



# **Análisis de Influencia Para el Modelo de Regresión Waring.**

por

**Luisa Isabel Rivas Calabrán**

Tesis presentada al Departamento de Estadística de la Pontificia Universidad Católica de Chile para optar al grado de Doctor en Estadística

Enero, 2017

Director de Tesis: **Dr. Manuel Galea Rojas**

©Copyright por **Luisa Isabel Rivas Calabrán**, 2017

FACULTAD DE MATEMÁTICAS  
PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

INFORME DE APROBACIÓN  
TESIS DE DOCTORADO

Se informa a la Facultad de Matemáticas que la Tesis de Doctorado presentada por la candidata

**Luisa Isabel Rivas Calabrán**

ha sido aprobada por la Comisión de Evaluación de la Tesis como requisito para optar al grado de Doctor en Estadística, en el examen de Defensa de Tesis rendido el día 24 de Enero de 2017.

**Director de Tesis**

Dr. Manuel Galea Rojas

\_\_\_\_\_

**Comisión de Evaluación de la Tesis**

Dr. Reinaldo Arellano Valle

\_\_\_\_\_

Dr. Mário de Castro Andrade Filho

\_\_\_\_\_

Dr. Wilfredo Palma Manríquez

\_\_\_\_\_

Dr. Miguel Uribe Opazo

\_\_\_\_\_

FACULTAD DE MATEMÁTICAS  
PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

Fecha:24 de Enero de 2017

Autor            :**Luisa Isabel Rivas Calabrán**  
Título            :**Análisis de Influencia Para  
el Modelo de Regresión Waring.**  
Departamento :**Estadística**  
Grado             :**Doctor**  
Convocación    :**Enero 2017**

Se le concede permiso para hacer circular y copiar, con propósitos no comerciales, el título ante dicho para los requerimientos de individuos y/o instituciones.

---

Firma del Autor

EL AUTOR SE RESERVA LOS DERECHOS DE OTRAS PUBLICACIONES, Y NI LA TESIS NI EXTRACTOS EXTENSOS DE ELLA, PUEDEN SER IMPRESOS O REPRODUCIDOS SIN EL PERMISO ESCRITO DEL AUTOR. EL AUTOR ATESTIGUA QUE EL PERMISO SE HA OBTENIDO PARA EL USO DE CUALQUIER MATERIAL COPYRIGHTED QUE APAREZCA EN ESTA TESIS (CON EXCEPCIÓN DE LOS BREVES EXTRACTOS QUE REQUIEREN SOLAMENTE EL RECONOCIMIENTO APROPIADO EN LA ESCRITURA DEL ESTUDIANTE) Y QUE TODO USO ESTÉ RECONOCIDO CLARAMENTE.

# Índice

<b>Lista de Figuras</b>	<b>VI</b>
<b>Lista de Tablas</b>	<b>X</b>
<b>Resumen</b>	<b>XI</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Definición de objetivos . . . . .	9
1.2. Contribuciones . . . . .	10
1.3. Estructura de la Tesis . . . . .	11
<b>2. Preliminares</b>	<b>13</b>
2.1. Modelo de mezcla Poisson-Gamma: Binomial Negativo . . . . .	13
2.2. Modelo de mezcla Beta-Geométrico: Waring . . . . .	18
2.2.1. Modelo de mezcla Beta-Binomial Negativo: Waring Generalizado	22
2.3. Relación entre modelos . . . . .	28
2.4. Estimación en los Modelos de Mezcla . . . . .	29
2.4.1. Método de Máxima Verosimilitud para modelos de mezcla: verosimilitud observada . . . . .	30
2.4.2. Método de Máxima Verosimilitud para modelos de mezcla: verosimilitud completa . . . . .	32

2.4.2.1. Estimación de los Errores Estándar en el Contexto EM . . . . .	34
2.5. Bondad del ajuste y selección del modelo . . . . .	35
2.6. Análisis de Influencia . . . . .	37
2.6.1. Influencia global: log-verosimilitud observada . . . . .	37
2.6.2. Influencia local: log-verosimilitud observada . . . . .	40
2.6.2.1. Algunos Tipos de Perturbaciones . . . . .	42
2.6.3. Elección del mejor esquema de perturbación . . . . .	44
2.6.4. Influencia global: $Q$ -función . . . . .	46
2.6.5. Influencia local: $Q$ -función . . . . .	48
<b>3. Modelo de Regresión Poisson-Gamma: Binomial Negativo</b>	<b>50</b>
3.1. Estimación en el modelo Poisson-Gamma . . . . .	52
3.2. Diagnóstico de Influencia para el modelo Binomial Negativo . . . . .	56
3.2.1. Influencia global . . . . .	56
3.2.2. Influencia local . . . . .	58
3.3. Aplicación . . . . .	61
3.3.1. Simulación . . . . .	61
3.3.2. Aplicación a datos reales: Cangrejos herradura . . . . .	67
3.3.3. Aplicación a datos reales: Visitas al médico . . . . .	74
3.3.4. Conclusión . . . . .	78
<b>4. Modelo de Regresión Beta-Geométrico: Waring</b>	<b>80</b>
4.1. Estimación en el modelo Beta-Geométrico . . . . .	81
4.1.1. Maximización directa de la función de log-verosimilitud Waring . . . . .	81
4.1.2. Maximización de la función de log-verosimilitud completa: Algoritmo EM . . . . .	83
4.2. Análisis de Influencia . . . . .	86

4.2.1. Influencia global . . . . .	86
4.2.1.1. Distancia de Cook generalizada . . . . .	86
4.2.1.2. Distancia de verosimilitud . . . . .	87
4.2.2. Influencia local . . . . .	87
4.2.2.1. Utilizando la función de log-verosimilitud observada . . . . .	88
4.2.2.2. Utilizando la función de log-verosimilitud completa . . . . .	91
4.3. Aplicación . . . . .	92
4.3.1. Visitas al médico . . . . .	92
4.3.1.1. Influencia global y local en el enfoque del algoritmo EM . . . . .	97
4.3.2. Conclusión . . . . .	99
<b>5. Modelo de Regresión Beta-Binomial Negativo: Waring Generalizado</b>	<b>101</b>
5.1. Estimación en el modelo Waring Generalizado . . . . .	102
5.2. Análisis de Influencia . . . . .	104
5.2.1. Influencia global . . . . .	104
5.2.1.1. Distancia de Cook generalizada . . . . .	104
5.2.1.2. Distancia de verosimilitud . . . . .	105
5.2.2. Influencia local . . . . .	105
5.3. Aplicación . . . . .	109
5.3.1. Aplicación a datos reales, visitas al médico . . . . .	109
5.3.1.1. Conclusión de la aplicación visitas al médico . . . . .	112
5.3.2. Aplicación a datos reales, fútbol . . . . .	115
5.3.2.1. Conclusión . . . . .	125
<b>6. Conclusiones y Trabajos Futuros</b>	<b>127</b>
<b>A. Apéndice</b>	<b>130</b>

**B. Apéndice**

**137**

**Referencias**

**140**

# Índice de figuras

2.1. Función de probabilidad de la distribución Waring para algunos valores del parámetro $\mu$ , manteniendo fijo $\phi$ . . . . .	20
2.2. Función de probabilidad de la distribución Waring para algunos valores del parámetro $\phi$ , manteniendo fijo $\mu$ . . . . .	21
2.3. Función de probabilidad para la distribución Geométrica versus la distribución Waring. . . . .	21
2.4. Función de probabilidad de la distribución Waring Generalizada para algunos valores del parámetro $\vartheta$ , manteniendo fijos $\kappa$ y $\rho$ . . . . .	26
2.5. Función de probabilidad de la distribución Waring Generalizada para algunos valores del parámetro $\rho$ , manteniendo fijos $\vartheta$ y $\kappa$ . . . . .	26
2.6. Función de probabilidad de la distribución Waring Generalizada para algunos valores del parámetro $\rho$ , manteniendo fijos $\vartheta$ y $\kappa$ . . . . .	27
2.7. Función de probabilidad para las distribuciones Binomial Negativa y Waring Generalizada. . . . .	27
2.8. Función de probabilidad para la distribución Waring y Waring Generalizada. . . . .	29
2.9. Relación entre los modelos. . . . .	30
3.1. Gráficos de dispersión conjunto 1 (a) y conjunto 451 (b). . . . .	63



3.2. Valores de $ h_{max} $ (a) y curvatura local total, $C_i$ (b) bajo el vector de perturbación $\omega$ y $\tilde{\omega}$ (c), (d) para el conjunto de datos 1. . . . .	64
3.3. Valores de $ h_{max} $ y curvatura local total, $C_i$ , bajo el vector de perturbación $\omega$ (a)-(b) y $\tilde{\omega}$ (c)-(d) para el conjunto de datos 451. . . . .	64
3.4. Residuos de Pearson para el conjunto de datos 1 con la observación 30 (izquierda) y sin ella (derecha). . . . .	66
3.5. Residuos de Pearson para el conjunto de datos 451 con la observación 30 (a) y sin ella (b). . . . .	66
3.6. Ajuste de los datos con y sin la observación 30 para el conjunto de datos 1 (a) y para el conjunto de datos 152 (b). . . . .	66
3.7. Gráfico de dispersión del ancho del caparazón <i>width</i> de los cangrejos versus el número de cangrejos satélites. . . . .	67
3.8. Estimaciones de los parámetros del modelo de mezcla Poisson-Gamma a través del algoritmo EM, sin la observación $i$ -ésima. . . . .	69
3.9. Residuos de Pearson (izquierda) y residuos Desvío (derecha). . . . .	70
3.10. Distancia de Cook generalizada a un paso para $\hat{\gamma}$ y $\hat{\beta}$ respectivamente. . . . .	70
3.11. Distancia de Cook generalizada a un paso para $\hat{\theta}$ . . . . .	71
3.12. Dirección de máxima curvatura, $( h_{max} )$ , y curvatura local total ( $C_i$ ), para el esquema de perturbación aditiva del predictor, bajo $\omega$ . . . . .	72
3.13. Dirección de máxima curvatura, $( h_{max} )$ , y curvatura local total ( $C_i$ ), para el esquema de perturbación aditiva del predictor, bajo $\tilde{\omega}$ . . . . .	72
3.14. Curvaturas locales para el esquema de perturbación aditiva del predictor bajo los vectores $\omega$ (izquierda) y $\tilde{\omega}$ (derecha). . . . .	73
3.15. Histograma del número de visitas al médico de 1755 mujeres alemanas durante 1987. . . . .	75

3.16. Diagrama de dispersión del número de visitas al médico versus Ingreso (a), Edad (b) y Años de estudio (c). . . . .	75
3.17. Gráfico índice de las distancia de Cook generalizada a un paso para $\hat{\theta}$ . . . . .	76
3.18. Dirección de máxima curvatura, ( $ h_{max} $ ), y curvatura local total, $C_i$ , para el esquema de perturbación aditiva de la variable ingreso. . . . .	76
3.19. Dirección de máxima curvatura, ( $ h_{max} $ ), y curvatura local total, $C_i$ , para el esquema de perturbación aditiva de la variable edad. . . . .	77
3.20. Dirección de máxima curvatura, ( $ h_{max} $ ), y curvatura local total, $C_i$ , para el esquema de perturbación aditiva de la variable años de estudio. . . . .	77
4.1. Envelope de los residuos de Pearson para el ajuste del modelo Geométrico (izquierda) y Waring (derecha). . . . .	94
4.2. Gráfico índice de las distancias de verosimilitud y de Cook generalizada para $\hat{\theta}$ . . . . .	95
4.3. Gráfico índice de $ h_{max} $ y curvatura local total, $C_i$ , para el esquema de perturbación aditiva del predictor Ingresos. . . . .	96
4.4. Gráfico índice para la $Q$ -distancia un paso para $\hat{\theta}$ . . . . .	98
4.5. Gráfico índice para $ h_{max} $ y curvatura local total, $C_i$ , bajo el esquema de perturbación aditiva del predictor Ingresos. . . . .	98
5.1. Envelope de los residuos de Pearson para el ajuste del modelo Binomial Negativo (izquierda) y Waring Generalizado (derecha). . . . .	111
5.2. Gráfico índice de la distancia de verosimilitud y distancia de Cook generalizada para $\hat{\theta}$ . . . . .	111
5.3. Gráfico índice de $ h_{max} $ y curvatura local total, $C_i$ , bajo el esquema de perturbación aditiva del predictor Edad. $C_i$ multiplicado por $10^{-9}$ . . . . .	112

5.4. Relación entre las fracciones de componente de varianza y número de visitas al médico. . . . .	115
5.5. Envelope de los residuos de Pearson para el ajuste del modelo Waring (izquierda), Binomial Negativo (centro) y Waring Generalizada (derecha).116	
5.6. Gráfico de caja número de goles según posición del futbolista en el campo de juego. . . . .	117
5.7. Número observado de goles ( $O_i$ ) y número esperado de goles ( $E_i$ ) por modelo ajustado. . . . .	120
5.8. Residuos de Pearson para el ajuste del modelo Waring Generalizado. . .	121
5.9. Envelope de los residuos de Pearson para el ajuste del modelo Binomial Negativo (izquierda) y Waring Generalizado (derecha). . . . .	121
5.10. Gráfico índice distancia generalizada de Cook y de verosimilitud para $\hat{\theta}$ .	122
5.11. Gráfico índice de $ h_{max} $ y curvatura local total, $C_i$ , para el esquema de ponderación de casos. $C_i$ multiplicado por $10^{-13}$ . . . . .	122
5.12. Gráfico índice de $ h_{max} $ y curvatura local total, $C_i$ , para el esquema de perturbación aditiva de la variable $\log(played)$ . . . . .	123
5.13. Gráfico de la razón de varianza para el conjunto de datos del fútbol. . .	125

# Índice de tablas

3.1. Porcentaje de detección observación 30 bajo $\omega$ y $\tilde{\omega}$ , $QD_i$ . . . . .	62
3.2. Resumen del ajuste del modelo Binomial Negativo para los conjuntos de datos 1 y 451, con y sin la observación 30. Las desviaciones estándar están expresadas entre paréntesis. . . . .	65
3.3. Resumen del ajuste de los modelos Poisson y Binomial Negativo. . . . .	69
3.4. Razón de cambio en los EMV al eliminar un subconjunto de observaciones en el modelo Poisson-Gamma. . . . .	73
3.5. Resumen del ajuste para el modelo de regresión Binomial Negativo de los datos de visitas médicas, eliminando potenciales observaciones influyentes. * señala que el parámetro es significativo al 5%. . . . .	77
4.1. Resumen de los ajustes del modelo de regresión Geométrico y Waring, para los datos de visitas al médico. . . . .	94
4.2. Razón de cambio de los EMV al eliminar un subconjunto de observaciones en el modelo Waring. . . . .	97
4.3. Razón de cambio de los EMV al eliminar un subconjunto de observaciones en el modelo Beta-Geométrico. . . . .	99
5.1. Resumen del ajuste del modelo de regresión Waring Generalizado, para los datos de visitas médicas. . . . .	110

5.2. Razón de cambio de los EMV al eliminar un subconjunto de observaciones en el modelo WG. . . . . 112

5.3. Resumen de los ajustes de los modelos Binomial Negativo y Waring Generalizado. . . . . 113

5.4. Varianza y proporción de varianza por componente para el conjunto de datos visitas al médico. . . . . 114

5.5. Análisis descriptivo de las variables de acuerdo a la posición del jugador en la cancha. . . . . 118

5.6. Resumen de los ajustes del modelo Waring Generalizado, Binomial Negativo, Waring y Geométrico. . . . . 118

5.7. Resumen de los ajustes del modelo Waring Generalizado para los datos de fútbol. . . . . 119

5.8. Tabla de frecuencias observadas ( $O_i$ ) y esperadas ( $E_i$ ) para el modelo Waring Generalizado, Binomial Negativo y Geométrico. . . . . 119

5.9. Tabla de frecuencias observadas ( $O_i$ ) y esperadas ( $E_i$ ) para el modelo Waring Generalizado, Binomial Negativo y Geométrico. . . . . 120

5.10. Varianza y proporción de varianza por componente para el conjunto de datos del fútbol. . . . . 124

5.11. Resumen descomposición de la varianza para las observaciones 19, 125, 344, 794, 804 y 979. . . . . 124

## Resumen

*”Lo mejor de ser estadístico es que podemos jugar en cualquier jardín”.*

*John W. Tukey*

Una de las cualidades de la Estadística es su amplia gama de aplicaciones en diversas áreas del conocimiento, tales como, biología, agronomía, economía, educación, etc, razón por la cual es necesario disponer de herramientas que permitan explicar los fenómenos provenientes de cada una estas áreas.

Cuando el objetivo se centra en estudiar y/o determinar la incidencia de una o más variables sobre otra, recurrimos a los modelos de regresión. A menudo y según sea el caso se deben utilizar modelos de mezclas, debido a que ellos proporcionan mayor flexibilidad, es decir, permiten explicar de mejor manera la relación existente entre las variables estudiadas.

En este trabajo se presenta una descripción de los modelos de mezclas, el estudio de los métodos de estimación de los parámetros, específicamente el método de máxima verosimilitud (MV), tanto para la maximización directa de la función de log-verosimilitud del modelo de mezcla (log-verosimilitud observada), como para la maximización de la función de log-verosimilitud de los datos completos; para ésta última se propone el uso del algoritmo EM.

El presente trabajo se centra en la aplicación de los modelos de mezclas a datos de conteo, específicamente se describen los modelos: Binomial Negativo, obtenido de la mezcla Poisson-Gamma; Waring, el cual surge de la mezcla Beta-Geométrica y el modelo Waring Generalizado, resultado de la mezcla Beta-Binomial Negativa. Para estos modelos se presentan los métodos de estimación, medidas de calidad del ajuste y herramientas de análisis de influencia, tanto global como local.

Por último cabe señalar que toda la implementación computacional fue desarrollada a través del software libre R project.

## Abstract

*"The best thing about being a statistician,  
is that you get to play in everyone's backyard".*

*John W. Tukey*

One of the aspects of Statistics is its wide range of applications on different areas of knowledge such as Biology, Agronomy, Economy and Education. It is for this reason that it is necessary to dispose of tools that allow us to explain the phenomena arising on each of these areas.

When the aim is centered on to study or determine the incidence of one or more variables over others, we resort to regression models. Frequently, and in accordance to the case, we must use mixture models, as they provide more flexibility, i.e. they allow us to capture the behavior of the data in a better way.

Whenever the aim is to study or determine the incidence of one or more variables over others we resort to regression models. Frequently it ought to use mixture models, this is because they provide more flexibility, that is, they allow to explain the relationship between the studied variables in a better way.

In this work we present a description of mixture models and the study of parameter estimation methods. Specifically, we study the maximum likelihood estimation method (ML) for maximization of the loglikelihood of the mixture model (observed loglikelihood), as well as for the maximization of the loglikelihood for the complete data. For the later it is proposed to use the EM algorithm.

This work is focused on the application of mixture models to counting data. Specifically, we describe the following models: Negative-Binomial obtained from the mixture of Poisson Gamma; Waring, which arises from the mixture of Beta-Geometric and the generalized Waring model as a result of the mixture of a Beta-Negative-



Binomial. For these models we present methods of parameter estimation, goodness of fit and tools for influence analysis as well as global and local for each model.

Finally, we remark that all the computational implementation was developed through the open software R project.

# Capítulo 1

## Introducción

En las últimas décadas la discusión acerca de qué modelo es más apropiado para describir un conjunto de datos ya no está centrada en las factibilidades técnicas y/o computacionales, sino más bien en capturar lo mejor posible la esencia de dichos datos, con tal propósito son variados los modelos que se pueden encontrar en la literatura que presentan mayor flexibilidad que los usuales, una clase de tales modelos se denominan modelos de mezcla.

Everitt and Hand (1981), Titterington et al. (1985) y McLachlan and Basford (1987), definen y discuten extensamente los modelos de mezcla, los cuales asumen que la distribución de la variable respuesta es generada por la mezcla de dos o más distribuciones. Este tipo de modelos ha sido utilizado ampliamente en muchas áreas del conocimiento dada su flexibilidad, convirtiéndolos en una herramienta estadística muy versátil, que permite obtener un mejor ajuste de los datos de un determinado fenómeno, principalmente en presencia de heterogeneidad.

En los modelos de mezcla se define una nueva componente, la cual recibe el nombre de variable latente o variable de datos faltantes, precisamente porque representa el conocimiento de los individuos y/o sujetos en estudio que no se puede medir, transformándola en el centro de los modelos de mezcla, por tal razón, la variable latente en muchas aplicaciones está relacionada con un parámetro del modelo, es decir, en

un modelo de mezcla el parámetro  $\varphi$  puede ser tratado como una variable aleatoria y tomar el lugar de la variable latente, y como tal tiene una distribución,  $f(\varphi|\theta)$ , donde  $\theta \in \Omega$  corresponde al vector de parámetros asociados a dicha distribución. Lo anteriormente expuesto se justifica debido a que pensar que una determinada población y/o individuos son homogéneos es muy restrictivo y como tal se puede suponer que el parámetro  $\varphi$  es una variable aleatoria cuya distribución dependerá de la heterogeneidad presente en el fenómeno a modelar.

Karlis (2005) denomina a estos modelos de mezcla como modelos de sobredispersión, porque permiten que la varianza sea mayor que la media, contrario a lo que ocurre en el modelo original.

En términos matemáticos, un modelo de mezcla está dado por,

$$P(y|\theta) = \int P(y|\varphi)f(\varphi|\theta)d\varphi, \quad (1.1)$$

donde  $f(\varphi|\theta)$  corresponde a la distribución de la variable latente (los datos faltantes) y es denominada distribución mezclante o latente,  $P(y|\varphi)$  representa al kernel de la mezcla y  $\theta$  es el vector de parámetros a estimar.

El modelo (1.1) puede ser expresado como,

$$\left. \begin{array}{l} y|\varphi \sim P(y|\varphi) \\ \varphi \sim f(\varphi|\theta) \end{array} \right\} \Rightarrow y \sim P(y|\theta) \quad (1.2)$$

Desde (1.1), es claro que la forma más natural de obtener la distribución de la mezcla es marginalizar la distribución conjunta del modelo de mezcla, sin embargo, estos cálculos podrían en ocasiones resultar muy tediosos y complicados.

Dentro de esta clase de modelos destacan los modelos de mezcla para datos de conteo, los cuales corresponden a aquella herramienta estadística que permite precisamente analizar fenómenos donde la variable respuesta toma valores en  $\{0, 1, 2, \dots\}$ . Para Hilbe (2011), los modelos cuya respuesta son conteos corresponden a un subconjunto de modelos de regresión con respuesta discreta, adicionalmente señala que

el objetivo de los modelos de conteo es explicar el número de ocurrencia o de conteo en cada evento. Menciona además que los modelos de datos de conteo se caracterizan por ser generalmente intrínsecamente heterocedásticos, por presentar asimetría a la derecha y por tener una varianza que se incrementa con la media de la distribución. La segunda característica se debe principalmente a la presencia abundante de ceros.

En la literatura existen muchas aplicaciones que consideran este tipo de modelos para representar fenómenos en diversas áreas, por ejemplo:

- En salud: el número de muertes por una determinada enfermedad por zona geográfica, el número de periodos menstruales hasta conseguir el embarazo, el número de visitas al médico durante un año, etc.
- En educación: el número de alumnos que aprueba una asignatura, el número de computadores disponibles en un colegio, el número de citas bibliográficas de un artículo científico, etc.

Para describir datos de conteo existe una variedad de modelos, donde destaca el modelo de regresión Poisson, en el cual la media y la varianza coinciden, condición que difícilmente pueden cumplir un amplio número de aplicaciones con datos reales. Una alternativa a estos modelos, cuando la varianza supera a la media, es decir, que permiten la sobredispersión, son los modelos de mezcla Poisson. Dentro de este enfoque destaca el modelo de mezcla Poisson-Gamma, más conocido como modelo de regresión Binomial Negativo, denotado en ocasiones por NB2, Cameron and Trivedi (1986). Para Hilbe (2011) una característica destacable de esta parametrización es que permite explicar la heterocedasticidad que presenta el modelo Poisson. El modelo Binomial Negativo ha sido discutido ampliamente en la literatura, por ejemplo, Lawless (1987) asigna a la variable offset de un modelo Poisson una distribución mezclante Gamma, de esta manera obtiene un modelo de regresión Binomial

Negativo. Lord (2006) utiliza datos de accidentes automovilísticos para exhibir sobredispersión, y con ello justificar el uso del modelo de mezcla Poisson Gamma por sobre el modelo Poisson, para lo cual examina el efecto en la estimación del parámetro de dispersión cuando el tamaño muestral es pequeño. Burrell (1990) discute lo apropiado del modelo mezcla Poisson Gamma para explicar la circulación de libros, particularmente en las bibliotecas de la Universidad de Saskatchewan, Canadá y de la Universidad de Sussex, Inglaterra. Van de Ven and Weber (1995) señalan lo adecuado de este modelo cuando el modelo Poisson presenta sobredispersión, para lo cual consideran que la sobredispersión presente es una variable aleatoria con distribución Gamma y proponen que la media de la distribución Poisson y la varianza de la distribución Gamma sean modeladas usando modelos log-lineales; para la estimación de los parámetros utilizan el método de máxima verosimilitud. El modelo también fue estudiado por Lee and Nelder (1996) y Chen and Ahn (1996), quienes utilizan el método de máxima verosimilitud para la estimación de los parámetros; sin embargo, dado el costo computacional que esto puede conllevar, también proponen el uso del método de quasi-verosimilitud. Autores como McNeney and Petkau (1994) también utilizan dicho método y lo aplican con el propósito de estudiar la relación entre la contaminación del aire y la salud. Jowaheer (2003) introduce la quasiverosimilitud generalizada para analizar un conjunto de datos, asume que los efectos aleatorios distribuyen Gamma. Siguiendo con métodos alternativos de estimación, Karlis (2001) propone el algoritmo EM para obtener los estimadores máximo verosímil del modelo de regresión mezcla Poisson, específicamente para el modelo de regresión Binomial Negativo y el modelo de regresión Poisson Gaussiana Inversa; los resultados son aplicados a datos reales concernientes a crímenes de Grecia. En cuanto al análisis de influencia del modelo Binomial Negativo, son variados los aportes realizados, tanto por aquellos autores que consideran la función de log-verosimilitud observada,

como por quienes trabajan con la función de log-verosimilitud completa; obtenida del modelo de mezcla Poisson-Gamma. Así por ejemplo, Zhu et al. (2001) generalizan el enfoque de Cook (1977), tomando como base la esperanza condicional de la función de log-verosimilitud de los datos completos en el algoritmo EM, obtienen algunas medidas de eliminación de caso y lo aplican al modelo de regresión lineal censurado. Zhu and Lee (2001) complementan el trabajo anterior derivando las medidas de influencia local para dicho modelo y lo ilustran, entre otros, mediante datos de penetración de aerosol, donde suponen que el efecto aleatorio sigue una distribución normal, y posteriormente lo extienden a modelos lineales mixtos generalizados. Svetliza (2003) discute la aplicación de influencia local y análisis residual a través de la deviance residual, particularmente del modelo Binomial Negativo no lineal. Por último, Xie and Wei (2008) realizan un análisis de influencia basado en el algoritmo EM para el modelo de regresión Poisson Gaussiana Inversa. Por otro lado, Li (2010) estudia la parte de inferencia de un modelo semiparamétrico con respuesta Binomial Negativa. Garay et al. (2011) estudian los métodos de estimación y las medidas de influencia para el modelo de regresión Binomial Negativo Zero Inflado. Hideto (2015), en su tesis de magíster, estudia las medidas de influencia local para el modelo semiparamétrico Binomial Negativo.

En la misma línea, un modelo muy utilizado para describir datos de conteo es el modelo Geométrico, el cual ha sido estudiado recurrentemente en aplicaciones de fecundidad, por ejemplo, Weinberg and Gladen (1986) asumieron que el número de ciclos menstruales necesarios hasta lograr la concepción, tiene distribución Geométrica con parámetro  $p$ , señalando que en esta situación de la vida diaria este parámetro no necesariamente es constante, es decir, varía de individuo en individuo, y con ello sugieren que el modelo de regresión Geométrico no es apropiado en presencia de heterogeneidad. Luego, asumen que el parámetro  $p$ ,  $0 < p < 1$ , sigue una distribución

Beta, generando de esta forma la distribución de mezcla Beta-Geométrica, conocida más tarde como distribución Waring, Irwin (1975). En este contexto,  $p$  tiene una interpretación muy clara, dado que representa la probabilidad de que ocurra la concepción para cualquier pareja escogida al azar; esta idea fue preconcebida por Henry (1957). Este modelo fue estudiado también con anterioridad por Miller (1961), quien propone que el número de vehículos en una cola (de tráfico) sigue una distribución Geométrica, pero que la tasa de adelantamiento tiene una distribución Beta. Pielou (1962) utiliza esta distribución para analizar el comportamiento de algunas especies de plantas, donde el parámetro  $p$  representa la proporción de segregación entre las especies. Posteriormente, Ridout and Morgan (1991) extienden el trabajo de Weinberg and Gladen (1986) y utilizan dicho modelo para separar a la muestra en mujeres fumadoras y no fumadoras, y según si usaban o no la píldora anticonceptiva; el principal objetivo fue reportar los valores esperados y concluir que las mujeres que fuman reducen su fertilidad. Esta conclusión es concordante con lo obtenido por Baird and Wilcox (1985), quienes realizaron un estudio epidemiológico con tal objetivo. Debido a que el modelo Waring posee un parámetro más que el modelo Geométrico, es necesario analizar si éste es significativo; con tal propósito, Paul (2005) propone un test que permite justificar el uso de una u otra distribución, conjuntamente con otras medidas de bondad de ajuste, dicho test está basado en el estadístico de razón de verosimilitud. Recientemente, Singh et al. (2014) obtienen los estimadores de momentos, de máxima verosimilitud y los estimadores bajo el punto de vista bayesiano, realizan una comparación de éstos, para lo cual presentan dos aplicaciones, la primera de ella a datos provenientes del Gobierno de India, relativo a las estadísticas de salud, y el segundo para un conjunto de datos simulados; ambas aplicaciones relacionadas a los estudios de fecundidad.

Otro modelo que surge con el propósito de modelar datos de conteo es el mode-

lo Waring Generalizado, que debe el adjetivo de Generalizado, al hecho de que corresponde a una mezcla de un modelo Beta-Binomial Negativo, tomando en cuenta que un modelo Binomial Negativo tiene como caso particular el modelo Geométrico. La distribución Waring Generalizada tiene su origen en trabajos de Edward Waring (1736-1798), quien fue el autor del conocido problema matemático denominado *el problema de Waring*. Waring afirmó que todo número natural puede expresarse como la suma de no más de 4 cuadrados, 9 cubos, 19 cuartas potencias y en general  $n$  potencias  $k$ -ésimas positivas, la demostración fue presentada por David Hilbert en 1909, Hilbe (2011). Newbold (1927) fue el primero en describir la distribución Waring Generalizada. Más tarde, Irwin (1968) le dio su representación actual, y considera una distribución que proporciona un modelo que permite que la varianza se pueda dividir en tres componentes, la aplica a la teoría de accidentes, en la cual esos tres elementos tienen una clara definición; el primero de ellos recoge la varianza debido al azar, explicando que ella representa toda la variabilidad que no puede ser, precisamente, explicada por las covariables introducidas en el modelo, mientras que las otras dos están relacionadas con las cualidades internas del individuo y las debidas al medio. En resumen, se puede decir que la distribución Waring Generalizada permite expresar la variabilidad de las observaciones en tres componentes:

- aleatoria
- diferencias internas entre los individuos
- y la presencia de otros factores externos que no son incluidos como covariables en el modelo.

Irwin (1968) prueba que la distribución Binomial Negativa es un caso límite de la distribución Waring Generalizada. Xekalaki (1983b) realiza una discusión acerca



del uso de la distribución Waring Generalizada, llamándola distribución Waring Generalizada Univariada (UGWD), y la aplica en teoría de accidentes poniendo énfasis en las componentes de la varianza y su identificación. También aplica esta distribución a los datos presentados por Irwin (1968) y los compara con las distribuciones relacionadas a contagios, en las que se asume que los individuos de una población están expuestos a cierto riesgo de accidentes. A lo largo de los años esta distribución ha recibido varios nombres, tales como, distribución Hipergeométrica Generalizada de Tipo B3 (o GHgB3), debido a su relación con las funciones hipergeométricas gaussianas. Johnson et al. (2005) mencionan que esta distribución también ha recibido el nombre de Markov-Pólya Inversa; sin embargo el nombre más natural que recibe es el de distribución Beta-Binomial Negativa. Las propiedades de la distribución Waring Generalizada Univariada han sido desarrolladas principalmente por contribuciones de Irwin (1968), Irwin (1975), Xekalaki (1983a), Xekalaki (1983b). Esta distribución ha sido utilizada en lexicología, Tesitelova (1967), para analizar el número de autores de artículos científicos. Ajiferuke (1991) compara la distribución Waring con otras 15 distribuciones discretas y utiliza un test de bondad de ajuste para elegir la que mejor describa a los datos, concluyendo que la mejor corresponde a la distribución Poisson-Gamma Inversa. Levene et al. (2002) estudian la evolución del número de links en la red informática mundial (www). Handcock et al. (2003) y Handcock and Jones (2004) la utilizan en estudios de comportamiento humano. Rodríguez-Avi et al. (2007) proponen su uso para estudiar datos provenientes del área del deporte, específicamente para modelar el número de tarjetas amarillas que recibieron los futbolistas, a excepción de los arqueros, durante la temporada 2001/2002 y 2002/2003 de la liga de fútbol española, donde es clara la descomposición de la varianza en sus tres componentes, la debida al azar propio del juego, la debida a las cualidades y características de cada individuo y las debidas a factores externos; además realizan una segunda aplicación

que tiene relación con describir el número de hoteles por municipio en una región de Andalucía, España, entre 1998 y 2003. Rodríguez-Avi et al. (2009) desarrollan un modelo de regresión para datos de conteo basada en la distribución Waring Generalizada y lo ilustran con datos provenientes del fútbol, donde la variable de respuesta es el número de goles convertidos por cada jugador; para la estimación de los parámetros del modelo desarrollan un paquete en R denominado *GWRM*. Recientemente, Rodríguez-Avi and Olmo-Jiménez (2015) señalan que dicho modelo es una opción para modelar datos que poseen sobredispersión, haciendo notar que el modelo proviene de una mezcla basada en la distribución Poisson. Harris et al. (2014) la consideran una extensión de la distribución Binomial Negativa y destacan la ventaja del modelo Waring Generalizado por sobre el primero, debido a que permite separar la heterogeneidad no observada de los factores o características internas de cada individuo de los factores externos. Ariza-López and Rodríguez-Avi (2015) utilizan el modelo de regresión Waring Generalizado para modelar datos provenientes de zonas geográficas y lo comparan con el modelo Poisson y Binomial Negativo, concluyen que el modelo de regresión Waring Generalizado es una herramienta poderosa dada su cualidad de dividir la varianza, pues proporciona información adicional y mejores predicciones. Peng et al. (2014) también señalan su importancia, al permitir separar la varianza, por sobre otros modelos más tradicionales como el Binomial Negativo; comparan dichos modelos utilizando datos provenientes de accidentes de tránsito.

## **1.1. Definición de objetivos**

El objetivo principal de este trabajo es desarrollar técnicas de influencia global y local, para los modelos de regresión Waring y Waring Generalizado, obtenidos a partir de mezcla de distribuciones.

Con el propósito anterior, los objetivos específicos se pueden resumir de la siguien-

te manera:

- Definir de manera general los modelos de mezcla.
- Estudiar algunas propiedades probabilísticas de las distribuciones propuestas en este trabajo. En particular, el cálculo de los momentos a partir de la Función Generadora de Momentos (FGM).
- Obtener las expresiones para el Análisis de Influencia, tanto global, como local de cada modelo propuesto.
- Implementar los algoritmos de estimación de parámetros, tales como, Newton Rapshon, Mínimos cuadrados reponderados iterativos (IRLS), algoritmo EM.
- Implementar las expresiones de Análisis de Influencia para cada modelo incluido en el trabajo.

## 1.2. Contribuciones

Los modelos de mezcla permiten analizar una gran variedad de fenómenos, la literatura al respecto es muy extensa, este trabajo colabora en esta línea dado, que tal como se aprecia en los objetivos propuestos, se desarrollan tópicos que hasta ahora no habían sido cubiertos, destacando:

- Obtención de las expresiones de influencia global y local de cada modelo en la tesis, y su respectiva implementación en R. En particular, para el modelo de regresión Binomial Negativa, a través del uso de la función de log-verosimilitud completa; para el modelo de regresión Waring, tanto a partir del enfoque de la función log-verosimilitud completa, como de la función de log-verosimilitud observada, y por último a través de la función de log-verosimilitud observada, para el modelo de regresión Waring Generalizado.

- Implementación del algoritmo EM para el modelo de regresión Waring, en el software libre R.
- Obtención de las Funciones Generadoras de Momentos, para la distribución de mezcla Beta-Geométrica: Waring y Beta-Binomial Negativa: Waring Generalizada.

### 1.3. Estructura de la Tesis

Este trabajo se encuentra organizado en 6 capítulos, de la siguiente forma. En el Capítulo 2 se presenta un conjunto de preliminares, en los cuales se entrega un detalle de los modelos de mezcla de datos de conteo incluidos en esta tesis, sus principales propiedades, los métodos requeridos para la estimación de sus parámetros. Adicionalmente, son presentadas algunas medidas que analizan la calidad del ajuste de un modelo, terminando el capítulo con la exposición de las herramientas de análisis de influencia, tanto global como local. En el Capítulo 3, se presenta el desarrollo del algoritmo EM y la obtención de las medidas de influencia a partir de la  $Q$ -función para el modelo de mezcla Poisson-Gamma: Binomial Negativo. En el Capítulo 4, se presentan los estimadores máximo verosímil obtenidos a partir de la maximización directa de la función de log-verosimilitud observada y a través del uso del algoritmo EM para el modelo de mezcla Beta-Geométrico: Waring y su respectivo análisis de influencia. En el Capítulo 5, se presenta el modelo de mezcla Beta-Binomial Negativo: Waring Generalizado, para el cual los estimadores máximo verosímil (EMV) son obtenidos maximizando la función de log-verosimilitud observada y el análisis de influencia es realizado a partir de este enfoque.

Los Capítulos 3, 4 y 5 contienen su respectiva aplicación a datos reales, adicionalmente, el Capítulo 3 incluye simulaciones. Por último, en el Capítulo 6 son expuestas

las principales conclusiones del trabajo y futuras líneas de investigación.

En los Apéndices se detallan algunos cálculos algebraicos desarrollados durante la elaboración de este trabajo.

# Capítulo 2

## Preliminares

En este capítulo se presenta un conjunto de preliminares, en los cuales se detallan los modelos de mezcla incluidos en este trabajo y sus principales propiedades. Además, se presenta una descripción del método de estimación a utilizar, finalizando el capítulo con una descripción de las herramientas de análisis de influencia, tanto global como local.

### 2.1. Modelo de mezcla Poisson-Gamma: Binomial Negativo

El modelo de regresión Poisson asume que cada observación  $y_i$ ,  $i = 1, \dots, n$ , tiene función de probabilidad dada por,

$$P(y_i|t_i, \lambda_i) = \frac{(t_i \lambda_i)^{y_i} e^{-t_i \lambda_i}}{y_i!}, \quad y_i \in \{0, 1, 2, \dots\}, \quad (2.1)$$

donde  $\lambda_i > 0$  y  $t_i > 0$  corresponde a una variable offset, la cual está asociada con área, tiempo de duración, etc., en que el evento de conteo se produce, Hilbe (2011).

Adicionalmente, asumiendo que la variable  $t_i \sim \text{Gamma}(\gamma, \delta)$ , con  $\gamma, \delta > 0$ , la que denota una distribución Gamma cuya función de densidad está dada por,

$$f(t_i|\gamma, \delta) = \frac{t_i^{\gamma-1} e^{-\delta t_i} \delta^\gamma}{\Gamma(\gamma)}, \quad t_i > 0,$$

donde  $\Gamma(\cdot)$  representa la función Gamma, Karlis (2001). En relación a (1.1), en este enfoque la variable  $t_i$  representa la variable latente del modelo de mezcla y como tal su densidad será la distribución mezclante.

Entonces, la distribución mezcla Poisson-Gamma, que da origen a la distribución Binomial Negativa se puede escribir, según (1.2), como,

$$\begin{cases} y_i|t_i, \lambda_i \sim \text{Poisson}(t_i \lambda_i) \\ t_i \sim \text{Gamma}(\gamma, \delta), \end{cases} \quad (2.2)$$

cuya función de probabilidad está dada por,

$$P(y_i|\lambda_i, \gamma, \delta) = \frac{\Gamma(y_i + \gamma)}{y_i! \Gamma(\gamma)} \left( \frac{\delta}{\lambda_i + \delta} \right)^\gamma \left( \frac{\lambda_i}{\lambda_i + \delta} \right)^{y_i}, \quad y_i \in \{0, 1, \dots\}. \quad (2.3)$$

Del modelo (2.1) se tiene que,

$$E(y_i|t_i, \lambda_i) = \text{Var}(y_i|t_i, \lambda_i) = t_i \lambda_i,$$

y por lo tanto,

$$E(y_i) = \frac{\gamma}{\delta} \lambda_i \text{ y } \text{Var}(y_i) = \lambda_i \frac{\gamma}{\delta} \left( 1 + \lambda_i \frac{1}{\delta} \right).$$

Dado que (2.3) presenta problemas de identificabilidad, Karlis (2005) asume que,  $E(t_i) = 1$ , lo que implica que  $\gamma = \delta$ . Con ello,

$$E(y_i) = \lambda_i \text{ y } \text{Var}(y_i) = \lambda_i + \frac{\lambda_i^2}{\gamma},$$

donde  $\gamma$  indica la presencia de sobredispersión, es decir, la distribución mezclante Gamma modela la sobredispersión de los conteos Poisson; cuando  $\gamma \rightarrow \infty$ , la distribución Binomial Negativa (BN) tiende a una distribución Poisson.

Sean  $y_1, y_2, \dots, y_n$  variables aleatorias independientes, donde  $y_i \sim \text{BN}(t_i \lambda_i, \gamma)$ , se puede asumir que existe una relación entre el parámetro  $\lambda_i$  con las covariables del modelo a través del predictor lineal dada por,

$$g(\lambda_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta},$$

donde  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)^T$  es el vector de parámetros desconocidos de la regresión de longitud  $(k + 1)$ , tal que,  $\boldsymbol{\beta} \in \mathbb{R}^{(k+1)}$  y  $\mathbf{x}_i^T$  representa la  $i$ -ésima fila de la matriz de diseño  $\mathbf{X}$  ( $k + 1 < n$ ), cuyos elementos se asumen fijos y conocidos. Por último,  $g(\cdot)$  corresponde a una función monótona y diferenciable denominada *función enlace*. En este trabajo se usa la función enlace log, es decir,  $g(\lambda) = \log(\lambda)$ , debido a que corresponde a la función de enlace más utilizada para el modelo Binomial Negativo, aunque no la canónica, luego,

$$\lambda_i = \exp\{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}\}, \quad i = 1, \dots, n.$$

Sea  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \gamma)^T$  el vector  $q$ -dimensional de parámetros desconocidos del modelo, por lo tanto, a partir de (2.3) la función de log-verosimilitud puede ser escrita como,

$$l(\boldsymbol{\theta}) = \sum_{i=1}^n \{ \log \Gamma(y_i + \gamma) + \gamma \log \gamma + y_i \log \lambda_i - \log y_i! - \log \Gamma(\gamma) - \gamma \log (\lambda_i + \gamma) - y_i \log (\lambda_i + \gamma) \}. \quad (2.4)$$

Note que en el modelo Binomial Negativo  $q = k + 2$ .

Por otro lado, al definir el modelo Binomial Negativo como mezcla Poisson-Gamma se debe definir un vector de datos observados,  $\mathbf{y}_0 = \{y_i, i = 1, \dots, n\}$  y un conjunto de datos faltantes, denotados por  $\mathbf{y}_m$ ; en la mezcla propuesta este rol lo cumple la secuencia  $\mathbf{t} = \{t_i, i = 1, \dots, n\}$  de variables offset. Luego, sea  $\mathbf{y}_c = (\mathbf{y}_0, \mathbf{t})$  el conjunto de datos completo, por lo tanto, la función de log-verosimilitud de los datos completos puede ser escrita como,

$$l_c(\boldsymbol{\theta} | \mathbf{y}_c) = \sum_{i=1}^n \{ -t_i \lambda_i + y_i \log \lambda_i + y_i \log t_i + (\gamma - 1) \log t_i - \log y_i! - \gamma t_i + \gamma \log \gamma - \log \Gamma(\gamma) \}. \quad (2.5)$$

Note que, al reescribir (2.2), asumiendo que  $E(t_i) = 1$ , se tiene que,

$$\begin{cases} y_i | t_i, \lambda_i & \sim \text{Poisson}(t_i \lambda_i) \\ t_i & \sim \text{Gamma}(\gamma, \gamma). \end{cases} \quad (2.6)$$



Luego, la distribución conjunta de  $(y_i, t_i)$  será,  $f(y_i|t_i)f_1(t_i)$ , con lo cual si  $y_i$  tiene esperanza finita, entonces  $E(y_i) = E(E(y_i|t_i))$ , por la ley de esperanza iterada. Siguiendo esta idea, Escobar and Villa (2006), afirman que si  $y_i$  tiene Función Generadora de Momentos (FGM) igual a  $M_{y_i}(t^*)$ , entonces,  $M_{y_i}(t^*) = E(e^{t^*y_i}) = E(M_{y_i|t_i}(t^*))$ , donde  $M_{y_i|t_i}(t^*)$ , corresponde a la FGM de  $y_i|t_i$ , y por lo tanto, su esperanza es calculada respecto a  $f_1(t_i)$ .

