W. (1999). *A Guide to L^AT_EX*, Third Edition, Addison-Wesley, England.
- [20] Le, N.D.; Martin, R.D. y Raftery, A.E. (1996). Modeling Flat Stretches, Burst and Outliers in Time Series Using Mixture Transition Distribution Models. *Journal of the American Statistical Association*, **91**, 1505-1515.
- [21] Liang, K.Y. y Zeger, S. L. (1986). Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, **73**, 13-22.
- [22] Ling, C.H. (1965). Representation of Associative Functions. *Publ Math Debrecen*, **12**, 189-212.
- [23] Little, R.J.A. y Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. Wiley, New York.
- [24] McCullagh, P. y Nelder, J.A. (1989). *Generalized Linear Models*, Second Edition, Chapman & Hall, London.
- [25] Muenz, L.R. y Rubinstein, L.V. (1985). Markov Models for Covariate Dependence of Binary Sequences. *Biometrics*, **41**, 91-101.
- [26] Nelder, J.A. y Wedderburn, R.W.M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society A*, **135**, 370-384.
- [27] Nelsen, R.B. (2006). *An Introduction to Copulas*, Second Edition, Springer, New York.
- [28] Norris, J.R. (1997). *Markov Chains*, Cambridge University Press, England.
- [29] Raftery, A.E. (1985). A Model for High-Order Markov Chains. *Journal of the Royal Statistical Society, Series B*, **47**, 528-539.

- [30] Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, **8**, 229-231.
- [31] Song, P.X.K. (2000). Multivariate Dispersion Models Generated from Gaussian Copula. *Scandinavian Journal of Statistics*, **27**, 305-320.
- [32] Tong, H. (1975). Determination of the Order of a Markov Chain by Akaike's Information Criterion. *Journal of Applied Probability*, **12**, 488-497.
- [33] Venables, W.N. y Ripley, B.D. (1994). *Modern Applied Statistics with S-Plus*, Springer-Verlag, New York.
- [34] Wedderburn, R.W.M. (1974). Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method. *Biometrika*, **61**, 439-447.
- [35] Yan, J. (2002). geepack: Yet Another Package for Generalized Estimating Equations. *R news*, **2**, 12-14.
- [36] Yan, J. (2006). Enjoy the Joy of Copulas. Reporte técnico #365, Department of Statistics and Actuarial Science, University of Iowa.
- [37] Yan, J. y Fine, J. (2004). Estimating Equations for Association Structures. *Statistics in Medicine*, **23**, 859-874.
- [38] Zeger, S.L. y Liang, K.Y. (1986). Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics*, **42**, 121-130.
- [39] Zeger, S.L. y Liang, K.Y. (1992). An Overview of Methods for the Analysis of Longitudinal Data. *Statistics in Medicine*, **11**, 1825-1839.
- [40] Zeger, S.L. y Qaqish, B. (1988). Markov Regression Models for Time Series: A Quasi-Likelihood Approach. *Biometrics*, **44**, 1019-1031.

UNIVERSIDAD AUTÓNOMA METROPOLITANA
UNIDAD IZTAPALAPA

MAESTRÍA EN CIENCIAS
(MATEMÁTICAS APLICADAS E INDUSTRIALES)

MODELADO DE DATOS DE TRANSICIÓN BINARIOS
A TRAVÉS DE UNA CÓPULA DISCRETA

T E S I S
QUE PARA OBTENER EL GRADO DE:
M A E S T R O E N C I E N C I A S

P R E S E N T A:
LUIS CARLOS PÉREZ RUIZ

ASESOR: DR. GABRIEL ESCARELA PÉREZ

A handwritten signature in dark ink, reading "Gabriel Escarela Pérez" with a stylized flourish at the end.

Diciembre de 2007.