Escobar and Villa (2006), presentan además dos importantes resultados que permiten obtener la FGM en modelos de mezcla:

- Suponga que existe,  $t_1 > 0$ , tal que,  $-t_1 < t^* < t_1$ , entonces,

$$M_{y_i|t_i}(t^*) = C_1(t^*)\exp\{C_2(t^*)t_i\},$$

donde  $C_1(t^*)$  y  $C_2(t^*)$ , son funciones finitas de  $t^*$  que no dependen de  $t_i$ .

- Suponga que la FGM,  $M_{t_i}(\cdot)$  existe, y que  $M_{t_i}(C_2(t^*))$  es finita para  $t^* \in (-t_1, t_1)$ , entonces la FGM de  $y_i$  estará dada por,

$$M_{y_i}(t^*) = C_1(t^*)M_{t_i}(C_2(t^*)), \quad -t_1 < t^* < t_1.$$

Aplicando estos resultados a (2.6) se obtiene,

$$\begin{aligned} M_{y_i}(t^*) &= E(\exp(t^*y_i)) \\ &= E(M_{y_i|t_i}(t^*)), \text{ como } y_i|t_i \sim \text{Poisson}(t_i\lambda_i) \Rightarrow M_{y_i|t_i}(t^*) = \exp\{t_i\lambda_i(e^{t^*} - 1)\} \\ &= E(\exp\{t_i\lambda_i(e^{t^*} - 1)\}). \end{aligned}$$

Luego, como  $C_1(t^*) = 1$  y  $C_2(t^*) = \lambda_i(e^{t^*} - 1)$  y tomando el primer resultado, se tiene que,

$$M_{y_i|t_i}(t^*) = C_1(t^*)\exp(C_2(t^*)t_i).$$

Por lo tanto, utilizando el segundo resultado,

$$M_{y_i}(t^*) = C_1(t^*)M_{t_i}(C_2(t^*)),$$

donde  $M_{t_i}(t^*) = \left(1 - \frac{t^*}{\gamma}\right)^{-\gamma}$ ,  $t^* < \gamma$ , pues  $t_i \sim \text{Gamma}(\gamma, \gamma)$ .

Entonces,

$$M_{y_i}(t^*) = \left(1 - \frac{\lambda_i(e^{t^*} - 1)}{\gamma}\right)^{-\gamma} = \left(\frac{\gamma}{\gamma + \lambda_i - \lambda_i e^{t^*}}\right)^{\gamma}, \quad (2.7)$$

que corresponde a la FGM de una distribución Binomial Negativa, a partir de la cual es posible obtener los momentos de dicha distribución, es decir,  $E(y_i)$ ,  $E(y_i^2)$ , etc. En particular con los dos primeros se obtiene la media y la varianza de la distribución, dado que,

$$E(y_i^k) = \frac{\partial^k M_{y_i}(t^*)}{\partial t^{*k}}, \text{ evaluado en } t^* = 0,$$

corresponde al  $k$ -ésimo momento.

Por otro lado, Barreto-Souza and Simas (2015) definen una clase de distribuciones de mezcla Poisson, la que denominan MP, que de acuerdo con (1.2) se puede escribir como,

$$\begin{cases} y_i | t_i, \lambda_i \sim \text{Poisson}(t_i \lambda_i) \\ t_i \sim \text{EF}(\gamma), \end{cases} \quad (2.8)$$

donde EF representa a cualquier distribución con soporte continuo, cuya función de densidad pertenece a la familia exponencial, dada por Nelder and Wedderburn (1972),

$$f(t) = \exp\{\gamma(t\xi_0 - b(t)) + c(t, \gamma)\}, \quad (2.9)$$

donde  $b(\cdot)$  es una función diferenciable tres veces y  $\xi_0$  es tal que  $b'(\xi_0) = 1$  y  $c(\cdot, \cdot)$  es una función conocida. A partir de estas condiciones se tiene que:  $E(t_i) = b'(\xi_0) = 1$  y  $\text{Var}(t_i) = \gamma^{-1}b''(\xi_0)$ . De acuerdo con (2.9), Barreto-Souza and Simas (2015) presentan una expresión general para la obtención de la FGM para la distribución MP, (2.8), dada por,

$$M(t^*) = \exp\{-\gamma[b(\xi_0) - b(\xi_0 + \lambda_i \gamma^{-1}(e^{t^*} - 1))]\}. \quad (2.10)$$

La expresión (2.10) es fácil de obtener a partir de los resultados dados por Escobar and Villa (2006). Específicamente para el modelo representado en la expresión (2.6), que corresponde a un caso particular de la distribución MP, se tiene que,  $\xi_0 = 1$ , lo que implica que  $b(\xi_0) = 0$ ,  $b'(\xi_0) = b''(\xi_0) = 1$  y que  $b(\xi_0 + \lambda_i \gamma^{-1}(e^{t^*} - 1)) = -\log(1 - \lambda_i \gamma^{-1}(e^{t^*} - 1))$ , con lo cual se obtiene la expresión (2.7).

## 2.2. Modelo de mezcla Beta-Geométrico: Waring

Los modelos de regresión Waring con parámetros  $\alpha$  y  $\beta > 0$  son usados para modelar variables discretas,  $y_i$ , que asumen valores enteros no negativos, y cuya función de probabilidades ha sido presentada por diversos autores, quienes utilizan distintas parametrizaciones. En este trabajo se utilizará la propuesta por Singh et al. (2014), la cual fue usada previamente por Crouchley and Dassios (1998), pero en el contexto de datos censurados. Luego, la función de probabilidad está dada por,

$$P(y_i|\alpha, \beta) = \frac{\alpha \Gamma(\alpha + \beta) \Gamma(y_i + \beta)}{\Gamma(\beta) \Gamma(y_i + \alpha + \beta + 1)}, \quad y_i \in \{0, 1, 2, \dots\}. \quad (2.11)$$

Singh et al. (2014) señalan además que (2.11) es una parametrización de la función de probabilidad propuesta originalmente por Weinberg and Gladen (1986).

Note que el modelo (2.11) no incluye covariables. La media y la varianza de la variable  $y_i \sim \text{Waring}(\alpha, \beta)$  están dadas respectivamente por,

$$\begin{aligned} \mu_i = \mu = \mathbf{E}(y_i) &= \frac{\beta}{\alpha - 1}, \quad \alpha > 1, \\ \text{Var}(y_i) &= \frac{\alpha \beta (\alpha + \beta - 1)}{(\alpha - 2)(\alpha - 1)^2}, \quad \alpha > 2. \end{aligned} \quad (2.12)$$

El modelo (2.11) en términos de mezcla de distribuciones se puede escribir como,

$$\begin{cases} y_i | p_i \sim \text{Geo}(p_i) \\ p_i \sim \text{Beta}(\alpha, \beta), \end{cases} \quad (2.13)$$

donde  $\text{Geo}(p_i)$  representa una distribución geométrica con función de probabilidad,

$$P(y_i | p_i) = p_i (1 - p_i)^{y_i}, \quad y_i \in \{0, 1, 2, \dots\},$$

con  $0 < p_i < 1$ , y cuya función de densidad está dada de acuerdo con Crouchley and Dassios (1998) por,

$$f(p_i|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p_i^{\alpha-1} (1 - p_i)^{\beta-1}, \quad \alpha, \beta > 0.$$

**Proposición 1:** La FGM de la distribución Waring( $\alpha, \beta$ ) está dada por,

$$M_y(t^*) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \frac{p^\alpha (1 - p)^{\beta-1}}{1 - (1 - p)e^{t^*}} dp. \quad (2.14)$$

A partir de (2.14) es posible obtener la media y varianza de la distribución Waring, dadas en (2.12), para demostración, ver Apéndice A.

En este trabajo se propone una reparametrización para (2.11) introduciendo una estructura de regresión, es decir,  $\mu \neq \mu_i \forall i$ , luego la función de probabilidades (2.11) reparametrizada en términos de  $\mu_i$  y  $\phi$  es,

$$P(y_i|\mu_i, \phi) = \frac{\phi_1 \Gamma(\phi_1 + \mu_i \phi_2) \Gamma(y_i + \mu_i \phi_2)}{\Gamma(\mu_i \phi_2) \Gamma(y_i + \phi_1 + \mu_i \phi_2 + 1)}, \quad y_i \in \{0, 1, 2, \dots\},$$

donde  $\phi_1 = \frac{2\phi}{\phi - 1}$  y  $\phi_2 = \frac{\phi + 1}{\phi - 1}$ , con  $\mu_i > 0$  y  $\phi > 1$ . A partir de lo anterior se puede escribir,  $y_i \sim \text{Waring}(\mu_i, \phi)$ , cuya media y varianza son,

$$E(y_i) = \mu_i \quad \text{y} \quad \text{Var}(y_i) = \phi(\mu_i^2 + \mu_i) = \phi V(\mu_i),$$

donde  $V(\mu_i) = \mu_i(\mu_i + 1)$  corresponde a la función de varianza, de este modo,  $\mu_i$  es la media de la variable de respuesta y  $\phi$  puede ser interpretado como un parámetro de dispersión, es decir, para un valor fijo de  $\mu_i$ , entre mayor es el valor de  $\phi$  mayor es la varianza de  $y_i$ .

La Figura 2.1 muestra el comportamiento de la función de probabilidades de la distribución Waring para diferentes valores del parámetro  $\mu$  manteniendo fijo el parámetro  $\phi$ ; se observa que a medida que  $\mu$  aumenta, la cola de la distribución es más pesada, este patrón se repite para los distintos valores del parámetro  $\phi$ .

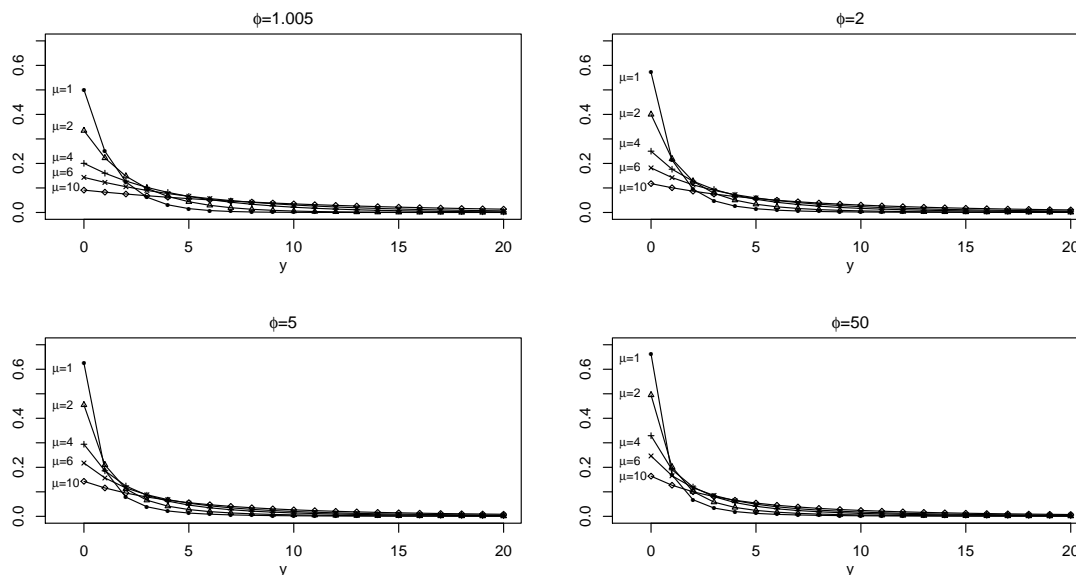


Figura 2.1: Función de probabilidad de la distribución Waring para algunos valores del parámetro  $\mu$ , manteniendo fijo  $\phi$ .

La Figura 2.2 muestra el comportamiento de la función de probabilidades de la distribución Waring cuando se mantiene fijo el parámetro  $\mu$ ; se observa que a medida que  $\phi$  disminuye las colas se hacen más pesadas, la disminución del valor de la probabilidad es más drástico a medida que  $\mu$  es más grande.

La Figura 2.3 (izquierda), muestra la función de probabilidades para la distribución Geométrica con parámetro  $p = 1/5$ , en que  $p = 1/(1 + \mu)$ ; comparada con la distribución Waring ( $\mu = 4, \phi$ ), se observa que a medida que  $\phi \rightarrow 1$ , la curva ajusta mejor a la distribución Geométrica. Análogamente, para la parte derecha de la figura, donde  $p = 1/11$ . En ambos lados de la figura se puede apreciar que la distribución Geométrica es un caso particular de la distribución Waring, cuando  $\phi \rightarrow 1$ .

Sean  $y_1, y_2, \dots, y_n$  variables aleatorias independientes, donde  $y_i \sim \text{Waring}(\mu_i, \phi)$ , se puede establecer al igual que el caso del modelo Binomial Negativo una relación que involucra a las variables explicativas, con la media de la variable respuesta en el

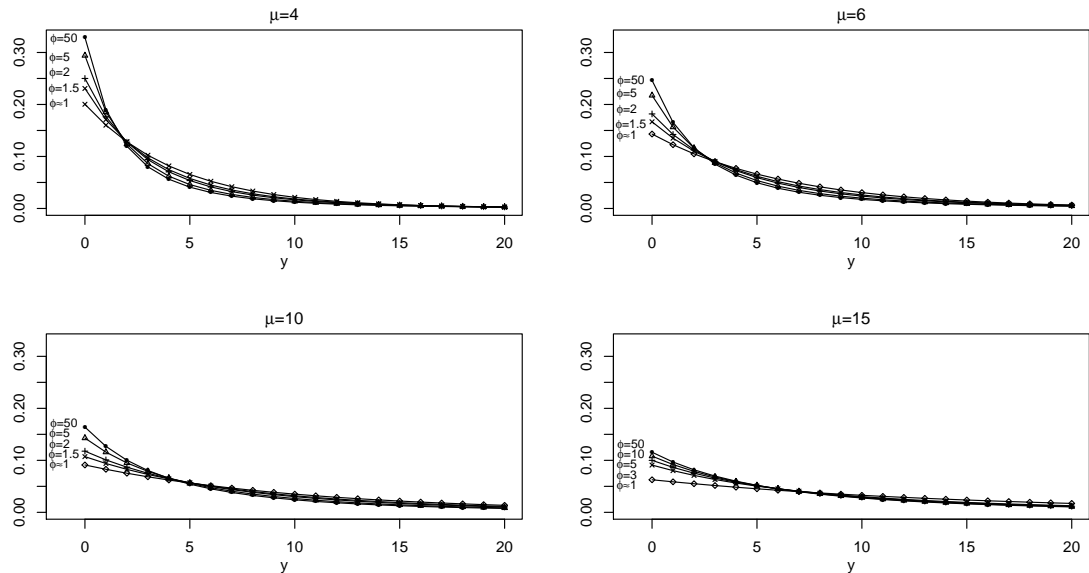


Figura 2.2: Función de probabilidad de la distribución Waring para algunos valores del parámetro  $\phi$ , manteniendo fijo  $\mu$ .

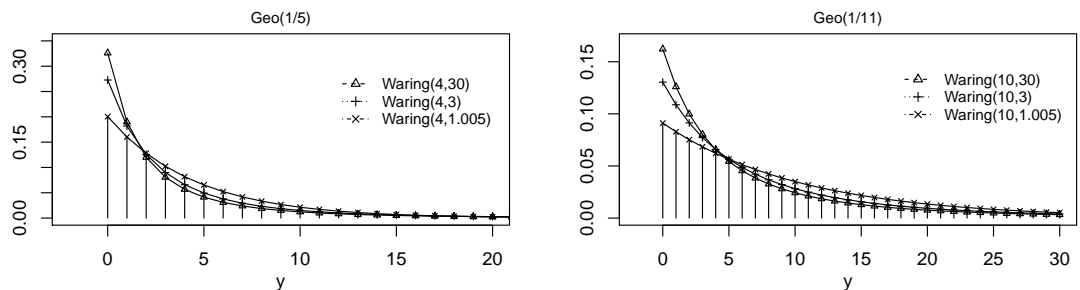


Figura 2.3: Función de probabilidad para la distribución Geométrica versus la distribución Waring.

modelo,

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} = \eta_i, \quad i = 1, \dots, n,$$

a partir de la cual, estableciendo como función enlace a log, se tiene que,  $\mu_i = \exp\{\mathbf{x}_i^T \boldsymbol{\beta}\}$ .

La función de log-verosimilitud basada en una muestra de  $n$  observaciones independientes está dada por,

$$l = l(\boldsymbol{\theta}) = \sum_{i=1}^n l_i(\mu_i, \phi), \quad (2.15)$$

donde  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \phi)^T$  es el vector de parámetros de longitud  $q$ , y la contribución de la

$i$ -ésima observación a la función de log-verosimilitud del modelo es,

$$l_i(\mu_i, \phi) = \log \phi_1 + \log \Gamma(\phi_1 + \mu_i \phi_2) + \log \Gamma(y_i + \mu_i \phi_2) - \log \Gamma(\mu_i \phi_2) - \log \Gamma(y_i + \phi_1 + \mu_i \phi_2 + 1), \quad (2.16)$$

para  $i = 1, \dots, n$ . En este caso,  $\mu_i = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta})$ , es una función de  $\boldsymbol{\beta}$ , el vector de parámetros de la regresión, cuya forma de estimación será descrita en el capítulo siguiente.

Análogamente a la expresión (2.5), la función de log-verosimilitud del conjunto de datos completos  $\mathbf{y}_c = (\mathbf{y}_0, \mathbf{y}_m)$  para el modelo de mezcla Beta-Geométrica se puede expresar como,

$$l(\boldsymbol{\theta} | \mathbf{y}_c) = \sum_{i=1}^n \{ \phi_1 \log p_i + (\mu_i \phi_2 + y_i - 1) \log (1 - p_i) + \log \Gamma(\phi_1 + \mu_i \phi_2) - \log \Gamma(\phi_1) - \log \Gamma(\mu_i \phi_2) \}, \quad (2.17)$$

donde  $y_i$  corresponde a los datos observados y  $p_i$  a los datos faltantes; luego el vector de datos faltantes es,  $\mathbf{y}_m = (p_1, p_2, \dots)$ .

### 2.2.1. Modelo de mezcla Beta-Binomial Negativo: Waring Generalizado

Los modelos de regresión Waring Generalizados son usados para modelar variables discretas,  $y_i$ , y su función de probabilidades de acuerdo con Rodríguez-Avi et al. (2009), está dada por,

$$P(y_i | \vartheta_i, \kappa, \rho) = \frac{\Gamma(\vartheta_i + \rho) \Gamma(\kappa + \rho)}{\Gamma(\rho) \Gamma(\vartheta_i + \kappa + \rho)} \frac{(\vartheta_i)_{y_i} (\kappa)_{y_i}}{(\vartheta_i + \kappa + \rho)_{y_i}} \frac{1}{y_i!}, \quad y_i \in \{0, 1, \dots\}, \quad (2.18)$$

y es denotada por  $\text{WG}(\vartheta_i, \kappa, \rho)$ , con  $\vartheta_i, \kappa, \rho > 0$ , donde  $(\alpha)_\beta$  corresponde al símbolo de Pochhmer definido por,

$$(\alpha)_\beta = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)} = \begin{cases} \alpha(\alpha + 1) \cdots (\alpha + \beta - 1), & \text{si } \beta = 1, 2, \dots \\ 1, & \text{si } \beta = 0. \end{cases}$$

El modelo dado en (2.18) en términos de mezcla de distribuciones puede escribirse como,

$$\begin{cases} y_i | v_i \sim \text{BN}(\vartheta_i, p_i), & \text{con } p_i = \frac{1}{1 + v_i}, \\ v_i \sim \text{BetaII}(\rho, \kappa). \end{cases}$$

En esta mezcla la función de probabilidad de la distribución Binomial Negativa,  $\text{BN}(\vartheta_i, p_i)$  es,

$$P(y_i|v_i) = \frac{\Gamma(y_i + \vartheta_i)}{y_i! \Gamma(\vartheta_i)} \left( \frac{1}{1 + v_i} \right)^{\vartheta_i} \left( \frac{v_i}{1 + v_i} \right)^{y_i} = \frac{\Gamma(y_i + \vartheta_i)}{y_i! \Gamma(\vartheta_i)} p_i^{\vartheta_i} (1 - p_i)^{y_i}, \quad y_i \in \{0, 1, \dots\},$$

la cual proviene de una mezcla Poisson-Gamma obtenida como,

$$\begin{cases} y_i | \lambda_i \sim \text{Poisson}(\lambda_i) \\ \lambda_i \sim \text{Gamma}(\vartheta_i, v_i), \end{cases} \quad (2.19)$$

donde la distribución Gamma tiene función de densidad dada por,

$$f(\lambda_i | \vartheta_i, v_i) = \frac{\lambda_i^{\vartheta_i - 1} \exp\{-\lambda_i/v_i\}}{v_i^{\vartheta_i} \Gamma(\vartheta_i)}, \quad \lambda_i > 0.$$

Por su parte, la distribución BetaII es conocida como distribución Beta Invertida y su función de densidad puede ser escrita como,

$$f(v_i | \kappa, \rho) = \frac{\Gamma(\kappa + \rho)}{\Gamma(\kappa) \Gamma(\rho)} v_i^{\kappa - 1} (1 + v_i)^{-(\kappa + \rho)}, \quad v_i > 0, \quad \kappa, \rho > 0,$$

donde la media y la varianza de  $v_i \sim \text{BetaII}(\rho, \kappa)$  son, respectivamente,

$$\mathbf{E}(v_i) = \frac{\kappa}{\rho - 1}, \quad \rho > 1, \quad \text{y} \quad \text{Var}(v_i) = \frac{\kappa(\kappa + \rho - 1)}{(\rho - 1)^2(\rho - 2)}, \quad \rho > 2.$$

Rodríguez-Avi et al. (2009) proponen el uso de la función enlace log, justifican su utilización por analogía con los modelos de regresión clásicos y por simplicidad, por lo tanto, se establece la relación,  $\mu_i = \exp\{\mathbf{x}_i^T \boldsymbol{\beta}\}$ . Luego, la media de una variable aleatoria distribuida Waring Generalizada (WG) está dada por,

$$\mathbf{E}(y_i) = \mathbf{E}(\mathbf{E}(y_i | v_i)) = \mathbf{E}(\vartheta_i v_i) = \frac{\vartheta_i \kappa}{\rho - 1}, \quad \rho > 1,$$

con  $\rho > 1$  para que la media no sea infinita. Finalmente, utilizando la función enlace, se tiene que,

$$\mathbf{E}(y_i) = \mu_i = \frac{\vartheta_i \kappa}{\rho - 1} \Rightarrow \vartheta_i = \frac{\mu_i(\rho - 1)}{\kappa}; \quad (2.20)$$



por lo tanto, para garantizar que  $\kappa > 0$  y que  $\rho > 1$ , Rodríguez-Avi et al. (2009) definen,

$$\kappa = \exp\{\kappa_0\}, \quad \rho = 1 + \exp\{\rho_0\}, \quad \text{con } \kappa_0 \text{ y } \rho_0 \in \mathbb{R}.$$

Siguiendo la misma idea se obtiene la varianza, dado que la particularidad de esta distribución es que permite identificar las tres componentes de la varianza,

$$\begin{aligned} \text{Var}(y_i) &= \mathbf{E}(\text{Var}(y_i|v_i)) + \text{Var}(\mathbf{E}(y_i|v_i)) \\ &= \mathbf{E}(\vartheta_i v_i (1 + v_i)) + \text{Var}(\vartheta_i v_i) \\ &= \mathbf{E}(\vartheta_i v_i) + \mathbf{E}(\vartheta_i v_i^2) + \text{Var}(\vartheta_i v_i) \\ &= \frac{\vartheta_i \kappa}{\rho - 1} + \frac{\vartheta_i \kappa (\kappa + 1)}{(\rho - 1)(\rho - 2)} + \frac{\vartheta_i^2 \kappa (\kappa + \rho - 1)}{(\rho - 1)^2 (\rho - 2)} \\ &= \underbrace{\mu_i}_{\text{azar}} + \underbrace{\frac{\kappa + 1}{\rho - 2} \mu_i}_{\text{factores externos}} + \underbrace{\frac{\kappa + \rho - 1}{\rho - 2} \frac{\mu_i^2}{\kappa}}_{\text{factores internos}}, \end{aligned} \tag{2.21}$$

con  $\vartheta_i, \kappa > 0$  y  $\rho > 2$ , esta última condición para garantizar que la varianza no sea infinita.

La expresión (2.21), puede ser reescrita como,

$$\text{Var}(y_i) = \pi \left( \mu_i + \frac{1}{\kappa} \mu_i^2 \right),$$

con  $\pi = (\kappa + \rho - 1)/(\rho - 2)$ .

**Proposición 2:** La Función Generadora de Momentos (FGM) para la distribución Waring Generalizada, en que  $y \sim \text{WG}(\vartheta, \kappa, \rho)$ , está dada por,

$$M_y(t) = \frac{\Gamma(\kappa + \rho)}{\Gamma(\kappa)\Gamma(\rho)} \int_0^\infty v^{\kappa-1} (1+v)^{-(\kappa+\rho)} \left( \frac{1}{1+v - ve^t} \right)^\vartheta dv. \tag{2.22}$$

A partir de (2.22) se puede obtener la media y varianza de la distribución Waring Generalizada, las que coinciden con (2.20) y (2.21), respectivamente, para demostración y mayores detalles, ver Apéndice A.

Rodríguez-Avi et al. (2009) discuten acerca de los problemas de identificabilidad del modelo cuando no existe información adicional, es decir, cuando el modelo no con-

sidera la inclusión de covariables, debido a que esto provoca que los parámetros  $\vartheta_i = \vartheta$  y  $\kappa$  sean intercambiables, con lo cual no se podría distinguir entre las componentes de la varianza relacionadas con los factores internos y/o externos (propios de la situación, individuos y medio). Para solucionar este problema, Irwin (1968) propone que sea el experto quien determine qué componente corresponde a cada división de la varianza, debido al conocimiento que él tiene del fenómeno. Xekalaki (1984) propone una solución menos subjetiva, para ello divide el periodo de observación en dos intervalos, no superpuestos, y desarrolla un modelo bivariado, así en cada subperiodo la componente debido a factores internos no cambia.

La Figura 2.4, muestra el comportamiento de la función de probabilidades de la distribución WG, para distintos valores del parámetro  $\vartheta$ , manteniendo fijos los restantes parámetros; se observa que a medida que  $\vartheta$  aumenta, la cola de la distribución se hace más pesada hacia la derecha, y cuando  $\kappa$  aumenta junto con  $\vartheta$  y  $\rho$  se mantiene fijo, la masa de la distribución se desplaza con mayor rapidez hacia la derecha. Además, se puede advertir que cuando  $\kappa \rightarrow 1$ , la distribución coincide con la presentada en la Figura 2.1, es decir, cuando  $\kappa \rightarrow 1$ , se obtiene la distribución Waring. Luego, la distribución Waring es un caso particular de la distribución WG; estas relaciones serán estudiadas con mayor detalle en la siguiente sección. Note que la escala de los gráficos es distinta, dado que las probabilidades disminuyen a medida que  $\kappa$  aumenta.

Al igual que lo observado en la Figura 2.4, la Figura 2.5 muestra el comportamiento de la función de probabilidades de la distribución WG; aumentando los parámetros  $\vartheta$  y  $\kappa$ , para distintos valores del parámetro  $\rho$ , se observa que entre menor sean los dos primeros parámetros y más grande  $\rho$ , la masa de la distribución está más cercana a cero. En la Figura 2.6 se observa un comportamiento similar al descrito previamente.

Por último, en la Figura 2.7, se observa que para valores grandes de  $\rho$ , la distribu-

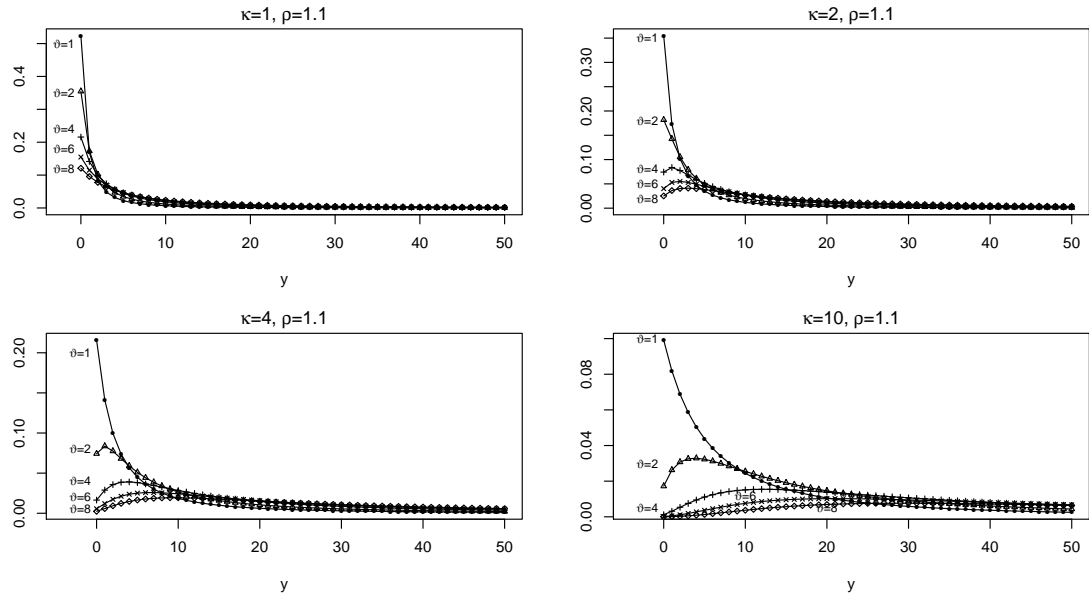


Figura 2.4: Función de probabilidad de la distribución Waring Generalizada para algunos valores del parámetro  $\vartheta$ , manteniendo fijos  $\kappa$  y  $\rho$ .

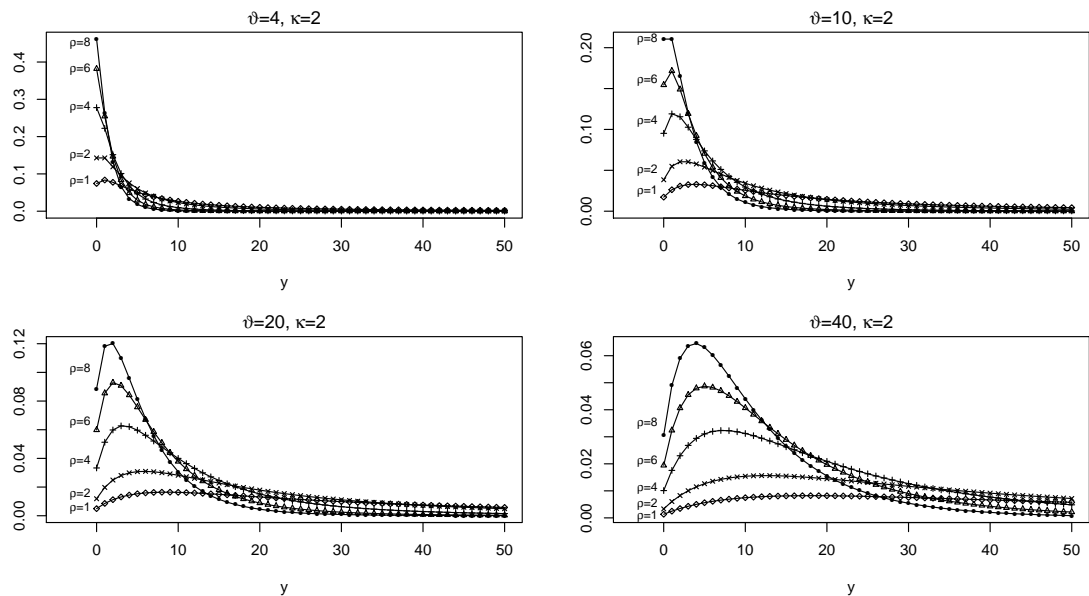


Figura 2.5: Función de probabilidad de la distribución Waring Generalizada para algunos valores del parámetro  $\rho$ , manteniendo fijos  $\vartheta$  y  $\kappa$ .

ción WG coincide con la distribución Binomial Negativa, razón por la cual se afirma que esta última es un caso límite de la distribución WG cuando  $\rho \rightarrow \infty$ .

Finalmente, para  $y_i \sim \text{WG}(\vartheta_i, \kappa, \rho)$ ,  $i = 1, \dots, n$ , se puede escribir su función de

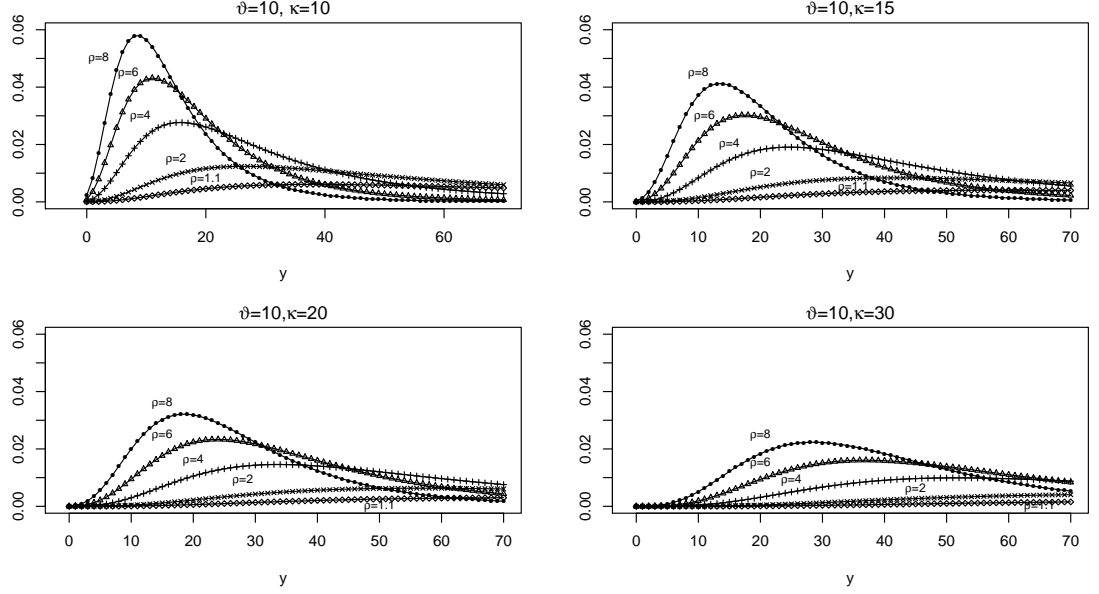


Figura 2.6: Función de probabilidad de la distribución Waring Generalizada para algunos valores del parámetro  $\rho$ , manteniendo fijos  $\vartheta$  y  $\kappa$ .

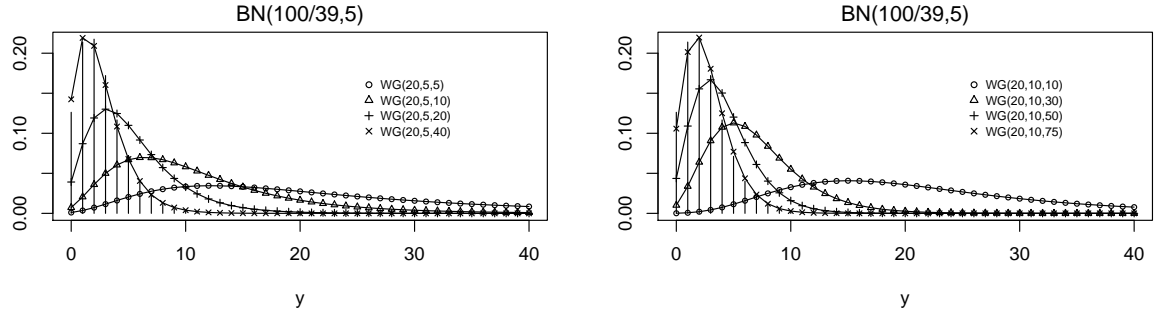


Figura 2.7: Función de probabilidad para las distribuciones Binomial Negativa y Waring Generalizada.

log-verosimilitud basada en una muestra de  $n$  variables aleatorias independientes, con vector de parámetros  $\theta = (\beta^T, \kappa, \rho)^T$ , de longitud  $q = k + 3$ , como,

$$l = l(\theta) = \sum_{i=1}^n l_i(\mu_i, \kappa, \rho), \quad (2.23)$$

donde la contribución de la  $i$ -ésima observación a esta cantidad está dada por,

$$l_i(\mu_i, \kappa, \rho) = \log \Gamma(\vartheta_i + \rho) + \log \Gamma(\kappa + \rho) + \log((\vartheta_i)_{y_i}) + \log((\kappa)_{y_i}) - \log \Gamma(\rho) - \log \Gamma(\vartheta_i + \kappa + \rho) - \log((\vartheta_i + \kappa + \rho)_{y_i}) - \log y_i!. \quad (2.24)$$

### 2.3. Relación entre modelos

En un modelo de regresión de Poisson, la heterogeneidad presente es explicada por cantidades y características observables de los individuos, debido a que la media es función de las covariables del modelo; sin embargo, cuando esta variabilidad excede la condición básica de un modelo Poisson (media = varianza), se está en presencia de sobredispersión y con ello se supone que la tasa de ocurrencia sigue una distribución Gamma y el resultado es un modelo de regresión Binomial Negativo. Luego, se ha creado una nueva fuente de variabilidad diferente a la atribuida al azar (aleatorio) y a las covariables, la cual está relacionada con los factores y características de cada individuo.

Siguiendo la idea y extendiendo el modelo Binomial Negativo, surge el modelo Waring Generalizado que distingue tres fuentes de variación: azar (inherente de un fenómeno aleatorio), la debida a factores externos y a factores internos.

A partir de 2.21, note que:

1. Para cumplir con el supuesto básico de un modelo Poisson,

$$\text{Var}(y_i) = \underbrace{\mu_i}_{\text{azar}} \Rightarrow \frac{\kappa + 1}{\rho - 2} \rightarrow 0 \text{ y } \Rightarrow \frac{\kappa + \rho - 1}{\kappa(\rho - 2)} \rightarrow 0 \quad \therefore \rho \text{ y } \kappa \rightarrow \infty.$$

2. En un modelo Binomial Negativo,

$$\text{Var}(y_i) = \underbrace{\mu_i}_{\text{azar}} + \underbrace{\frac{\mu_i^2}{\kappa}}_{\text{factores internos}}.$$

Para que esto ocurra,

$$\frac{\kappa + 1}{\rho - 2} \rightarrow 0 \text{ y } \Rightarrow \frac{\kappa + \rho - 1}{\kappa(\rho - 2)} \rightarrow \frac{1}{\kappa} \quad \therefore \rho \rightarrow \infty.$$

Entonces, es claro, que un modelo Poisson y Binomial Negativo son casos límites de un modelo Waring Generalizado, en que un modelo Poisson es a su vez un caso particular de un modelo Binomial Negativo.

Por otro lado, en la Figura 2.8 se observa que la distribución Waring es un caso particular de la distribución WG, cuando  $\kappa \rightarrow 1$ , a partir de lo cual se obtiene que,

$$\text{Var}(y_i) = \frac{\rho}{\rho - 2}(\mu_i + \mu_i^2) \Rightarrow \frac{\rho}{\rho - 2} = \phi, \rho = \phi_1.$$

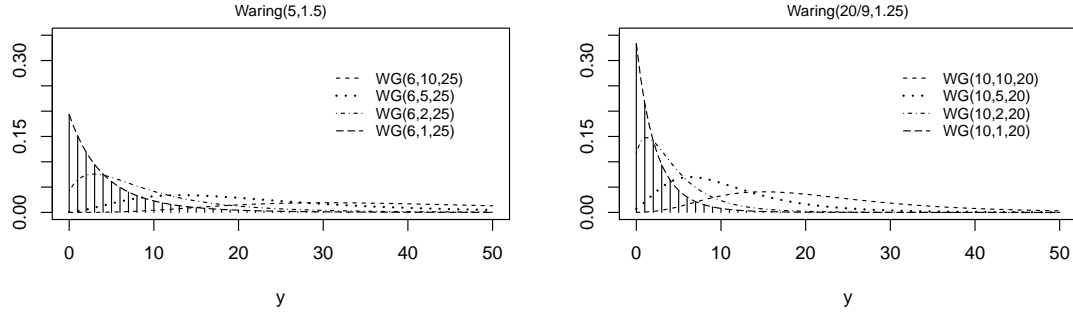


Figura 2.8: Función de probabilidad para la distribución Waring y Waring Generalizada.

Finalmente, si  $\phi \rightarrow 1 \Leftrightarrow \rho \rightarrow \infty$ , se obtiene la distribución Geométrica, correspondiente a un caso particular de la distribución Waring. Note que la distribución Geométrica surge también como un caso particular de la distribución Binomial Negativa, cuando  $\gamma \rightarrow 1$ , dado que si,  $y_i \sim \text{Geo}(p_i)$ , donde  $p_i = \frac{1}{1 + \mu_i}$ , entonces,

$$\text{Var}(y_i) = \frac{1 - p_i}{p_i^2} = \mu_i(1 + \mu_i).$$

La Figura 2.9 muestra un resumen de las relaciones establecidas entre las distribuciones presentadas en este trabajo.

## 2.4. Estimación en los Modelos de Mezcla

El planteamiento y/o obtención de modelos más flexibles y por ende más cercanos a la realidad de los datos, conlleva un aumento del número de parámetros a estimar, y con ello a un problema computacional más complejo.

En este trabajo se obtienen los estimadores máximo verosímil de los parámetros a través de la maximización directa de la función de log-verosimilitud de los datos observados o de la función de log-verosimilitud de los datos completos.

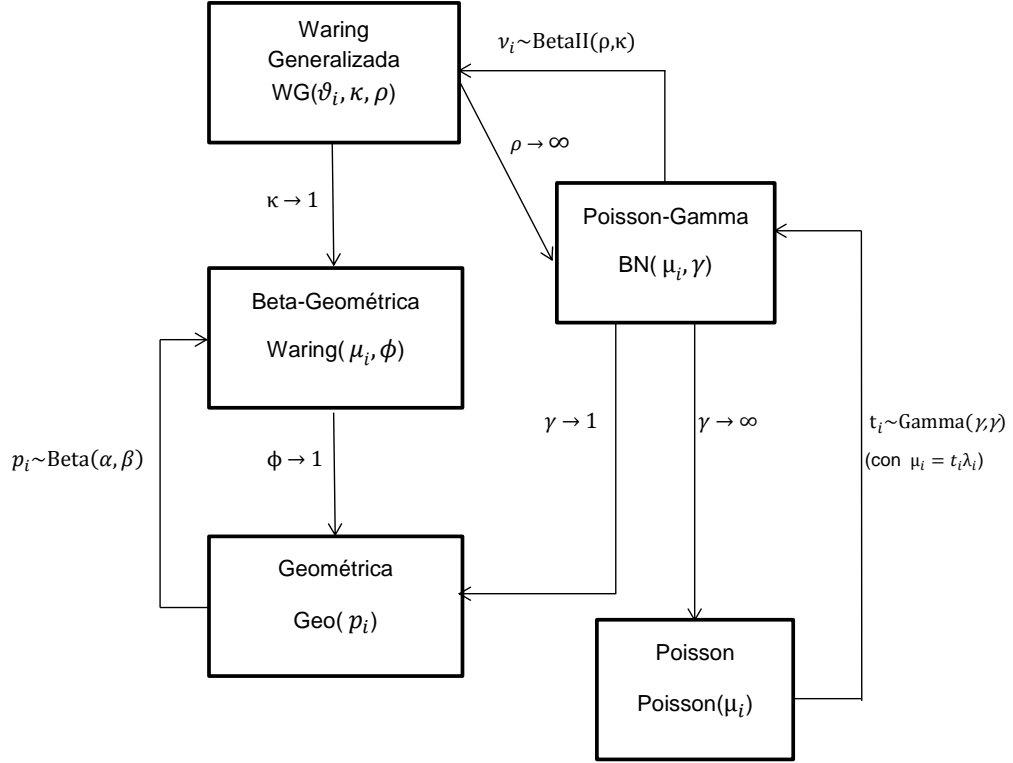


Figura 2.9: Relación entre los modelos.

#### 2.4.1. Método de Máxima Verosimilitud para modelos de mezcla: verosimilitud observada

La estimación de máxima verosimilitud consiste, en primer lugar en encontrar la función de distribución conjunta de las observaciones, asumiendo la independencia de éstas. A partir de la expresión (1.1), suponiendo que se dispone de una muestra aleatoria de tamaño  $n$ , donde los valores observados son fijos y los parámetros pueden variar libremente, las funciones de verosimilitud y log-verosimilitud se pueden escribir respectivamente como,

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n \int_0^{\infty} P(y_i|\varphi_i) f(\varphi_i|\boldsymbol{\theta}) d\varphi_i. \quad (2.25)$$

$$l = l(\boldsymbol{\theta}) = \sum_{i=1}^n \log \int_0^{\infty} P(y_i|\varphi_i) f(\varphi_i|\boldsymbol{\theta}) d\varphi_i. \quad (2.26)$$

A partir de (2.26) se puede concluir que la dificultad para obtener las estimaciones de los parámetros del modelo está directamente relacionada con la integral involucrada, si dicha integral no tiene solución explícita se deben usar métodos numéricos, complicando el proceso de estimación de los parámetros. Por otro lado, si es posible llevar a cabo el proceso de integración se puede trabajar directamente con la función de verosimilitud resultante de la mezcla, es decir,  $P(y_i|\boldsymbol{\theta})$  tiene una forma explícita, en cuyo caso las expresiones (2.25) y (2.26) se reducen a,

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n P(y_i|\boldsymbol{\theta}),$$

$$l = l(\boldsymbol{\theta}) = \sum_{i=1}^n \log P(y_i|\boldsymbol{\theta}).$$

Note que para los tres modelos presentados en este trabajo la integral involucrada en (2.26) tiene solución explícita.

Bajo determinadas condiciones de regularidad, ver Casella and Berger (2002), los candidatos a estimador máximo verosímil (EMV) de  $\boldsymbol{\theta}$  se encuentran resolviendo el sistema de ecuaciones,

$$\mathbf{U}_{\boldsymbol{\theta}} = \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}, \quad (2.27)$$

donde  $\mathbf{U}_{(\cdot)}$  se denomina función score.

En muchos casos, las ecuaciones involucradas en (2.27) no son lineales, lo que conlleva a que no se pueda obtener una solución explícita, por lo se requiere del uso de métodos numéricos para determinar dichas soluciones. Entre las alternativas disponibles para estimar los parámetros destacan los métodos iterativos, tales como algoritmo de Newton-Raphson, Fisher Scoring, Mínimos cuadrados reponderados iterativos (IRLS) y el algoritmo EM.

En cuanto a los errores estándar asintóticos del EMV de  $\boldsymbol{\theta}$ , éstos se pueden obtener directamente de la matriz de información de Fisher esperada,  $\mathbf{K}_{\boldsymbol{\theta}} = E(-\mathbf{I}_{\boldsymbol{\theta}}) =$



$E(\mathbf{U}_\theta \mathbf{U}_\theta^T)$ , donde  $\mathbf{I}_\theta$  es la matriz de información de Fisher observada.

Sin embargo, en ocasiones el cálculo de esta matriz involucra funciones y/o expresiones no lineales para las cuales no es posible obtener expresiones analíticas, en dichos casos una opción es aproximar la matriz de varianzas covarianzas con  $(-\mathbf{I}_\theta)^{-1}$ . En este trabajo se usarán independientemente de acuerdo a la que sea más razonable desde el punto de vista del cálculo.

#### **2.4.2. Método de Máxima Verosimilitud para modelos de mezcla: verosimilitud completa**

A continuación se describe un método alternativo de estimación, el cual puede simplificar el cálculo de los estimadores máximo verosímil, denominado algoritmo EM, del inglés Expectation Maximization.

El algoritmo EM es un método iterativo, desarrollado por Dempster et al. (1977), que permite estimar los parámetros de un modelo por medio de máxima verosimilitud en presencia de información faltante o no observada.

Intuitivamente este algoritmo se puede resumir como:

- i) Sustituir los valores no observados por valores estimados.
- ii) Con lo obtenido en i) estimar los valores de los parámetros del modelo.
- iii) • Repetir el paso i) usando lo obtenido en ii) asumiendo que ellos corresponden a los verdaderos valores.  
 • Repetir el paso ii) usando los valores estimados como valores observados. Iterar hasta obtener la convergencia.

A partir de lo anterior, el conjunto de datos completos  $\mathbf{y}_c$ , está dado por  $\mathbf{y}_c = (\mathbf{y}_0, \mathbf{y}_m)$ , donde  $\mathbf{y}_0$  contiene la información observada (incompleta) y  $\mathbf{y}_m$  los datos faltantes. El algoritmo EM considera una primera aproximación o valor inicial de los

valores de los parámetros,  $\theta^{(0)}$ , y debido a su naturaleza iterativa cuenta con dos etapas:

- **Paso E** (Esperanza): Se define la  $Q$ -función como,

$$Q(\theta|\theta^{(m)}) = E\{l_c(\theta|\mathbf{y}_c)|\mathbf{y}_0, \theta^{(m)}\}, \quad (2.28)$$

donde  $\theta^{(m)}$  corresponde al estimador del vector de parámetros en la  $m$ -ésima iteración,  $m = 0, 1, 2, \dots$  y  $l_c(\cdot)$  corresponde a la función de log-verosimilitud de los datos completos, la cual se puede escribir como,

$$\begin{aligned} l_c(\theta|\mathbf{y}_c) &= \log \prod_{i=1}^n P(y_{ic}|\theta) \\ &= \log \prod_{i=1}^n P(y_{i0}|y_{im}, \theta) f(y_{im}|\theta) \\ &= \sum_{i=1}^n l_{ic}(\theta|y_{ic}), \end{aligned}$$

donde  $l_{ic}(\theta|y_{ic}) = \log(P(y_{i0}|y_{im}, \theta) f(y_{im}|\theta)) = \log(P(y_{i0}|y_{im}, \theta)) + \log(f(y_{im}|\theta))$  corresponde a la contribución de la  $i$ -ésima observación a la función de log-verosimilitud completa.

- **Paso M** (Maximización): Consiste en encontrar el valor de  $\theta^{(m+1)} \in \Theta$  que maximice la cantidad  $Q(\theta|\theta^{(m)})$ , es decir,

$$\theta^{(m+1)} = \underset{\theta}{\text{ArgMax}} Q(\theta|\theta^{(m)}),$$

de modo que,  $Q(\theta^{(m+1)}|\theta^{(m)}) \geq Q(\theta^{(m)}|\theta^{(m)})$ ,  $\forall \theta \in \Theta$ , a partir de lo cual se establece el criterio de convergencia para el proceso iterativo. Los pasos E y M son repetidos alternadamente hasta que se alcance dicho criterio.

El algoritmo EM debe su nombre al hecho que está basado en un procedimiento que maximiza la esperanza de la función log-verosimilitud de los datos completos.

### 2.4.2.1. Estimación de los Errores Estándar en el Contexto EM

Una opción para calcular los errores estándar asintóticos de  $\theta$  es a partir de la función de log-verosimilitud de los datos completos, McLachlan and Krishnan (1997), donde la función de densidad de los datos completos se puede escribir como,

$$P(\mathbf{y}_c|\theta) = P(\mathbf{y}_0|\mathbf{y}_m, \theta)P(\mathbf{y}_m|\theta),$$

o equivalentemente,

$$P(\mathbf{y}_c|\theta) = P(\mathbf{y}_m|\mathbf{y}_0, \theta)P(\mathbf{y}_0|\theta). \quad (2.29)$$

Luego, la función de log-verosimilitud de los datos completos a partir de (2.29) está dada por,

$$\begin{aligned} l_c(\theta|\mathbf{y}_c) &= l_{\mathbf{y}_m|\mathbf{y}_m}(\theta) + l(\theta), \\ \Rightarrow l(\theta) &= l_c(\theta|\mathbf{y}_c) - l_{\mathbf{y}_m|\mathbf{y}_0}(\theta). \end{aligned} \quad (2.30)$$

Diferenciando dos veces a (2.30) con respecto a  $\theta$  se tiene que,

$$\mathbf{I}_\theta(\mathbf{y}_0) = \mathbf{I}_\theta^c(\mathbf{y}_0) - \frac{\partial^2 l_{\mathbf{y}_m|\mathbf{y}_0}(\theta)}{\partial \theta \partial \theta^T}, \quad (2.31)$$

donde

- $\mathbf{I}_\theta(\mathbf{y}_0) = \frac{\partial^2 l(\theta)}{\partial \theta \partial \theta^T}$  es denominada matriz de información observada (hessiano) de los datos incompletos (observados).
- $\mathbf{I}_\theta^c(\mathbf{y}_0) = \frac{\partial^2 l_c(\theta|\mathbf{y}_c)}{\partial \theta \partial \theta^T}$  es la matriz de información observada (hessiano) de los datos completos.

Tomando esperanza en ambos lados de (2.31), con respecto a la distribución condicional de los datos faltantes dado los valores observados, se tiene,

$$\mathbf{I}_\theta(\mathbf{y}_0) = \mathbf{K}_\theta^c(\mathbf{y}_0) - \mathbf{K}_\theta^m(\mathbf{y}_0), \quad (2.32)$$

donde  $\mathbf{K}_\theta^c(\mathbf{y}_0)$  corresponde a la matriz de información de Fisher esperada de los datos incompletos (observados), y  $\mathbf{K}_\theta^m(\mathbf{y}_0)$  es denominada matriz de información pérdida, y está dada por,

$$\mathbf{K}_\theta^m(\mathbf{y}_0) = -E \left( \frac{\partial^2 l_{\mathbf{y}_m|\mathbf{y}_0}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \middle| \mathbf{y}_0 \right).$$

Por último, tomando esperanza en (2.32), con respecto a los datos observados se tiene que,

$$\mathbf{K}_\theta = \mathbf{K}_\theta^c - E(\mathbf{K}_\theta^m | \boldsymbol{\theta}). \quad (2.33)$$

Luego, los errores estándar asintóticos de  $\hat{\boldsymbol{\theta}}$  se pueden obtener a partir de la matriz  $\mathbf{K}_\theta^{-1}$  evaluada en el EMV de  $\boldsymbol{\theta}$ . Por otro lado, si las esperanzas involucradas en (2.33) son difíciles de obtener, se puede obtener una aproximación de los errores estándar asintóticos utilizando  $(-\mathbf{I}_\theta)^{-1}$ , evaluada en el EMV de  $\boldsymbol{\theta}$ .

## 2.5. Bondad del ajuste y selección del modelo

Una vez ajustado el modelo, es importante realizar un análisis acerca de la bondad del ajuste del modelo estimado. Un método utilizado ampliamente con dicho propósito es el análisis de residuos.

En este trabajo se utilizan los residuos de Pearson, dados por,

$$r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\widehat{Var}(y_i)}}, \quad i = 1, \dots, n,$$

donde  $\hat{\mu}_i = \exp\{x_i^T \hat{\boldsymbol{\beta}}\}$ . En el gráfico de  $r_i$  versus  $\hat{y}_i$  se espera no observar patrones, y en el correspondiente gráfico índice de  $r_i$ , si  $|r_i| > 3$  se puede asumir que la observación  $i$  es discrepante. Luego, si los residuos están concentrados entorno al cero, se considera que el modelo es adecuado para el conjunto de datos.

Es común que los residuos no tengan distribución normal estándar, lo cual se puede concluir a partir de un gráfico denominado Cuantil-Cuantil, o simplemente

*qqplot*, que consiste en comparar los cuantiles teóricos con los cuantiles muestrales. Debido a esto y dada la importancia de analizar lo adecuado de un modelo, Barreto-Souza and Simas (2015) sugieren construir los residuos envelope. La idea a grandes rasgos consiste en:

- Simular varios conjuntos de datos utilizando los parámetros estimados (del modelo de interés).
- Realizar el respectivo ajuste para cada uno de estos conjuntos.
- Obtener los residuos de Pearson.
- Calcular los cuantiles teóricos a partir de los residuos mínimos y máximos por individuo, ordenados de manera creciente.
- Graficar las bandas generadas por dichos valores.

Para mayores detalles ver, Barreto-Souza and Simas (2015).

Luego, si el modelo es adecuado, se espera que los cuantiles de los  $r_i$  estén contenidos en las bandas construidas, Atkinson (1985).

Dado que la variable de respuesta considerada en este trabajo corresponde a conteos, otro método útil para analizar lo adecuado de un modelo para describir un determinado conjunto de datos es la prueba de bondad de ajuste, la cual mide la discrepancia entre la frecuencia observada ( $O_i$ ) y la frecuencia esperada o teórica ( $E_i$ ); de acuerdo con esto, la hipótesis nula será,  $H_0$  : los datos siguen la distribución de interés. El estadístico de prueba asociado está dado por,

$$\chi^2 = \sum_{i=1}^{k_1} \frac{(O_i - E_i)^2}{E_i},$$

donde  $k_1$  corresponde al número de clases construidas, tal que,  $E_i = nP(y_i|\boldsymbol{\theta}) \geq 5$ . El estadístico  $\chi^2$  sigue asintóticamente una distribución chi-cuadrado con  $k_1 - q - 1$

grados de libertad, con  $q$  el número de parámetros estimados. La hipótesis nula es rechazada si  $\chi_c^2 > \chi_{1-\alpha}^2(k_1 - q - 1)$ .

En cuanto a la selección del modelo, los criterios usuales para seleccionar entre varios modelos plausibles para un conjunto de datos, están precisamente relacionados con la información disponible una vez que se han ajustados dichos modelos a comparar, destacándose el criterio de información de Akaike, *AIC*, dado por Akaike (1974),

$$AIC = -2l(\hat{\theta}) + 2q,$$

con  $q$  el número de parámetros estimados. Este criterio de selección consiste en escoger el modelo que presente menor *AIC*. Otra alternativa es escoger aquel modelo cuya función de log-verosimilitud sea mayor.

## 2.6. Análisis de Influencia

Otro punto fundamental en el análisis de un modelo es la detección de observaciones influyentes, ya que éstas, si es que existen, pueden afectar las conclusiones extraídas de los resultados del modelo ajustado. En esta sección se evalúa dicha influencia estudiando el conjunto de datos sin una determinada observación, o introduciendo pequeñas perturbaciones, ya sea en el modelo y/o en los datos. Se presentan dichas herramientas, tanto para el ámbito de la función de log-verosimilitud observada, como también para la *Q*-función.

### 2.6.1. Influencia global: log-verosimilitud observada

Dada la importancia que tienen los estimadores de los parámetros de un modelo en el análisis de regresión, es común estudiar el cambio que se puede producir en dichos estimadores al eliminar un caso o un subconjunto de ellos del conjunto de datos. Sean  $l(\theta) = l(\theta|\mathbf{y})$  y  $l(\theta|\mathbf{y}_{[i]})$ , la función de log-verosimilitud de todo el conjunto

y del conjunto de datos sin la  $i$ -ésima observación, respectivamente, donde  $[i]$  denota la eliminación del caso  $i$ .

Con el propósito de evaluar la influencia de la  $i$ -ésima observación, la idea básica es utilizar una medida que permita comparar la distancia entre el estimador de máxima verosimilitud obtenido luego de eliminar un subconjunto de observaciones,  $\hat{\theta}_{[i]}$  y el correspondiente estimador sin eliminar observaciones,  $\hat{\theta}$ . En este trabajo se consideran la distancia de Cook generalizada ( $GD_i$ ) y la distancia de verosimilitud ( $DV_i$ ).

### 1.- Distancia de Cook generalizada

Zhao and Lee (1998) utilizan un estadístico para analizar la distancia entre  $\hat{\theta}$  y  $\hat{\theta}_{[i]}$  denominado distancia de Cook generalizada, dicha distancia es denotada por  $GD_i$ , y está basada en los estimadores máximo verosímiles de  $\theta$ , y es calculada mediante la siguiente expresión,

$$GD_i = \frac{(\hat{\theta} - \hat{\theta}_{[i]})^T \hat{\Sigma}_{\theta}^{-1} (\hat{\theta} - \hat{\theta}_{[i]})}{q}, \quad \forall i = 1, \dots, n,$$

donde  $\hat{\Sigma}_{\theta}$  corresponde a la matriz de covarianzas asintótica estimada de  $\hat{\theta}$  y  $q$  el número de parámetros del modelo. Una elección posible para esta matriz es  $(-\ddot{l}(\theta))^{-1} = (-\mathbf{I}_{\theta})^{-1}$ , evaluada en  $\hat{\theta}$ , dado que corresponde a un estimador consistente de  $\Sigma_{\theta}$ , luego,

$$GD_i = \frac{(\hat{\theta} - \hat{\theta}_{[i]})^T \{-\mathbf{I}_{\hat{\theta}}\} (\hat{\theta} - \hat{\theta}_{[i]})}{q}, \quad \forall i = 1, \dots, n, \quad (2.34)$$

donde  $\mathbf{I}_{\hat{\theta}} = \frac{\partial^2 l(\theta)}{\partial \theta \partial \theta^T} \Big|_{\theta=\hat{\theta}}$ . Para estimar el vector de parámetros  $\theta_{[i]}$  se maximiza la función  $l_{[i]}(\theta)$  para  $i = 1, \dots, n$ , sin embargo, este procedimiento puede resultar computacionalmente muy intenso, especialmente si  $n$  es grande. Pregibon (1981), sugiere usar la aproximación a un paso,  $\hat{\theta}_{[i]}^1$  de  $\hat{\theta}_{[i]}$ , dada por,

$$\hat{\theta}_{[i]}^{(1)} = \hat{\theta} + \{-\mathbf{I}_{\hat{\theta}}\}^{-1} \dot{l}_{[i]}(\hat{\theta}), \quad (2.35)$$

donde  $\dot{l}_{[i]}(\hat{\theta}) = \frac{\partial l_{[i]}(\theta)}{\partial \theta} \Big|_{\theta=\hat{\theta}}$  y la matriz  $\ddot{l}_{[i]}(\theta) \Big|_{\theta=\hat{\theta}}$  es reemplazada por  $\mathbf{I}_{\hat{\theta}}$ , debido a que  $\ddot{l}_{[i]}(\hat{\theta}) \approx \ddot{l}(\hat{\theta})$ .

Sustituyendo (2.35) en (2.34) se obtiene la aproximación a un paso de  $GD_i$ ,

$$\begin{aligned} GD_i^{(1)} &= \frac{(\dot{l}_{[i]}(\hat{\theta}))^T \{-\mathbf{I}_{\hat{\theta}}\}^{-1} (\dot{l}_{[i]}(\hat{\theta}))}{q} \\ &\approx (\dot{l}_{[i]}(\hat{\theta}))^T \{-\mathbf{I}_{\hat{\theta}}\}^{-1} (\dot{l}_{[i]}(\hat{\theta})) \end{aligned} \quad (2.36)$$

Luego, los puntos que tienen valores grandes de  $GD_i^{(1)}$  se espera que presenten un gran impacto sobre las estimaciones de los parámetros del modelo.

## 2.- Distancia de verosimilitud

La distancia de verosimilitud corresponde a otra medida para analizar la distancia entre  $\hat{\theta}$  y  $\hat{\theta}_{[i]}$ , y se define como,

$$DV_i = 2\{l(\hat{\theta}) - l(\hat{\theta}_{[i]})\}, \quad i = 1, \dots, n. \quad (2.37)$$

Al igual que (2.34), el cálculo de (2.37) puede resultar computacionalmente muy intenso, por lo tanto, es útil considerar la aproximación a un paso dada en (2.35), con lo cual se obtiene la distancia de verosimilitud a un paso  $DV_i^{(1)}$ . Cuando los contornos de la función de log-verosimilitud tienen forma elíptica,  $DV_i$  puede ser aproximada usando expansión de Taylor de  $l(\theta_{[i]})$  alrededor de  $\hat{\theta}$ , es decir,

$$l(\hat{\theta}_{[i]}) \approx l(\hat{\theta}) + (\hat{\theta} - \hat{\theta}_{[i]})^T \mathbf{U}_{\hat{\theta}} + \frac{1}{2} (\hat{\theta} - \hat{\theta}_{[i]})^T \mathbf{I}_{\hat{\theta}} (\hat{\theta} - \hat{\theta}_{[i]}).$$

Como,  $\mathbf{U}_{\hat{\theta}} = 0$ , al sustituir en (2.37), se obtiene que,

$$\begin{aligned} DV_i &\approx 2\{l(\hat{\theta}) - l(\hat{\theta}) - \frac{1}{2} (\hat{\theta} - \hat{\theta}_{[i]})^T \mathbf{I}_{\hat{\theta}} (\hat{\theta} - \hat{\theta}_{[i]})\} \\ &\approx (\hat{\theta} - \hat{\theta}_{[i]})^T \{-\mathbf{I}_{\hat{\theta}}\} (\hat{\theta} - \hat{\theta}_{[i]}). \end{aligned} \quad (2.38)$$

Utilizando la aproximación a un paso de  $\hat{\theta}_{[i]}$  dada en (2.35) y dado que  $\ddot{l}_{[i]}(\hat{\theta}) \approx \mathbf{I}_{\hat{\theta}}$ , la expresión (2.38) puede ser reescrita como,

$$DV_i^{(1)} \approx (\dot{l}_{[i]}(\hat{\theta}))^T \{-\mathbf{I}_{\hat{\theta}}\}^{-1} (\dot{l}_{[i]}(\hat{\theta})). \quad (2.39)$$



Puede apreciarse que la expresión (2.39) es proporcional a la distancia de Cook generalizada aproximada a un paso. En este trabajo según sea el caso y/o lo exhaustivo del cálculo computacional se usan indistintamente la estimación de los parámetros sin una observación con o sin la aproximación a un paso.

### 2.6.2. Influencia local: log-verosimilitud observada

En esta sección se presenta una metodología para evaluar el cambio generado en los estimadores de máxima verosimilitud al realizar pequeñas perturbaciones, las cuales pueden ser aplicadas al vector de respuesta (si este es continuo), a las variables explicativas o introduciendo ponderación de casos. Este método recibe el nombre de influencia local, fue propuesto por Cook (1986) y difiere del método de influencia global, de eliminación de casos, pues en todo momento se trabaja con el conjunto de datos con todas las observaciones, con lo cual es posible estudiar los cambios sobre los EMV, con respecto a perturbaciones en los supuestos del modelo y/o en los datos sin eliminar casos.

Sea  $\omega = (\omega_1, \omega_2, \dots, \omega_n)^T$  un vector de longitud  $n$  de perturbaciones variando en una región abierta  $\Omega \subset \mathbb{R}^n$ , denominado espacio de perturbaciones. Se define un modelo estadístico perturbado,  $\mathcal{M} = \{P(\mathbf{y}, \theta, \omega) : \omega \in \Omega\}$ , tal que  $P(\mathbf{y}, \theta, \omega)$  corresponde a la función de densidad conjunta de la variable aleatoria  $\mathbf{y}$ , donde cada elemento del vector de respuesta tiene función de densidad dada en (1.1) perturbada por  $\omega$ . Adicionalmente, se define  $l(\theta|\omega) = \log P(\mathbf{y}, \theta, \omega)$  como la función de log-verosimilitud de los datos perturbada por  $\omega$ , y se asume que existe un vector nulo  $\omega_0 \in \Omega$ , tal que,  $l(\theta) = l(\theta|\omega_0)$ ,  $\forall \theta$ . Sean  $\hat{\theta}$  y  $\hat{\theta}_\omega$  los estimadores máximo verosímil bajo la función de log-verosimilitud original  $l(\theta)$  y perturbada  $l(\theta|\omega)$ , respectivamente. Con el propósito de comparar a  $\hat{\theta}$  y  $\hat{\theta}_\omega$  cuando  $\omega$  varía en el espacio de perturbaciones, Cook (1986) propone el uso del desplazamiento de verosimilitudes, definido precisamente

como la diferencia entre  $l(\boldsymbol{\theta})$  y  $l(\boldsymbol{\theta}|\boldsymbol{\omega})$ , evaluadas en sus respectivos EMV,

$$DV(\boldsymbol{\omega}) = 2\{l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})\}, \quad (2.40)$$

donde  $l(\hat{\boldsymbol{\theta}}) = l(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)$  y  $l(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}) = l(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}|\boldsymbol{\omega}_0)$ . Note que  $DV(\boldsymbol{\omega}) \geq 0 \forall \boldsymbol{\omega} \in \Omega$  y  $DV(\boldsymbol{\omega}_0) = 0$ , y  $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$  denota el EMV bajo el modelo  $\mathcal{M}$ .

El gráfico de  $DV(\boldsymbol{\omega})$  proporciona información relevante respecto a la influencia del esquema de perturbación, y se define como,

$$\alpha(\boldsymbol{\omega}) = \begin{pmatrix} \boldsymbol{\omega} \\ DV(\boldsymbol{\omega}) \end{pmatrix},$$

cuando  $\boldsymbol{\omega}$  varía en  $\Omega$ .

Sea  $C_{\mathbf{h}}(\boldsymbol{\theta})$  la curvatura normal de la superficie  $\alpha(\boldsymbol{\omega})$ , en  $\boldsymbol{\omega}_0$ , en la dirección del vector unitario  $\mathbf{h}$  ( $\|\mathbf{h}\| = 1$ ). Cook (1986) muestra que,

$$C_{\mathbf{h}}(\boldsymbol{\theta}) = 2|\mathbf{h}^T \Delta^T (-\mathbf{I}_{\boldsymbol{\theta}})^{-1} \Delta \mathbf{h}|,$$

donde  $\mathbf{I}_{\boldsymbol{\theta}}$ , es evaluada en  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$  y  $\Delta = \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^T}$ , es evaluada en  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$  y en  $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ . Valores grandes de  $C_{\mathbf{h}}(\boldsymbol{\theta})$  indican sensibilidad al esquema de perturbación en la dirección  $\mathbf{h}$ .  $C_{\mathbf{h}}(\boldsymbol{\theta})$  es denominada influencia local sobre la estimación de  $\boldsymbol{\theta}$ , del esquema de perturbación respectivo en la dirección  $\mathbf{h}$ .

Sea  $\mathbf{h}_{max}$  la dirección de máxima curvatura normal, luego,  $C_{max} = 2\lambda_{max}$ , donde  $\lambda_{max}$  es el mayor valor propio de la matriz  $\mathbf{B} = \Delta^T (-\mathbf{I}_{\boldsymbol{\theta}})^{-1} \Delta$  y  $\mathbf{h}_{max}$  es el correspondiente vector propio, el cual muestra como perturbar el modelo postulado para obtener los mayores cambios locales en el desplazamiento de verosimilitudes. Así, por ejemplo, si la  $i$ -ésima componente del vector  $\mathbf{h}_{max}$  es relativamente grande, implica que modificaciones del peso  $\omega_i$  puede generar cambios relevantes en las conclusiones de un análisis. Otra dirección de interés es el  $i$ -ésimo vector unitario de  $\mathbb{R}^n$ , que corresponde a perturbar sólo el peso  $i$ -ésimo, es decir,  $\omega_i \neq \omega_{0i}$  para algún  $i$ , en

cuyo caso se obtiene  $\mathbf{h} = \mathbf{h}_i$ , bajo este escenario la curvatura normal está dada por  $C_i = 2|B_{ii}|$ , donde  $B_{ii}$  corresponde al  $i$ -ésimo elemento de la diagonal de la matriz  $\mathbf{B}$ , para  $i = 1, \dots, n$ , Galea (2013). Gráficos índices de  $C_i$ ,  $\mathbf{h}_{max}$  son elementos importantes del método de influencia local. En este trabajo se utilizará como criterio para detectar posibles observaciones influyentes  $C_i \geq \bar{C} + 2ds(C)$  o  $C_i \geq \bar{C} + 3ds(C)$ , donde  $\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i$  y  $ds(C)$  corresponde a la desviación estándar de los elementos del vector  $C = (C_1, \dots, C_n)^T$ .

Debido a que  $C_{\mathbf{h}}(\theta)$  no es invariante bajo cambio de escala, Poon and Poon (1999), redefinen la curvatura local por  $C_{\mathbf{h}}^*(\theta) = \frac{C_{\mathbf{h}}(\theta)}{tr(2\mathbf{B})}$ , donde  $tr$  representa la traza de la matriz. Una propiedad interesante de esta nueva curvatura es que para cualquier vector unitario  $\mathbf{h}$ , se tiene que  $0 \leq C_{\mathbf{h}}^*(\theta) \leq 1$ . El principal uso de esta redefinición es que permite comparar curvaturas de diferentes modelos. De Bastiani et al. (2015) denotan a  $C_i^* = \frac{2|B_{ii}|}{tr(2\mathbf{B})}$ , a la curvatura local en la dirección unitaria con  $i$ -ésimo elemento 1 y los restantes 0. Zhu and Lee (2001) consideran que una observación es potencialmente influyente si  $C_i^* \geq \bar{C}^* + 2ds(C^*)$ , donde  $\bar{C}^* = \frac{1}{n} \sum_{i=1}^n C_i^*$  y  $ds(C^*)$  corresponde a la desviación estándar de los elementos del vector  $C^* = (C_1^*, \dots, C_n^*)^T$ .

### 2.6.2.1. Algunos Tipos de Perturbaciones

En general, se entiende por perturbación a cualquier cambio, ya sea en los supuestos del modelo postulado o en los datos; el objetivo es evaluar algún cambio sustancial en los resultados del análisis estadístico a través de esquemas de perturbación. Entre los más comunes se encuentran, ponderación de casos, perturbación de la respuesta y perturbación de los predictores.

#### 1.- Ponderación de casos

Con este esquema se desea evaluar si el estimador máximo verosímil de  $\theta$  se ve afectado cuando las observaciones tienen distintas ponderaciones en el modelo.

Entonces, la función de log-verosimilitud del modelo perturbado estará dado por,

$$l(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^n w_i l_i(\boldsymbol{\theta}),$$

donde  $\omega_i > 0$ ,  $i = 1, \dots, n$ , representa una ponderación fija para el  $i$ -ésimo caso. Es fácil ver que cuando  $\boldsymbol{\omega} = \boldsymbol{\omega}_0 = (1, 1, \dots, 1)^T$ , el modelo perturbado se reduce al modelo original, es decir,  $l(\boldsymbol{\theta}) = l(\boldsymbol{\theta}|\boldsymbol{\omega}_0)$ .

## 2.- Perturbación de la variable de Respuesta

Este tipo de perturbación se puede realizar de dos maneras,

$$y_{i\boldsymbol{\omega}} = y_i + \omega_i \quad \text{o} \quad y_{i\boldsymbol{\omega}} = y_i \times \omega_i, \quad i = 1, \dots, n,$$

denominadas perturbaciones aditivas y multiplicativas, respectivamente.

Note que, dada la naturaleza discreta de la variable de respuesta de este trabajo, este tipo de perturbación no es adecuado.

## 3.- Perturbación de los predictores

Este tipo de perturbación también se puede realizar de dos maneras, tanto aditiva como multiplicativa,

$$x_{ij^*\boldsymbol{\omega}} = x_{ij^*} + \omega_i, \quad \text{o} \quad x_{ij^*\boldsymbol{\omega}} = x_{ij^*} \times \omega_i,$$

para  $i = 1, \dots, n$ , y  $j^*$  de acuerdo a la cantidad de variables predictoras, las cuales están relacionadas con  $\mu_{i\boldsymbol{\omega}} = \exp\{\mathbf{x}_{i\boldsymbol{\omega}}^T \boldsymbol{\beta}\}$ .

Note que para algún esquema de perturbación, la densidad del modelo perturbado,  $\mathcal{M}$ , está dado por,

$$P(\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\omega}) = \prod_{i=1}^n P(y_i, \boldsymbol{\theta}, \omega_i), \quad \text{para } i = 1, \dots, n.$$

### 2.6.3. Elección del mejor esquema de perturbación

A partir del trabajo de Cook (1986), basado en la curvatura normal para evaluar la influencia local de pequeñas perturbaciones de un modelo estadístico, Zhu et al. (2007) y Chen et al. (2009) discuten la selección de esquemas de perturbación apropiados. La propuesta de Zhu et al. (2007) está basada en la matriz de información de Fisher esperada, con respecto al vector de perturbación, dada por,

$$\mathbf{G}(\boldsymbol{\omega}) = (G_{ir}(\boldsymbol{\omega})) = E_{\boldsymbol{\omega}}(\mathbf{U}_{\boldsymbol{\theta}}(\boldsymbol{\omega})\mathbf{U}_{\boldsymbol{\theta}}^T(\boldsymbol{\omega})) = -E_{\boldsymbol{\omega}}\left(\frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^T}\right),$$

donde  $E_{\boldsymbol{\omega}}$  denota la esperanza con respecto a la densidad del modelo perturbado,  $P(\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\omega})$ , y  $\mathbf{U}_{\boldsymbol{\theta}}(\boldsymbol{\omega})$  corresponde a la función score del modelo perturbado.

Los elementos de la diagonal principal de la matriz  $\mathbf{G}(\boldsymbol{\omega})$  indican la cantidad de perturbación introducida por la  $i$ -ésima componente del vector  $\boldsymbol{\omega}$ , mientras que los elementos fuera de ella corresponden a las covarianzas de la función score con respecto a las componentes de  $\boldsymbol{\omega}$ , es decir, representan la asociación lineal entre diferentes componentes de  $\boldsymbol{\omega}$ . Por otra parte, cuando  $\mathbf{G}(\boldsymbol{\omega})$  no es definida positiva para algún esquema de perturbación, implica que algunas componentes del vector de perturbaciones son redundantes, ésto debido a que las filas de la matriz  $\mathbf{G}(\boldsymbol{\omega})$  no son linealmente independientes, lo cual debe ser solucionado. Basado en ésto Zhu et al. (2007) indican que una perturbación es adecuada para un modelo estadístico si satisface las condiciones siguientes:

- $\mathbf{G}(\boldsymbol{\omega})$  es definida positiva en una vecindad de  $\boldsymbol{\omega}_0$ .
- Los elementos fuera de la diagonal principal de  $\mathbf{G}(\boldsymbol{\omega})$ , evaluada en  $\boldsymbol{\omega}_0$ , deben ser lo más pequeños posible.

Luego, una perturbación  $\boldsymbol{\omega}$  es considerada adecuada si al evaluar la matriz de

información de Fisher esperada en  $\omega_0$  se obtiene que,

$$\mathbf{G}(\omega_0) = c\mathbf{I}_n,$$

donde  $c = E_{\omega}(\partial \log P(y_i, \boldsymbol{\theta}, \omega_i) / \partial \omega_i)^2 > 0$  evaluada en  $\omega = \omega_0$ , lo que garantiza que los elementos de  $\omega$  son asintóticamente independientes. Además, si la matriz  $\mathbf{G}(\omega_0) = \text{Diag}(G_{11}(\omega_0), \dots, G_{nn}(\omega_0))$ , es siempre posible escoger un nuevo esquema de perturbación  $\tilde{\omega}$ , definido por,

$$\tilde{\omega} = \omega_0 + c^{-1/2} \mathbf{G}(\omega_0)^{1/2} (\omega - \omega_0), \quad (2.41)$$

tal que,  $\mathbf{G}(\tilde{\omega})$  evaluado en  $\omega_0$  sea igual a  $c\mathbf{I}_n$ .

**Lema:** Si  $\mathbf{G}(\omega) \neq c\mathbf{I}_n$ , entonces  $\mathbf{G}_1(\tilde{\omega})$  evaluado en  $\omega_0$  es igual a  $c\mathbf{I}_n$ , con  $\mathbf{G}(\cdot)$  y  $\mathbf{G}_1(\cdot)$ , la matriz de información de Fisher de  $\omega$  y  $\tilde{\omega}$ , respectivamente.

**Dem:** Sea  $\tilde{\omega} = g(\omega)$ , donde  $g$  es una función 1 a 1 de manera que su inversa exista, tal que  $h(\tilde{\omega}) = \omega$ .

Por definición,

$$\mathbf{G}_1(\tilde{\omega}) = E(\mathbf{U}_{\theta}(\tilde{\omega})\mathbf{U}_{\theta}^T(\tilde{\omega})), \quad (2.42)$$

donde  $\mathbf{U}_{\theta}(\tilde{\omega}) = \frac{\partial l(\boldsymbol{\theta}|\tilde{\omega})}{\partial \tilde{\omega}}$ ; utilizando regla de la cadena se obtiene,

$$\mathbf{U}_{\theta}(\tilde{\omega}) = \frac{\partial h(\tilde{\omega})}{\partial \tilde{\omega}} \frac{\partial l(\boldsymbol{\theta}|\omega)}{\partial \omega}.$$

Por lo tanto, se puede escribir, la expresión (2.42) evaluada en  $\omega = h(\tilde{\omega})$  como,

$$\mathbf{G}_1(\tilde{\omega}) = E \left( \frac{\partial h(\tilde{\omega})}{\partial \tilde{\omega}} \frac{\partial l(\boldsymbol{\theta}|h(\tilde{\omega}))}{\partial h(\tilde{\omega})} \frac{\partial l(\boldsymbol{\theta}|h(\tilde{\omega}))^T}{\partial h(\tilde{\omega})} \frac{\partial h(\tilde{\omega})}{\partial \tilde{\omega}} \right) = \frac{\partial h(\tilde{\omega})}{\partial \tilde{\omega}} \mathbf{G}(h(\tilde{\omega})) \frac{\partial h(\tilde{\omega})}{\partial \tilde{\omega}}.$$

Por otro lado, a partir de (2.41) se tiene que,

$$h(\tilde{\omega}) = \omega = \omega_0 + c^{1/2} \mathbf{G}(\omega_0)^{-1/2} (\tilde{\omega} - \omega_0).$$

Luego,  $h(\omega_0) = \omega_0 \Rightarrow \mathbf{G}(h(\omega_0)) = \mathbf{G}(\omega_0)$  y  $\frac{\partial h(\tilde{\omega})}{\partial \tilde{\omega}} = c^{1/2} \mathbf{G}(\omega_0)^{-1/2}$ ,

$$\begin{aligned} \mathbf{G}_1(\tilde{\omega}_0) &= [c^{1/2} \mathbf{G}(\omega_0)^{-1/2}] \mathbf{G}(\omega_0) [c^{1/2} \mathbf{G}(\omega_0)^{-1/2}]^T \\ &= c \mathbf{G}(\omega_0)^{-1/2} \mathbf{G}(\omega_0) \mathbf{G}^T(\omega_0)^{-1/2}. \end{aligned} \quad (2.43)$$

Como  $\mathbf{G}(\omega)$  es la matriz de información de Fisher esperada para  $\omega$ , y como tal, es simétrica y definida positiva, utilizando la descomposición de Cholesky se tiene que  $\mathbf{G}(\omega) = \mathbf{G}(\omega)^{1/2} \mathbf{G}^T(\omega)^{1/2}$ , la cual al ser sustituida y evaluada en  $\omega_0$  en (2.43), lleva a,

$$\mathbf{G}_1(\tilde{\omega}_0) = \mathbf{c} \mathbf{G}(\omega_0)^{-1/2} \mathbf{G}(\omega_0)^{1/2} (\mathbf{G}(\omega_0)^{-1/2} \mathbf{G}(\omega_0)^{1/2})^T = \mathbf{c} \mathbf{I}_n. \quad \#$$

Zhu et al. (2007) presentan variados ejemplos destinados a discutir y encontrar, bajo diversos escenarios, la perturbación más adecuada. Giménez and Galea (2013) analizan tres esquemas de perturbación propios del modelo de calibración comparativa: ponderación de caso, perturbación aditiva del instrumento de referencia y perturbación aditiva del instrumento alternativo; muestran que el primero de estos esquemas cumple con (2.6.3), sin necesidad de elegir nuevo esquema  $\tilde{\omega}$ , mientras que el segundo y tercer tipo de perturbación requiere de esta nueva elección.

#### 2.6.4. Influencia global: Q-función

Por otro lado, si la estimación de los parámetros es realizada mediante la maximización de la  $Q$ -función, el análisis de influencia será realizado a partir de dicha función. Sean  $l(\theta|\mathbf{y}_c)$  y  $l(\theta|\mathbf{y}_{c[i]})$ , la función de log-verosimilitud del vector de parámetros  $q$  dimensional,  $\theta$ , para los datos completos y para los datos completos sin la  $i$ -ésima observación, respectivamente, donde  $[i]$  denota la eliminación del caso  $i$ .

Luego, la función  $Q(\cdot)$  para el conjunto de datos sin la  $i$ -ésima observación estará dada por,

$$Q_{[i]}(\theta|\hat{\theta}) = E\{l(\theta|\mathbf{y}_{c[i]})|\mathbf{y}_{0[i]}, \hat{\theta}\},$$

cuyo máximo será denotado por  $\hat{\theta}_{[i]}$ ,  $i = 1, \dots, n$  y  $\hat{\theta}$  corresponde al estimador de máxima verosimilitud de  $\theta$ , utilizando los datos completos.

Zhu et al. (2001) proponen el uso de una versión alternativa de la distancia de Cook generalizada dada en (2.34) a partir de la  $Q$ -función, en cuyo caso la distancia

entre  $\hat{\theta}$  y  $\hat{\theta}_{[i]}$  está dada por,

$$GD_i = \frac{(\hat{\theta} - \hat{\theta}_{[i]})^T \{-\ddot{Q}(\hat{\theta}|\hat{\theta})\}(\hat{\theta} - \hat{\theta}_{[i]})}{q}.$$

Realizando la aproximación a un paso  $\hat{\theta}_{[i]}^{(1)}$  de  $\hat{\theta}_{[i]}$  propuesta por Pregibon (1981), se obtiene,

$$\hat{\theta}_{[i]}^{(1)} = \hat{\theta} + \{-\ddot{Q}(\hat{\theta}|\hat{\theta})\}^{-1} \dot{Q}_{[i]}(\hat{\theta}|\hat{\theta}), \quad (2.44)$$

donde  $\dot{Q}_{[i]}(\hat{\theta}|\hat{\theta}) = \frac{\partial Q_{[i]}(\theta|\hat{\theta})}{\partial \theta} \Big|_{\theta=\hat{\theta}}$  y  $\ddot{Q}_{[i]}(\theta|\hat{\theta}) \Big|_{\theta=\hat{\theta}}$  del algoritmo de estimación es reemplazada por  $\ddot{Q}(\hat{\theta}|\hat{\theta})$ .

Realizando la sustitución correspondiente, se obtiene la aproximación a un paso de  $GD_i$  en el contexto EM,

$$\begin{aligned} GD_i^{(1)} &= \frac{(\dot{Q}_{[i]}(\hat{\theta}|\hat{\theta}))^T \{-\ddot{Q}(\hat{\theta}|\hat{\theta})\}^{-1} (\dot{Q}_{[i]}(\hat{\theta}|\hat{\theta}))}{q} \\ &\approx (\dot{Q}_{[i]}(\hat{\theta}|\hat{\theta}))^T \{-\ddot{Q}(\hat{\theta}|\hat{\theta})\}^{-1} (\dot{Q}_{[i]}(\hat{\theta}|\hat{\theta})) \end{aligned} \quad (2.45)$$

Luego, los puntos que tienen valores grandes de  $GD_i^{(1)}$  presentan un gran impacto sobre las estimaciones de los parámetros del modelo.

Análogamente, Zhu et al. (2001) proponen un resultado homólogo a la distancia de verosimilitud ( $DV_i$ ), basado en la  $Q$ -función,

$$QD_i = 2\{Q(\hat{\theta}|\hat{\theta}) - Q(\hat{\theta}_{[i]}|\hat{\theta})\};$$

por (2.44) se obtiene la aproximación de  $QD_i$  a un paso:

$$QD_i^{(1)} = 2\{Q(\hat{\theta}|\hat{\theta}) - Q(\hat{\theta}_{[i]}^{(1)}|\hat{\theta})\}. \quad (2.46)$$

Note que la obtención de las correspondientes medidas de influencia local en el contexto del algoritmo EM, consisten en sustituir  $l(\theta|\omega)$  por  $l_c(\theta|\omega, \mathbf{y}_c)$  y definir al modelo estadístico perturbado  $\mathcal{M}$  como,

$$\mathcal{M} = \{P(\mathbf{y}_c, \theta, \omega) : \omega \in \Omega\},$$



donde  $P(\mathbf{y}_c, \boldsymbol{\theta}, \boldsymbol{\omega})$  corresponde a la función de densidad conjunta de los datos completos perturbados por  $\boldsymbol{\omega}$ .

Zhu et al. (2001) demuestran que bajo ciertas condiciones de regularidad, existe una relación entre  $GD_i^{(1)}$  en el enfoque de la log-verosimilitud completa y la  $Q$ -distancia aproximada a un paso,  $QD_i^{(1)}$ , concluyendo que;  $GD_i^{(1)} \approx QD_i^{(1)}$ .

### 2.6.5. Influencia local: $Q$ -función

Sea  $\boldsymbol{\omega}$  el vector de perturbaciones,  $l_0(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_0)$  y  $l_c(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)$  la función de log-verosimilitud de los datos observados y la función log-verosimilitud de los datos completos perturbados por  $\boldsymbol{\omega}$ , respectivamente, donde  $\boldsymbol{\theta}$  corresponde al vector de parámetros  $q$  dimensional desconocido. Se asume que existe un vector nulo  $\boldsymbol{\omega}_0 \in \Omega$  tal que  $l_0(\boldsymbol{\theta}|\mathbf{y}_0) = l_0(\boldsymbol{\theta}|\boldsymbol{\omega}_0, \mathbf{y}_0)$  y  $l_c(\boldsymbol{\theta}|\mathbf{y}_c) = l_c(\boldsymbol{\theta}|\boldsymbol{\omega}_0, \mathbf{y}_c)$  para todo  $\boldsymbol{\theta}$ . Sean  $\hat{\boldsymbol{\theta}}$  y  $\hat{\boldsymbol{\theta}}_\omega$  los estimadores máximo verosímil bajo las log-verosimilitudes original y perturbada, respectivamente, que maximizan a  $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = E(l_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}_0, \hat{\boldsymbol{\theta}})$  y a  $Q(\boldsymbol{\theta}|\boldsymbol{\omega}, \hat{\boldsymbol{\theta}}) = E(l_c(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)|\mathbf{y}_0, \hat{\boldsymbol{\theta}})$ , basadas en el algoritmo EM. De acuerdo al método descrito en la Sección 2.6.2, la idea es comparar  $\hat{\boldsymbol{\theta}}$  y  $\hat{\boldsymbol{\theta}}_\omega$ , cuando  $\boldsymbol{\omega}$  varía en  $\Omega$ ; estimaciones similares indican que la perturbación realizada tiene poco efecto sobre los estimadores, mientras que, si las estimaciones son muy diferentes, hay claros indicios que el proceso de estimación es altamente sensible a tal perturbación. Para medir esta distancia, Cook(1986) propuso el desplazamiento de verosimilitudes definido en (2.40); siguiendo esta idea, Zhu et al. (2001) y Zhu and Lee (2001) proponen el desplazamiento de la  $Q$ -función, la cual mide la diferencia entre  $\hat{\boldsymbol{\theta}}$  y  $\hat{\boldsymbol{\theta}}_\omega$ , definida como,

$$f_Q(\boldsymbol{\omega}) = 2[Q(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) - Q(\hat{\boldsymbol{\theta}}_\omega|\hat{\boldsymbol{\theta}})],$$

donde  $Q(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = Q(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0, \hat{\boldsymbol{\theta}})$  y  $Q(\hat{\boldsymbol{\theta}}_\omega|\hat{\boldsymbol{\theta}}) = Q(\hat{\boldsymbol{\theta}}_\omega|\boldsymbol{\omega}_0, \hat{\boldsymbol{\theta}})$ . Note que,  $f_Q(\boldsymbol{\omega}) \geq 0 \forall \boldsymbol{\omega} \in \Omega$  y que  $f_Q(\boldsymbol{\omega}_0) = 0$ .

Un gráfico de  $f_Q(\omega)$  versus  $\omega$ , contiene información esencial respecto a la influencia del esquema de perturbación; siguiendo el esquema planteado por Cook (1986), Zhu and Lee (2001) consideran el gráfico definido como:  $\alpha(\omega) = (\omega^T, f_Q(\omega))^T$ .

Luego,  $C_{Q,h}(\theta)$  denota la curvatura normal de la superficie  $\alpha(\omega)$ , en  $\omega_0$ , en la dirección del vector unitario  $h$  ( $\|h\| = 1$ ). Así también, extienden la definición de curvatura normal al ámbito de la  $Q$ -función, la cual queda expresada por,

$$C_{Q,h}(\theta) = 2|h^T \ddot{Q}_{\omega_0} h|,$$

donde  $\ddot{Q}_{\omega_0} = \Delta_{\omega_0}^T \{-\ddot{Q}(\theta|\hat{\theta})\}^{-1} \Delta_{\omega_0}$ ,  $\ddot{Q}(\theta|\hat{\theta}) = \frac{\partial^2 Q(\theta|\hat{\theta})}{\partial \theta \partial \theta^T} \Big|_{\theta=\hat{\theta}}$  y  $\Delta_{\omega} = \frac{\partial^2 Q(\theta, \omega|\hat{\theta})}{\partial \theta \partial \omega^T} \Big|_{\theta=\hat{\theta} \text{ y } \omega=\omega_0}$ .

## Capítulo 3

# Modelo de Regresión Poisson-Gamma: Binomial Negativo

En la literatura se presenta una gran variedad de modelos para describir datos de conteo, dentro de los cuales destaca el modelo de regresión Poisson, el cual se caracteriza por la condición de que la media y la varianza coinciden, la que difícilmente se cumple en aplicaciones con datos reales. Una variante a estos modelos, son aquellos que permiten la existencia de sobredispersión, es decir, que la varianza supere a la media, dichos modelos reciben el nombre de modelos de regresión Poisson Mixtos o modelos de mezcla Poisson. Entre los que destaca el modelo de regresión Binomial Negativo, discutido ampliamente en la literatura. Lord (2006) utiliza datos de accidentes automovilísticos para exhibir sobredispersión, con los cuales justifica el uso del modelo de mezcla Poisson-Gamma y examina el efecto en la estimación del parámetro de dispersión cuando el tamaño muestral es pequeño. Burrell (1990) discute lo apropiado del modelo mezcla Poisson Gamma para explicar la circulación de libros, en las bibliotecas de la Universidad de Saskatchewan, Canadá y de la Universidad de Sussex, Inglaterra. Jowaheer (2003) introduce la quasiverosimilitud generalizada para analizar un conjunto de datos, asume que los efectos aleatorios se distribuyen

Gamma. Karlis (2001) obtiene los EMV de los parámetros de los modelos Poisson-Gamma y Poisson-Gaussiano Inverso utilizando el algoritmo EM, los resultados son aplicados a datos reales concernientes a crímenes de Grecia. En cuanto al análisis de influencia el aporte realizados por diversos autores es extenso, entre ellos destaca el trabajo de Zhu et al. (2001) quienes generalizan el enfoque de Cook (1977) tomando como base la esperanza condicional de la función de log-verosimilitud de los datos completos en el algoritmo EM, obtienen algunas medidas de eliminación de caso y lo aplican al modelo de regresión lineal censurado. Zhu and Lee (2001) complementan el trabajo anterior derivando las medidas de influencia local y lo ilustran, entre otros, mediante datos de penetración de aerosol, donde suponen que el efecto aleatorio sigue una distribución normal, y posteriormente lo extienden a modelos lineales mixtos generalizados. Svetliza (2003) discute la aplicación de influencia local y análisis residual a través de la deviance residual al modelo Binomial Negativo no lineal. Xie and Wei (2008) realizan un análisis de influencia basado en el algoritmo EM para el modelo de regresión Poisson Gaussiana Inversa. Por último, Barreto-Souza and Simas (2015) introducen una clase más general de modelos de regresión mezcla Poisson, a la que denominan MP. Como su nombre lo indica, esta clase de modelos está basada en una mezcla de la distribución Poisson y una distribución que pertenece a la familia exponencial. Para la estimación de los parámetros utilizan el algoritmo EM, y basados en dicho algoritmo, obtienen las medidas de diagnóstico, tanto de influencia global como local. Ilustran estos resultados a un conjunto de datos reales, donde la variable de respuesta corresponde al número de ausencias a clases de 314 alumnos, y las variables explicativas son: el sexo, el puntaje estandarizado obtenido en matemáticas y una variable dicotómica que indica el tipo de programa escolar al que pertenece el alumno. En particular, analizan estos datos bajo el modelo de regresión Binomial Negativo, el cual pertenece a la clase de modelos MP. Para el ajuste del

modelo introducen una estructura de regresión a la media y al parámetro de dispersión. En cuanto a la influencia local consideran los esquemas de ponderación de caso, perturbación aditiva de las variables explicativas y perturbación de la variable no observada (datos faltantes). Sin embargo, en ningún momento hacen referencia a la elección adecuada del esquema de perturbación, lo cual es considerado un aspecto relevante en esta tesis, discutido en la Sección 2.6.3.

En este capítulo se presentan las medidas de influencia para el modelo de regresión Binomial Negativo, como mezcla Poisson-Gamma. Se utiliza el algoritmo EM, basado en la esperanza condicional de la función log-verosimilitud de los datos completos; tanto como método de estimación de los parámetros y como punto de partida para derivar algunas medidas de influencia del modelo, tales como, eliminación de caso y análisis de influencia local. Con el propósito de ilustrar los resultados se presentan simulaciones y aplicaciones a datos reales.

### 3.1. Estimación en el modelo Poisson-Gamma

Para obtener la estimación máximo verosímil de  $\theta$  se utiliza una combinación del método de mínimos cuadrados reponderados (IRLS) y el algoritmo EM. El algoritmo IRLS, según Nelder and Wedderburn (1972) es utilizado para estimar el vector de medias de un modelo, cuando dicho modelo pertenece a la clase de los modelos lineales generalizados y dado que el modelo Poisson pertenece a dicha familia es posible su uso. Así el valor de los parámetros una vez realizadas las  $m$  iteraciones será actualizado en,

$$\beta^{(m+1)} = [\mathbf{X}^T \mathbf{A}^{(m)} \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{A}^{(m)} \mathbf{z}^{(m)}, \quad (3.1)$$

donde  $\mathbf{z}$  es un vector, denominado variable dependiente ajustada (o modificada), dado por,

$$\mathbf{z} = \boldsymbol{\eta} + \mathbf{A}^{1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu}),$$

con  $\mathbf{V}$  una matriz diagonal, cuyos elementos corresponden a la función varianza del modelo, y  $\mathbf{A}$  una matriz diagonal denominada matriz de pesos, que cambia en cada paso del proceso iterativo al igual que el vector  $\mathbf{z}$ , y cuyos elementos están dados por,

$$A_i = \frac{1}{V(y_i)} \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2. \quad (3.2)$$

Por su parte, el algoritmo EM, se puede resumir para este modelo como,

- **Paso E:** Calcular la  $Q$ -función de acuerdo con la expresión (2.28) para el modelo de mezcla Poisson-Gamma. De acuerdo a ella y a (2.5), es necesario calcular la esperanza de la variable offset y su logaritmo:

- Por lemma de Sapatinas (1995) con  $r = 1$ , se tiene de manera general de acuerdo con (1.2) que,

$$E(\varphi | \mathbf{y}_0, \boldsymbol{\theta}^{(m)}) = \frac{P(y+1) \mathbf{a}_y}{P(y) \mathbf{a}_{y+1}}, \quad (3.3)$$

donde  $P$  corresponde a la distribución resultante de la mezcla. Aplicando (3.3) al modelo Poisson-Gamma se obtiene,

$$E(t_i | \mathbf{y}_0, \boldsymbol{\theta}^{(m)}) = \frac{y_i + \gamma^{(m)}}{\lambda_i^{(m)} + \gamma^{(m)}} = e_i,$$

dado que en el modelo Binomial Negativo  $\mathbf{a}_y = y!$ .

- Y, como  $t_i \sim \text{Gamma}(\gamma, \gamma)$  implica que  $t_i | \mathbf{y}_0, \lambda_i \sim \text{Gamma}(y_i + \gamma, \lambda_i + \gamma)$ , por lo tanto se tiene que,

$$E(\log t_i | \mathbf{y}_0, \boldsymbol{\theta}^{(m)}) = \psi(y_i + \gamma^{(m)}) - \log(\lambda_i^{(m)} + \gamma^{(m)}) = s_i,$$

para  $i = 1, \dots, n$ , donde  $\psi(\cdot)$  denota la función digamma.

Sustituyendo los pseudovalores  $e_i$  y  $s_i$  en (2.5) se obtiene la  $Q$ -función,

$$Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(m)}) = \sum_{i=1}^n \{-e_i \lambda_i + y_i \log \lambda_i + y_i s_i + (\gamma - 1) s_i - \log y_i! - \gamma e_i + \gamma \log \gamma - \log \Gamma(\gamma)\}. \quad (3.4)$$

■ **Paso M:** Encontrar el valor de  $\theta$  que maximiza la  $Q$ -función dada en (3.4).

- Actualizar los estimadores  $\beta^{(m)}$ , relativos al modelo Poisson, usando los seudovalores  $e_i$  como valores de la variable offset ( $t_i$ ). Para la estimación de  $\beta$  en este trabajo se utilizó el método IRLS dado en (3.1), donde  $\mathbf{X}$  representa la matriz de diseño del modelo y los elementos de la matriz diagonal  $\mathbf{A}$ , cuyos elementos son calculados a partir de la expresión (3.2) y están dados por,

$$A_i = \mu_i,$$

donde  $\mu_i = t_i \lambda_i = t_i \exp\{\mathbf{x}_i^T \beta\}$ , mientras que los elementos del vector  $\mathbf{z}$  serán:  $z_i = \eta_i + \frac{y_i - \mu_i}{\mu_i} + \log(t_i)$ , con  $\eta_i = \log(t_i) + \exp\{\mathbf{x}_i^T \beta\}$ .

- Actualizar  $\gamma^{(m)}$  utilizando el algoritmo de Newton-Raphson (Hilbe (2011)), donde,

$$\mathbf{U}_\gamma(\mathbf{Q}) = \frac{\partial \mathbf{Q}(\theta | \hat{\theta})}{\partial \gamma} = \frac{\partial \mathbf{E}(l_c(\theta | \mathbf{y}_c) | \mathbf{y}_0, \hat{\theta})}{\partial \gamma} = -n(\psi(\gamma) - \log \gamma - \bar{s} + \bar{e} - 1), \quad (3.5)$$

$$\mathbf{I}_{\gamma\gamma}(\mathbf{Q}) = \frac{\partial^2 \mathbf{Q}(\theta | \hat{\theta})}{\partial \gamma^2} = \frac{\partial^2 (\mathbf{E}(l_c(\theta | \mathbf{y}_c) | \mathbf{y}_0, \hat{\theta}))}{\partial \gamma^2} = -n \left( \psi'(\gamma) - \frac{1}{\gamma} \right). \quad (3.6)$$

Con lo cual,

$$\gamma^{(m+1)} = \gamma^{(m)} - \frac{\psi(\gamma^{(m)}) - \log \gamma^{(m)} - \bar{s} + \bar{e} - 1}{\psi'(\gamma^{(m)}) - \frac{1}{\gamma^{(m)}}},$$

donde  $\psi'(\cdot)$  denota la función trigamma.

Una vez que se satisface el criterio de convergencia, se detienen las iteraciones y se regresa al paso **E**.

Para la obtención de los errores estándar asintóticos de  $\hat{\theta}$ , dado que es conocido que la mezcla Poisson-Gamma da origen a una distribución Binomial Negativa, se utilizará directamente la matriz de información observada, donde  $l(\theta)$  está dada por (2.4), luego, los elementos de  $\mathbf{I}_\theta$  son:

- $\mathbf{I}_{\beta\beta} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \beta_k \partial \beta_j} = -\sum_{i=1}^n \gamma \lambda_i \frac{(\gamma + y_i)}{(\gamma + \lambda_i)^2} x_{ik} x_{ij},$
- $\mathbf{I}_{\gamma\gamma} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \gamma^2} = \sum_{i=1}^n \left( \psi'(y_i + \gamma) - \psi'(\gamma) + \frac{1}{\gamma} - \frac{2}{(\gamma + \lambda_i)} + \frac{\gamma + y_i}{(\gamma + \lambda_i)^2} \right),$
- $\mathbf{I}_{\beta\gamma} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \beta_j \partial \gamma} = \sum_{i=1}^n \frac{\lambda_i (y_i - \lambda_i) x_{ij}}{(\gamma + \lambda_i)^2}.$

Tomando esperanza a cada término de la matriz  $\mathbf{I}_{\boldsymbol{\theta}}$  se obtiene:

- $\mathbf{K}_{\beta\beta} = -\mathbf{E}(\mathbf{I}_{\beta\beta}) = \gamma \mathbf{X}^T \boldsymbol{\Upsilon} \mathbf{X},$
- $\mathbf{K}_{\gamma\gamma} = -\mathbf{E}(\mathbf{I}_{\gamma\gamma}) = -\mathbf{E} \left[ \sum_{i=1}^n \left( \psi'(y_i + \gamma) - \psi'(\gamma) + \frac{1}{\gamma} - \frac{2}{(\gamma + \lambda_i)} + \frac{\gamma + y_i}{(\gamma + \lambda_i)^2} \right) \right],$
- $\mathbf{K}_{\beta\gamma} = -\mathbf{E}(\mathbf{I}_{\beta\gamma}) = \mathbf{0},$

donde  $\boldsymbol{\Upsilon}$  es una matriz diagonal cuyos elementos están dados por  $\frac{\lambda_i}{\gamma + \lambda_i}$ . Debido a que  $\mathbf{K}_{\beta\gamma} = \mathbf{0}$ ,  $\beta$  y  $\gamma$  son ortogonales, luego la matriz  $\mathbf{K}_{\boldsymbol{\theta}}$  es diagonal por bloque; por lo tanto, para determinar los errores estándar asintóticos de  $\hat{\beta}$  se utilizará  $\mathbf{K}_{\beta\beta}^{-1}$ . Ahora bien, dada la dificultad para determinar  $\mathbf{E}(\psi'(y_i + \gamma))$ , se usará  $(-\mathbf{I}_{\gamma\gamma})^{-1}$  para obtener el correspondiente error estándar asintótico de  $\gamma$ . Note que una manera de resolver dicha dificultad es utilizar el método de Monte Carlo, el cual consiste, a grandes rasgos, en simular una muestra aleatoria de tamaño  $n$  (grande) proveniente de la distribución de  $y_i$ , utilizando los EMV de  $\boldsymbol{\theta}$  ya encontrados. Es decir, sea  $\mathbf{y} = (y_1, \dots, y_n)^T$  una muestra aleatoria de tamaño  $n$ , dado que  $\psi'(\cdot)$  es una función medible y  $\mathbf{E}(\mathbf{y}) < \infty$ , se tiene por Ley de los Grandes Números que,

$$\frac{\sum_{i=1}^n y_i}{n} \xrightarrow{a.s.} \mathbf{E}(y),$$

entonces,

$$\frac{\sum_{i=1}^n \psi'(y_i + \gamma)}{n} \xrightarrow{a.s.} \mathbf{E}(\psi'(y + \gamma)).$$



$\mathbf{K}_{\gamma\gamma}$  puede ser calculado (aproximadamente) además a través de la definición de esperanza, en cuyo caso,

$$\begin{aligned} \mathbf{K}_{\gamma\gamma} &= -\sum_{i=1}^n \mathbf{E} \left( \psi'(y_i + \gamma) - \psi'(\gamma) + \frac{1}{\gamma} - \frac{2}{(\gamma + \lambda_i)} + \frac{\gamma + y_i}{(\gamma + \lambda_i)^2} \right) \\ &= -\sum_{i=1}^n \left[ \sum_{y_i^*=0}^{\infty} \left( \psi'(y_i^* + \gamma) - \psi'(\gamma) + \frac{1}{\gamma} - \frac{2}{(\gamma + \lambda_i)} + \frac{\gamma + y_i^*}{(\gamma + \lambda_i)^2} \right) P(Y_i = y_i^*) \right], \end{aligned} \quad (3.7)$$

donde  $P(Y_i = y_i^*) = P(y_i^* | \lambda_i, \gamma)$  corresponde a la probabilidad para cada  $y_i^* \in \{0, 1, \dots\}$ , calculadas a partir de la distribución Binomial Negativa dada en (2.3), asumiendo que  $E(t_i) = 1$ .

## 3.2. Diagnóstico de Influencia para el modelo Binomial Negativo

En esta sección se presentan las expresiones que permiten detectar potenciales observaciones influyentes, tanto del punto de vista global como local para el modelo Binomial Negativo a través de la  $Q$ -función.

### 3.2.1. Influencia global

Para el modelo de regresión Binomial Negativo, como mezcla Poisson-Gamma dado en (2.3), bajo el supuesto que  $E(t_i) = 1$ , condición impuesta por Karlis (2001) para evitar problemas de identificabilidad, se tiene que la primera derivada de la función  $Q_{[i]}(\theta | \hat{\theta})$  con respecto a  $\gamma$  y  $\beta$ , evaluado en  $\hat{\theta}$ , están dadas respectivamente por,

$$\begin{aligned} \frac{\partial Q_{[i]}(\hat{\theta} | \hat{\theta})}{\partial \gamma} &= -\sum_{j \neq i} (\psi(\gamma) - \log \gamma - s_j + e_j - 1), \\ \frac{\partial Q_{[i]}(\hat{\theta} | \hat{\theta})}{\partial \beta} &= \sum_{j \neq i} (y_j x_j - c_j x_j \lambda_j) = \mathbf{X}_{[i]}^T (\mathbf{y}_{[i]} - \mathbf{e}_{[i]}^*), \end{aligned}$$

donde  $\mathbf{e}_{[i]}^* = \mathbf{e}_{[i]} \odot \lambda_{[i]}$  es un vector de longitud  $(n - 1)$ , con  $\odot$  denotando el producto de Hadamard.

Mientras que las derivadas correspondientes respecto a  $\gamma$  de la función  $Q(\theta|\hat{\theta})$  están dadas por (3.5) y (3.6) y las derivadas cruzadas por,

$$\frac{\partial^2 Q(\theta|\hat{\theta})}{\partial \gamma \partial \beta^T} = \mathbf{0}.$$

Adicionalmente, el vector score y la matriz hessiana para  $\beta$  obtenidos a partir de la  $Q$ -función son,

$$\mathbf{U}_\beta(Q) = \frac{\partial Q(\theta|\hat{\theta})}{\partial \beta} = \frac{\partial E(l_c(\theta|\mathbf{y}_c)|\mathbf{y}_0, \hat{\theta})}{\partial \beta} = \mathbf{X}^T(\mathbf{y} - \mathbf{e}^*),$$

donde  $\mathbf{e}^* = \mathbf{e} \odot \lambda$ , es un vector de longitud  $n$  y,

$$\mathbf{I}_{\beta\beta}(Q) = \frac{\partial^2 Q(\theta|\hat{\theta})}{\partial \beta \partial \beta^T} = \frac{\partial^2 E(l_c(\theta|\mathbf{y}_c)|\mathbf{y}_0, \hat{\theta})}{\partial \beta \partial \beta^T} = -\mathbf{X}^T \mathbf{J}_1 \mathbf{X},$$

donde  $\mathbf{J}_1$  corresponde a una matriz diagonal de dimensión  $n \times n$ , cuyos elementos son  $e_i^*$ ,  $i = 1, \dots, n$ . Luego,

$$\ddot{Q}(\theta|\hat{\theta}) = \begin{pmatrix} -\mathbf{X}^T \mathbf{J}_1 \mathbf{X} & \mathbf{0} \\ \mathbf{0}^T & -n \left( \psi'(\gamma) - \frac{1}{\gamma} \right) \end{pmatrix}; \text{ evaluada en } \theta = \hat{\theta}.$$

Note que,  $\ddot{Q}(\hat{\theta}|\hat{\theta})$  corresponde a la matriz hessiana de  $\theta$  en el contexto del algoritmo EM.

Sustituyendo estas expresiones en (2.44) se obtiene la relación entre los parámetros estimados de los datos completos y los datos sin el  $i$ -ésimo caso del modelo de regresión Binomial Negativo con respecto a  $\gamma$  y a  $\beta$  a un paso, las que están dadas respectivamente por,

$$\hat{\gamma}_{[i]}^{(1)} = \hat{\gamma} + \left\{ n \left( \psi'(\hat{\gamma}) - \frac{1}{\hat{\gamma}} \right) \right\}^{-1} \left\{ - \sum_{j \neq i} (\psi(\hat{\gamma}) - \log \hat{\gamma} - s_j + e_j - 1) \right\}, \quad (3.8)$$

$$\hat{\beta}_{[i]}^{(1)} = \hat{\beta} + (\mathbf{X}^T \mathbf{J}_1 \mathbf{X})^{-1} \{ \mathbf{X}_{[i]}^T (\mathbf{y}_{[i]} - \mathbf{e}_{[i]}^*) \}. \quad (3.9)$$

Para este modelo se pueden evaluar las observaciones influyentes a través de la distancia de Cook generalizada a un paso de acuerdo con (2.45) para  $\gamma$  y  $\beta$ ,

$$GD_i^{(1)}(\gamma) = \left\{ n \left( \psi'(\gamma) - \frac{1}{\gamma} \right) \right\}^{-1} \left\{ \sum_{j \neq i} (\psi(\gamma) - \log \gamma - s_j + e_j - 1) \right\}^2 \quad \text{y}$$

$$GD_i^{(1)}(\boldsymbol{\beta}) = (\mathbf{y}_{[i]} - \mathbf{e}_{[i]}^*)^T \mathbf{X}_{[i]} (\mathbf{X}^T \mathbf{J}_1 \mathbf{X})^{-1} \mathbf{X}_{[i]}^T (\mathbf{y}_{[i]} - \mathbf{e}_{[i]}^*).$$

Ambas expresiones están evaluadas en el estimador máximo verosímil de  $\boldsymbol{\theta}$ . Luego, la distancia de Cook generalizada total quedará expresada como,

$$GD_i^{(1)} = GD_i^{(1)}(\gamma) + GD_i^{(1)}(\boldsymbol{\beta}), \text{ evaluada en } \boldsymbol{\theta} = \hat{\boldsymbol{\theta}}.$$

Para calcular la  $Q$ -distancia a un paso dada en (2.46) para el modelo Binomial Negativo se deben considerar las estimaciones de los parámetros a un paso sin la  $i$ -ésima observación, dadas por (3.8) y (3.9).

### 3.2.2. Influencia local

A continuación se presentan las expresiones que permitirán analizar los cambios sobre EMV al aplicar pequeñas perturbaciones, en particular a los predictores del modelo mezcla Poisson-Gamma. De dicho modelo se tiene que,

$$f(y_{ic}) = f(y_i, t_i) = \underbrace{P(y_i | t_i, \lambda_i)}_{\text{Poisson}} \underbrace{f(t_i | \gamma)}_{\text{Gamma}}.$$

Luego, la densidad conjunta bajo el modelo perturbado  $\mathcal{M}$ , se puede escribir como,

$$f(\mathbf{y}_c) = \prod_{i=1}^n f(y_{ic}, \boldsymbol{\theta}, \omega_i), \quad (3.10)$$

donde  $f(y_{ic}, \boldsymbol{\theta}, \omega_i)$  tienen la misma forma funcional  $\forall i = 1, \dots, n$ . A partir de (3.10) la función de log-verosimilitud de los datos completos estará dada por,

$$\begin{aligned} l_c(\boldsymbol{\theta} | \boldsymbol{\omega}, \mathbf{y}_c) &= \sum_{i=1}^n \log f(y_{ic}, \boldsymbol{\theta}, \omega_i) \\ &= \sum_{i=1}^n [\log P(y_{i\omega} | t_i, \boldsymbol{\beta}) + \log f(t_i | \gamma)]. \end{aligned} \quad (3.11)$$

De acuerdo con (3.11) la función de log-verosimilitud del modelo completo bajo el esquema de perturbación aditiva del predictor estará dada por,

$$\begin{aligned} l_c(\boldsymbol{\theta} | \boldsymbol{\omega}, \mathbf{y}_c) &= \sum_{i=1}^n \{-t_i \lambda_{i\omega} + y_i \log \lambda_{i\omega} + y_i \log t_i + (\gamma - 1) \log t_i - \\ &\quad \log y_i! - \gamma t_i + \gamma \log \gamma - \log \Gamma(\gamma)\}, \end{aligned} \quad (3.12)$$

donde  $\lambda_{i\omega} = \exp\{\mathbf{x}_{i\omega}^T \boldsymbol{\beta}\} = \exp\{\mathbf{x}_i^T \boldsymbol{\beta} + \omega_i \beta_j^*\}$ , con  $j^*$  representando al  $j^*$ -ésimo predictor perturbado y haciendo  $\boldsymbol{\omega} = \boldsymbol{\omega}_0 = (0, 0, \dots, 0)^T$  el modelo perturbado se reduce al modelo original, es decir,  $l_c(\boldsymbol{\theta}|\mathbf{y}_c) = l_c(\boldsymbol{\theta}|\boldsymbol{\omega}_0, \mathbf{y}_c)$ .

Al perturbar de manera aditiva a un predictor se tiene que la matriz  $\Delta_{\boldsymbol{\omega}_0}$  tiene la forma,

$$\Delta_{\boldsymbol{\omega}_0} = \begin{pmatrix} \boldsymbol{\tau}_{\boldsymbol{\beta}}^T \\ \boldsymbol{\tau}_{\boldsymbol{\gamma}}^T \end{pmatrix}_{q \times n}, \quad (3.13)$$

donde  $\boldsymbol{\tau}_{\boldsymbol{\gamma}}^T = \mathbf{0}$  corresponde a un vector de longitud  $n$ ; mientras que  $\boldsymbol{\tau}_{\boldsymbol{\beta}}^T$  es una matriz de dimensiones  $(p+1) \times n$ , cuyo  $i$ -ésimo elemento con respecto al parámetro  $\beta_j$ , con  $j = 0, \dots, p$  está dado por,

$$\tau_{\beta_{ij}} = \begin{cases} -e_i \lambda_{i\omega} x_{ij} \beta_{j^*} & j \neq j^*; \\ -e_i \lambda_{i\omega} (x_{ij^*} + \omega_i) \beta_{j^*} - e_i \lambda_{i\omega} + y_i, & j = j^*. \end{cases}$$

para  $i = 1, \dots, n$  y todas las expresiones son evaluadas en  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$  y  $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ .

De acuerdo con lo expuesto la Sección 2.6.3, un esquema de perturbación es adecuado si  $\mathbf{G}(\boldsymbol{\omega}_0) = c\mathbf{I}_n$ , a continuación se analiza si para el modelo de mezcla Poisson-Gamma la perturbación aditiva de uno de los predictores es adecuada.

A partir de (3.12), se tiene que,

$$\begin{aligned} \frac{\partial l_c(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)}{\partial \omega_i} &= -t_i \lambda_{i\omega} \beta_j^* + y_i \beta_j^*, \\ \frac{\partial^2 l_c(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)}{\partial \omega_i^2} &= -t_i \lambda_{i\omega} \beta_j^{*2}, \\ \frac{\partial^2 l_c(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)}{\partial \omega_i \partial \omega_r} &= 0 \quad i \neq r. \end{aligned}$$

Entonces los elementos de la matriz  $\mathbf{G}$  son,

$$\begin{aligned} G_{ii}(\boldsymbol{\omega}) &= -\mathbf{E} \left( \frac{\partial^2 l_c(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)}{\partial \omega_i^2} \right) = \mathbf{E}(t_i) \lambda_{i\omega} \beta_j^{*2} = \beta_j^{*2} \lambda_{i\omega}, \\ G_{ir}(\boldsymbol{\omega}) &= \mathbf{E} \left( \frac{\partial l_c(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)}{\partial \omega_i} \frac{\partial l_c(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)}{\partial \omega_r} \right) = 0, \quad i \neq r, \end{aligned}$$

dado que  $t_i \sim \text{Gamma}(\gamma, \gamma)$ . Por lo tanto, la matriz de información de Fisher esperada

con respecto a  $\omega$  bajo este esquema de perturbación es,

$$\mathbf{G}(\omega) = \beta_j^{*2} \begin{pmatrix} \lambda_{1\omega} & 0 & \cdots & 0 \\ 0 & \lambda_{2\omega} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_{n\omega} \end{pmatrix},$$

y evaluada en  $\omega_0$  se tiene que,

$$\mathbf{G}(\omega_0) = \beta_j^{*2} \text{Diag}(\lambda_1, \dots, \lambda_n) = \beta_j^{*2} \text{Diag}(\boldsymbol{\lambda}) \neq \mathbf{cI}_n. \quad (3.14)$$

con  $\mathbf{c} = \beta_j^{*2} > 0$  y  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)^T$ , por lo cual es necesario elegir una nueva perturbación (parametrización)  $\tilde{\omega}$ , definida por,

$$\tilde{\omega} = \omega_0 + \mathbf{c}^{-1/2} \mathbf{G}(\omega_0)^{1/2} (\omega - \omega_0).$$

Reemplazando (3.14),

$$\tilde{\omega} = \omega_0 + (\beta_j^{*2})^{-1/2} (\beta_j^{*2} \text{Diag}(\boldsymbol{\lambda}))^{1/2} (\omega - \omega_0);$$

como  $\omega_0 = \mathbf{0}^T$ , se obtiene  $\tilde{\omega} = \text{Diag}(\boldsymbol{\lambda})^{1/2} \omega \implies \omega = \text{Diag}(\boldsymbol{\lambda})^{-1/2} \tilde{\omega}$ .

Por lo tanto, en el modelo perturbado (3.11) se debe definir  $x_{ij^*\omega} = x_{ij^*} + \lambda_i^{-1/2} \tilde{\omega}_i$ .

Repitiendo para verificar que este esquema de perturbación es adecuado se tiene,

$$\begin{aligned} \frac{\partial l_c(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)}{\partial \tilde{\omega}_i} &= -t_i \lambda_{i\tilde{\omega}} \lambda_i^{-1/2} \beta_j^* + y_i \beta_j^* \lambda_i^{-1/2}, \\ \frac{\partial^2 l_c(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)}{\partial \tilde{\omega}_i^2} &= -t_i \lambda_{i\tilde{\omega}} \lambda_i^{-1} \beta_j^{*2}, \\ \frac{\partial^2 l_c(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)}{\partial \tilde{\omega}_i \partial \tilde{\omega}_r} &= 0 \quad i \neq r. \end{aligned}$$

Recalculando la matriz de información de Fisher esperada,

$$\mathbf{G}_1(\tilde{\omega}) = \beta_j^{*2} \lambda_{i\tilde{\omega}} \lambda_i^{-1} \implies \mathbf{G}_1(\tilde{\omega}_0) = \beta_j^{*2} \mathbf{I}_n = \mathbf{cI}_n, \quad (3.15)$$

con  $\mathbf{c} = \beta_j^{*2} > 0$ .

En general, se puede escribir la perturbación aditiva de un predictor como,

$$\lambda_{i\omega} = \exp\{\mathbf{x}_i^T \boldsymbol{\beta} + S_i \omega_i \beta_j^*\};$$

en adelante, a partir de dicha expresión se analizará lo adecuado de la perturbación aditiva para un modelo.

En particular, según lo obtenido en (3.15),  $S_i = (\lambda_i)^{-1/2}$  para un modelo de mezcla Poisson-Gamma: Binomial Negativo, el esquema de perturbación aditiva del predictor es adecuado.

Redefiniendo la  $Q$ -función perturbada utilizando,  $\lambda_{i\omega} = \exp\{\mathbf{x}_i^T \boldsymbol{\beta} + (\lambda_i)^{-1/2} \omega_i \beta_j^*\}$ , se obtiene,

$$Q(\boldsymbol{\theta}|\boldsymbol{\omega}, \boldsymbol{\theta}^{(m)}) = \sum_{i=1}^n \{-e_i \lambda_{i\omega} + y_i \log \lambda_{i\omega} + y_i s_i + (\gamma - 1) s_i - \log y_i! - \gamma e_i + \gamma \log \gamma - \log \Gamma(\gamma)\}.$$

Luego, los elementos de la matriz  $\tau_{\beta}^T$ , dada en (3.13), necesaria para calcular la matriz  $\Delta_{\omega_0}$  están dados por,

$$\tau_{\beta_{ij}} = \begin{cases} -\beta_{j^*} e_i \lambda_{i\omega} x_{ij} \lambda_i^{-1/2} + \frac{1}{2} \beta_{j^*} (e_i \lambda_{i\omega} - y_i) x_{ij} \lambda_i^{-3/2} & , j \neq j^*; \\ -\beta_{j^*} e_i \lambda_{i\omega} x_{ij} \lambda_i^{-1/2} + \frac{1}{2} \beta_{j^*} (e_i \lambda_{i\omega} - y_i) x_{ij} \lambda_i^{-3/2} - (e_i \lambda_{i\omega} - y_i) \lambda_i^{-1/2} & , j = j^*. \end{cases}$$

### 3.3. Aplicación

En esta sección se presenta la ilustración de la metodología expuesta en las secciones precedentes. En primer lugar, se realiza un estudio de simulación que tiene como finalidad mostrar la importancia para el análisis de influencia local, la correcta elección del esquema de perturbación y con ello el vector que permitirá realizarla. Además, se analizan dos conjuntos de datos reales, disponibles en la literatura.

#### 3.3.1. Simulación

Con el propósito de comparar el vector de perturbación  $\boldsymbol{\omega}$  con  $\tilde{\boldsymbol{\omega}}$  bajo el esquema aditivo de perturbación del predictor, donde  $\tilde{\boldsymbol{\omega}}$  corresponde al vector de perturbación más adecuado de acuerdo con lo expuesto en la Sección 3.2.2; y verificar la importancia y eficiencia que la elección de dicho vector tiene para la detección de

potenciales observaciones influyentes, se realizó la simulación de 500 conjuntos de datos de longitud 100, con parámetros  $\theta = (\beta_0 = 3, \beta_1 = -3, \gamma = 2)^T$ , y covariable  $x \sim \text{Uniforme}(0, 2)$ , fija para cada conjunto. A cada conjunto de datos se le agregó una observación atípica, correspondiente a la número 30 (elegida de forma arbitraria), a la cual se le sumó  $\zeta$  veces la desviación estándar de la correspondiente covariable, donde  $\zeta = \{0.5, 1, 1.5, 2, 2.5, 3, 4, 5, 6\}$ . La finalidad de este estudio de simulación es comparar el número de veces que dicha observación es detectada como posible caso influyente, según el esquema de perturbación aditiva del predictor, bajo los vectores  $\omega$  y  $\tilde{\omega}$ , los resultados se encuentran en la Tabla 3.1. Adicionalmente, se contabilizó la cantidad de veces que la observación 30 era considerada potencialmente influyente de acuerdo a la  $Q$ -distancia (análisis de influencia global), resultados que también son incluidos en la Tabla siguiente.

vector	Medida	0.5	1	1.5	2	2.5	3	4	5	6
$\omega$	$ h_{max} $	0.0	0.0	0.4	2.2	3.8	6.8	12.0	18.0	21.8
	$C_i$	1.6	6.0	9.2	13.6	16.6	20.2	28.0	35.2	41.8
$\tilde{\omega}$	$ h_{max} $	0.2	10.8	47.0	78.8	93.4	96.0	97.2	100.0	100.0
	$C_i$	0.2	7.8	41.6	74.6	91.6	96.2	96.6	99.8	100.0
	$QD_i$	20.2	55.0	76.6	87.4	90.2	92.6	96.0	96.0	96.0

Tabla 3.1: Porcentaje de detección observación 30 bajo  $\omega$  y  $\tilde{\omega}$ ,  $QD_i$ .

A partir de la Tabla 3.1 se puede ver que bajo el vector de perturbación  $\omega$  la observación 30 fue detectada un porcentaje muy pequeño de veces, mientras que bajo el vector perturbación  $\tilde{\omega}$  el porcentaje de detección es considerablemente mayor, como era de esperar dado que  $\tilde{\omega}$  corresponde al vector de perturbaciones más adecuado como se mostró anteriormente. Para llevar acabo ésto se estableció que la observación sería considerada potencialmente influyente si:  $|h_{max_i}| \geq \overline{|h_{max}|} + 3ds(|h_{max}|)$ ,  $C_i \geq \bar{C} + 3ds(C)$  y  $QD_i \geq \overline{QD_i} + 3ds(QD_i)$ ,  $\forall i = 1, \dots, n$  en cada muestra generada. La Figura 3.1 muestra la dispersión de dos de estos conjuntos de datos, 1 (a) y 451

(b) (elegidos de manera arbitraria, de entre los 500 conjuntos de datos simulados); en ellos se observa que para un valor  $\zeta = 4$ , la observación 30 está alejada del comportamiento de las restantes observaciones (en ambos conjuntos de datos), por lo tanto, se espera que ella sea detectada como posible caso influyente, con ambos vectores de perturbación.

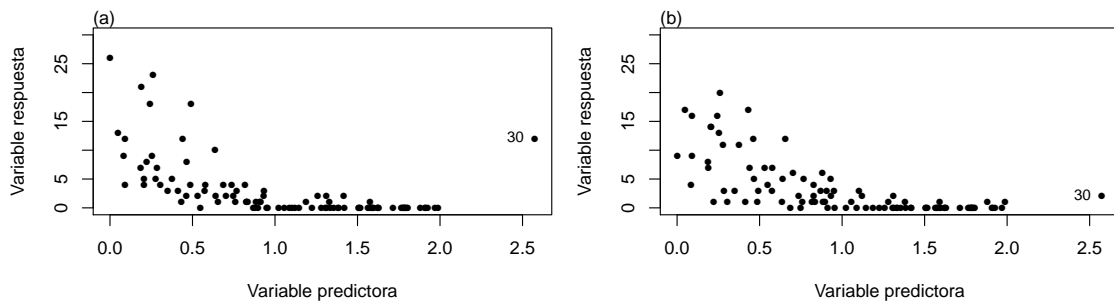


Figura 3.1: Gráficos de dispersión conjunto 1 (a) y conjunto 451 (b).

En la Figura 3.2(a) y (c) se presentan los gráficos índice, correspondiente a la dirección unitaria  $|\mathbf{h}_{max}|$ , bajo el vector de perturbaciones  $\omega$  y  $\tilde{\omega}$ , respectivamente; se observa que en (a) la observación 30 no es detectada como potencial valor influyente, pero si la observación 74, que corresponde a aquella que toma mayor valor en la variable respuesta, mientras que en (c) sí es detectada la observación 30, como era de esperar, destacando de manera considerable por sobre las restantes observaciones. Así también, en la Figura 3.2(b) y (d) se muestran los gráficos índice de la curvatura local total,  $C_i$ , considerando los dos vectores de perturbación ( $\omega$  y  $\tilde{\omega}$ ), nuevamente el patrón de detección de la observación 30 es el mismo.

La Figura 3.3(a)-(c) muestra los gráficos correspondientes de  $|\mathbf{h}_{max}|$  y curvatura local total,  $C_i$  (b)-(d), bajo el vector de perturbación  $\omega$  y  $\tilde{\omega}$ , respectivamente, para el conjunto de datos 451; de las figuras (a) y (b) se puede concluir la observación 30 no es detectada como posible caso influyente, sin embargo, y como se esperaba el vector



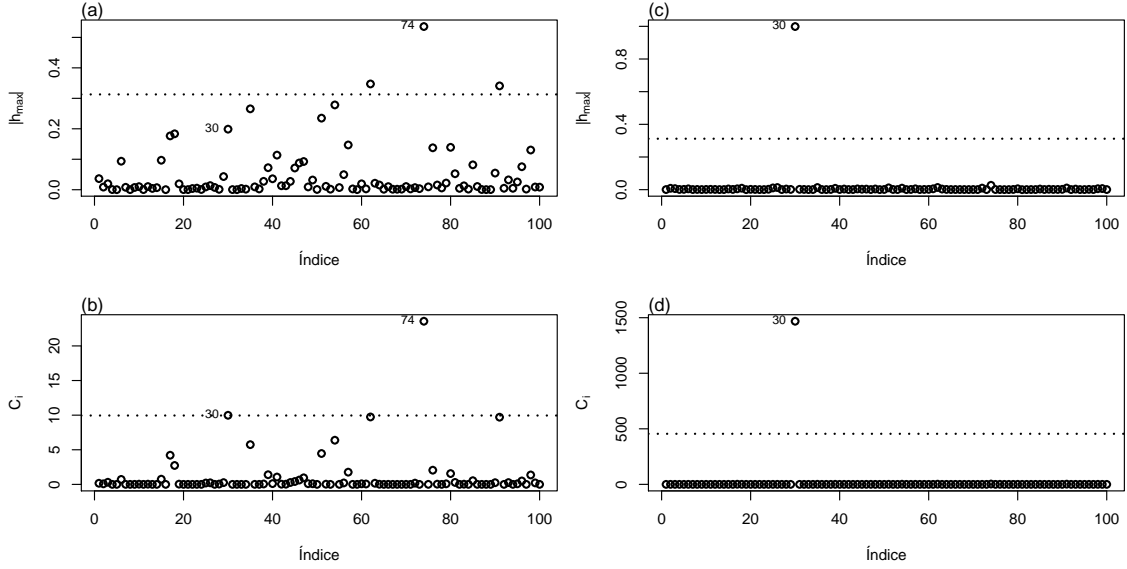


Figura 3.2: Valores de  $|h_{max}|$  (a) y curvatura local total,  $C_i$  (b) bajo el vector de perturbación  $\omega$  y  $\tilde{\omega}$  (c), (d) para el conjunto de datos 1.

$\tilde{\omega}$  si la detectó (c)-(d).

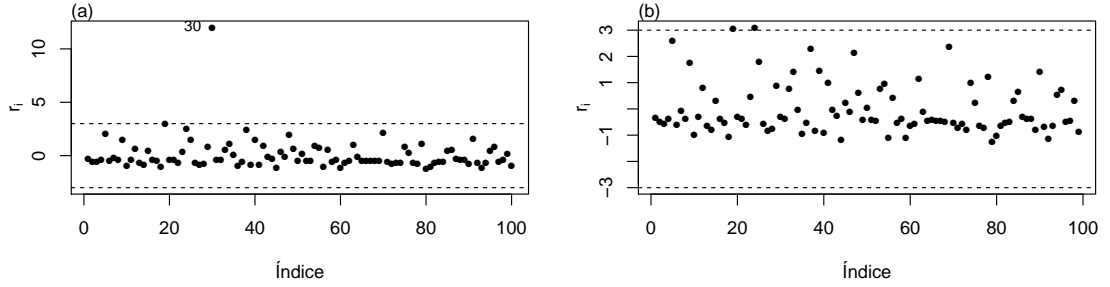


Figura 3.3: Valores de  $|h_{max}|$  y curvatura local total,  $C_i$ , bajo el vector de perturbación  $\omega$  (a)-(b) y  $\tilde{\omega}$  (c)-(d) para el conjunto de datos 451.

La Tabla 3.2 muestra las estimaciones respectivas por conjunto de datos con y sin la observación 30, además de la razón de cambio ( $RC$ ) en las estimaciones, que en general está dada por,

$$RC_j = \left| \frac{\hat{\theta}_j - \hat{\theta}_{j(I)}}{\hat{\theta}_j} \right| \times 100, \quad \forall j, \quad (3.16)$$

donde  $\hat{\theta}_{j(I)}$  denota el EMV cuando se elimina un subconjunto de datos.

Adicionalmente, a lo presentado en la Tabla 3.2, se estudió la significancia de los

Conjunto	Parámetro	Estimación con caso 30	Estimación sin caso 30	RC
1	$\beta_0$	2.0964 (0.2297)	3.0361 (0.1869)	44.82 %
	$\beta_1$	-1.4591 (0.2487)	-3.1700 (0.2820)	117.26 %
	$\gamma$	0.7333 (0.2196)	2.6749 (0.8607)	264.78 %
	$l$	-193.0602	-148.6691	
451	$\beta_0$	2.8750 (0.1836)	2.9999 (0.1919)	4.34 %
	$\beta_1$	-2.5247 (0.2381)	-2.7542 (0.2562)	9.09 %
	$\gamma$	2.2853 (0.7070)	2.2901 (0.6976)	0.21 %
	$l$	-175.7843	-167.7108	

Tabla 3.2: Resumen del ajuste del modelo Binomial Negativo para los conjuntos de datos 1 y 451, con y sin la observación 30. Las desviaciones estándar están expresadas entre paréntesis.

parámetros y ésta nunca cambió, pese a esto se puede observar como en el conjunto de datos 1 la razón de cambio de las estimaciones es muy grande, además este conjunto presenta un aumento considerable en la función de log-verosimilitud al extraer la observación 30, esto hace pensar que la observación 30 si provoca alguna influencia en los EMV de los parámetros, a pesar que la significancia de éstos nunca cambió. En el conjunto de datos 451, los cambios son menores al 10 % al extraer el caso 30 y llevarían a concluir que la observación 30 no es influyente, a pesar de haber sido detectada de manera gráfica como una potencial observación influyente.

La Figura 3.4(a) muestra el comportamiento de los residuos de Pearson para el conjunto de datos 1, se puede ver que el residuo de la observación 30 sobresale de los restantes. Cuando dicha observación es extraída del conjunto de datos (Figura 3.4(b)) se puede ver que los residuos están en un 99 % dentro del criterio usual  $|r_i| < 3$ .

Situación análoga a la anterior se puede observar en la Figura 3.5.

Finalmente, en la Figura 3.6 (a), se puede observar que para el conjunto 1 existen notorias diferencias en las curvas ajustadas, permitiendo afirmar que la curva ajustada sin la observación 30 describe mejor el fenómeno de interés, sin embargo, para el conjunto 451 (b), esa diferencia es prácticamente nula.

A partir de este estudio de simulación se puede concluir que:

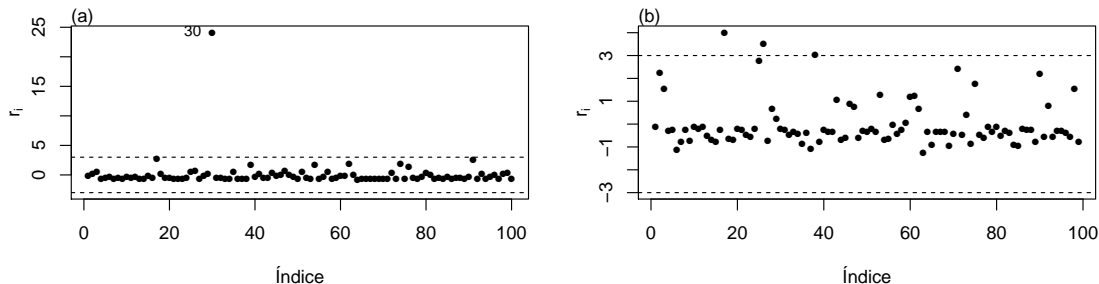


Figura 3.4: Residuos de Pearson para el conjunto de datos 1 con la observación 30 (izquierda) y sin ella (derecha).

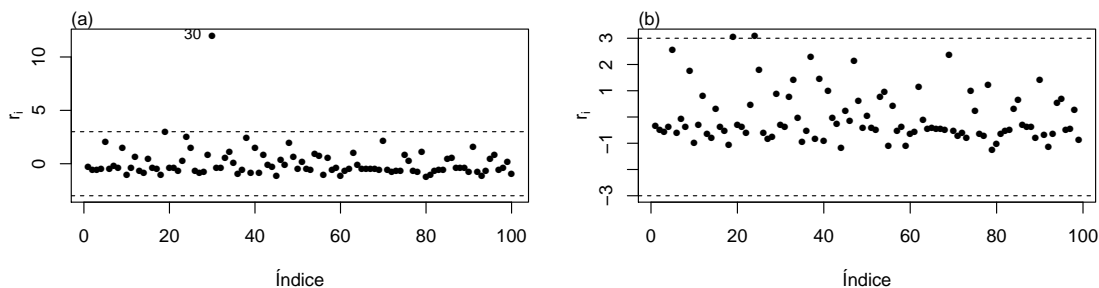


Figura 3.5: Residuos de Pearson para el conjunto de datos 451 con la observación 30 (a) y sin ella (b).

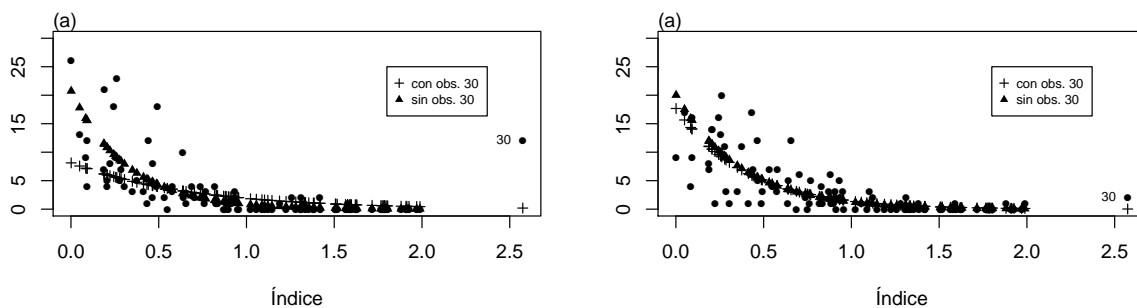


Figura 3.6: Ajuste de los datos con y sin la observación 30 para el conjunto de datos 1 (a) y para el conjunto de datos 152 (b).

- La elección del esquema de perturbación debe ser riguroso, pues permite identificar de mejor manera posibles observaciones influyentes.
- El esquema de perturbación bajo  $\tilde{\omega}$  detectó sin problemas a la observación 30, la cual debe ser analizada, y su influencia estudiada en ambos conjuntos de datos.
- De la Tabla 3.2 se puede afirmar que si bien la observación 30 no provoca cam-

bios en la significancia de los parámetros, en el caso particular del primer conjunto de datos, al extraer dicha observación el ajuste de las restantes observaciones mejoró considerablemente. Esto no sucede en el conjunto de datos 451, dado que de acuerdo con la Figura 3.1(b) la observación 30 sigue la forma natural (evolución) del resto de las observaciones.

### 3.3.2. Aplicación a datos reales: Cangrejos herradura

Los datos corresponden a una muestra de 173 cangrejos hembras del tipo herradura, científicamente conocidos como *Limulus polyphemus*, cada cangrejo herradura tenía un cangrejo macho en su nido, la base de datos se encuentra disponible en Agresti (1996), pág. 127, Tabla 4.3 . La variable de respuesta  $y_i$  corresponde al número de machos, llamados *individuos satélites*, que vivían cerca de ella y la perseguían, se desea estudiar si el ancho del caparazón de la hembra (*width*) expresada en centímetros, puede explicar el número de machos a su alrededor o cangrejos "satélites".

La Figura 3.7 muestra el comportamiento de la variable respuesta con respecto a la variable explicativa *width*, en ella se observa la presencia de al menos 4 observaciones (48, 101, 165 y 167) que están más alejadas de la nube de puntos.

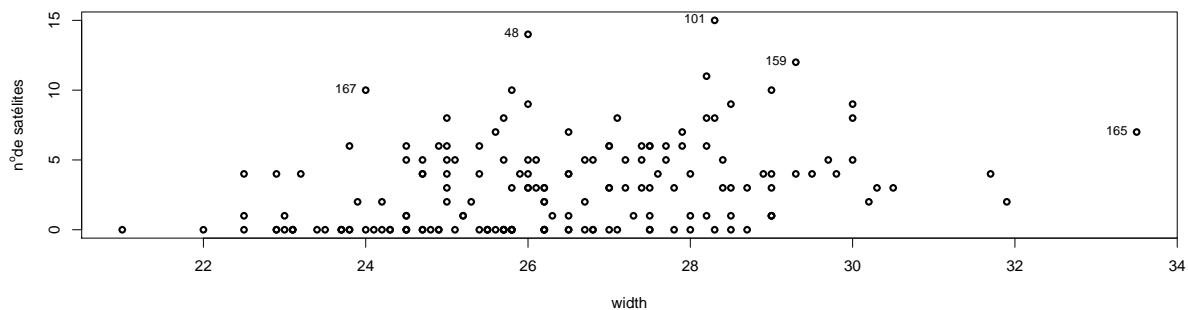


Figura 3.7: Gráfico de dispersión del ancho del caparazón *width* de los cangrejos versus el número de cangrejos satélites.

Dado que la variable de respuesta corresponde a datos de conteo, se propone el uso

del modelo Poisson y el modelo Binomial Negativo. La Tabla 3.3 muestra el resumen del ajuste de los modelos, en la cual se puede ver la estimación de los parámetros para el modelo Binomial Negativo a través del algoritmo EM. Para la estimación de los errores estándar asintóticos se utilizó  $(-\mathbf{I}_\theta)^{-1}$ , debido a que los parámetros son ortogonales; sin embargo, también se calculó  $\mathbf{K}_\theta^{-1}$ , de acuerdo con (3.7), ambas evaluadas en  $\theta = \hat{\theta}$ , en cuyo caso el error estándar asintótico de  $\hat{\gamma}$  es 0.1512. En cuanto al ajuste de los modelos, se puede observar que el modelo Binomial Negativo presenta mayor log-verosimilitud ( $l$ ), menor  $AIC$  y  $AIC_c$  ( $AIC$  corregido), con respecto al modelo Poisson. Adicionalmente, dado que en el modelo Poisson se debe cumplir que la media y la varianza sean iguales, se evaluó el test de multiplicadores de Lagrange para probar dicha condición, Hilbe (2011), dado por,

$$LM = \frac{\left( \sum_{i=1}^n \hat{y}_i^2 - \sum_{i=1}^n \hat{y}_i \right)^2}{2 \sum_{i=1}^n \hat{y}_i^2} \sim \chi^2(1),$$

donde la hipótesis nula indica que el modelo Poisson no presenta sobredispersión, es decir,  $H_0 : \gamma^{-1} = 0$ , el rechazo de  $H_0$  implica que el modelo de Poisson no es adecuado para describir los datos. Para el conjunto de datos estudiado, el valor calculado es  $LM_c = 413.3694$ , con un valor-p asociado  $\ll 0.05$ , dando claros indicios de sobredispersión en los datos, lo cual apoya la elección del modelo de regresión Binomial Negativo como el más adecuado para esta aplicación, lo que puede justificar que el modelo Poisson presenta menor  $l$  y mayor  $AIC$ .

Adicionalmente, se ajustó el modelo usando la maximización directa de la log-verosimilitud de los datos observados, a través del uso de la función `glm.nb` disponible en el software R, este tipo de estimación arrojó, como era de esperar, valores muy próximos a los presentados en la Tabla 3.3.

En la Figura 3.8 se observa el comportamiento de las estimaciones de los paráme-

Modelo	$\hat{\theta}$	ds	$q$	$l$	$AIC$	$AIC_c$
Poisson	-3.3048 0.1640	0.5422 0.0200	2	-461.5881	927.1762	927.2468
Binomial Negativo(EM)	-4.0529 0.1921 0.9046	1.1714 0.0441 0.1612	3	-375.6455	757.2910	757.4330

Tabla 3.3: Resumen del ajuste de los modelos Poisson y Binomial Negativo.

tros al eliminar el  $i$ -ésimo caso. Se puede afirmar que cuando se elimina la observación 167, la estimación de  $\beta$  difiere del comportamiento de la nube de puntos, dicha observación también destaca en el gráfico que muestra las estimaciones del parámetro  $\gamma$ , pero en menor grado, así como también se puede destacar una pequeña diferencia de la observación 48.

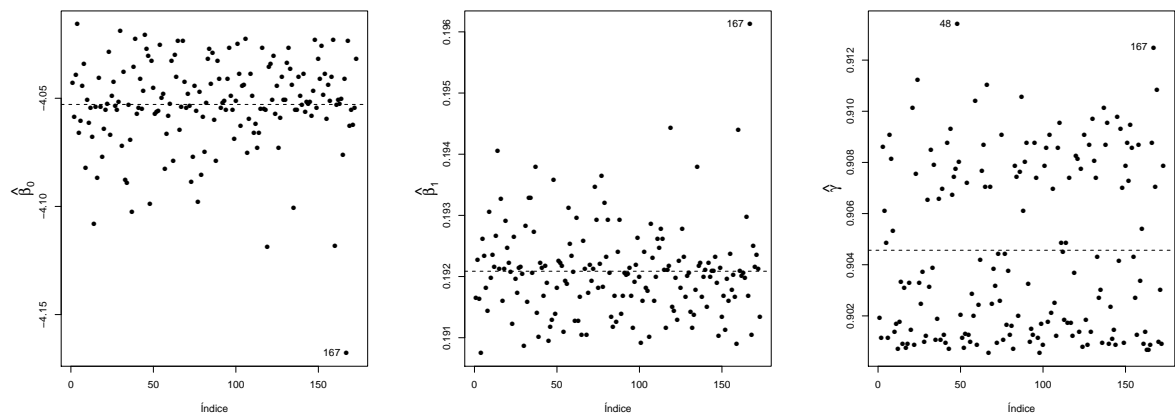


Figura 3.8: Estimaciones de los parámetros del modelo de mezcla Poisson-Gamma a través del algoritmo EM, sin la observación  $i$ -ésima.

Por otro lado, la Figura 3.9 muestra el comportamiento de los residuos de Pearson (izquierda) y de los residuos Desvío (derecha), definidos como,  $r_i^D = \text{signo}(y_i - \hat{\mu}_i) \sqrt{D_i}$ ,  $i = 1, \dots, n$ , donde  $D_i$  representa la contribución de la  $i$ -ésima observación a la deviance, posible de calcular dado que el modelo BN pertenece a la familia de modelos lineales generalizados. En dicha figura se puede ver que sólo las observaciones 48 y 167 son mayores que 3 en el primer gráfico, mientras que en el segundo de ellos ninguna observación supera dicho valor.

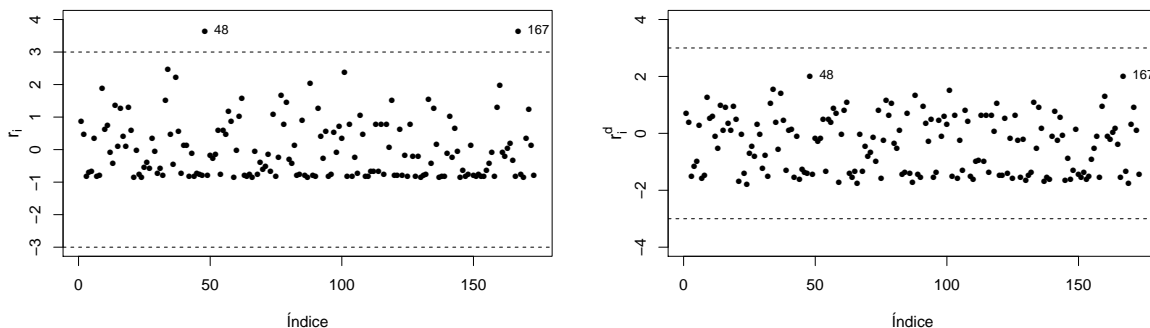


Figura 3.9: Residuos de Pearson (izquierda) y residuos Desvío (derecha).

La Figura 3.10 muestra los resultados obtenidos del análisis de influencia global para el modelo BN a partir del enfoque de la  $Q$ -función, dichas figuras corresponden a la distancia de Cook generalizada aproximada a un paso para  $\hat{\gamma}$  y  $\hat{\beta}$ , respectivamente; en ambas figuras destaca la posible influencia de las observaciones 48 y 167 en la estimación de los parámetros, aunque en el segundo de éstos la observación 48 no difiere tanto.

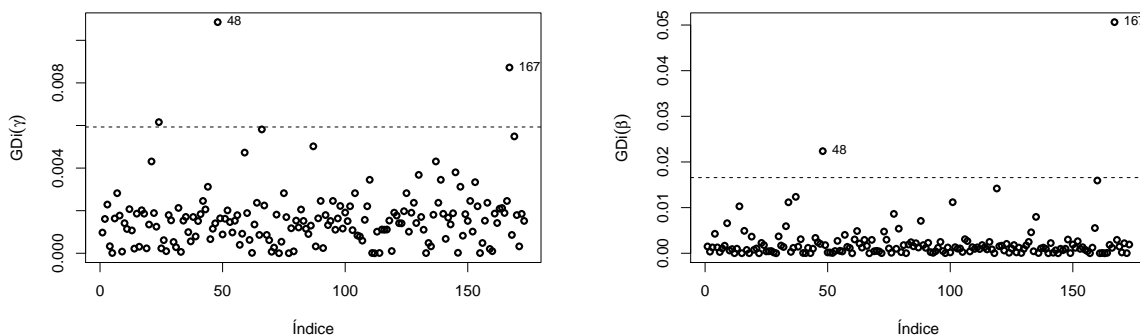


Figura 3.10: Distancia de Cook generalizada a un paso para  $\hat{\gamma}$  y  $\hat{\beta}$  respectivamente.

La Figura 3.11 muestra la distancia de Cook generalizada a un paso para  $\hat{\theta}$ ,  $GD_i^{(1)} = GD_i^{(1)}(\hat{\gamma}) + GD_i^{(1)}(\hat{\beta})$ ,  $i = 1, \dots, n$ , en ella se observa nuevamente la posible influencia de las observaciones 48 y 167, sobresaliendo la segunda de ellas. Adicionalmente, se observa que ambas sobrepasan la línea de corte,  $GD_i^{(1)} \geq \overline{GD}^{(1)} + 3ds(GD^{(1)})$ , donde  $\overline{GD}^{(1)} = \frac{1}{n} \sum_{i=1}^n GD_i^{(1)}$  y  $ds$  corresponde a la desviación estándar del

vector  $GD^{(1)}$ . Se omitió el gráfico relativo a la  $Q$ -distancia aproximada a 1 paso, debido a que dichos valores son iguales a los anteriores. Note que las líneas de corte señaladas serán utilizadas en adelante para todo el documento.

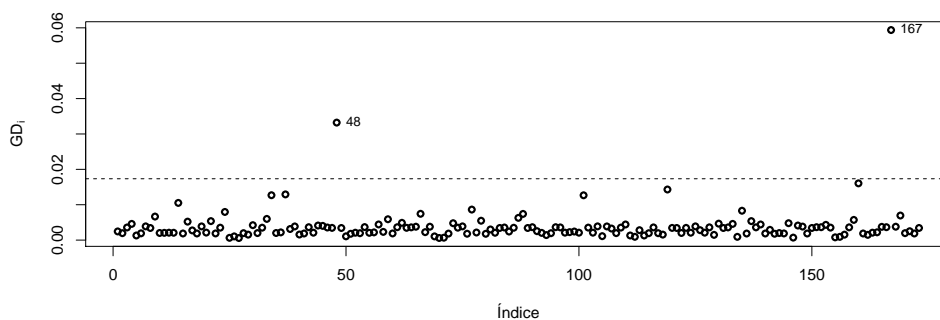


Figura 3.11: Distancia de Cook generalizada a un paso para  $\hat{\theta}$ .

En el ámbito de la influencia local se realizó el esquema de perturbación del predictor, de manera aditiva. La Figura 3.12 muestra, en su parte izquierda, la dirección de máxima curvatura local en valor absoluto,  $|\mathbf{h}_{max}|$ , en ella se observa que bajo este esquema de perturbación, en la dirección del vector  $\omega$ , las observaciones 48, 165 y 167 son las que más afectan al EMV de  $\theta$ ; dado que ellas sobrepasan la línea de corte, es decir,  $|\mathbf{h}_{max_i}| \geq \bar{\mathbf{h}} + 3ds(|\mathbf{h}_{max}|)$ , donde  $\bar{\mathbf{h}} = \frac{1}{n} \sum_{i=1}^n |\mathbf{h}_{max_i}|$  y  $ds$  es la correspondiente desviación estándar del vector  $\mathbf{h}_{max}$ . Note que la potencial influencia de la observación 48 es de menor grado. Una conclusión semejante se puede obtener de la parte derecha de la figura, la que muestra la curvatura local total en la dirección de máxima curvatura,  $C_i$ , en que la línea de corte respectiva está dada por:  $C_i \geq \bar{C} + 3ds(C)$ , donde  $\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i$  y  $ds$  la desviación estándar de los elementos del vector  $C = (C_1, C_2, \dots)^T$ .

Sin embargo, tal y como se mostró en la sección anterior, este esquema debe ser reparametrizado por el vector de perturbaciones  $\tilde{\omega}$ . Lo obtenido se encuentra en la Figura 3.13, en ella se puede observar que las observaciones que presentan una posible mayor influencia en la estimación de los parámetros son también la 48 y 167.



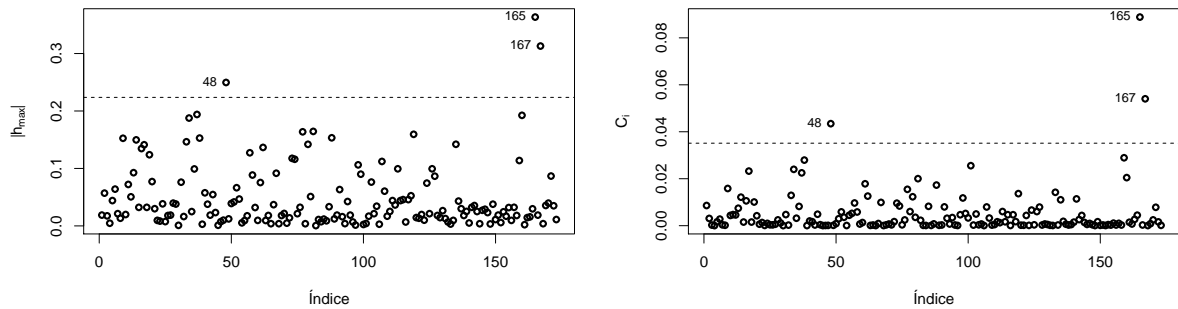


Figura 3.12: Dirección de máxima curvatura,  $(|h_{max}|)$ , y curvatura local total  $(C_i)$ , para el esquema de perturbación aditiva del predictor, bajo  $\omega$ .

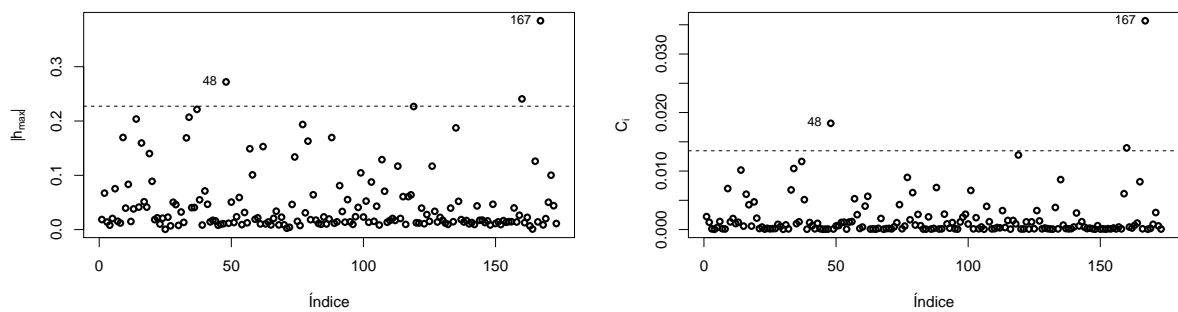


Figura 3.13: Dirección de máxima curvatura,  $(|h_{max}|)$ , y curvatura local total  $(C_i)$ , para el esquema de perturbación aditiva del predictor, bajo  $\tilde{\omega}$ .

La Figura 3.14, muestra una comparación de las curvaturas locales, usando  $C_i^*$ , la primera de ellas bajo el esquema de perturbación dado por el vector  $\omega$ , que como se discutió en (3.14) no genera una perturbación adecuada, y la segunda bajo el vector  $\tilde{\omega}$ , donde se puede apreciar que existen diferencias en la detección de posibles observaciones influyentes; sin embargo, en ambos esquemas las observaciones 48, 165 y 167 siguen destacando.

Dado que las observaciones 48, 165 y 167 parecen ser las potencialmente más influyentes bajo este esquema de perturbación, se concluye esta aplicación calculando las razones de cambio correspondientes, con el propósito de ver cuál es el efecto que tiene sobre los EMV eliminar una o más observaciones. La Tabla 3.4 presenta los resultados obtenidos para los tres parámetros involucrados en esta aplicación.

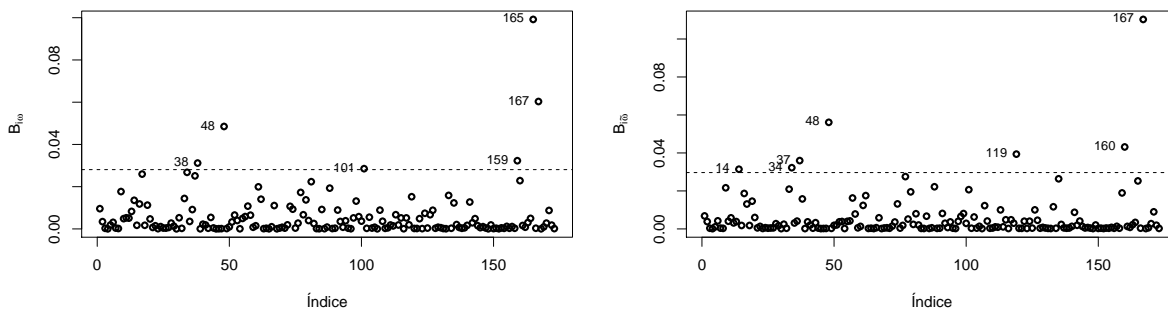


Figura 3.14: Curvaturas locales para el esquema de perturbación aditiva del predictor bajo los vectores  $\omega$  (izquierda) y  $\tilde{\omega}$  (derecha).

Obs.Eliminada	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\gamma}$	$RC_1$	$RC_2$	$RC_3$
Ninguna	-4.0529	0.1921	0.9046	-	-	-
48	-4.1594	0.1951	0.9391	2.6293	1.5819	3.8150
165	-4.1868	0.1972	0.8910	3.3052	2.6876	1.4987
167	-4.5118	0.2085	0.9345	11.3231	8.5269	3.3122
48,165	-4.2913	0.2002	0.9245	5.8834	4.2277	2.1993
48,167	-4.6295	0.2119	0.9755	14.2270	10.3019	7.8395
165,167	-4.6741	0.2147	0.9208	15.3284	11.7784	1.7957
48,165,167	-4.7902	0.2181	0.9606	18.1937	13.5213	6.1963

Tabla 3.4: Razón de cambio en los EMV al eliminar un subconjunto de observaciones en el modelo Poisson-Gamma.

En la Tabla 3.4 se puede ver que los mayores cambios relativos en las estimaciones de  $\beta$  son aquellas en que se ha eliminado la observación 167. Para complementar esto, se obtuvo el valor de la función de log-verosimilitud Binomial Negativa resultante de la mezcla en cada caso de eliminación, la cual fue mayor en presencia de todas las observaciones, y por último se estudió la significancia de los parámetros, la cual nunca cambió bajo ningún esquema de eliminación. Lo anterior permite concluir que este conjunto de datos no presenta observaciones influyentes, contrariamente a la conclusión que se podía desprender de sólo considerar los gráficos; es decir, a pesar de que la observación 167 destaca durante todo el análisis como un posible valor influyente sobre la estimación de los parámetros, ésta no resultó ser tal, pues la significancia del parámetro asociado al ancho del caparazón nunca cambió. Las razones

para éstos pueden ser variadas, la primera, el modelo propuesto, si bien es mejor que el modelo Poisson, no es adecuado para describir el fenómeno; conjuntamente, se puede deber a que no se han considerado otras variables explicativas. Sin embargo, es destacable que la inclusión de esta observación provoque cambios tan importantes en la estimación de los parámetros. Note además que esta aplicación pone nuevamente en manifiesto la importancia de la elección adecuada del vector de perturbación, por lo cual de ahora en adelante sólo se procederá directamente a su obtención.

### 3.3.3. Aplicación a datos reales: Visitas al médico

Los datos corresponden al número de visitas al médico, *NVIS*, de 1755 mujeres Alemanas de entre 25 y 64 años, durante 1987, explicada por el ingreso familiar expresado en marcos dividida por 10, *ING*, la edad en años de las pacientes, *ED*, y sus años de educación, *EDUC*. Los datos se encuentran disponibles en la librería *COUNT*, del paquete del mismo nombre del software *R*, y denominada *rwm5yr*, ella proporciona información de 6127 pacientes Alemanes de entre 25 y 64 años de edad, durante los años 1984-1988, algunos de estos pacientes presentan información durante todos los años y otros sólo en algunos de estos años, por lo cual la base tiene en total 19609 datos. Sin embargo, con el propósito de ejemplificar la metodología, se seleccionó a los pacientes del sexo femenino que tuvieran información durante el año 1987. La Figura 3.15 muestra el comportamiento de la variable respuesta por individuo, en ella se observa el alto número de mujeres que no acudió al médico durante ese año y como la frecuencia disminuye a medida que el número de visitas aumenta. Note que existen dos pacientes que acudieron al médico, 82 y 90 veces durante el año 1987, respectivamente, cantidades muy altas en relación a las restantes, dichas pacientes corresponden a las número 37 y 291.

La Figura 3.16 muestra el comportamiento del número de visitas al médico, con

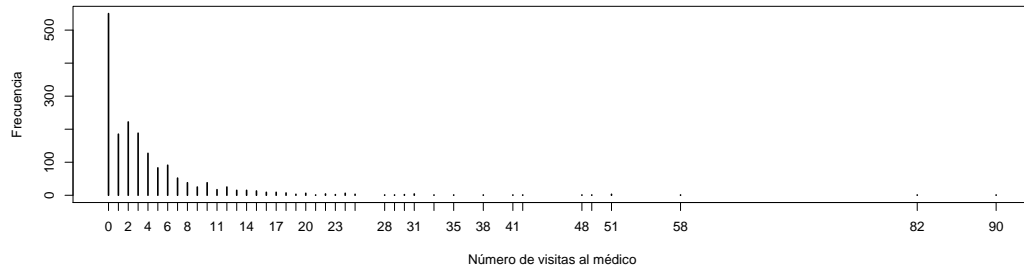


Figura 3.15: Histograma del número de visitas al médico de 1755 mujeres alemanas durante 1987.

respecto a cada una de las covariables mencionadas. En el primero (a) de ellos se ve que existen tres observaciones alejadas de la nube de puntos, dos de ellas, la 37 y 291, corresponden a observaciones con alto número de visitas al médico y bajo ingreso familiar, mientras que la tercera observación, 285, corresponde a una mujer que presenta alto ingreso familiar y alto número de visitas. Análogamente, en el segundo (b) y tercer gráfico (c) se ve que las mujeres que presentan mayor número de visitas han pasado los 50 años de edad y tienen 9 años de educación. Luego, es posible concluir que los pacientes 37 y 291 presentan un alto número de visitas al médico durante 1987, tienen bajos ingresos per-cápita, superan los 50 años y sólo acudieron a educación formal durante 9 años.

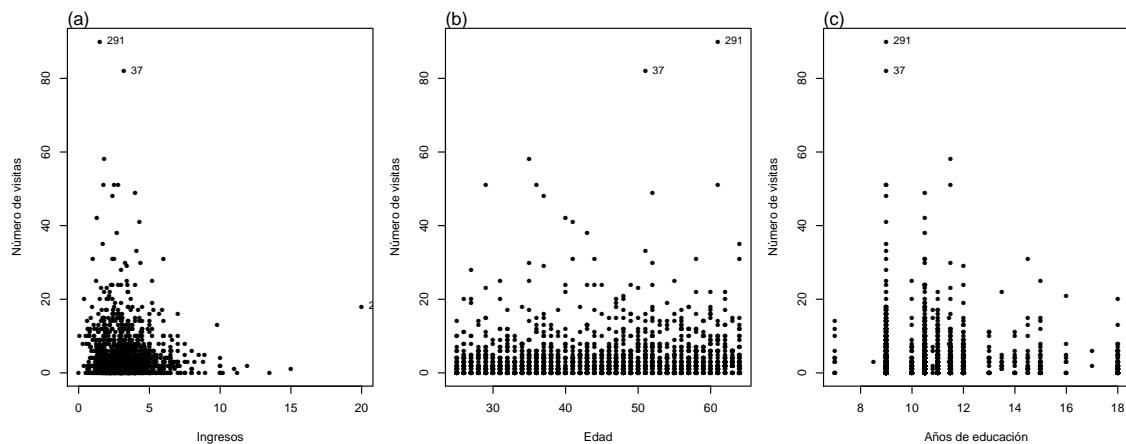


Figura 3.16: Diagrama de dispersión del número de visitas al médico versus Ingreso (a), Edad (b) y Años de estudio (c).

La Figura 3.17 muestra la distancia de Cook generalizada a un paso para  $\hat{\theta}$ , en ella destaca la observación 285, es decir, al eliminarla del conjunto del datos se produce la mayor variación en la estimación de los parámetros.

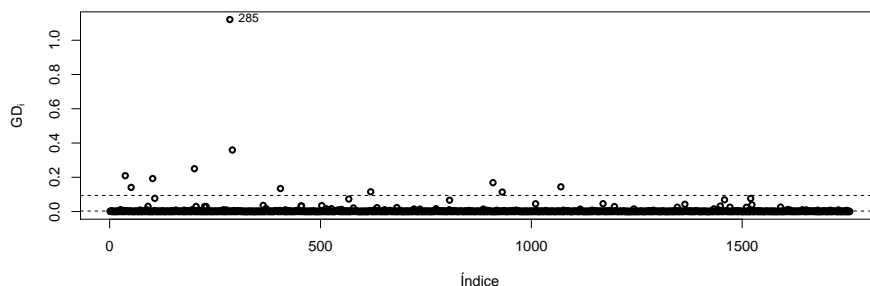


Figura 3.17: Gráfico índice de la distancia de Cook generalizada a un paso para  $\hat{\theta}$ .

En la Figura 3.18 se puede ver que la observación que presenta una mayor influencia sobre los EMV, bajo el esquema de perturbación aditiva de la covariable ingresos es la 285.

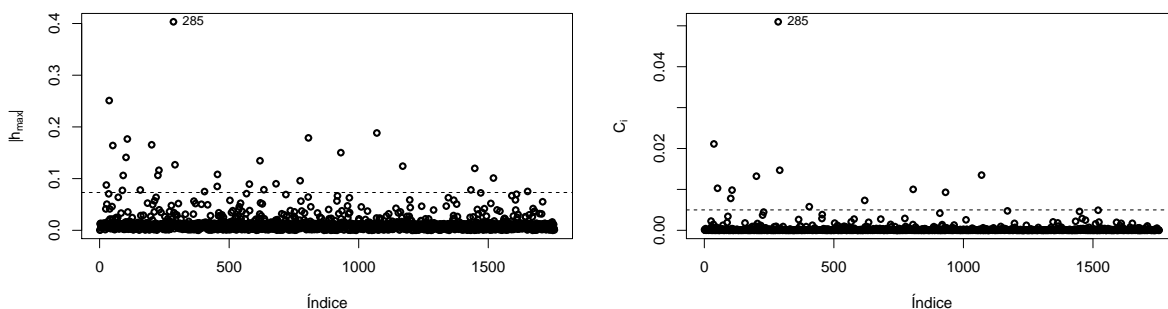


Figura 3.18: Dirección de máxima curvatura,  $(|h_{max}|)$ , y curvatura local total,  $C_i$ , para el esquema de perturbación aditiva de la variable ingreso.

De las Figuras 3.19 y 3.20 se puede concluir que la observación 285 es la que presenta mayor influencia sobre los EMV, tanto bajo el esquema de perturbación aditiva de la variable  $ED$ , como  $EDUC$ , sin embargo, en el primer caso destacan también las observaciones 102 y 201, mientras que en el segundo, las pacientes 102 y 291.

La Tabla 3.5 muestra las estimaciones obtenidas al ajustar el modelo de regresión Binomial Negativo con todas las observaciones y al eliminar las potenciales observa-

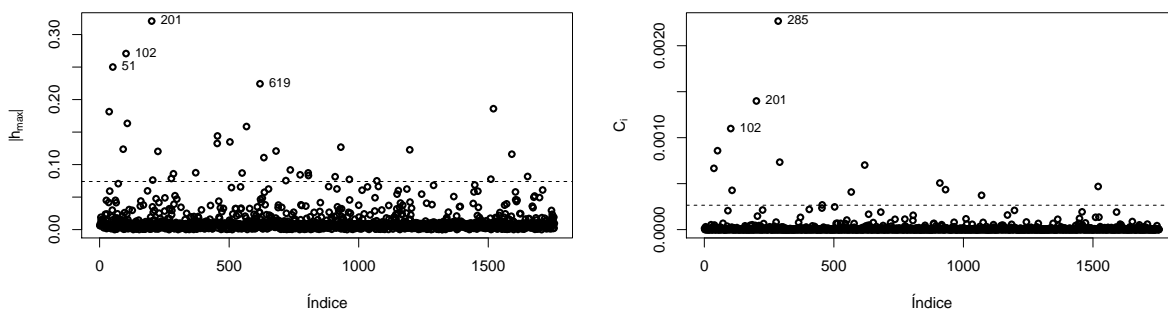


Figura 3.19: Dirección de máxima curvatura,  $(|h_{max}|)$ , y curvatura local total,  $C_i$ , para el esquema de perturbación aditiva de la variable edad.

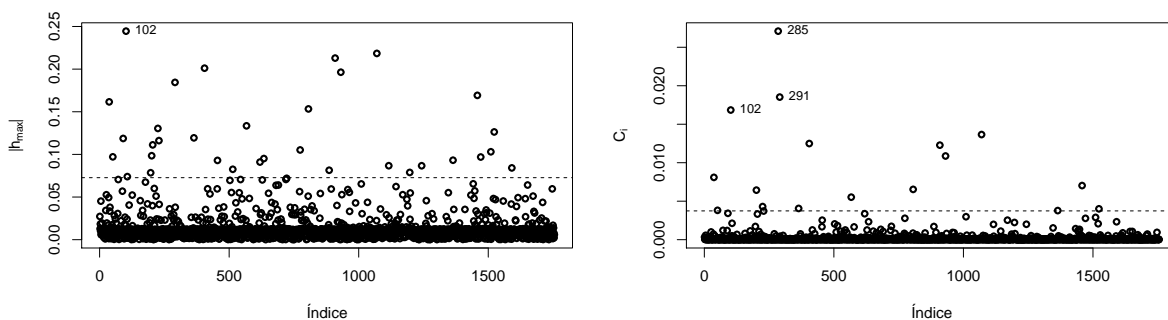


Figura 3.20: Dirección de máxima curvatura,  $(|h_{max}|)$ , y curvatura local total,  $C_i$ , para el esquema de perturbación aditiva de la variable años de estudio.

ciones influyentes, 37, 285 y 291, y la combinación de ellas.

Obs.Eliminada	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\gamma}$	$l$
Ninguna	1.4154*	-0.0355	0.0099*	-0.0355*	0.5840*	-4267.953
37	1.4142*	-0.0359	0.0098*	-0.0353	0.5917*	-4273.024
285	1.4379*	-0.0639*	0.0099*	-0.0296	0.5865*	-4289.466
291	1.4164*	-0.0309	0.0088*	-0.0335	0.5913*	-4275.595
37 y 285	1.3986*	-0.0650*	0.0096*	-0.0253	0.5944*	-4271.864
37 y 291	1.3765*	-0.0314	0.0085*	-0.0293	0.5945*	-4259.220
285 y 291	1.4381*	-0.0577*	0.0088*	-0.0279	0.5936*	-4269.885
37, 285 y 291	1.3979*	-0.0587*	0.0085*	-0.0236	0.6020*	-4257.538

Tabla 3.5: Resumen del ajuste para el modelo de regresión Binomial Negativo de los datos de visitas médicas, eliminando potenciales observaciones influyentes. \* señala que el parámetro es significativo al 5 %.

A partir de la Tabla 3.5 se puede afirmar que la función de log-verosimilitud es mayor al eliminar las tres observaciones, también es la que presenta mayores razones de cambio simultáneas, 1.24 %, 65.35 %, 14.14 %, 33.52 % y 3.08 %, respectivamente.

te; mientras que, al eliminar sólo las observaciones 37 y 285, se genera la mayor  $RC$  individual del parámetro relacionado a la variable  $ING$ , 83.10 %. Note que al eliminar las tres observaciones también cambió la significancia de los parámetros, dado que al considerar todo el conjunto de datos el parámetro correspondiente a la variable  $ING$  resultó no significativo, mientras que los restantes si lo son (indicado con un \* en la Tabla 3.5); esta conclusión cambia al eliminar el conjunto de observaciones mencionadas, donde  $\beta_1$  se vuelve significativo, pero el parámetro asociado  $EDUC$  ( $\beta_3$ ) deja de serlo. Adicionalmente, se puede ver que este cambio sucede siempre cuando en la eliminación se considera a la observación 285. Cabe señalar que también se analizó la eliminación de la observación 102, debido a que ésta destacó en los gráficos relativos a  $|h_{max}|$ , cuando la perturbación es realizada con respecto a la variable  $ED$  y  $EDUC$ , y también en el gráfico de curvatura local; sin embargo, su eliminación no provocó cambios en la significancia de los parámetros; lo mismo fue realizado con la observación 201, para la cual se obtuvo una conclusión semejante. Es de destacar que estas dos observaciones corresponden a pacientes con 58 y 51 visitas al médico durante 1987, respectivamente; sin embargo, a nivel de covariables no destacan.

### 3.3.4. Conclusión

A partir de los resultados obtenidos desde el estudio de simulación y de las aplicaciones de datos reales presentadas, puede concluirse que tanto los métodos de influencia global como los de influencia local son efectivos para detectar potenciales observaciones influyentes sobre la estimación de los parámetros, o al menos para fijar la atención del investigador en dichas observaciones, cambien o no éstas la significancia de los parámetros asociados a las covariables introducidas en el modelo.

Finalmente, dado que estas observaciones potencialmente influyentes pueden generar efectivamente modificaciones sobre las estimaciones de los parámetros, permi-

tiendo o no una mejora en el ajuste, su detección puede ayudar a la toma de decisiones y/o a la obtención de conclusiones que permiten entender mejor el fenómeno de interés. Por lo tanto, la elección del esquema de perturbación adecuado es fundamental, debido a que permite que la detección de esas potenciales observaciones influyentes sea la apropiada.

En particular en la aplicación de visitas al médico, se debe prestar una atención especial a la observación 285, pues ella provoca grandes cambios en la estimación de los parámetros del modelo (altas razones de cambio), y por sobre todo, genera un cambio en la significancia del parámetro asociado a la variable *ING*. Sin embargo, a partir de la Tabla 3.5, se observa que al eliminar esta observación por si sola del conjunto de datos, el valor de la función de log-verosimilitud disminuye, contrariamente a lo que sucede si la observación 285 es extraída del conjunto de datos de manera simultánea con la observación 37 y 291. Cabe señalar, que en ausencia de las tres observaciones mencionadas, ya sea extraídas individualmente o como combinación de ellas, el parámetro asociado a la variable *EDUC* no es significativo. Para complementar el análisis sobre la posible influencia de la observación 285, se sugiere la utilización de métodos robustos de estimación, los cuales escapan del objetivo de este trabajo.



## Capítulo 4

# Modelo de Regresión Beta-Geométrico: Waring

En diversas aplicaciones estadísticas es común que la variable de respuesta corresponda a conteo. Variadas son las posibilidades de modelos que pueden describir el comportamiento de este tipo de datos, uno de ellos corresponde a la distribución Geométrica, el cual ha sido ampliamente estudiado en aplicación de fecundidad. El trabajo pionero en esta línea es el de Weinberg and Gladen (1986), quienes consideraron que el número de ciclos menstruales necesarios hasta lograr la concepción, seguía una distribución geométrica de parámetro  $p$ , donde queda de manifiesto que este parámetro no necesariamente es constante, pues varía de pareja en pareja que desea concebir, y con ello sugieren que el modelo de regresión Geométrico no es apropiado en presencia de heterogeneidad. Asumen que el parámetro  $0 < p < 1$  sigue una distribución Beta, generando de esta forma la distribución de mezcla Beta-Geométrica, conocida más tarde como distribución Waring. En esta misma línea, Ridout and Morgan (1991), utilizan dicho modelo para separar a la muestra en mujeres fumadoras y no fumadoras, y si usaban o no la píldora anticonceptiva; el principal objetivo fue reportar los valores esperados y concluir que las mujeres que fuman reducen su fertilidad, conclusión concordante con el estudio de Baird and Wilcox (1985), quienes

realizaron un análisis epidemiológico con tal objetivo. Paul (2005) propone un test de bondad de ajuste que permite justificar el uso de la distribución Waring por sobre la distribución geométrica o viceversa, conjuntamente con otras medidas de bondad de ajuste; dicho test está basado en el estadístico de razón de verosimilitud.

En este capítulo se discute un modelo de regresión donde la variable de respuesta corresponde a datos de conteo que siguen una distribución Waring; se presenta una nueva parametrización de ella, la cual depende de la media y de un parámetro de dispersión. Este modelo de regresión es comparado con el modelo de regresión geométrico.

Este capítulo se centra en la obtención de la estimación máxima verosímil de los parámetros del modelo de regresión Waring, a través de la maximización directa de la función de log-verosimilitud resultante de la mezcla Beta-Geométrica, para lo cual se usaron las funciones `nlm` y `optim` disponibles en el software R. También, se realiza la estimación de los parámetros usando la maximización de la  $Q$ -función definida en (2.28) utilizando el algoritmo EM. Adicionalmente, se presentan las medidas de influencia global y local para este modelo, usando los dos enfoques mencionados, dichas medidas son aplicadas a un subconjunto de datos reales presentados por Hilbe (2011) y ya analizados en el capítulo anterior, denominado visitas al médico. Algunos detalles específicos son entregados en los Anexos del documento.

## **4.1. Estimación en el modelo Beta-Geométrico**

### **4.1.1. Maximización directa de la función de log-verosimilitud Waring**

Este enfoque tiene como objetivo maximizar la expresión (2.15). Luego, la función score obtenida de diferenciar (2.15) con respecto al vector de parámetros desconocidos

$\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \phi)^T$ , respectivamente, está dada por,

$$\mathbf{U}_{\boldsymbol{\theta}} = \begin{pmatrix} U_{\boldsymbol{\beta}}^T \\ U_{\phi} \end{pmatrix},$$

donde

$$\mathbf{U}_{\boldsymbol{\beta}} = \phi_2 \mathbf{X}^T \mathbf{J}_2 (\mathbf{a} - \mathbf{b}), \quad (4.1)$$

con  $\mathbf{a}$  y  $\mathbf{b}$  vectores de longitud  $n$  cuyos elementos están dados respectivamente por,  $\mathbf{a}_i = \psi(\phi_1 + \mu_i \phi_2) - \psi(y_i + \phi_1 + \mu_i \phi_2 + 1)$ ,  $\mathbf{b}_i = \psi(\mu_i \phi_2) - \psi(y_i + \mu_i \phi_2)$ , y  $\mathbf{J}_2$  una matriz diagonal, con elementos,  $J_{2i} = \frac{1}{g'(\mu_i)}$ ,  $i = 1, \dots, n$ . Ver Apéndice B, para mayores detalles.

Análogamente, derivando la log-verosimilitud con respecto a  $\phi$ ,

$$\mathbf{U}_{\phi} = \frac{\partial l(\boldsymbol{\theta})}{\partial \phi} = \sum_{i=1}^n \left[ \frac{1}{\phi} - \frac{1}{\phi - 1} - 2\mathbf{a}_i \frac{\mu_i + 1}{(\phi - 1)^2} + 2\mathbf{b}_i \frac{\mu_i}{(\phi - 1)^2} \right]. \quad (4.2)$$

La matriz de información de Fisher observada está dada por,

$$\mathbf{I} = \mathbf{I}_{\boldsymbol{\theta}} = \begin{pmatrix} \mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \mathbf{I}_{\boldsymbol{\beta}\phi} \\ \mathbf{I}_{\phi\boldsymbol{\beta}} & \mathbf{I}_{\phi\phi} \end{pmatrix}, \quad (4.3)$$

donde

$$\mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = \phi_2^2 \mathbf{X}^T \text{Diag}(\mathfrak{R}^*) \mathbf{X} + \phi_2 \mathbf{X}^T \text{Diag}(\mathfrak{R}) \mathbf{X}, \quad (4.4)$$

y los elementos de los vectores  $\mathfrak{R}$  y  $\mathfrak{R}^*$  están dados respectivamente por,

$$\mathfrak{R}_i = (\mathbf{a}_i - \mathbf{b}_i) \frac{1}{g'(\mu_i)}, \quad \mathfrak{R}_i^* = (\mathbf{a}_i^* - \mathbf{b}_i^*) \frac{1}{g'(\mu_i)^2},$$

con  $\mathbf{a}_i^* = \psi'(\phi_1 + \mu_i \phi_2) - \psi'(y_i + \phi_1 + \mu_i \phi_2 + 1)$ ,  $\mathbf{b}_i = \psi'(\mu_i \phi_2) - \psi'(y_i + \mu_i \phi_2)$  y  $\psi'(\cdot)$  corresponde a la función trigamma,

$$\mathbf{I}_{\boldsymbol{\beta}\phi} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \phi} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \phi \partial \boldsymbol{\beta}^T} = \mathbf{I}_{\phi\boldsymbol{\beta}}^T.$$

Luego,

$$\mathbf{I}_{\boldsymbol{\beta}\phi} = \frac{-2}{(\phi - 1)^2} \mathbf{X}^T \mathbf{J}_3, \quad (4.5)$$

con  $\mathbf{J}_3$  un vector de longitud  $n$ , cuyos elementos están dados por,

$$\mathbf{J}_{3i} = [(\mathbf{a}_i - \mathbf{b}_i) + \phi_2(\mathbf{a}_i^*(1 + \mu_i) - \mathbf{b}_i^*\mu_i)] \frac{1}{g'(\mu_i)},$$

$$\mathbf{I}_{\phi\phi} = \frac{\partial^2 l(\boldsymbol{\beta}, \phi)}{\partial \phi^2} = \sum_{i=1}^n \left( -\frac{1}{\phi^2} + \frac{1}{(\phi - 1)^2} + 4\frac{(\mu_i + 1)^2}{(\phi - 1)^4} \mathbf{a}_i^* + 4\frac{\mu_i + 1}{(\phi - 1)^3} \mathbf{a}_i \right. \\ \left. - 4\frac{\mu_i^2}{(\phi - 1)^4} \mathbf{b}_i^* - 4\frac{\mu_i}{(\phi - 1)^3} \mathbf{b}_i \right). \quad (4.6)$$

Para mayores detalles, Ver Apéndice B. Note que a partir de estas expresiones se puede obtener la matriz de información de Fisher esperada. Para determinar los errores estándar asintóticos de los EMV se puede usar  $(-\mathbf{I}_{\theta})^{-1}$  o  $\mathbf{K}_{\theta}^{-1}$ , para la estimación de los parámetros se propone el uso de métodos numéricos, particularmente el uso de las funciones `nlm` o `optim` disponibles en el software R.

#### 4.1.2. Maximización de la función de log-verosimilitud completa: Algoritmo EM

Una opción para obtener la estimación de  $\boldsymbol{\theta}$  es utilizar el algoritmo EM. En efecto desde (2.17) se tiene que,

$$\mathbf{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}) = \sum_{i=1}^n \{ \phi_1 \mathbf{E}(\log p_i | \mathbf{y}_0, \boldsymbol{\theta}^{(m)}) + (\mu_i \phi_2 + y_i - 1) \mathbf{E}(\log(1 - p_i) | \mathbf{y}_0, \boldsymbol{\theta}^{(m)}) \\ + \log \Gamma(\phi_1 + \mu_i \phi_2) - \log \Gamma(\phi_1) - \log \Gamma(\mu_i \phi_2) \}. \quad (4.7)$$

Los pasos del algoritmo EM son:

- **Paso E:** Dado que,  $p_i \sim \text{Beta}(\phi_1, \mu_i \phi_2)$ , entonces,  $p_i | \mathbf{y}_0, \boldsymbol{\theta} \sim \text{Beta}(\phi_1 + 1, y_i + \mu_i \phi_2)$ , luego utilizando las propiedades de la distribución Beta, Johnson et al. (1970), se obtiene que,

$$\mathbf{E}(\log p_i | \mathbf{y}_0, \boldsymbol{\theta}^{(m)}) = \psi(\phi_1^{(m)} + 1) - \psi(y_i + \mu_i^{(m)} \phi_2^{(m)} + \phi_1^{(m)} + 1) = e_i, \quad i = 1, \dots, n. \quad (4.8)$$

Análogamente,

$$\mathbf{E}(\log(1 - p_i) | \mathbf{y}_0, \boldsymbol{\theta}^{(m)}) = \psi(y_i + \mu_i^{(m)} \phi_2^{(m)}) - \psi(y_i + \mu_i^{(m)} \phi_2^{(m)} + \phi_1^{(m)} + 1) = s_i, \quad (4.9)$$

para  $i = 1, \dots, n$ . Sustituyendo (4.8) y (4.9) en (4.7) se tiene,

$$\mathbf{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}) = \sum_{i=1}^n \{ \phi_1 e_i + (\mu_i \phi_2 + y_i - 1) s_i + \log \Gamma(\phi_1 + \mu_i \phi_2) - \log \Gamma(\phi_1) - \log \Gamma(\mu_i \phi_2) \}. \quad (4.10)$$

- **Paso M:** Se debe actualizar  $\boldsymbol{\theta}^{(m)}$  maximizando  $\mathbf{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})$ , es decir,

$$\boldsymbol{\theta}^{(m+1)} = \underset{\boldsymbol{\theta}}{\text{ArgMax}} \mathbf{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}).$$

Luego, se actualiza  $\boldsymbol{\theta}^{(m)}$  utilizando los pseudovalores  $e_i$  y  $s_i$ , donde los parámetros  $\boldsymbol{\theta}$  son estimados a través del algoritmo de Newton Rapshon, donde la función score está dada por,

$$\mathbf{U}_{\boldsymbol{\theta}}(\mathbf{Q}) = \frac{\partial \mathbf{Q}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \mathbf{U}_{\beta}^T(\mathbf{Q}) \\ \mathbf{U}_{\phi}(\mathbf{Q}) \end{pmatrix}, \quad (4.11)$$

con

$$\mathbf{U}_{\beta}(\mathbf{Q}) = \frac{\partial \mathbf{Q}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \beta} = \phi_2 \mathbf{X}^T \mathbf{R}_1,$$

donde  $\mathbf{R}_1$  es un vector de longitud  $n$ , cuyos elementos están dados por,

$$\mathbf{R}_{1i} = (s_i + \psi(\phi_1 + \mu_i \phi_2) - \psi(\mu_i \phi_2)) \mu_i, \quad i = 1, \dots, n, \text{ y}$$

$$\mathbf{U}_{\phi}(\mathbf{Q}) = \frac{\partial \mathbf{Q}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \phi} = \frac{-2}{(\phi - 1)^2} \mathbf{F}_1,$$

$$\text{donde } \mathbf{F}_1 = \sum_{i=1}^n [e_i + \mu_i s_i + \psi(\phi_1 + \mu_i \phi_2)(1 + \mu_i) - \psi(\phi_1) - \psi(\mu_i \phi_2) \mu_i].$$

Mientras que la matriz de segundas derivadas está dada por,

$$\ddot{\mathbf{Q}}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \frac{\partial^2 \mathbf{Q}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = \begin{pmatrix} \mathbf{I}_{\beta\beta}(\mathbf{Q}) & \mathbf{I}_{\beta\phi}(\mathbf{Q}) \\ \mathbf{I}_{\phi\beta}(\mathbf{Q}) & \mathbf{I}_{\theta\theta}(\mathbf{Q}) \end{pmatrix}, \quad (4.12)$$

donde

$$\mathbf{I}_{\beta\beta}(\mathbf{Q}) = \frac{\partial^2 \mathbf{Q}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \beta \partial \beta^T} = \phi_2 \mathbf{X}^T \text{Diag}(\mathbf{R}_1) \mathbf{X} + \phi_2 \mathbf{X}^T \mathbf{R}_2 \mathbf{X}, \quad (4.13)$$

con  $\mathbf{R}_2$  una matriz diagonal de elementos,  $\mathbf{R}_{2i} = (\psi'(\phi_1 + \mu_i\phi_2) - \psi'(\mu_i\phi_2))\mu_i$ , y  $\text{Diag}(\mathbf{R}_1)$  representa la matriz diagonal cuyos elementos son precisamente las componentes del vector  $\mathbf{R}_1$ .

$$\begin{aligned}\mathbf{I}_{\phi\phi}(\mathbf{Q}) &= \frac{\partial^2 \mathbf{Q}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \phi^2} \\ &= \frac{4}{(\phi-1)^3} \mathbf{F}_1 + \frac{4}{(\phi-1)^4} \sum_{i=1}^n [\psi'(\phi_1 + \mu_i\phi_2)(1 + \mu_i)^2 - \psi'(\phi_1) - \psi'(\mu_i\phi_2)\mu_i^2] \\ &= \frac{4}{(\phi-1)^3} \mathbf{F}_1 + \frac{4}{(\phi-1)^4} \mathbf{F}_2,\end{aligned}\tag{4.14}$$

$$\text{con } \mathbf{F}_2 = \sum_{i=1}^n [\psi'(\phi_1 + \mu_i\phi_2)(1 + \mu_i)^2 - \psi'(\phi_1) - \psi'(\mu_i\phi_2)\mu_i^2].$$

Y,

$$\mathbf{I}_{\beta\phi}(\mathbf{Q}) = \frac{\partial \mathbf{Q}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \beta \partial \phi} = -\frac{2}{(\phi-1)^2} \mathbf{X}^T \mathbf{R}_1 + \phi_2 \mathbf{X}^T \mathbf{R}_3,\tag{4.15}$$

donde los elementos del vector  $\mathbf{R}_3$  son,  $\mathbf{R}_{3i} = (\psi'(\phi_1 + \mu_i\phi_2)(1 + \mu_i) - \psi'(\mu_i\phi_2)\mu_i)\mu_i$ ,  $i = 1, \dots, n$ .

Sustituyendo en (4.12) las expresiones (4.13), (4.14) y (4.15), se tiene,

$$\ddot{\mathbf{Q}}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \begin{pmatrix} \phi_2 \mathbf{X}^T \text{Diag}(\mathbf{R}_1) \mathbf{X} + \phi_2 \mathbf{X}^T \mathbf{R}_2 \mathbf{X} & -\frac{2}{(\phi-1)^2} \mathbf{X}^T \mathbf{R}_1 + \phi_2 \mathbf{X}^T \mathbf{R}_3 \\ -\frac{2}{(\phi-1)^2} \mathbf{R}_1^T \mathbf{X} + \phi_2 \mathbf{R}_3^T \mathbf{X} & \frac{4}{(\phi-1)^3} \mathbf{F}_1 + \frac{4}{(\phi-1)^4} \mathbf{F}_2 \end{pmatrix},\tag{4.16}$$

evaluada en  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ .

Luego, (4.11) y (4.16) son utilizados en el algoritmo de Newton-Raphson para obtener las estimaciones del vector de parámetros  $\boldsymbol{\theta}$ . Cuando se ha alcanzado la convergencia, se retorna al paso **E**.

Note que al igualar  $\mathbf{U}_{\boldsymbol{\theta}}(\mathbf{Q})$  a cero no se obtiene una expresión explícita para la estimación de los parámetros, razón por la cual se optó por utilizar el método descrito.

Para obtener los errores asintóticos de los EMV de  $\boldsymbol{\theta}$  se utilizará la matriz  $(-\mathbf{I}_{\boldsymbol{\theta}})^{-1}$ , con  $\mathbf{I}_{\boldsymbol{\theta}}$  dada en (4.3).

## 4.2. Análisis de Influencia

En esta sección son presentadas las expresiones que permitirán detectar posibles observaciones influyentes en la estimación de los parámetros del modelo de regresión Waring, utilizando (2.15) y (4.10), respectivamente.

### 4.2.1. Influencia global

El objetivo de esta herramienta es detectar qué ocurre con la estimación de los parámetros al eliminar un subconjunto de observaciones. A continuación se exponen las expresiones que permitirán concluir al respecto.

#### 4.2.1.1. Distancia de Cook generalizada

- a) **Utilizando la función de log-verosimilitud observada.** Para obtener la distancia de Cook generalizada ( $GD_i$ ) se debe disponer de la primera derivada de la función de log-verosimilitud con respecto a  $\theta$  sin la  $i$ -ésima observación,  $\dot{l}_{[i]}(\theta|\hat{\theta})$ , y la segunda derivada, pero con todas las observaciones,  $\mathbf{I}_\theta$ , ambas evaluadas en  $\hat{\theta}$ . Las expresiones correspondientes son,

$$\frac{\partial l_{[i]}(\theta)}{\partial \beta} = \phi_2 \mathbf{X}_{[i]}^T \mathbf{J}_{2[i]} (\mathbf{a}_{[i]} - \mathbf{b}_{[i]}),$$

y

$$\frac{\partial^2 l_{[i]}(\theta)}{\partial \phi^2} = \sum_{i \neq j} \left[ \frac{1}{\phi} - \frac{1}{\phi - 1} - 2\mathbf{a}_j \frac{\mu_j + 1}{(\phi - 1)^2} + 2\mathbf{b}_j \frac{\mu_j}{(\phi - 1)^2} \right].$$

Por lo tanto,

$$\dot{l}_{[i]}(\hat{\theta}) = \left( \begin{array}{c} \phi_2 \mathbf{X}_{[i]}^T \mathbf{J}_{2[i]} (\mathbf{a}_{[i]} - \mathbf{b}_{[i]}) \\ \sum_{i \neq j} \left[ \frac{1}{\phi} - \frac{1}{\phi - 1} - 2\mathbf{a}_j \frac{\mu_j + 1}{(\phi - 1)^2} + 2\mathbf{b}_j \frac{\mu_j}{(\phi - 1)^2} \right] \end{array} \right) \Bigg|_{\theta = \hat{\theta}}, \quad (4.17)$$

que es un vector de longitud  $q$ , mientras que  $\mathbf{I}_\theta$  corresponde a la matriz hessiana (Fisher observada) cuyos elementos están dados por (4.4), (4.5) y (4.6).

Sustituyendo las expresiones (4.17) y la matriz  $\mathbf{I}_\theta$  en (2.35), se obtienen las aproximaciones a un paso de  $\hat{\theta}_{[i]}$ , las que permiten calcular la distancia de Cook generalizada a un paso tal y como lo indica (2.36).

b) **Utilizando la función de log-verosimilitud completa.** En este enfoque se tiene,

$$\frac{\partial Q_{[i]}(\theta|\hat{\theta})}{\partial \beta} = \phi_2 \mathbf{X}_{[i]}^T \mathbf{R}_{[i]},$$

donde  $\mathbf{R}_{[i]}$  es un vector de longitud  $(n-1)$ , cuyos elementos están dados por,

$$\mathbf{R}_{[i]} = (s_j + \psi(\phi_1 + \mu_j \phi_2) - \psi(\mu_j \phi_2)) \mu_j, \text{ con } \forall j \neq i,$$

$$\frac{\partial Q_{[i]}(\theta|\hat{\theta})}{\partial \phi} = \frac{-2}{(\phi-1)^2} \sum_{i \neq j} [e_j + \mu_j s_j + \psi(\phi_1 + \mu_j \phi_2)(1 + \mu_j) - \psi(\phi_1) - \psi(\mu_j \phi_2) \mu_j].$$

Luego,

$$\dot{Q}_{[i]}(\hat{\theta}|\hat{\theta}) = \left( \begin{array}{c} \phi_2 \mathbf{X}_{[i]}^T \mathbf{R}_{[i]} \\ \frac{-2}{(\phi-1)^2} \sum_{i \neq j} [e_j + \mu_j s_j + \psi(\phi_1 + \mu_j \phi_2)(1 + \mu_j) - \psi(\phi_1) - \psi(\mu_j \phi_2) \mu_j] \end{array} \right) \Bigg|_{\theta=\hat{\theta}}, \quad (4.18)$$

es un vector de longitud  $q$ .

Las expresiones (4.18) y (4.16) permiten obtener las estimaciones a un paso de  $\theta_{[i]}$  como en (2.44), y con ello la distancia de Cook generalizada a 1 paso en el contexto del algoritmo EM, dada en (2.45).

#### 4.2.1.2. Distancia de verosimilitud

Análogamente a lo anterior, se calcula la distancia de verosimilitudes ( $DV_i$ ) y la  $Q$ -distancia ( $QD_i$ ) dadas por (2.37) y (2.46), respectivamente.

#### 4.2.2. Influencia local

En esta sección se presentan las expresiones a utilizar en el análisis de influencia local para el modelo Waring, utilizando la perturbación aditiva adecuada del predic-



tor.

#### 4.2.2.1. Utilizando la función de log-verosimilitud observada

Bajo el esquema de perturbación aditiva del predictor se tiene que,  $l(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^n l_i(\boldsymbol{\theta}|\omega_i)$ , donde  $\mu_{i\omega} = \exp\{\mathbf{x}^T\boldsymbol{\beta} + S_i\omega_i\beta_j^*\}$ , la cual corresponde a la perturbación aditiva del  $j^*$ -ésimo predictor. En primer lugar, se determina  $S_i$ , para que la perturbación sea adecuada, recordando que una perturbación es adecuada si la matriz de información de Fisher esperada evaluada en  $\boldsymbol{\omega}_0$  es igual a  $c\mathbf{I}_n$ , luego,

$$\frac{\partial l(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_i} = \phi_2\beta_j^*(\mathbf{a}_{i\omega} - \mathbf{b}_{i\omega})\mu_{i\omega}S_i,$$

$$\frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_i^2} = \phi_2\beta_j^{*2}[\phi_2(\mathbf{a}_{i\omega}^* - \mathbf{b}_{i\omega}^*)\mu_{i\omega} + (\mathbf{a}_{i\omega} - \mathbf{b}_{i\omega})]\mu_{i\omega}S_i^2, \quad (4.19)$$

$$\frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_i \partial \omega_r} = 0, \quad i \neq r, \quad (4.20)$$

donde  $\mathbf{a}_{i\omega} = \psi(\phi_1 + \mu_{i\omega}\phi_2) - \psi(y_i + \phi_1 + \mu_{i\omega}\phi_2 + 1)$ ,  $\mathbf{b}_{i\omega} = \psi(\mu_{i\omega}\phi_2) - \psi(y_i + \mu_{i\omega}\phi_2)$ ,  $\mathbf{a}_{i\omega}^* = \psi'(\phi_1 + \mu_{i\omega}\phi_2) - \psi'(y_i + \phi_1 + \mu_{i\omega}\phi_2 + 1)$  y  $\mathbf{b}_{i\omega}^* = \psi'(\mu_{i\omega}\phi_2) - \psi'(y_i + \mu_{i\omega}\phi_2)$ .

Tomando esperanza en (4.19) y (4.20) se obtiene, respectivamente,

$$G_{ii}(\boldsymbol{\omega}) = -\mathbf{E}\left(\frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_i^2}\right) = -\phi_2\beta_j^{*2}[\phi_2\mathbf{E}(\mathbf{a}_{i\omega}^* - \mathbf{b}_{i\omega}^*)\mu_{i\omega} + \mathbf{E}(\mathbf{a}_{i\omega} - \mathbf{b}_{i\omega})]\mu_{i\omega}S_i^2, \quad (4.21)$$

$$G_{ir}(\boldsymbol{\omega}) = \mathbf{E}\left(\frac{\partial l(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_i} \frac{\partial l(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_r}\right) = 0, \quad i \neq r. \quad (4.22)$$

Al evaluar (4.21) y (4.22) en  $\boldsymbol{\omega}_0$ , sigue que la matriz  $\mathbf{G}(\boldsymbol{\omega})$  es diagonal; pero  $\mathbf{G}(\boldsymbol{\omega}_0) \neq c\mathbf{I}_n$ , es decir, cada elemento de la diagonal principal de  $\mathbf{G}(\boldsymbol{\omega}_0)$  está dado por,

$$G_{ii}(\boldsymbol{\omega}) = -\phi_2\beta_j^{*2}[\phi_2\mathbf{E}(\mathbf{a}_i^* - \mathbf{b}_i^*)\mu_i + \mathbf{E}(\mathbf{a}_i - \mathbf{b}_i)]\mu_i S_i^2.$$

Luego, para que la perturbación sea adecuada se debe cumplir que,

$$S_i = |(-\phi_2\mathbf{E}(\mathbf{a}_i^* - \mathbf{b}_i^*)\mu_i - \mathbf{E}(\mathbf{a}_i - \mathbf{b}_i))\mu_i|^{-1/2},$$

con  $\mathbf{a}_i$  y  $\mathbf{b}_i$ , los correspondientes valores de  $\mathbf{a}_{i\omega}$  y  $\mathbf{b}_{i\omega}$  evaluados en  $\omega_0$ , definidos en la Sección 4.1.1.

Para determinar los elementos de la matriz  $\Delta$  se define,

$$\mu_{i\omega} = \exp\{\mathbf{x}_i^T \boldsymbol{\beta} + |(-\phi_2 \mathbf{E}(\mathbf{a}_i^* - \mathbf{b}_i^*) \mu_i - \mathbf{E}(\mathbf{a}_i - \mathbf{b}_i)) \mu_i|^{-1/2} \omega_i \beta_j^*\}.$$

Para este esquema de perturbación los elementos de la matriz  $\Delta$  son,

$$\begin{aligned} \tau_{\phi_i} = & -\frac{2}{(\phi-1)^2} \beta_j^* [\mathbf{a}_{i\omega} - \mathbf{b}_{i\omega}] \mu_{i\omega} S_i - \frac{2\phi_2}{(\phi-1)^2} \beta_j^* [\mathbf{a}_{i\omega}^* (1 + \mu_{i\omega}) - \mathbf{b}_{i\omega}^* \mu_{i\omega}] \mu_{i\omega} S_i \\ & + \phi_2 \beta_j^{*2} [\mathbf{a}_{i\omega} - \mathbf{b}_{i\omega}] \mu_{i\omega} S_{i\phi} S_i \omega_i + \phi_2 \beta_j^* [\mathbf{a}_{i\omega} - \mathbf{b}_{i\omega}] \mu_{i\omega} S_{i\phi}, \end{aligned}$$

donde  $S_{i\phi} = \frac{\partial S_i}{\partial \phi}$  está dado por,

$$\begin{aligned} S_{i\phi} = & -\frac{1}{2} S_i^3 \left( \frac{2}{(\phi-1)^2} \mathbf{E}(\mathbf{a}_i^* - \mathbf{b}_i^*) \mu_i - \phi_2 \frac{\partial}{\partial \phi} \mathbf{E}(\mathbf{a}_i^* - \mathbf{b}_i^*) \mu_i - \frac{\partial}{\partial \phi} \mathbf{E}(\mathbf{a}_i - \mathbf{b}_i) \right) \mu_i \\ & \times \frac{(-\phi_2 \mathbf{E}(\mathbf{a}_i^* - \mathbf{b}_i^*) \mu_i - \mathbf{E}(\mathbf{a}_i - \mathbf{b}_i)) \mu_i}{|(-\phi_2 \mathbf{E}(\mathbf{a}_i^* - \mathbf{b}_i^*) \mu_i - \mathbf{E}(\mathbf{a}_i - \mathbf{b}_i)) \mu_i|}, i = 1, \dots, n. \end{aligned}$$

Para el cálculo de esta expresión se utilizó el resultado  $\frac{\partial}{\partial x} |f(x)| = \frac{f(x)}{|f(x)|} f'(x)$ , Bartle and Sherbert (2000). Luego, bajo determinadas condiciones de regularidad que permiten intercambiar la derivada con la suma, se tiene que,

$$\begin{aligned} \frac{\partial \mathbf{E}(\mathbf{a}_i - \mathbf{b}_i)}{\partial \phi} &= \frac{\partial}{\partial \phi} \sum_{y_i^*=0}^{\infty} (\mathbf{a}_i - \mathbf{b}_i) P(Y_i = y_i^*) \\ &= -\frac{2}{(\phi-1)^2} \sum_{y_i^*=0}^{\infty} (\mathbf{a}_i^* (1 + \mu_i) - \mathbf{b}_i^* \mu_i) P(Y_i = y_i^*) + \sum_{y_i^*=0}^{\infty} (\mathbf{a}_i - \mathbf{b}_i) \frac{\partial P(Y_i = y_i^*)}{\partial \phi}, \end{aligned}$$

$$\text{con } \frac{\partial P(Y_i = y_i^*)}{\partial \phi} = -\frac{2}{(\phi-1)^2} \left( \frac{1}{\phi} + \mathbf{a}_i (1 + \mu_i) - \mathbf{b}_i \mu_i \right) P(Y_i = y_i^*). \mathbf{Y},$$

$$\begin{aligned} \frac{\partial \mathbf{E}(\mathbf{a}_i^* - \mathbf{b}_i^*)}{\partial \phi} &= \frac{\partial}{\partial \phi} \sum_{y_i^*=0}^{\infty} (\mathbf{a}_i^* - \mathbf{b}_i^*) P(Y_i = y_i^*) \\ &= -\frac{2}{(\phi-1)^2} \sum_{y_i^*=0}^{\infty} (\mathbf{a}_i^{**} (1 + \mu_i) - \mathbf{b}_i^{**} \mu_i) P(Y_i = y_i^*) + \sum_{y_i^*=0}^{\infty} (\mathbf{a}_i^* - \mathbf{b}_i^*) \frac{\partial P(Y_i = y_i^*)}{\partial \phi}, \end{aligned}$$

donde  $\mathbf{a}_i^{**} = \psi''(\phi_1 + \mu_i \phi_2) - \psi''(y_i^* + \phi_1 + \mu_i \phi_2 + 1)$ ,  $\mathbf{b}_i^{**} = \psi''(\mu_i \phi_2) - \psi''(y_i^* + \mu_i \phi_2)$  y  $\psi''(\cdot)$  corresponde a la función poligamma. Note además que  $\mathbf{a}_i$ ,  $\mathbf{b}_i$ ,  $\mathbf{a}_i^*$  y  $\mathbf{b}_i^*$  se evalúan en  $y_i^*$ , cuando están involucradas en el cálculo de esperanzas.

Por último, también se tiene que,

$$\mathbf{E}(\mathbf{a}_i^* - \mathbf{b}_i^*) = \sum_{y_i^*=0}^{\infty} (\mathbf{a}_i^* - \mathbf{b}_i^*) P(Y_i = y_i^*).$$

Análogamente, se calcula  $\tau_{\beta_{ij}}$ . En primer lugar para  $j = j^*$ ,

$$\begin{aligned} \tau_{\beta_{ij}} &= \phi_2(\mathbf{a}_{i\omega} - \mathbf{b}_{i\omega})\mu_{i\omega}S_i + \phi_2\beta_j^*[\phi_2(\mathbf{a}_{i\omega}^* - \mathbf{b}_{i\omega}^*)\mu_{i\omega} + (\mathbf{a}_{i\omega} - \mathbf{b}_{i\omega})]\mu_{i\omega}S_i(x_{ij} + S_{i\beta_j}\omega_i\beta_j^* + S_i\omega_i) \\ &\quad + \phi_2\beta_j^*(\mathbf{a}_{i\omega} - \mathbf{b}_{i\omega})\mu_{i\omega}S_{i\beta_j}, \end{aligned}$$

con

$$\begin{aligned} S_{i\beta_j} &= \frac{\partial S_i}{\partial \beta_j} = -\frac{1}{2}S_i^3 \left[ \left\{ -\phi_2 \frac{\partial}{\partial \beta_j} \mathbf{E}(\mathbf{a}_i^* - \mathbf{b}_i^*)\mu_i - \phi_2 \mathbf{E}(\mathbf{a}_i^* - \mathbf{b}_i^*)\mu_i x_{ij} - \frac{\partial}{\partial \beta_j} \mathbf{E}(\mathbf{a}_i - \mathbf{b}_i) \right\} \mu_i \right. \\ &\quad \left. + \{ -\phi_2 \mathbf{E}(\mathbf{a}_i^* - \mathbf{b}_i^*)\mu_i - \mathbf{E}(\mathbf{a}_i - \mathbf{b}_i) \} \mu_i x_{ij} \right] \times \frac{(-\phi_2 \mathbf{E}(\mathbf{a}_i^* - \mathbf{b}_i^*)\mu_i - \mathbf{E}(\mathbf{a}_i - \mathbf{b}_i))\mu_i}{|(-\phi_2 \mathbf{E}(\mathbf{a}_i^* - \mathbf{b}_i^*)\mu_i - \mathbf{E}(\mathbf{a}_i - \mathbf{b}_i))\mu_i|}, \end{aligned}$$

donde

$$\begin{aligned} \frac{\partial \mathbf{E}(\mathbf{a}_i - \mathbf{b}_i)}{\partial \beta_j} &= \frac{\partial}{\partial \beta_j} \sum_{y_i^*=0}^{\infty} (\mathbf{a}_i - \mathbf{b}_i) P(Y_i = y_i^*) \\ &= \phi_2 \sum_{y_i^*=0}^{\infty} (\mathbf{a}_i^* - \mathbf{b}_i^*)\mu_i x_{ij} P(Y_i = y_i^*) + \sum_{y_i^*=0}^{\infty} (\mathbf{a}_i - \mathbf{b}_i) \frac{\partial P(Y_i = y_i^*)}{\partial \beta_j}, \end{aligned}$$

con  $\frac{\partial P(Y_i = y_i^*)}{\partial \beta_j} = \phi_2 \mu_i x_{ij} (\mathbf{a}_i - \mathbf{b}_i) P(Y_i = y_i^*)$ . Por otro lado,

$$\begin{aligned} \frac{\partial \mathbf{E}(\mathbf{a}_i^* - \mathbf{b}_i^*)}{\partial \beta_j} &= \frac{\partial}{\partial \beta_j} \sum_{y_i^*=0}^{\infty} (\mathbf{a}_i^* - \mathbf{b}_i^*) P(Y_i = y_i^*) \\ &= \phi_2 \sum_{y_i^*=0}^{\infty} (\mathbf{a}_i^{**} - \mathbf{b}_i^{**})\mu_i x_{ij} P(Y_i = y_i^*) + \sum_{y_i^*=0}^{\infty} (\mathbf{a}_i^* - \mathbf{b}_i^*) \frac{\partial P(Y_i = y_i^*)}{\partial \beta_j}. \end{aligned}$$

Y,  $\mathbf{E}(\mathbf{a}_i^* - \mathbf{b}_i^*) = \sum_{y_i^*=0}^{\infty} (\mathbf{a}_i^* - \mathbf{b}_i^*) P(Y_i = y_i^*)$ .

Repitiendo para  $j \neq j^*$ ,

$$\tau_{\beta_{ij}} = \phi_2\beta_j^*[\phi_2(\mathbf{a}_{i\omega}^* - \mathbf{b}_{i\omega}^*)\mu_{i\omega} + (\mathbf{a}_{i\omega} - \mathbf{b}_{i\omega})]\mu_{i\omega}S_i(x_{ij} + S_{i\beta_j}\omega_i\beta_j^*) + \phi_2\beta_j^*(\mathbf{a}_{i\omega} - \mathbf{b}_{i\omega})\mu_{i\omega}S_{i\beta_j}.$$

Note que todas las expresiones anteriores deben ser evaluadas en el EMV de  $\theta$  y en

$$\omega = \omega_0 = (0, \dots, 0)^T.$$

#### 4.2.2.2. Utilizando la función de log-verosimilitud completa

Bajo la perturbación aditiva del predictor, la función de log-verosimilitud de los datos completos se puede escribir como,

$$l(\boldsymbol{\theta}|\mathbf{y}_c) = \sum_{i=1}^n \{ \log p_{i\omega} + y_i \log(1 - p_{i\omega}) + \log \Gamma(\phi_1 + \mu_i \phi_2) + (\phi_1 - 1) \log p_i \\ + (\mu_i \phi_2 - 1) \log(1 - p_i) - \log \Gamma(\phi_1) - \log \Gamma(\mu_i \phi_2) \},$$

donde  $p_{i\omega} = \frac{1}{1 + \mu_{i\omega}}$ , con  $\mu_{i\omega} = \exp\{\mathbf{x}_i^T \boldsymbol{\beta} + S_i \omega_i \beta_j^*\}$ , para  $i = 1, \dots, n$ . Primero se determina a  $S_i$  para que la perturbación propuesta sea adecuada. En efecto se tiene que,

$$\begin{aligned} \frac{\partial l_c(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)}{\partial \omega_i} &= - (y_i + 1) \left( \frac{\mu_{i\omega}}{1 + \mu_{i\omega}} \right) S_i \beta_j^* + y_i S_i \beta_j^*, \\ \frac{\partial^2 l_c(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)}{\partial \omega_i^2} &= - (y_i + 1) \left( \frac{\mu_{i\omega}}{(1 + \mu_{i\omega})^2} \right) S_i^2 \beta_j^{*2}, \\ \frac{\partial^2 l_c(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)}{\partial \omega_i \partial \omega_r} &= 0, \quad i \neq r. \end{aligned} \quad (4.23)$$

Tomando esperanza a los últimos dos términos en (4.23), se obtiene, respectivamente,

$$\begin{aligned} G_{ii}(\boldsymbol{\omega}) &= - \mathbf{E} \left( \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)}{\partial \omega_i^2} \right) = \left( \frac{2 + \mu_{i\omega}}{1 + \mu_{i\omega}} \right) \frac{\mu_{i\omega}}{(1 + \mu_{i\omega})^2} S_i^2 \beta_j^{*2}, \\ G_{ir}(\boldsymbol{\omega}) &= \mathbf{E} \left( \frac{\partial l(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)}{\partial \omega_i} \frac{\partial l(\boldsymbol{\theta}|\boldsymbol{\omega}, \mathbf{y}_c)}{\partial \omega_r} \right) = 0. \end{aligned}$$

Al evaluar cada elemento en  $\boldsymbol{\omega}_0$  se obtiene,

$$\mathbf{G}(\boldsymbol{\omega}_0) = \beta_j^{*2} \mathbf{Diag} \left( \frac{2 + \mu_1}{(1 + \mu_1)^3} S_1^2, \dots, \frac{2 + \mu_n}{(1 + \mu_n)^3} S_n^2 \right) \neq \mathbf{cI}_n.$$

Por lo tanto, para que la perturbación sea adecuada se toma a,

$$S_i = \left( \frac{2 + \mu_i}{(1 + \mu_i)^3} \right)^{-1/2}, \quad \text{para } i = 1, \dots, n.$$

Reescribiendo  $\mu_{i\omega}$ , se tiene,

$$\mu_{i\omega} = \exp \left\{ \mathbf{x}_i^T \boldsymbol{\beta} + \left( \frac{2 + \mu_i}{(1 + \mu_i)^3} \right)^{-1/2} \omega_i \beta_j^* \right\}.$$

Por lo tanto, los elementos de la matriz  $\Delta$  están dados por,  $\tau_{\phi_i} = 0$ ,  $i = 1, \dots, n$ , mientras que  $\tau_{\beta_{ij}}$  para  $j = j^*$  se tiene que,

$$\tau_{\beta_{ij}} = -\frac{(y_i + 1)}{(1 + \mu_{i\omega})^2} \mu_{i\omega} [\beta_j^* S_i (x_{ij} + S_{i\beta_j} \omega_i \beta_j^* + S_i \omega_i) + \beta_j^* S_{i\beta_j} (1 + \mu_{i\omega} + y_i) + S_i (1 + \mu_{i\omega} + y_i)].$$

Y para  $j \neq j^*$ ,

$$\tau_{\beta_{ij}} = -\frac{(y_i + 1)}{(1 + \mu_{i\omega})^2} \mu_{i\omega} [\beta_j^* S_i (x_{ij} + S_{i\beta_j} \omega_i \beta_j^*) + \beta_j^* S_{i\beta_j} (1 + \mu_{i\omega})] + \beta_j^* y_i S_{i\beta_j}.$$

Note que en este contexto,  $S_{i\beta_j}$  está dado por,

$$S_{i\beta_j} = \frac{\partial S_i}{\partial \beta_j} = -\frac{1}{2} S_i^3 \left( \frac{-2\mu_i^2 x_{ij} - 5\mu_i x_{ij}}{(1 + \mu_i)^4} \right), \quad \forall i = 1, \dots, n, j = 0, \dots, k.$$

Todas las expresiones deben ser evaluadas en el EMV de  $\theta$  y en  $\omega = \omega_0 = (0, \dots, 0)^T$ .

### 4.3. Aplicación

En esta sección se presenta una aplicación del modelo discutido y las herramientas de diagnóstico presentadas en la sección anterior, para un conjunto de datos reales. En primer lugar se muestra la estimación de los parámetros, por un lado a través de la maximización directa de la función de log-verosimilitud observada, y por otro de la función de log-verosimilitud completa por medio del algoritmo EM. Finalmente, se realiza un análisis de influencia global y local a estos datos para estudiar la presencia de datos influyentes en el estimador máximo verosímil.

#### 4.3.1. Visitas al médico

Los datos corresponden a los ya presentados y estudiados en el capítulo anterior, donde la variable de respuesta es el número de visitas al médico, *NVIS*, de 1755 mujeres en Alemania durante 1987, y las covariables a introducir en el modelo son:

el ingreso familiar per-cápita expresado en marcos dividido en 10, *ING*, la edad en años de las pacientes *ED* y sus años de educación *EDUC*.

En la Figura 3.16 se observó el comportamiento de la variable respuesta con respecto a cada una de las covariables, a partir de la cual se concluyó que existen tres observaciones alejadas de la nube de puntos, dos ellas, 37 y 291, corresponden a observaciones con alto número de visitas al médico y bajo ingreso familiar, mientras que la tercera observación 285 corresponde a una mujer que presenta alto ingreso familiar y alto número de visitas. Se puede señalar además que las pacientes 37 y 291, son mujeres mayores a 50 años con sólo 9 años de educación.

La Tabla 4.1 muestra las estimaciones obtenidas al ajustar el modelo de regresión Geométrico y Waring, respectivamente. En ella se observa que, a pesar de que las estimaciones de los parámetros no son muy distintas en ambos modelos, el valor de la función de log-verosimilitud aumenta, mientras que el AIC se reduce en el modelo de regresión Waring, dando indicios que el modelo de regresión Waring es más adecuado que el modelo de regresión Geométrico para describir el comportamiento de estos datos. Adicionalmente, al realizar el test de razón de verosimilitud sobre los modelos a comparar se obtiene un valor de  $LR_c = 166.0824$ , con un *valor-p* asociado  $< 0.05$ , en que el estadístico correspondiente a la prueba sigue una distribución  $\chi^2$  con un grado de libertad, dado que el modelo Geométrico está anidado en el modelo Waring. Ésto apoya la conclusión anterior. En el contexto del problema, cobra relevancia entonces que la probabilidad,  $p$ , de que cada una de estas mujeres acuda al médico, cambie de individuo en individuo, como es natural pensar, justificando así el uso de la regresión Waring. Note que los parámetros  $\beta_1$  y  $\beta_3$ , asociados a *ING* y *EDUC*, respectivamente, no son significativos.

En la Tabla 4.1 no se incluyeron las estimaciones obtenidas a través del algoritmo EM, dado que éstas coinciden con lo obtenido a través de la maximización directa

Modelo	Estimación	
Geométrico	$\hat{\beta}_0 = 1.4142 (0.3017)$	$l = -4345.632$
	$\hat{\beta}_1 = -0.0364 (0.0246)$	$AIC = 8699.263$
	$\hat{\beta}_2 = 0.0099 (0.0035)$	
	$\hat{\beta}_3 = -0.0352 (0.0198)$	
Waring	$\hat{\beta}_0 = 1.1092 (0.2662)$	$l = -4265.269$
	$\hat{\beta}_1 = -0.0258 (0.0216)$	$AIC = 8540.538$
	$\hat{\beta}_2 = 0.0130 (0.0031)$	
	$\hat{\beta}_3 = -0.0225 (0.0174)$	
	$\hat{\phi} = 2.3701 (0.3552)$	
$LR_c = 2(l_{waring} - l_{geo}) = 166.0824$ , valor-p < 0.05 $LR \sim \chi^2(1)$		

Tabla 4.1: Resumen de los ajustes del modelo de regresión Geométrico y Waring, para los datos de visitas al médico.

de la función de log-verosimilitud observada. Para el cálculo de los errores estándar asintóticos se utilizó la matriz de información de Fisher esperada,  $\mathbf{K}_\theta^{-1}$ .

En la Figura 4.1 se presentan las simulaciones del envelope para los residuos de Pearson, tanto para el modelo de regresión Geométrico (izquierda), como para el modelo de regresión Waring (derecha), se observa que éste último muestra un mejor ajuste, conclusión concordante con lo expuesto anteriormente, en el sentido de que el modelo de regresión Waring es más adecuado para describir estos datos que el modelo de regresión Geométrico.

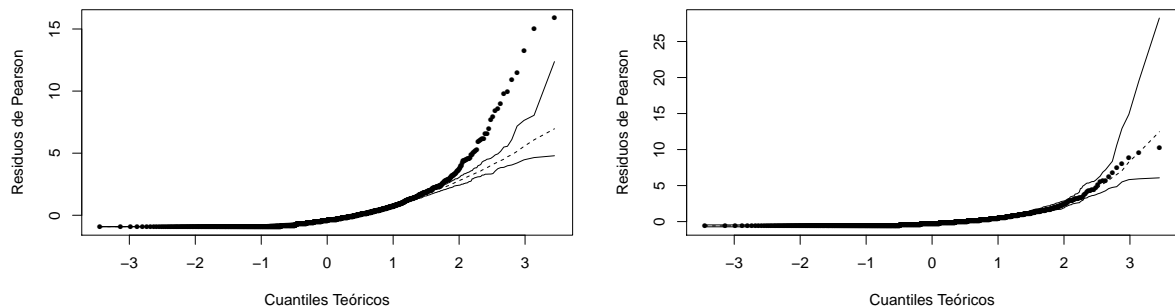


Figura 4.1: Envelope de los residuos de Pearson para el ajuste del modelo Geométrico (izquierda) y Waring (derecha).

La Figura 4.2 muestra el comportamiento de las medidas de influencia global, los

resultados al calcular la distancia de verosimilitud ( $DV_i$ ), señala a la observación 285 como la que presenta una mayor influencia sobre el EMV de  $\theta$ , con un valor aproximado de 450, por encima del resto de los valores y sobre la línea de corte propuesta en este trabajo. La misma conclusión se obtiene del gráfico de la distancia de Cook generalizada ( $GD_i$ ).

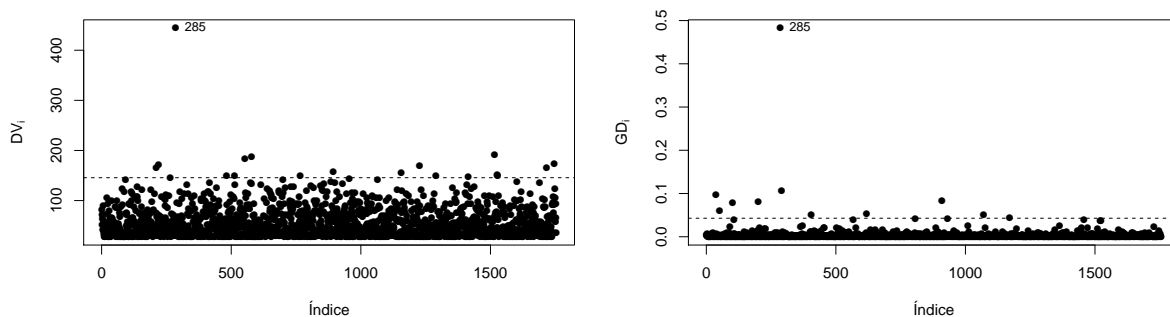


Figura 4.2: Gráfico índice de las distancias de verosimilitud y de Cook generalizada para  $\hat{\theta}$ .

En cuanto al análisis de influencia local, la Figura 4.3 considera el esquema de perturbación aditiva del predictor de la variable *ING*, si bien en el gráfico de  $|h_{max}|$  ninguna observación destaca por sobre el resto, se puede ver que los casos 37 y 291 son levemente superiores que los demás, comportamiento que se mantiene en el gráfico relativo a la curvatura local; mismo patrón presentan las observaciones 102 y 201. Similar comportamiento se puede observar cuando se perturban de manera aditiva las variables *ED* y *EDUC*, por lo cual no se incluyen dichos gráficos en este trabajo.

Finalmente, se puede notar que los casos más sensibles, tanto del punto de vista de la influencia global, como de la local, fueron las observaciones 37, 101, 201, 285 y 291. La Tabla 4.2 muestra los resultados del análisis al eliminar algunos subconjuntos de datos, note que no fueron incluidas todas las combinaciones posibles dadas por estas cinco observaciones, debido a su extensión (31 combinaciones). Por lo tanto, fueron incluidas sólo aquellas combinaciones cuya función de log-verosimilitud re-



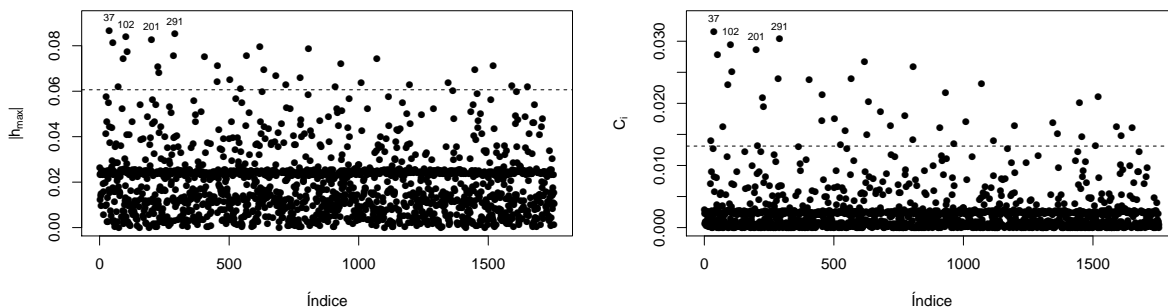


Figura 4.3: Gráfico índice de  $|h_{max}|$  y curvatura local total,  $C_i$ , para el esquema de perturbación aditiva del predictor Ingresos.

sultó mayor que la respectiva log-verosimilitud al considerar todos los datos, pero que por sobre todo, provocan un cambio en la significancia de los parámetros. Se puede observar que todos los cambios de significancia del parámetro asociado a la variable Ingresos ( $\beta_1$ ), son producto de la eliminación de combinaciones que incluyen a la observación 285, cabe señalar además que dicha observación corresponde a la única que provocó cambio al ser eliminados subconjuntos de tamaño 1. Note además que es el único subconjunto en el cual se logra rechazar la hipótesis nula,  $H_0 : \beta_j = 0$ , con un nivel de significancia del 5%, las restantes combinaciones rechazan dicha hipótesis a un nivel de significancia del 10% (ésto es indicado con un \* en la Tabla 4.2). Paralelamente, se puede concluir que la significancia del parámetro asociado a la variable Educación de la paciente nunca cambió.

La Tabla 4.2, también incluye las razones de cambio de los EMV, se puede observar que precisamente al ser eliminada la observación 285 del conjunto de datos, la  $RC$  del parámetro asociado a la variable  $ING$  presenta un mayor cambio; los restantes subconjuntos también presentan una  $RC$  superior al 50% (en relación a esta variable), pero dado que dichas combinaciones no arrojaron cambios en la significancia cuando no contenían a la observación 285, se puede concluir que es precisamente la eliminación conjunta de la paciente 285 la que provoca tales resultados. Por lo tan-

Observación Eliminada	$l$	$RC$					Significancia
		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\phi}$	
Ninguna	-4265.3	-	-	-	-	-	$\beta_2$ significativo
285	-4258.7	13.76	65.44	9.04	17.79	0.21	$\beta_1, \beta_2$ significativos
37,285	-4248.2	3.62	54.74	1.66	0.24	6.25	$\beta_1^*, \beta_2$ significativos
102,285	-4249.0	0.32	54.69	1.33	11.98	5.27	$\beta_1^*, \beta_2$ significativos
201,285	-4249.3	3.03	59.60	3.09	22.00	4.97	$\beta_1^*, \beta_2$ significativos
285,291	-4264.5	1.07	55.50	2.97	16.01	5.74	$\beta_1^*, \beta_2$ significativos
37,285,291	-4237.7	0.17	58.24	4.01	21.18	9.88	$\beta_1^*, \beta_2$ significativos
37,102,201,285	-4228.8	5.86	55.62	3.61	24.97	12.37	$\beta_1^*, \beta_2$ significativos
37,102,201,285,291	-4218.2	6.08	49.47	0.70	27.13	15.94	$\beta_1^*, \beta_2$ significativos

Tabla 4.2: Razón de cambio de los EMV al eliminar un subconjunto de observaciones en el modelo Waring.

to, se puede concluir que la observación que presenta una potencial influencia, es la 285 por sí sola, debido a que produce cambios en la inferencia, principalmente, en la significancia del parámetro asociado a la variable *ING*.

Por otro lado, ya que el parámetro asociado a la variable *EDUC* nunca cambió su significancia, fue extraída del modelo y se repitió el análisis anterior, para las variables *ING* y *ED*, las conclusiones sobre la influencia de 285 se conservan, dado que cada vez que dicha observación es extraída del conjunto de datos la significancia del parámetro asociado a la variable *ING* cambia, y nuevamente, el nivel de significancia para la hipótesis nula es un 5% en el caso de la eliminación individual de 285, y un 10% en el caso de las restantes combinaciones que incluyen la eliminación de la observación 285.

#### 4.3.1.1. Influencia global y local en el enfoque del algoritmo EM

En esta sección, se presentan los resultados del análisis de influencia desde el enfoque del algoritmo EM, aplicado al ejemplo de visitas al médico.

La Figura 4.4 muestra el comportamiento de la  $Q$ -distancia a un paso para  $\hat{\theta}$ , en ella se observa claramente que la observación 285 es la que presenta mayor influencia

en la estimación de los parámetros.

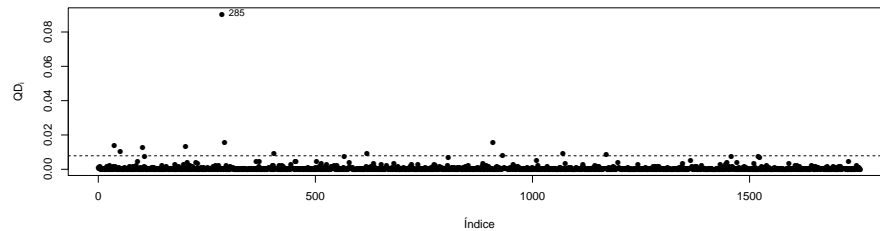


Figura 4.4: Gráfico índice para la  $Q$ -distancia un paso para  $\hat{\theta}$ .

La Figura 4.5 permite observar como las observaciones 37 y 291 son las que mayor influencia presentan en los EMV de los parámetros, bajo el esquema de perturbación aditiva de la variable Ingresos, tanto en el gráfico de  $|h_{max}|$ , como para el gráfico de curvatura local total. Los correspondientes gráficos para las variables  $ED$  y  $EDUC$  no fueron incluidos, dado que presentan un comportamiento similar.

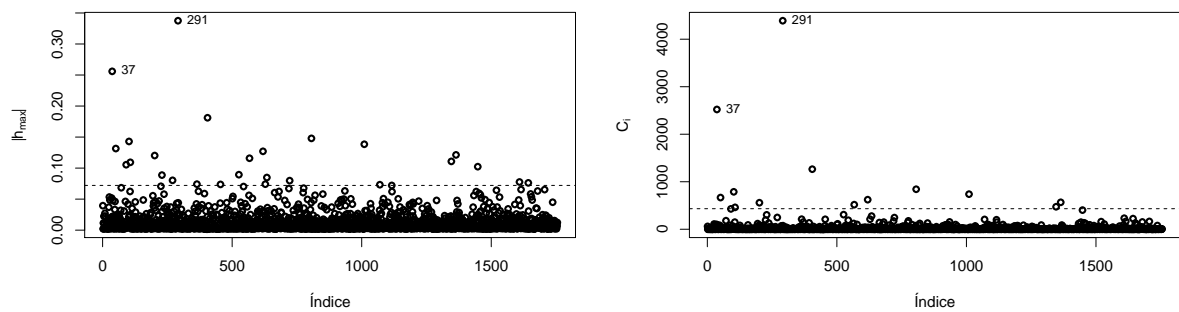


Figura 4.5: Gráfico índice para  $|h_{max}|$  y curvatura local total,  $C_i$ , bajo el esquema de perturbación aditiva del predictor Ingresos.

La Tabla 4.3 muestra la razón de cambio ( $RC$ ) de los EMV, para el ajuste de la mezcla Beta-Geométrico, utilizando para dicha estimación el algoritmo EM, al ser eliminadas las posibles observaciones influyentes detectadas en los gráficos previos (37, 285 y 291) y sus combinaciones. A partir de dicha tabla se puede concluir que el parámetro  $\beta_2$  en todos los casos resultó ser significativo y el parámetro  $\beta_3$  siempre se mantuvo no significativo para el modelo; sin embargo, el parámetro  $\beta_1$  cambió su significancia cuando estaba involucrada la eliminación de la observación 285, mismo

comportamiento que se observó en la Tabla 4.2.

Observación Eliminada	RC					Significancia
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\phi}$	
Ninguna	-	-	-	-	-	$\beta_2$ significativo
37	1.21	1.55	0.77	4.44	4.34	$\beta_2$ significativo
285	1.13	60.85	0.00	14.22	1.79	$\beta_1, \beta_2$ significativos
291	0.15	6.20	3.08	1.78	4.13	$\beta_2$ significativo
37,285	0.06	63.95	1.54	19.11	6.07	$\beta_1^*, \beta_2$ significativos
37,291	1.34	4.26	3.85	6.22	8.32	$\beta_2$ significativo
285,291	0.99	55.04	3.08	16.00	5.74	$\beta_1^*, \beta_2$ significativos
37,285,291	0.17	57.75	3.85	20.89	9.88	$\beta_1^*, \beta_2$ significativos

Tabla 4.3: Razón de cambio de los EMV al eliminar un subconjunto de observaciones en el modelo Beta-Geométrico.

Análogamente, se puede observar en la Tabla 4.3, que las razones de cambio de los EMV con respecto a  $\beta_1$  son mayores cuando del conjunto de datos ha sido eliminado un subconjunto que contiene a la observación 285. Además, cuando dicha observación es eliminada, provoca un cambio en la significancia del parámetro  $\beta_1$ , es decir, cuando esta observación está incluida en el conjunto de datos la hipótesis  $H_0 : \beta_1 = 0$  no es rechazada, sin embargo, cuando es eliminada, y la hipótesis es nuevamente contrastada, ésta es rechazada con un nivel de significancia del 5%. En los restantes casos que incluyen la eliminación de la observación 285, también existe un cambio en la significancia del parámetro, pero con un nivel de significancia del 10%.

Nuevamente, cabe destacar que se repitió el análisis sólo con las variables *ING* y *ED*, dado que el parámetro  $\beta_3$  nunca resultó significativo; las conclusiones son análogas a lo presentado en la Tabla 4.3, por lo tanto, la variable *EDUC* puede ser extraída del modelo.

### 4.3.2. Conclusión

A partir de las Tablas 4.2 y 4.3 se puede afirmar que la observación 285 provoca cambios en los EMV de los parámetros, es decir, tanto el uso de la función de log-

verosimilitud observada como el de la función log-verosimilitud completa detectaron a dicha observación como un caso potencialmente influyente, la cual corresponde a una mujer con altos Ingresos, de 42 años de Edad y con sólo 9 años de Educación.

Se puede afirmar que, para esta aplicación, se justifica la utilización del modelo Waring, por sobre el modelo Geométrico, dado que es natural que la probabilidad asociada a acudir al médico de cualquier persona es propia del individuo. Por lo tanto, es lógico pensar que ésta varíe de sujeto a sujeto, y como tal dicha probabilidad puede tener una distribución asociada, en este caso Beta.

En cuanto a la aplicación misma, y dado que ambos enfoques detectan a la observación 285 como una potencial observación influyente, es recomendable estudiar su naturaleza.

## Capítulo 5

# Modelo de Regresión Beta-Binomial Negativo: Waring Generalizado

El modelo Waring Generalizado (WG) es utilizado para modelar datos de conteo, debe su adjetivo de Generalizado a que se obtiene de la mezcla de un modelo Beta con un modelo Binomial Negativo, tomando en cuenta el hecho que un modelo Geométrico es un caso límite (particular) del modelo Binomial Negativo. La distribución Waring Generalizada debe su nombre a Edward Waring (1736-1798). Fue descrita inicialmente por Newbold (1927), sin embargo, fue Irwin (1968) quien dio a conocer su representación actual, destacando que su particularidad es que permite que la varianza se pueda dividir en tres componentes, ejemplifica lo anterior con un conjunto de datos relativo al riesgo de sufrir accidentes, donde la primera componente recoge la varianza debida al azar, mientras que las otras dos están relacionados con las cualidades internas del individuo y las debidas al medio. La distribución Waring Generalizada permite descomponer la variabilidad de las observaciones en tres componentes: las netamente relacionadas con el azar (aleatorias), las debidas a las diferencias internas entre los individuos y la relacionada con la presencia de otros factores externos, que no son incluidos como covariables en el modelo.

Adicionalmente, Irwin (1968) prueba que un caso límite de la distribución Waring Generalizada es la distribución Binomial Negativa. Más adelante, Xekalaki (1983b) la bautiza como distribución Waring Generalizada Univariada (UGWD), y la aplica en el contexto del riesgo de sufrir accidentes. Poniendo énfasis en las componentes de la varianza y su identificación, este autor aplica esta distribución a los datos presentados por Irwin (1968) y los compara con las distribuciones de contagio, donde se asume que los individuos de una población están expuestos a cierto riesgo de accidentes.

A partir de entonces son variados los análisis que se han presentado para dicha distribución, todos ellos coinciden en que la ventaja radica en la posibilidad de poder descomponer la varianza, de manera tal de obtener más información del comportamiento de los datos. Así mismo destacan que es un modelo más general, que tiene como casos particulares al modelo Binomial Negativo, Waring y el modelo de Poisson, y como tal es una excelente alternativa para analizar datos que poseen heterogeneidad.

## 5.1. Estimación en el modelo Waring Generalizado

Para la estimación de los parámetros del modelo WG se usó la función *GWRM.fit* disponible en el paquete *GWRM*-versión 1.1 o la función *gw* en la versión 2.1.0.2 de dicho paquete, desarrollado por Rodríguez-Avi et al. (2009), el que hace uso, entre otras, de las funciones `nlm` y `optim`.

A continuación se presentan los elementos que componen la función score y la matriz de información de Fisher observada o hessiano, necesarios para maximizar la función de log-verosimilitud del modelo, donde la contribución de la  $i$ -ésima observación está dada por (2.24).

En primer lugar, la función score obtenida de diferenciar (2.23) con respecto al

vector de parámetros desconocido,  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \kappa, \rho)^T$  está dada por,

$$\mathbf{U}_{\boldsymbol{\theta}} = \begin{pmatrix} U_{\boldsymbol{\beta}}^T \\ U_{\kappa} \\ U_{\rho} \end{pmatrix},$$

donde

$$\mathbf{U}_{\kappa} = \sum_{i=1}^n \left[ -\frac{\vartheta_i}{\kappa} Z_{1i} + \psi(\kappa + \rho) + \psi(\kappa + y_i) - \psi(\kappa) - \psi(\vartheta_i + \kappa + \rho + y_i) \right],$$

$$\mathbf{U}_{\rho} = \sum_{i=1}^n \left[ \frac{\mu_i}{\kappa} Z_{1i} + \psi(\kappa + \rho) - \psi(\rho) + \psi(\vartheta_i + \rho) - \psi(\vartheta_i + \kappa + \rho + y_i) \right],$$

y

$$\mathbf{U}_{\boldsymbol{\beta}} = \frac{\rho - 1}{\kappa} \mathbf{X}^T Z_1 \boldsymbol{\mu},$$

con  $Z_1$  un vector de longitud  $n$ , cuyos elementos están dados por,

$$Z_{1i} = \psi(\vartheta_i + \rho) + \psi(\vartheta_i + y_i) - \psi(\vartheta_i) - \psi(\vartheta_i + \kappa + \rho + y_i) \text{ y } \vartheta_i = \mu_i \frac{\rho - 1}{\kappa}.$$

De manera análoga, se calcula la matriz de información de Fisher observada, dada por,

$$\mathbf{I}_{\boldsymbol{\theta}} = \begin{pmatrix} \mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \mathbf{I}_{\boldsymbol{\beta}\kappa} & \mathbf{I}_{\boldsymbol{\beta}\rho} \\ \mathbf{I}_{\kappa\boldsymbol{\beta}} & \mathbf{I}_{\kappa\kappa} & \mathbf{I}_{\kappa\rho} \\ \mathbf{I}_{\rho\boldsymbol{\beta}} & \mathbf{I}_{\rho\kappa} & \mathbf{I}_{\rho\rho} \end{pmatrix},$$

donde

$$\mathbf{I}_{\kappa\kappa} = \sum_{i=1}^n \left[ 2 \frac{\vartheta_i}{\kappa^2} Z_{1i} + \frac{\vartheta_i^2}{\kappa^2} Z_{1i}' + \psi'(\kappa + \rho) + \psi'(\kappa + y_i) - \psi'(\kappa) - \psi'(\vartheta_i + \kappa + \rho + y_i) (-2\vartheta_i \kappa + 1) \right],$$

$$\mathbf{I}_{\rho\rho} = \sum_{i=1}^n \left[ \frac{\mu_i^2}{\kappa^2} Z_{1i}' + \psi'(\kappa + \rho) - \psi'(\rho) + [\psi'(\vartheta_i + \rho) - \psi'(\vartheta_i + \kappa + \rho + y_i)] \left( 2 \frac{\mu_i}{\kappa} + 1 \right) \right],$$

$$\mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \frac{(\rho - 1)^2}{\kappa^2} \mathbf{X}^T \text{Diag}(Z_1' \odot \boldsymbol{\mu}^2) \mathbf{X} + \frac{\rho - 1}{\kappa} \mathbf{X}^T \text{Diag}(Z_1 \odot \boldsymbol{\mu}) \mathbf{X},$$

$$\mathbf{I}_{\boldsymbol{\beta}\kappa} = -\frac{(\rho - 1)^2}{\kappa^3} \mathbf{X}^T Z_1 \odot \boldsymbol{\mu}^2 - \frac{\rho - 1}{\kappa^2} \mathbf{X}^T Z_1 \odot \boldsymbol{\mu} - \frac{\rho - 1}{\kappa} \mathbf{X}^T \psi'(\boldsymbol{\vartheta} + \kappa + \rho + y) \odot \boldsymbol{\mu},$$

$$\mathbf{I}_{\boldsymbol{\beta}\rho} = \frac{\rho - 1}{\kappa^2} \mathbf{X}^T Z_1' \odot \boldsymbol{\mu}^2 + \frac{1}{\kappa} \mathbf{X}^T Z_1 \boldsymbol{\mu} + \frac{\rho - 1}{\kappa} \mathbf{X}^T (\psi'(\boldsymbol{\vartheta} + \rho) - \psi'(\boldsymbol{\vartheta} + \kappa + \rho + y) \odot \boldsymbol{\mu}),$$

y

$$\mathbf{I}_{\kappa\rho} = \sum_{i=1}^n \left[ -\frac{\vartheta_i}{\kappa^2} \mu_i Z_{1i}' - \frac{\mu_i}{\kappa^2} Z_{1i} + \psi'(\kappa + \rho) - \frac{\vartheta_i}{\kappa} \psi'(\vartheta_i + \rho) - \psi'(\vartheta_i + \kappa + \rho + y_i) \left( \frac{\mu_i}{\kappa} - \frac{\vartheta_i}{\kappa} + 1 \right) \right],$$



con  $Z'_{1i} = \psi'(\vartheta_i + \rho) + \psi'(\vartheta_i + y_i) - \psi'(\vartheta_i) - \psi'(\vartheta_i + \kappa + \rho + y_i)$  y  $\boldsymbol{\mu}^2 = \boldsymbol{\mu} \odot \boldsymbol{\mu}$ , donde  $\odot$  representa el producto Hadamard.

Utilizando estas expresiones se obtiene la matriz de Fisher esperada y los errores estándar asintóticos de los EMV.

## 5.2. Análisis de Influencia

En esta sección se presentan las expresiones, tanto para influencia global, como para influencia local, que permitirán detectar posibles observaciones influyentes en la estimación de los parámetros del modelo de regresión WG.

### 5.2.1. Influencia global

Debido a que esta herramienta permite detectar qué observación o subconjunto de ellas provoca mayores cambios en la estimación de los parámetros cuando han sido eliminadas del proceso de estimación, a continuación se presentan las expresiones relativas a influencia global para el modelo WG.

#### 5.2.1.1. Distancia de Cook generalizada

De acuerdo con (2.36), la distancia de Cook generalizada a un paso requiere el cálculo de la matriz de información de Fisher observada,  $\mathbf{I}_\theta$ , evaluada en  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ , cuyos elementos fueron presentados en la sección anterior, y de la función score sin la  $i$ -ésima observación  $\dot{l}_{[i]}(\boldsymbol{\theta})$ , con componentes dadas por,

$$\frac{\partial l_{[i]}(\boldsymbol{\theta})}{\partial \kappa} = \sum_{i \neq j} \left[ -\frac{\vartheta_j}{\kappa} Z_{1j} + \psi(\kappa + \rho) + \psi(\kappa + y_j) - \psi(\kappa) - \psi(\vartheta_j + \kappa + \rho + y_j) \right],$$

$$\frac{\partial l_{[i]}(\boldsymbol{\theta})}{\partial \rho} = \sum_{i \neq j} \left[ \frac{\mu_j}{\kappa} Z_{1j} + \psi(\kappa + \rho) - \psi(\rho) + \psi(\vartheta_j + \rho) - \psi(\vartheta_j + \kappa + \rho + y_j) \right],$$

y

$$\frac{\partial l_{[i]}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \frac{\rho - 1}{\kappa} \mathbf{X}_{[i]}^T Z_{1[i]} \odot \boldsymbol{\mu}_{[i]}.$$

Luego,

$$\dot{l}_{[i]}(\hat{\boldsymbol{\theta}}) = \left( \begin{array}{c} \frac{\partial l_{[i]}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta}} \\ \frac{\partial l_{[i]}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \kappa} \\ \frac{\partial l_{[i]}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \rho} \end{array} \right) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad i = 1, \dots, n.$$

### 5.2.1.2. Distancia de verosimilitud

Para complementar la distancia de Cook generalizada a un paso, también se presenta la distancia de verosimilitudes a un paso ( $DV_i$ ), dada por,

$$DV_i^{(1)} = 2\{l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}_{[i]}^{(1)})\}, \quad i = 1, \dots, n,$$

donde  $l(\hat{\boldsymbol{\theta}}_{[i]}^{(1)})$  corresponde a la función de log-verosimilitud sin la  $i$ -ésima observación, evaluada en  $\hat{\boldsymbol{\theta}}_{[i]}^{(1)}$ , calculada a partir de (2.35).

### 5.2.2. Influencia local

En esta sección se obtienen las expresiones que permiten estudiar la influencia de determinados esquemas de perturbación en la estimación de máxima verosimilitud de los parámetros del modelo. En particular, para el modelo de regresión WG se presenta la perturbación aditiva de un predictor.

La función de log-verosimilitud del modelo WG bajo este esquema de perturbación está dado por,

$$\begin{aligned} l(\boldsymbol{\theta}|\boldsymbol{\omega}) &= \sum_{i=1}^n \{ \log \Gamma(\vartheta_{i\omega} + \rho) + \log \Gamma(\kappa + \rho) + \log((\vartheta_{i\omega})_{y_i}) + \log((\kappa)_{y_i}) - \log \Gamma(\rho) \\ &\quad - \log \Gamma(\vartheta_{i\omega} + \kappa + \rho) - \log((\vartheta_{i\omega} + \kappa + \rho)_{y_i}) - \log y_i! \} \\ &= \sum_{i=1}^n \{ \log \Gamma(\vartheta_{i\omega} + \rho) + \log \Gamma(\kappa + \rho) + \log \Gamma(\vartheta_{i\omega} + y_i) + \log \Gamma(\kappa + y_i) - \log \Gamma(\rho) \\ &\quad - \log \Gamma(\vartheta_{i\omega}) - \log \Gamma(\kappa) - \log \Gamma(\vartheta_{i\omega} + \kappa + \rho + y_i) - \log y_i! \}, \end{aligned} \quad (4.1)$$

donde  $\vartheta_{i\omega} = \mu_{i\omega} \frac{(\rho - 1)}{\kappa}$ , con  $\mu_{i\omega} = \exp\{\mathbf{x}_i^T \boldsymbol{\beta} + S_i \omega_i \beta_j^*\}$ , y  $j^*$  representa al predictor perturbado.

En primer lugar, se determina el vector de perturbación adecuado bajo este esquema, es decir, tal que  $\mathbf{G}(\omega_0) = \mathbf{cI}_n$ . En efecto a partir de (4.1) se tiene,

$$\begin{aligned}\frac{\partial l(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_i} &= \beta_j^* Z_{1i\omega} \vartheta_{i\omega} S_i, \\ \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_i^2} &= \beta_j^{*2} [Z'_{1i\omega} \vartheta_{1i\omega} + Z_{1i\omega}] \vartheta_{1i\omega} S_i^2, \\ \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_i \partial \omega_r} &= 0, \quad i \neq r,\end{aligned}$$

donde  $Z_{1i\omega} = \psi(\vartheta_{i\omega} + \rho) + \psi(\vartheta_{i\omega} + y_i) - \psi(\vartheta_{i\omega}) - \psi(\vartheta_{i\omega} + \kappa + \rho + y_i)$  y

$$Z'_{i\omega} = \psi'(\vartheta_{i\omega} + \rho) + \psi'(\vartheta_{i\omega} + y_i) - \psi'(\vartheta_{i\omega}) - \psi'(\vartheta_{i\omega} + \kappa + \rho + y_i), \quad i = 1, \dots, n.$$

Entonces, los elementos de la matriz  $\mathbf{G}$  están dados por,

$$\begin{aligned}G_{ii}(\boldsymbol{\omega}) &= -\beta_j^{*2} [\mathbf{E}(Z'_{1i\omega}) \vartheta_{1i\omega} + \mathbf{E}(Z_{1i\omega})] \vartheta_{1i\omega} S_i^2, \\ G_{ir}(\boldsymbol{\omega}) &= \mathbf{E} \left( \frac{\partial l(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_i} \frac{\partial l(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_r} \right) = 0, \quad i \neq r.\end{aligned}\tag{4.2}$$

Evaluando a (4.2) en  $\omega_0$ , la matriz de información de Fisher esperada con respecto a  $\omega$ , bajo este esquema de perturbación es,

$$\begin{aligned}G(\omega_0) &= \beta_j^{*2} \begin{pmatrix} -[\mathbf{E}(Z'_{11}) \vartheta_1 + \mathbf{E}(Z_{11})] \vartheta_1 S_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & -[\mathbf{E}(Z'_{1n}) \vartheta_n + \mathbf{E}(Z_{1n})] \vartheta_n S_n^2 \end{pmatrix} \\ &\neq \mathbf{cI}_n.\end{aligned}$$

Por lo tanto, para que la perturbación sea adecuada, se define  $S_i$  como,

$$S_i = | -[\mathbf{E}(Z'_{1i}) \vartheta_i + \mathbf{E}(Z_{1i})] \vartheta_i |^{-1/2}, \quad i = 1, \dots, n.\tag{4.3}$$

Luego, para la perturbación aditiva de una de las variables predictoras, donde  $\mu_{i\omega} = \exp\{\mathbf{x}_i^T \boldsymbol{\beta} + S_i \omega_i \beta_j^*\}$ , con  $S_i$  definido en (4.3), la matriz  $\Delta_{\omega_0}$  tiene la forma,

$$\Delta_{\omega_0} = \begin{pmatrix} \boldsymbol{\tau}_{\boldsymbol{\beta}}^T \\ \boldsymbol{\tau}_{\kappa}^T \\ \boldsymbol{\tau}_{\rho}^T \end{pmatrix}_{q \times n},$$

donde  $\boldsymbol{\tau}_{\kappa}^T$  y  $\boldsymbol{\tau}_{\rho}^T$  corresponden a vectores de longitud  $n$  cuyo  $i$ -ésimo elemento está dado respectivamente por,

$$\begin{aligned}\tau_{\kappa i} &= \beta_j^{*2} \left[ -Z'_{1i\omega} \frac{\vartheta_{i\omega}}{\kappa} + Z_{1i\omega} \vartheta_{i\omega} \right] S_{i\kappa} \omega_i S_i + \beta_j^* Z_{1i\omega} \vartheta_{i\omega} S_{i\kappa} - \beta_j^* Z_{1i\omega} \frac{\vartheta_{i\omega}}{\kappa} S_i \\ &\quad - \beta_j^* \psi'(\vartheta_{i\omega} + \kappa + \rho + y_i) \vartheta_{i\omega} S_i,\end{aligned}$$

donde

$$\begin{aligned} S_{i\kappa} &= \frac{\partial S_i}{\partial \kappa} = -\frac{1}{2} S_i^3 \frac{\partial}{\partial \kappa} | - [\mathbf{E}(Z'_{1i})\vartheta_i^2 + \mathbf{E}(Z_{1i}\vartheta_i)] | \\ &= \frac{1}{2} (S_i)^3 \left[ \frac{\partial \mathbf{E}(Z'_{1i})}{\partial \kappa} \vartheta_i^2 - 2\mathbf{E}(Z'_{1i}) \frac{\vartheta_i^2}{\kappa} + \frac{\partial \mathbf{E}(Z_{1i})}{\partial \kappa} \vartheta_i - \mathbf{E}(Z_{1i}) \frac{\vartheta_i}{\kappa} \right] \\ &\quad \times \frac{-[\mathbf{E}(Z'_{1i})\vartheta_i^2 + \mathbf{E}(Z_{1i}\vartheta_i)]}{| - [\mathbf{E}(Z'_{1i})\vartheta_i^2 + \mathbf{E}(Z_{1i}\vartheta_i)] |}. \end{aligned}$$

Luego, bajo ciertas condiciones de regularidad es posible intercambiar la suma por la derivada, con lo cual se tiene que,

$$\begin{aligned} \frac{\partial}{\partial \kappa} \mathbf{E}(Z_{1i}) &= \frac{\partial}{\partial \kappa} \sum_{y_i^*=0}^{\infty} Z_{1i} P(Y_i = y_i^*) \\ &= \sum_{y_i^*=0}^{\infty} \left[ Z'_{1i} \frac{-\vartheta_i}{\kappa} - \psi'(\vartheta_i + \kappa + \rho + y_i) \right] P(Y_i = y_i^*) + \sum_{y_i^*=0}^{\infty} Z_{1i} \frac{\partial P(Y_i = y_i^*)}{\partial \kappa}, \end{aligned}$$

donde

$$\frac{\partial}{\partial \kappa} P(Y_i = y_i^*) = \left[ Z_{1i} \frac{-\vartheta_i}{\kappa} + \psi(\kappa + \rho) + \psi(\kappa + y_i) - \psi(\kappa) - \psi(\vartheta_i + \kappa + \rho + y_i) \right] P(Y_i = y_i^*).$$

Análogamente, se obtiene  $\frac{\partial}{\partial \kappa} \mathbf{E}(Z'_{1i})$ ,

$$\begin{aligned} \frac{\partial}{\partial \kappa} \mathbf{E}(Z'_{1i}) &= \frac{\partial}{\partial \kappa} \sum_{y_i^*=0}^{\infty} Z'_{1i} P(Y_i = y_i^*) \\ &= \sum_{y_i^*=0}^{\infty} \left[ Z''_{1i} \frac{-\vartheta_i}{\kappa} - \psi''(\vartheta_i + \kappa + \rho + y_i) \right] P(Y_i = y_i^*) + \sum_{y_i^*=0}^{\infty} Z'_{1i} \frac{\partial P(Y_i = y_i^*)}{\partial \kappa}, \end{aligned}$$

con  $Z''_{1i} = \psi''(\vartheta_i + \rho) + \psi''(\vartheta_i + y_i) - \psi''(\vartheta_i) - \psi''(\vartheta_i + \kappa + \rho + y_i)$ ,  $i = 1, \dots, n$ , y  $\psi''(\cdot)$

denota la función poligamma.

Repetiendo para  $\rho$  se tiene,

$$\begin{aligned} \tau_{\rho i} &= \beta_j^{*2} [Z'_{1i\omega} \vartheta_{i\omega}^2 + Z_{1i\omega} \vartheta_{i\omega}] S_{i\rho} \omega_i S_i + \beta_j^* Z_{1i\omega} \vartheta_{i\omega} S_{i\rho} + \beta_j^* Z_{1i\omega} \frac{\vartheta_{i\omega}}{\kappa} S_i \\ &\quad + \beta_j^* [\psi'(\vartheta_{i\omega} + \rho) - \psi'(\vartheta_{i\omega} + \kappa + \rho + y_i^*)] \vartheta_{i\omega} S_i, \end{aligned}$$

donde

$$\begin{aligned} S_{i\rho} &= \frac{\partial S_i}{\partial \rho} = -\frac{1}{2} S_i^3 \frac{\partial}{\partial \rho} | - [\mathbf{E}(Z'_{1i})\vartheta_i^2 + \mathbf{E}(Z_{1i}\vartheta_i)] | \\ &= \frac{1}{2} S_i^3 \left[ \frac{\partial \mathbf{E}(Z'_{1i})}{\partial \rho} \vartheta_i^2 + \mathbf{E}(Z'_{1i}) \frac{\mu_i^2}{\kappa^2} + \frac{\partial \mathbf{E}(Z_{1i})}{\partial \rho} \vartheta_i + \mathbf{E}(Z_{1i}) \frac{\mu_i}{\kappa} \right] \\ &\quad \times \frac{-[\mathbf{E}(Z'_{1i})\vartheta_i^2 + \mathbf{E}(Z_{1i}\vartheta_i)]}{| - [\mathbf{E}(Z'_{1i})\vartheta_i^2 + \mathbf{E}(Z_{1i}\vartheta_i)] |}, \end{aligned}$$

con

$$\begin{aligned}\frac{\partial \mathbf{E}(Z_{1i})}{\partial \rho} &= \frac{\partial}{\partial \rho} \sum_{y_i^*=0}^{\infty} Z_{1i} P(Y_i = y_i^*) \\ &= \sum_{y_i^*=0}^{\infty} \left[ Z'_{1i} \frac{\mu_i}{\kappa} + \psi'(\vartheta_i + \rho) - \psi'(\vartheta_i + \kappa + \rho + y_i^*) \right] P(Y_i = y_i^*) + \sum_{y_i^*=0}^{\infty} Z_{1i} \frac{\partial P(Y_i = y_i^*)}{\partial \rho},\end{aligned}$$

donde

$$\frac{\partial}{\partial \rho} P(Y_i = y_i^*) = \left[ Z_{1i} \frac{\mu_i}{\kappa} + \psi(\vartheta_i + \rho) - \psi(\vartheta_i + \kappa + \rho + y_i^*) + \psi(\kappa + \rho) - \psi(\rho) \right] P(Y_i = y_i^*).$$

Del mismo modo, se obtiene  $\frac{\partial \mathbf{E}(Z'_{1i})}{\partial \rho}$ ,

$$\begin{aligned}\frac{\partial \mathbf{E}(Z'_{1i})}{\partial \rho} &= \frac{\partial}{\partial \rho} \sum_{y_i^*=0}^{\infty} Z'_{1i} P(Y_i = y_i^*) \\ &= \sum_{y_i^*=0}^{\infty} \left[ Z''_{1i} \frac{\mu_i}{\kappa} + \psi''(\vartheta_i + \rho) - \psi''(\vartheta_i + \kappa + \rho + y_i^*) \right] P(Y_i = y_i^*) + \sum_{y_i^*=0}^{\infty} Z'_{1i} \frac{\partial P(Y_i = y_i^*)}{\partial \rho}.\end{aligned}$$

Note además que usando definición de esperanza se tiene que,

$$\mathbf{E}(Z_{1i}) = \sum_{y_i^*=0}^{\infty} Z_{1i} P(Y_i = y_i^*) \quad \text{y} \quad \mathbf{E}(Z'_{1i}) = \sum_{y_i^*=0}^{\infty} Z'_{1i} P(Y_i = y_i^*).$$

Por último, los elementos de  $\tau_{\beta_{ji}}$  están dados, en primer lugar para  $j = j^*$  por,

$$\begin{aligned}\tau_{\beta_{ij}} &= Z_{1i\omega} \vartheta_{i\omega} S_i + \beta_j^* \frac{\partial Z_{1i\omega}}{\partial \beta_j} \vartheta_{i\omega} S_i + \beta_j^* Z_{1i\omega} \frac{\partial \vartheta_{i\omega}}{\partial \beta_j} S_i + \beta_j^* Z_{1i\omega} \vartheta_{i\omega} S_{i\beta_j} \\ &= Z_{1i\omega} \vartheta_{i\omega} S_i + \beta_j^* Z'_{1i\omega} \vartheta_{i\omega}^2 x_{ij} S_i + \beta_j^* Z_{1i\omega} (x_{ij} + S_{i\beta_j} \omega_i \beta_j + S_i \omega_i) \vartheta_{i\omega} S_i + \beta_j^* Z_{1i\omega} \vartheta_{i\omega} S_{i\beta_j},\end{aligned}$$

donde

$$\begin{aligned}S_{i\beta_j} &= \frac{\partial S_i}{\partial \beta_j} = -\frac{1}{2} S_i^3 \frac{\partial}{\partial \beta_j} \left[ -[\mathbf{E}(Z'_{1i}) \vartheta_i^2 + \mathbf{E}(Z_{1i} \vartheta_i)] \right] \\ &= \frac{1}{2} S_i^3 \left[ \frac{\partial \mathbf{E}(Z'_{1i})}{\partial \beta_j} \vartheta_i^2 + \mathbf{E}(Z'_{1i}) \frac{2\vartheta_i(\rho-1)}{\kappa} x_{ij} + \frac{\partial \mathbf{E}(Z_{1i})}{\partial \beta_j} \vartheta_i + \mathbf{E}(Z_{1i}) \vartheta_i x_{ij} \right] \\ &\quad \times \frac{-[\mathbf{E}(Z'_{1i}) \vartheta_i^2 + \mathbf{E}(Z_{1i} \vartheta_i)]}{\left| -[\mathbf{E}(Z'_{1i}) \vartheta_i^2 + \mathbf{E}(Z_{1i} \vartheta_i)] \right|},\end{aligned}$$

con

$$\begin{aligned}\frac{\partial \mathbf{E}(Z_{1i})}{\partial \beta_j} &= \frac{\partial}{\partial \beta_j} \sum_{y_i^*=0}^{\infty} Z_{1i} P(Y_i = y_i^*) \\ &= \sum_{y_i^*=0}^{\infty} \left[ Z_{1i} \frac{\mu_i}{\kappa} + \psi'(\vartheta_i + \rho) - \psi'(\vartheta_i + \kappa + \rho + y_i^*) \right] P(Y_i = y_i^*) \\ &\quad + \sum_{y_i^*=0}^{\infty} Z_{1i} \frac{\partial P(Y_i = y_i^*)}{\partial \beta_j}, \quad i = 1, \dots, n \text{ y } j = 0, \dots, k,\end{aligned}$$

y,

$$\begin{aligned}\frac{\partial \mathbf{E}(Z'_{1i})}{\partial \beta_j} &= \frac{\partial}{\partial \beta_j} \sum_{y_i^*=0}^{\infty} Z'_{1i} P(Y_i = y_i^*) \\ &= \sum_{y_i^*=0}^{\infty} \left[ Z'_{1i} \frac{\mu_i}{\kappa} + \psi''(\vartheta_i + \rho) - \psi''(\vartheta_i + \kappa + \rho + y_i^*) \right] P(Y_i = y_i^*) \\ &\quad + \sum_{y_i^*=0}^{\infty} Z'_{1i} \frac{\partial P(Y_i = y_i^*)}{\partial \beta_j}, \quad i = 1, \dots, n \text{ y } j = 0, \dots, k.\end{aligned}$$

Finalmente, para  $j \neq j^*$

$$\tau_{\beta_{ij}} = \beta_j^* Z'_{1i\omega} \vartheta_{i\omega}^2 x_{ij} S_i + \beta_j^* Z_{1i\omega} (x_{ij} + S_{i\beta_j} \omega_i \beta_j + S_i \omega_i) \vartheta_{i\omega} S_i + \beta_j^* Z_{1i\omega} \vartheta_{i\omega} S_{i\beta_j}.$$

Todas las expresiones anteriores deben ser evaluadas en el EMV de  $\theta$  y en  $\omega = \omega_0 = (0, \dots, 0)^T$ .

### 5.3. Aplicación

En esta sección se ilustran los resultados presentados a dos conjuntos de datos reales, el primero de ellos relativo a las visitas al médico y el segundo relacionado con el fútbol.

#### 5.3.1. Aplicación a datos reales, visitas al médico

Los datos corresponden a los ya presentados y estudiados en los capítulos anteriores, donde la variable de respuesta es el número de visitas al médico,  $NVIS$ , de 1755

mujeres en Alemania durante 1987, y las covariables a introducir en el modelo son: *ING*, *ED* y *EDUC*.

La Tabla 5.1 muestra las estimaciones obtenidas al ajustar el modelo de regresión WG, en ella se observa que los parámetros asociados a las variables *ING* y *EDUC* no son significativos, contrario a lo que pasa con el parámetro relativo a la variable *ED*. También, se presentan los errores estándar asintóticos, los que fueron calculados a partir de la matriz de Fisher esperada,  $\mathbf{K}_\theta^{-1}$ , evaluada en  $\theta = \hat{\theta}$ . También se presenta el valor de la función de log-verosimilitud evaluada en los EMV. Al compararla con lo obtenido para el modelo BN en la Tabla 3.5, se puede observar una diferencia de alrededor 10 puntos en favor del modelo de regresión WG, con lo cual este último modelo es más adecuado que el modelo BN para describir los datos, misma conclusión que se obtiene cuando es comparado con el modelo Waring en la Tabla 4.1.

Parámetro	$\hat{\theta}$	ds
$\beta_0$	1.2102	0.2758
$\beta_1$	-0.0300	0.0225
$\beta_2$	0.0117	0.0032
$\beta_3$	-0.0263	0.0180
$\kappa$	0.7110	0.0521
$\rho$	5.9608	1.5890
$l=-4257.091$		$AIC=8526.182$

Tabla 5.1: Resumen del ajuste del modelo de regresión Waring Generalizado, para los datos de visitas médicas.

En la Figura 5.1 se presentan las simulaciones envelope para los residuos de Pearson, tanto para el modelo BN (izquierda), como para el modelo WG (derecha). Dado que el segundo de estos gráficos muestra un mejor ajuste, es razonable elegir al modelo WG por sobre el modelo BN, conclusión que apoya lo previamente expuesto.

La Figura 5.2 muestra el comportamiento de la distancia de verosimilitud y la distancia de Cook generalizada para  $\hat{\theta}$ , en ambas se puede apreciar que la observación 285 es la que presenta una potencial influencia en la estimación de los parámetros.

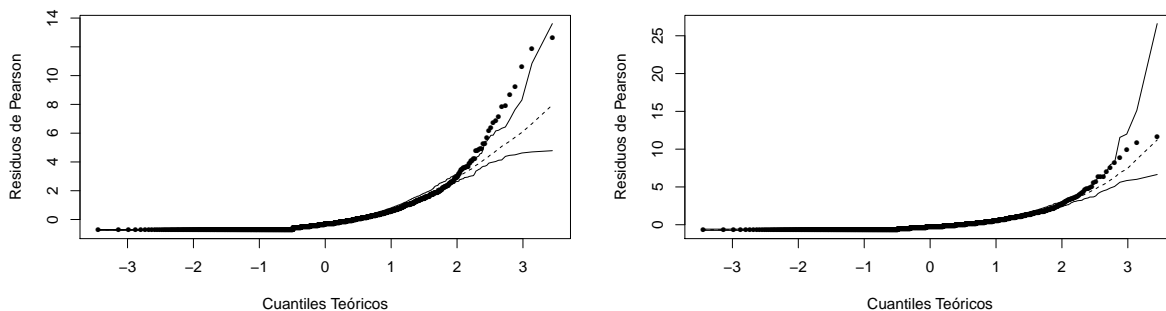


Figura 5.1: Envelope de los residuos de Pearson para el ajuste del modelo Binomial Negativo (izquierda) y Waring Generalizado (derecha).

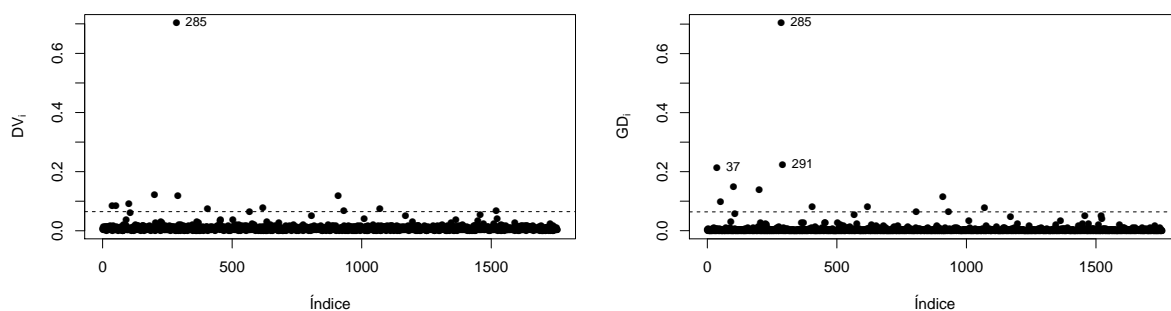


Figura 5.2: Gráfico índice de la distancia de verosimilitud y distancia de Cook generalizada para  $\hat{\theta}$ .

La Figura 5.3 muestra que bajo el esquema de perturbación aditiva del predictor  $ED$ , la observación 291, en la dirección  $|\mathbf{h}_{max}|$ , presenta un potencial efecto en el EMV de  $\theta$ . Conclusión similar se obtiene del gráfico relativo a la curvatura local total,  $C_i$ . Los gráficos relativos a las variables  $ING$  y  $EDUC$ , no fueron incluidos en este trabajo, dadas sus similitudes con lo descrito para la variable  $ED$ .

A partir de lo obtenido del análisis de influencia global y local, las potenciales observaciones influyentes en los EMV de los parámetros del modelo WG son las pacientes 285 y 291. La Tabla 5.2 muestra los resultados del análisis al eliminar los subconjuntos de datos que se pueden formar con estas dos observaciones. Se puede observar que cuando se elimina un subconjunto que incluye a la observación 285 se produce un cambio en la significancia de  $\beta_1$  y precisamente al quitar dicha observa-



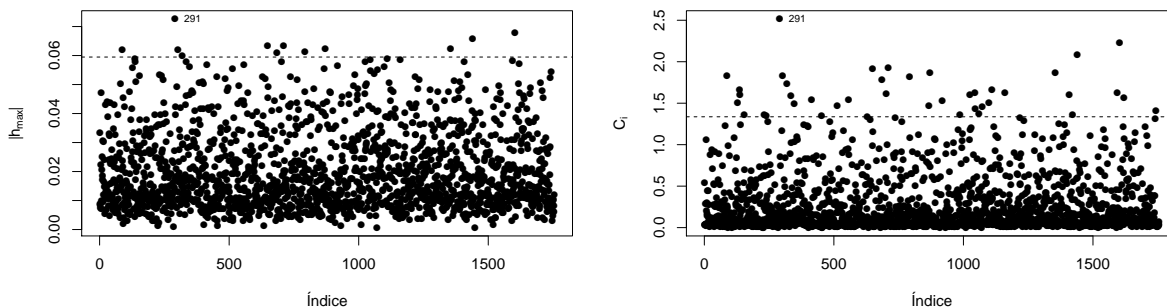


Figura 5.3: Gráfico índice de  $|h_{max}|$  y curvatura local total,  $C_i$ , bajo el esquema de perturbación aditiva del predictor Edad.  $C_i$  multiplicado por  $10^{-9}$

ción se produce un cambio mayor en la estimación de dicho parámetro, ésto se ve reflejado en el valor de la razón de cambio ( $RC$ ). Note que la conclusión acerca de la no significancia del parámetro  $\beta_3$  nunca cambió (al 5%), es decir,  $\beta_3$  siempre se mantuvo no significativo; por lo tanto la variable fue extraída del modelo, y se repitió el análisis anterior para las variables  $ING$  y  $ED$ ; las conclusiones sobre la influencia de 285 se conservan, es decir, al extraer esta observación del conjunto de datos la significancia de la variable  $ING$  cambia. Lo anterior nos lleva a concluir que la variable  $EDUC$  puede ser extraída del modelo.

Observación Eliminada	$l$	$RC$						Significancia
		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\kappa}$	$\hat{\rho}$	
Ninguna	-4257.1	-	-	-	-	-	-	$\beta_2$ significativo
285	-4250.2	1.68	69.71	0.22	15.86	0.31	3.47	$\beta_1, \beta_2$ significativos
291	-4246.0	0.74	7.38	5.56	1.50	1.02	10.78	$\beta_2$ significativo
285,291	-4239.1	2.44	62.70	5.79	17.51	1.24	14.56	$\beta_1, \beta_2$ significativos

Tabla 5.2: Razón de cambio de los EMV al eliminar un subconjunto de observaciones en el modelo WG.

### 5.3.1.1. Conclusión de la aplicación visitas al médico

De acuerdo con lo obtenido de la Tabla 5.2, se puede afirmar que la observación 285 provoca cambios en los EMV y por tanto se le puede considerar una observación potencialmente influyente en la estimación de los parámetros.

En cuanto a la comparación de los modelos WG con el modelo BN, se optará por el primero, dado que aumenta, aunque levemente, el valor de la log-verosimilitud, y además, recordando que el segundo es un caso particular del primero cuando  $\rho \rightarrow \infty$ . Conjuntamente, el valor del estadístico de razón de verosimilitud es 21.724, con un valor-p asociado  $\ll 0.05$ , con un grado de libertad, con lo cual se concluye que el modelo WG es mejor que un modelo BN para describir estos datos.

En la Tabla 5.3 se observa que existen diferencias entre los coeficientes estimados de  $\theta$  y sus respectivos errores estándar asintóticos, particularmente en el modelo BN los parámetros  $\beta_2$  y  $\beta_3$  son significativos, mientras que en el modelo Waring Generalizado, sólo resulta ser significativo el parámetro  $\beta_2$ .

Modelo	$\hat{\theta}$	ds	$l$
Binomial Negativo	$\hat{\beta}_0 = 1.4154$	0.2641	-4267.9
	$\hat{\beta}_1 = -0.0355$	0.0215	
	$\hat{\beta}_2 = 0.0099$	0.0031	
	$\hat{\beta}_3 = -0.0355$	0.0173	
	$\hat{\gamma} = 0.5840$	0.0253	
Waring Generalizado	$\hat{\beta}_0 = 1.2102$	0.2758	-4257.1
	$\hat{\beta}_1 = -0.0300$	0.0225	
	$\hat{\beta}_2 = 0.0117$	0.0032	
	$\hat{\beta}_3 = -0.0263$	0.0180	
	$\hat{\kappa} = 0.7110$	0.0521	
	$\hat{\rho} = 5.9608$	1.5890	

Tabla 5.3: Resumen de los ajustes de los modelos Binomial Negativo y Waring Generalizado.

Note que el parámetro  $\gamma$ , que representa la heterogeneidad en el modelo BN, es dividido en dos fuentes de sobredispersión en el modelo WG,  $\kappa$  y  $\rho$ . Dado que la principal característica del modelo de regresión WG es que permite dicha descomposición, la Tabla 5.4 muestra las componentes de la varianza para el conjunto de datos analizados, y además contempla las razones de varianza dadas por Rodríguez-Avi et al. (2009) como:

- Aleatoria:  $\frac{\kappa(\rho - 2)}{(\kappa + \rho - 1)(\kappa + \mu_i)}$ .
- Factores Externos:  $\frac{\kappa(\kappa + 1)}{(\kappa + \rho - 1)(\kappa + \mu_i)}$ .
- Factores Internos:  $\frac{\mu_i}{\kappa + \mu_i}$ .

Fuente de variación	Varianza	Proporción de varianza
Aleatorio	$\mu_i$	$\frac{0.4965}{0.7110 + \mu_i}$
Factores Externos	$0.4320\mu_i$	$\frac{0.2145}{0.7110 + \mu_i}$
Factores Internos	$2.0140\mu_i^2$	$\frac{\mu_i}{0.7110 + \mu_i}$
Total	$1.4320 \left( \mu_i + \frac{\mu_i^2}{0.7110} \right)$	1

Tabla 5.4: Varianza y proporción de varianza por componente para el conjunto de datos visitas al médico.

La Figura 5.4 muestra la relación entre las razones de las tres componentes de varianza y el valor esperado del número de visitas al médico por cada individuo, de acuerdo las expresiones presentadas en la Tabla 5.4. En dicha figura se observa que la mayor proporción de varianza se debe a los factores internos de cada paciente, es decir, la variabilidad de individuo a individuo depende mayormente de los factores inherentes precisamente de cada persona, entre los cuales se pueden considerar, si la paciente presenta alguna enfermedad que necesita constante control médico, es una persona hipocondriaca, etc., las que no fueron introducidas como covariables en el modelo. Por otro lado, a pesar que las dos restantes componentes explican muy poco de la variación total, es interesante entender qué factores pueden explicar dicho fenómeno, por ejemplo, dentro de los factores externos se puede mencionar la existencia de un plan de salud que requería que las personas se hicieran un chequeo,

entre otros. Cabe señalar también que, a medida que el valor esperado aumenta, la proporción de varianza asociada a los factores internos aumenta levemente.

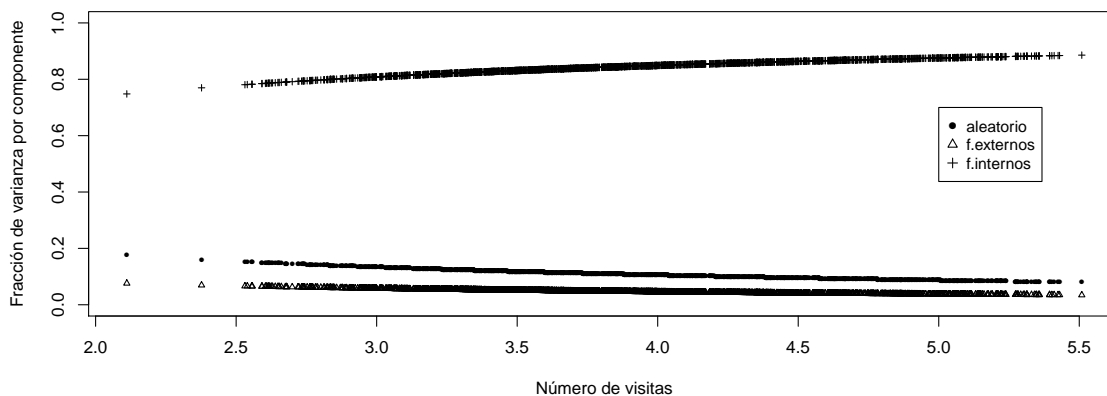


Figura 5.4: Relación entre las fracciones de componente de varianza y número de visitas al médico.

Finalmente, con el propósito de comparar los tres modelos presentados en este trabajo, la Figura 5.5 muestra las simulaciones envelope para los residuos de Pearson, para el modelo Waring (izquierda), para el modelo BN (centro) y para el modelo WG (derecha). En ellas se observa que los residuos están dentro de las bandas de confianza, tanto para el modelo Waring como para el modelo WG (siendo mejor el primero de éstos), no así para el modelo BN. Por otro lado, los valores de la función de log-verosimilitud entre los tres modelos son muy cercanas, sin embargo, un poco mayor la del modelo WG, eso adicionado a que este modelo permite la descomposición de la varianza en tres componentes, lo convierten en un modelo más atractivo, desde el punto de vista de la comprensión e interpretación de los resultados.

### 5.3.2. Aplicación a datos reales, fútbol

La variable de respuesta,  $y_i$ , corresponde al número de goles convertidos por 1224 futbolistas de la primera división de la Liga Española. Los datos fueron analizados previamente por Rodríguez-Avi et al. (2009) y se encuentran disponibles en el paquete

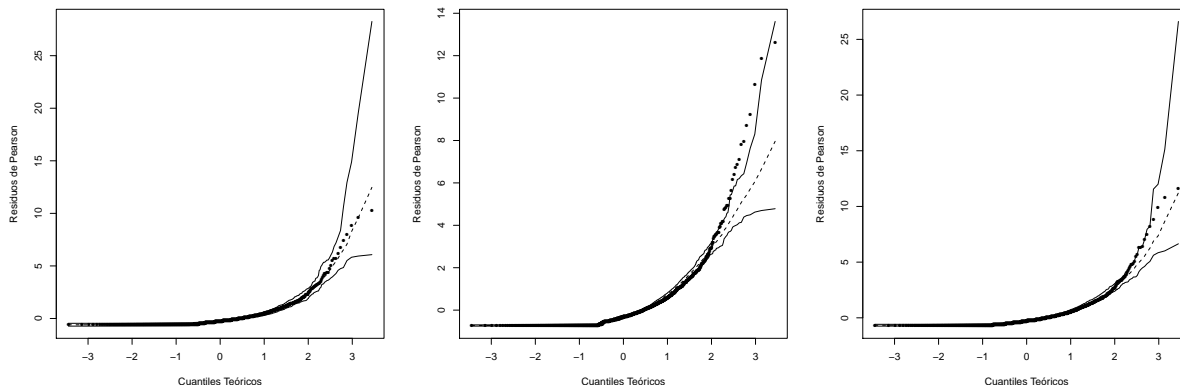


Figura 5.5: Envelope de los residuos de Pearson para el ajuste del modelo Waring (izquierda), Binomial Negativo (centro) y Waring Generalizada (derecha).

*GWRM* de R. En dicho trabajo, se propone el modelo de regresión Waring Generalizada, destacan que su principal cualidad es que permite analizar y obtener información más precisa acerca de la variabilidad observada, específicamente señalan que este modelo es útil cuando se puede justificar y/o se tiene claridad acerca de quiénes son los factores que pertenecen a la componente externa o a la interna, es decir, es posible separar dichos factores.

La información utilizada en este trabajo proviene de la revista deportiva *MARCA*, en particular se consideraron las temporadas 2000/2001 y 2006/2007. Como variables explicativas se incluyen:

- *clasif*: el puesto de clasificación del equipo respectivo por jugador, variable que toma valores entre 1 y 20.
- *position*: la posición en el campo de juego, defensa, delantero y mediocampista, donde claramente se deja afuera a los arqueros. Esta variable es considerada categórica.
- $\log(\textit{played})$ : el logaritmo del número de partidos jugados por cada deportista, el cual es ingresado adicionalmente al modelo como una variable offset.

De acuerdo a lo anterior, la media del modelo estará dada por,

$$\mu_i = played \times \exp\{\mathbf{x}_i^T \beta\} = \exp\{\mathbf{x}_i^T \beta + \log(played)\}$$

Es necesario señalar que si un jugador en particular jugó ambas temporadas, sólo se le contabiliza aquella donde convirtió más goles.

La Figura 5.6 muestra el número de goles convertidos durante las temporadas consideradas de acuerdo a la posición de los futbolistas en la cancha, en ella se observan varios puntos atípicos, entre ellos destacan las observaciones: 729, correspondiente a un defensa que convirtió 8 goles en 27 partidos jugados; 344, delantero que convirtió 29 goles en 38 partidos (en que 38 es el máximo de partidos jugados) y por último la observación 19, la cual corresponde a un mediocampista que convirtió 24 goles jugando 35 partidos.

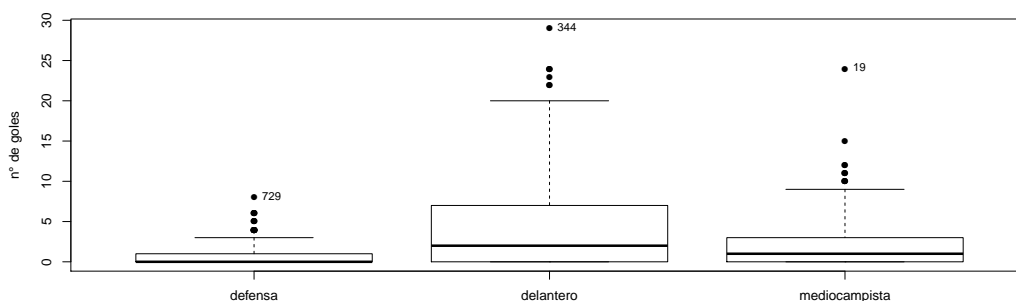


Figura 5.6: Gráfico de caja número de goles según posición del futbolista en el campo de juego.

En la Tabla 5.5 se presenta la información resumida por variable, de acuerdo a la posición del futbolista en la cancha, en que la media global (sin importar la posición) del número de goles convertidos es 2.21 con una desviación estándar de 3.71 goles.

En la Tabla 5.6 se observa la comparación entre los modelos Waring Generalizado, Binomial Negativo, Waring y Geométrico. En adelante no se considera el modelo Waring, pues de su ajuste se obtuvo un valor estimado para  $\phi$  próximo a 1 ( $\hat{\phi} = 1.000041$ ), con una desviación estándar de 0.2740, luego al contrastar la hipótesis nula  $H_0 : \phi = 1$ ,

Posición	Variable	Mínimo	Mediana	Media	Máximo	ds
defensa( $n=410$ )	<i>clasif</i>	1.00	12.00	11.87	20.00	6.11
	<i>played</i>	1.00	22.00	20.78	37.00	11.74
	<i>goals</i>	0.00	0.00	0.85	8.00	1.32
delantero( $n=281$ )	<i>clasif</i>	1.00	13.00	12.35	20.00	5.63
	<i>played</i>	1.00	22.00	20.32	38.00	12.11
	<i>goals</i>	0.00	2.00	4.78	29.00	5.81
mediocampo( $n=533$ )	<i>clasif</i>	1.00	13.00	12.02	20.00	5.76
	<i>played</i>	1.00	23.00	20.74	38.00	12.19
	<i>goals</i>	0.00	1.00	1.90	24.00	2.73

Tabla 5.5: Análisis descriptivo de las variables de acuerdo a la posición del jugador en la cancha.

el valor del estadístico asociado es aproximadamente 0, en cuyo caso el valor-p de la prueba es  $\approx 0.5$ . Por lo tanto, no existe evidencia en contra de la hipótesis nula, y dado que si  $\phi \rightarrow 1$  se obtiene el modelo Geométrico, como un caso particular del modelo Waring, se optará por el primero, que corresponde al modelo con menos parámetros. Conclusión concordante con la obtenida al aplicar el test de razón de verosimilitud, donde  $LR_c = 0.092$ , con valor-p=  $0.9866 > 0.05$ . También se desprende que el modelo que presenta menor AIC y mayor función de log-verosimilitud es el modelo WG. En particular, al aplicar el test de razón de verosimilitud ( $LR$ ) entre el modelo WG y BN se obtiene que el estadístico en cuestión es igual a 7.462, con valor-p=  $0.0063 < 0.05$ , por lo tanto se rechaza la hipótesis nula, lo que implica que el modelo WG es mejor que el modelo BN para describir los datos en considerados.

Modelo	n° de parámetros	$l$	AIC
Waring Generalizado	7	-1819.833	3653.667
Binomial Negativo	6	-1823.564	3659.129
Waring	6	-1871.179	3754.357
Geométrico	5	-1871.133	3752.267

Tabla 5.6: Resumen de los ajustes del modelo Waring Generalizado, Binomial Negativo, Waring y Geométrico.

La Tabla 5.7 muestra los resultados obtenidos del ajuste del modelo WG a los

datos, se observa que todas las covariables incluidas en el modelo son significativas, pues el valor-p para cada contraste individual,  $H_0 : \beta_j = 0$ , es mucho menor a 0.05.

parámetro	$\hat{\theta}$	ds	$z$ -valor	$l$	AIC
$\beta_0$	-5.7030	0.2979	-19.2455	-1819.833	3653.667
$\beta_1$	-0.0277	0.0052	-5.3477		
$\beta_2$	1.7259	0.0856	20.3221		
$\beta_3$	0.7713	0.0826	9.2779		
$\beta_4$	0.8555	0.0836	10.3383		
$\kappa$	4.5283	1.1773			
$\rho$	19.1885	4.9501			

Tabla 5.7: Resumen de los ajustes del modelo Waring Generalizado para los datos de fútbol.

La Tabla 5.8 muestra el número de goles observados y el número de goles esperados obtenidos al ajustar los modelos Waring Generalizado ( $E_i(a)$ ), Binomial Negativo ( $E_i(b)$ ) y Geométrico ( $E_i(c)$ ), respectivamente.

$y$	$O_i$	$E_i(a)$	$E_i(b)$	$E_i(c)$	$y$	$O_i$	$E_i(a)$	$E_i(b)$	$E_i(c)$
0	550	544.6889	536.6097	614.7931	15	6	3.6243	3.5198	3.8204
1	205	206.7991	217.8767	200.0766	16	2	2.9786	2.9074	3.2398
2	143	132.9946	134.8052	112.4232	17	4	2.4582	2.4140	2.7650
3	80	89.4972	89.3137	72.2837	18	1	2.0358	2.0130	2.3734
4	56	61.8855	61.2349	49.7048	19	3	1.6909	1.6845	2.0479
5	39	43.8468	43.1476	35.6971	20	2	1.4081	1.4137	1.7754
6	32	31.7919	31.1537	26.4646	21	0	1.1751	1.1893	1.5459
7	18	23.5550	22.9961	20.1144	22	2	0.9827	1.0023	1.3513
8	21	17.7997	17.3152	15.6023	23	1	0.8232	0.8461	1.1854
9	20	13.6883	13.2713	12.3113	24	4	0.6907	0.7152	1.0434
10	9	10.6878	10.3333	9.8579	25	0	0.5804	0.6052	0.9211
11	12	8.4536	8.1580	7.9946	26	0	0.4884	0.5125	0.8154
12	5	6.7592	6.5189	6.5562	27	0	0.4116	0.4344	0.7236
13	4	5.4532	5.2637	5.4299	28	0	0.3472	0.3684	0.6437
14	4	4.4322	4.2882	4.5368	29	1	1.9718	2.0881	5.9017

Tabla 5.8: Tabla de frecuencias observadas ( $O_i$ ) y esperadas ( $E_i$ ) para el modelo Waring Generalizado, Binomial Negativo y Geométrico.



Juntando las filas, para que se cumpla  $n * p_i \geq 5$ , con  $p_i = P(y_i|\theta)$ .

$O_i$	$E_i(a)$	$E_i(b)$	$E_i(c)$	$O_i$	$E_i(a)$	$E_i(b)$	$E_i(c)$
550	544.6889	536.6097	614.7931	20	13.6883	13.2713	12.3113
205	206.7991	217.8767	200.0766	9	10.6878	10.3333	9.8579
143	132.9946	134.8052	112.4232	12	8.4536	8.1580	7.9946
80	89.4972	89.3137	72.2837	13	16.6445	16.0708	16.5229
56	61.8855	61.2349	49.7048	6	3.6243	3.5198	3.8204
39	43.8468	43.1476	35.6972	6	5.4368	5.3214	6.0049
32	31.7919	31.1537	26.4646	6	6.3099	6.3005	7.7427
18	23.5550	22.9961	20.1144	8	6.296	6.5723	12.5857
21	17.7997	17.3152	15.6023				

Tabla 5.9: Tabla de frecuencias observadas ( $O_i$ ) y esperadas ( $E_i$ ) para el modelo Waring Generalizado, Binomial Negativo y Geométrico.

Los valores de los estadísticos de Pearson a partir de la Tabla 5.9 son:  $\chi_c^2 = 12.3690$ , con 9 g.l, con valor-p= 0.1933, para el modelo WG,  $\chi_c^2 = 13.4417$ , con 10 g.l, con valor-p=0.2000, para el modelo BN y  $\chi_c^2 = 31.3810$ , con 11 g.l, con valor-p=0.0010 para el modelo Geométrico. De acuerdo con ésto, resultan apropiados los modelos WG y BN; sin embargo, de acuerdo con el test de razón de verosimilitud, el modelo WG es más adecuado para describir a estos datos.

En la Figura 5.7, se pueden apreciar los valores observados versus los valores esperados de cada modelo presentados en la Tabla 5.9, en ella se observa como los correspondientes al modelo Waring Generalizado ajustan de mejor manera a los valores observados que los restantes modelos.

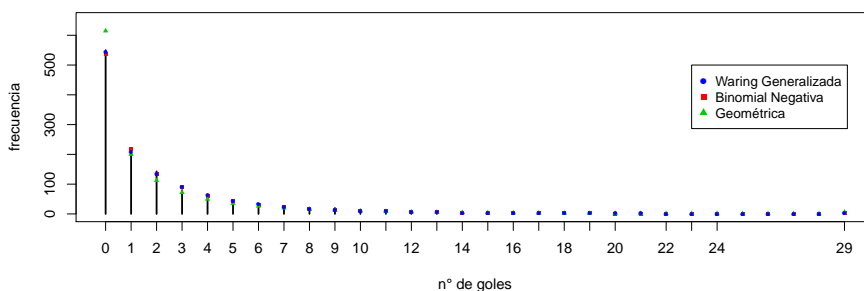


Figura 5.7: Número observado de goles ( $O_i$ ) y número esperado de goles ( $E_i$ ) por modelo ajustado.

La Figura 5.8 muestra los residuos de Pearson para el modelo WG, definidos a partir del estadístico de Pearson como,  $SR_i = \frac{O_i - E_i(a)}{\sqrt{E_i(a)}}$ ,  $i = 1, \dots, 17$ , en ella se observa que todos los valores están comprendidos dentro de las bandas -3 y 3 de manera aleatoria.

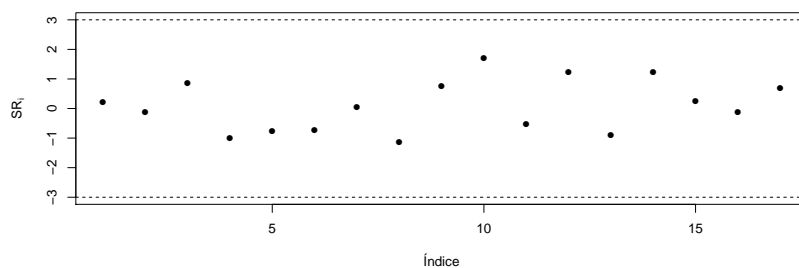


Figura 5.8: Residuos de Pearson para el ajuste del modelo Waring Generalizado.

En la Figura (5.9) se presentan las simulaciones envelope para los residuos de Pearson, tanto para el modelo BN, como para el modelo GW, donde se observa que ambos modelos ajustan bien; sin embargo el modelo WG tiene una leve ventaja, conclusión que apoya la decisión previa de que este modelo es mejor para describir a estos datos.

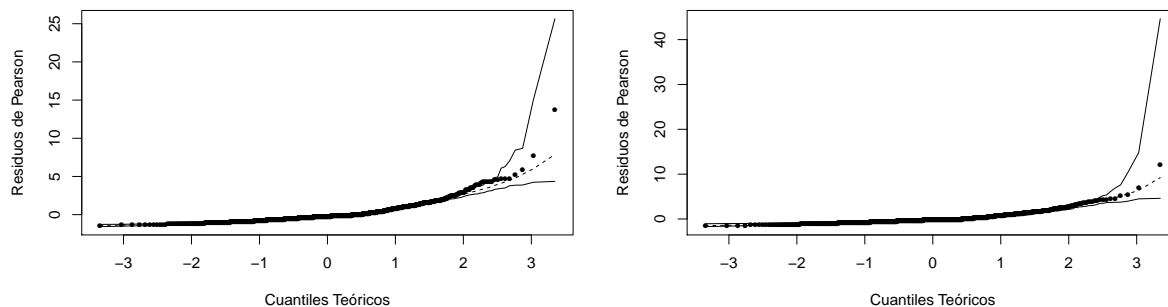


Figura 5.9: Envelope de los residuos de Pearson para el ajuste del modelo Binomial Negativo (izquierda) y Waring Generalizado (derecha).

Desde el punto de vista de la influencia global, se observa a partir de la Figura 5.10 que las potenciales observaciones influyentes son: 19, 125, 794, 804 y 979, donde el jugador 19 corresponde a un mediocampista que convirtió 24 goles en 35 partidos,

en que la media de goles convertidos en dicha posición alcanza sólo 1.90 goles. Los jugadores 125, 804 y 979 son delanteros que jugaron un alto número de partidos, 38, 36 y 34 respectivamente, pero sólo convirtieron 1, 1 y 0 goles, donde la media es de 4.78 goles. Mientras que el jugador 794 corresponde a un defensa que convirtió 5 goles en 15 partidos, en que la media de goles de este grupo es de sólo 0.85 goles.

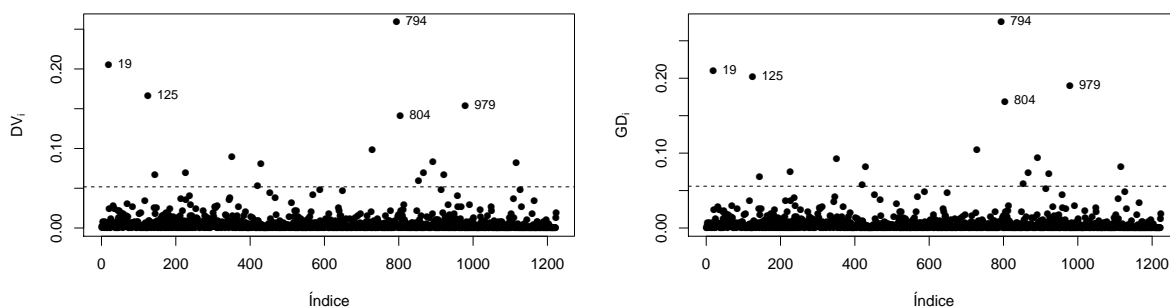


Figura 5.10: Gráfico índice distancia generalizada de Cook y de verosimilitud para  $\hat{\theta}$ .

La Figura 5.11 muestra el comportamiento de las observaciones bajo el esquema de ponderación de casos, a partir de ella se puede concluir que nuevamente los jugadores 19, 125, 794, 804 y 979 son lo que presentan una posible influencia en los EMV de los parámetros.

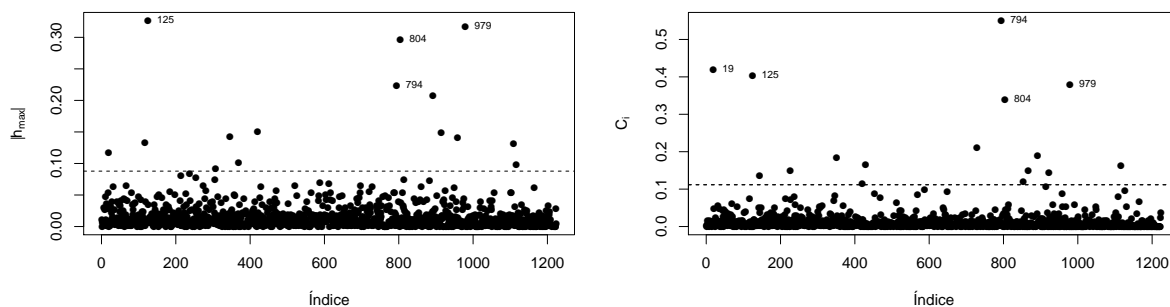


Figura 5.11: Gráfico índice de  $|h_{max}|$  y curvatura local total,  $C_i$ , para el esquema de ponderación de casos.  $C_i$  multiplicado por  $10^{-13}$ .

En cuanto al esquema de perturbación aditivo de la variable  $\log(\text{played})$ , se observa en la Figura 5.12 que el jugador 344 presenta una posible influencia en los EMV del modelo, dicho jugador corresponde a un delantero que jugó 38 partidos y

convirtió 29 goles.

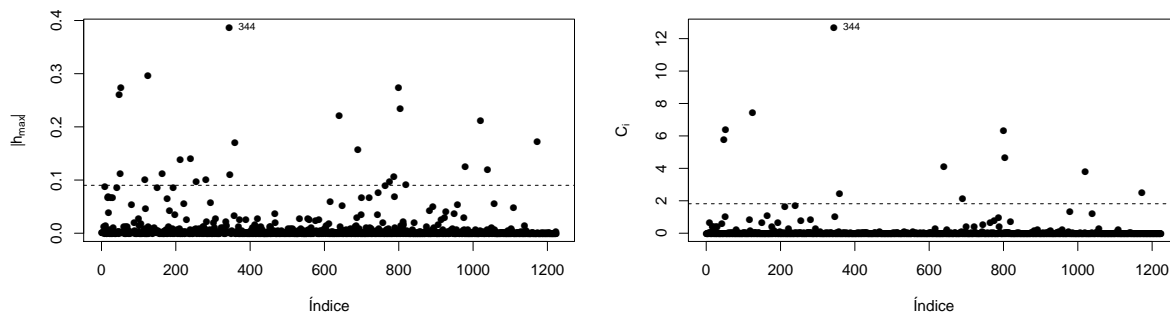


Figura 5.12: Gráfico índice de  $|h_{max}|$  y curvatura local total,  $C_i$ , para el esquema de perturbación aditiva de la variable  $\log(\text{played})$ .

De acuerdo a lo anterior, se consideró la eliminación de estas observaciones y de sus combinaciones del conjunto de datos; sin embargo, nunca hubo cambio en la significancia de los parámetros. Los principales cambios relativos en las estimaciones, siempre estuvieron relacionadas al parámetro  $\kappa$ , no a los parámetros involucrados en la media.

Adicionalmente, se calcularon las razones de cambio al eliminar a las potenciales observaciones influyentes del conjunto de datos, tanto de forma individual como la combinaciones de ellas. El mayor cambio de  $\kappa$  se produjo cuando se eliminó la observación 125 (16.5803%), mientras que el mayor cambio de  $\rho$  se produjo cuando se eliminó la observación 794 (19.6194%), cuando se analizó la eliminación de ambas observaciones, las respectivas razones de cambio disminuyeron considerablemente a 0.0721% y 9.5292%, respectivamente. Dado que la eliminación de estos subconjuntos de observaciones no arrojó cambios en la significancia de los parámetros no son presentados en este trabajo.

En la Tabla 5.10 se presenta la descomposición de la varianza y sus correspondientes proporciones para este conjunto de datos.

En la Tabla 5.11 se encuentra la descomposición de la varianza en proporciones

Fuente de variación	Varianza	Proporción de varianza
Aleatorio	$\mu_i$	$\frac{3.4263}{4.5283 + \mu_i}$
Factores Externos	$0.3216\mu_i$	$\frac{1.1020}{4.5283 + \mu_i}$
Factores Internos	$0.2919\mu_i^2$	$\frac{\mu_i}{4.5283 + \mu_i}$
Total	$1.3216 \left( \mu_i + \frac{\mu_i^2}{4.5283} \right)$	1

Tabla 5.10: Varianza y proporción de varianza por componente para el conjunto de datos del fútbol.

para las observaciones que fueron consideradas potencialmente influyentes, 19, 125, 344, 794, 804 y 979, coincidentemente las observaciones 125, 344 y 804 son las que presentan un mayor valor de la componente factores internos, es decir, la mayor contribución a la varianza lo ejercen las cualidades propias del individuo, se observa además, que ésta es mayor a medida que el número de partidos jugados aumenta.

	<i>clasif</i>	<i>played</i>	<i>position</i>	Aleatorio	F. Externos	F. Internos
19	4	35	mediocampista	0.37	0.12	0.51
125	8	38	delantero	0.20	0.06	0.74
344	3	38	delantero	0.18	0.06	0.76
794	17	15	defensa	0.71	0.23	0.07
804	7	36	delantero	0.21	0.07	0.72
979	12	34	delantero	0.25	0.08	0.67

Tabla 5.11: Resumen descomposición de la varianza para las observaciones 19, 125, 344, 794, 804 y 979.

Finalmente, en la Figura 5.13 se muestra la proporción de varianza atribuida a las covariables incluidas en el modelo, a los factores externos y a lo factores internos, de manera general y de acuerdo a la posición del jugador en la cancha. A partir de ella se puede concluir que:

- Sin importar la posición del jugador la proporción de varianza atribuida a la

aleatoriedad del juego y los factores externos decrece a medida que el número de goles aumenta, con lo cual aumenta la proporción relativa a los factores internos.

- Al comparar la proporción de aleatoriedad, se observa que ésta es mayor en los defensas, seguido por los mediocampistas y finalmente los delanteros.
- Apoyando lo presentado en la Tabla 5.11, se puede ver que los delanteros presentan una mayor proporción de la varianza relativa a los factores internos, luego los mediocampistas y por último los defensas.

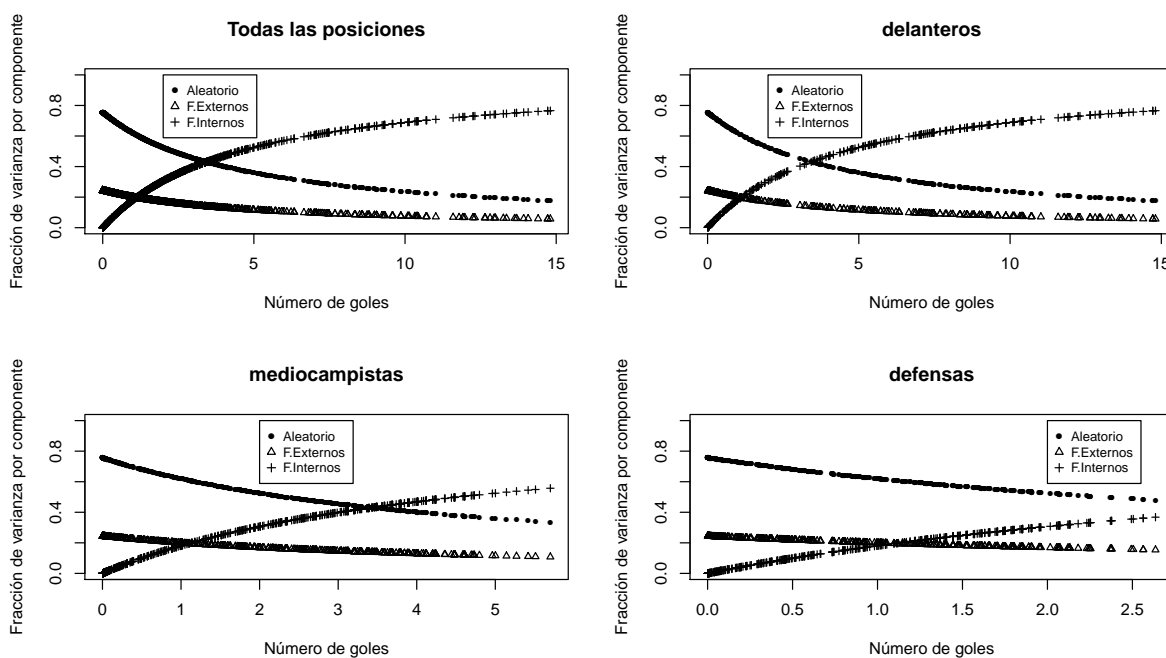


Figura 5.13: Gráfico de la razón de varianza para el conjunto de datos del fútbol.

### 5.3.2.1. Conclusión

A partir de estos datos fue posible analizar la descomposición de la varianza para cada individuo, ventaja principal del modelo Waring Generalizado, que considera los factores que no se pueden medir u observar, pero que se sabe influyen en la respuesta. Por ejemplo, para los factores internos es sabido que, además de la posición de jugador

en la cancha, su habilidad para convertir goles es un factor relevante; otro factor que no se puede medir, pero que se cree puede ser importante al momento de jugar es el estado de la cancha y la habilidad para el juego del equipo contrario (factores externos al individuo).

En el ámbito del análisis de influencia, si bien las observaciones detectadas no generaron cambios en la significancia de los EMV, se observó que la mayor proporción de varianza está relacionada con los factores internos del jugador y que ésta era mayor en los delanteros, la cual era precisamente, la posición de la mayoría de la observaciones detectadas.

Finalmente, con esta aplicación se quería mostrar el efecto de descomponer la varianza y qué sentido tiene aquello, así como la relación con las covariables involucradas.

## Capítulo 6

# Conclusiones y Trabajos Futuros

El objetivo central de esta tesis fue obtener expresiones de influencia global y local para el modelo de mezcla Waring y Waring Generalizado. Con tal propósito se presentó una descripción de cada modelo a estudiar, incluido el modelo Poisson-Gamma (Binomial Negativo). Se obtuvo la función generadora de momentos para cada modelo usado en este trabajo. Se establecieron las relaciones entre dichos modelos, notando que algunos correspondían a casos particulares de otros, destacando que el modelo más general es el modelo Waring Generalizado de parámetros  $(\vartheta_i, \kappa, \rho)$ , de manera tal que, si  $\kappa \rightarrow 1$ , se obtiene el modelo Waring  $(\mu_i, \phi)$ . Mientras que si  $\rho \rightarrow \infty$  se recupera el modelo Binomial Negativo  $(\mu_i, \gamma)$ . Por otro lado, si  $\gamma \rightarrow 1$  en un modelo BN, se obtiene como caso particular el modelo Geométrico  $(p_i)$ , el cual también se puede obtener cuando el parámetro  $\phi \rightarrow 1$  en el modelo Waring.

En particular, en el Capítulo 1 se analizó el modelo BN, el cual proviene de la mezcla Poisson-Gamma, donde el modelo Poisson es utilizado para modelar datos de conteo en que la tasa de ocurrencia es igual para todos los individuos; luego, en presencia de heterogeneidad en los datos, la que no puede ser explicada por las características observables e incluidas en el modelo provocan que dicho modelo no sea adecuado. Cuando ésto ocurre es necesario modelar la heterogeneidad, en caso que la variabilidad sea mayor que la media (sobredispersión) una manera de solucionarlo



es asumir que la tasa de ocurrencia tiene distribución Gamma. Con ello se ha creado una nueva fuente de variación, diferente a la que puede ser explicada por las covariables observadas e introducidas en el modelo. Luego, es lógico tratar de explicar a qué se debe esta nueva fuente de variación, Irwin (1968) y Xekalaki (1983b), dan respuesta a esto introduciendo el modelo WG, dicho modelo considera las tres fuentes de variación mencionadas en este trabajo, la primera explicada por las covariables (azar propio del fenómeno), los factores internos y los factores externos, estas dos últimas fuentes corresponden a la subdivisión de la fuente adicional creada para el modelo BN. En el modelo WG los factores internos corresponden a aquellos inherentes a cada individuo, mientras que los factores externos tienen relación con la exposición del individuo a cierto fenómeno, pero que no dependen de él.

En cuanto a la estimación de máxima verosimilitud de los parámetros se optó y se utilizaron algoritmos, tales como, Newton-Rapshon, IRLS y algoritmo EM, los cuales se implementaron en el software libre R. Para el modelo de regresión Binomial Negativo y Waring, la estimación de los parámetros se obtuvo maximizando la función de log-verosimilitud completa, a través del algoritmo EM. Mientras que, para la obtención de las estimaciones a través de la maximización de la función log-verosimilitud observada para el modelo Waring, se modificó la función propuesta por Rodríguez-Avi et al. (2009), *GWRM.fit*, para el modelo de regresión Waring Generalizada.

En esta tesis se presentan expresiones de influencia global y local para cada uno de los modelos presentados. Particularmente, en el enfoque local, se discute la importancia de obtener un esquema de perturbación adecuado, es decir, que la matriz de información de Fisher esperada en el modelo perturbado, evaluada en  $\omega_0$ , sea diagonal, es decir,  $\mathbf{G}(\omega_0) = c\mathbf{I}_n$ . La metodología desarrollada en este trabajo fue implementada en el software libre R.

La aplicación transversal de este trabajo se denominó visitas al médico, se con-

sideraron 3 variables explicativas, Ingreso per-cápita, dividido en 10, expresado en marcos, la Edad y los años de Educación, mientras que la variable de respuesta es el número de visitas al médico durante el año 1987. Se observó que las medidas de influencia coinciden con señalar a la paciente 285 como potencial observación influyente. En particular, los modelos Waring y WG generan cambios en la significancia de la variable Ingresos cuando esta observación es extraída del conjunto de datos. En cuanto a las medidas de ajuste, comparativamente con el modelo BN y Waring, el modelo WG presenta mayor valor de la función de log-verosimilitud.

Adicionalmente a todo lo desarrollado en este trabajo, quedan abiertas las posibilidades de complementar la comparación de estos modelos con:

- Repetir la aplicación del número de visitas al médico, considerando covariables dicotómicas, tales como, si la paciente trabaja o no, si tiene o no hijos, etc., dado que en particular en el modelo de regresión Waring Generalizado se genera una división en tres fuentes de variación, permitiendo con ésto obtener el porcentaje de variación respectiva a cada fuente.
- Obtener e implementar el algoritmo EM y, a partir de este enfoque obtener las expresiones de las herramientas de influencia global y local para el modelo WG.
- Obtener la versión Waring Zero Inflado y Waring Generalizado Zero Inflado y compararlo con el modelo Binomial Negativo Zero Inflado y Poisson Zero Inflado, dada la alta frecuencia observada que alcanza la variable respuesta cuando toma el valor 0.

## Apéndice A

# Función Generadora de Momentos

### Función Generadora de Momentos de la distribución Waring

En esta sección se demuestra la **Proposición 1** y a partir de ella se obtienen la media y varianza de la distribución Waring.

**Dem. Proposición 1:** La función generadora de momentos (FGM) de la distribución Waring por definición está dada por,

$$M_y(t^*) = E(e^{t^*y}) = \sum_{y=0}^{\infty} e^{t^*y} P(y|\alpha, \beta),$$

donde  $P(y|\alpha, \beta)$  corresponde a la función de probabilidad de la distribución Waring, que es obtenida de la mezcla dada en (2.13), es decir,

$$P(y|\alpha, \beta) = \int_0^1 p(1-p)^y \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1} dp,$$

con lo cual,

$$M_y(t^*) = \sum_{y=0}^{\infty} e^{t^*y} \int_0^1 p(1-p)^y \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1} dp,$$

de donde,

$$\begin{aligned} M_y(t^*) &= \lim_{n \rightarrow \infty} \sum_{y=0}^n e^{t^*y} \int_0^1 p(1-p)^y \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1} dp \\ &= \int_0^1 \lim_{n \rightarrow \infty} \sum_{y=0}^n e^{t^*y} p(1-p)^y \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1} dp. \end{aligned}$$

Dado que,

$$e^{t^*y} p(1-p)^y \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1} \geq 0,$$

es posible definir una sucesión creciente, tal que,

$$A_n = \sum_{y=0}^n e^{t^*y} p(1-p)^y \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1},$$

con lo cual se tiene que,  $0 \leq A_n \leq A_{n+1}$ , por lo tanto, por teorema de convergencia monótona,

$$M_y(t^*) = \int_0^1 \sum_{y=0}^{\infty} e^{t^*y} p(1-p)^y \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1} dp.$$

Luego,

$$\begin{aligned} M_y(t^*) &= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1} \left( \sum_{y=0}^{\infty} e^{t^*y} p(1-p)^y \right) dp \\ &= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1} \frac{p}{1-(1-p)e^{t^*}} dp, \end{aligned}$$

lo anterior, debido a que,

$$M_{y|p}(t^*) = \frac{p}{1-(1-p)e^{t^*}}, \text{ ya que } y|p \sim \text{Geo}(p).$$

Finalmente,

$$M_y(t^*) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \frac{p^\alpha (1-p)^{\beta-1}}{1-(1-p)e^{t^*}} dp. \quad \# \quad (\text{A.1})$$

A partir de (A.1) es posible obtener la media y varianza de la distribución Waring,

dadas en (2.12). Note que,

$$\begin{aligned}
 \int_0^1 \frac{\partial}{\partial t^*} \frac{p^\alpha (1-p)^{\beta-1}}{1-(1-p)e^{t^*}} \Big|_{t^*=0} dp &= \int_0^1 \frac{p^\alpha (1-p)^\beta e^{t^*}}{(1-(1-p)e^{t^*})^2} \Big|_{t^*=0} dp \\
 &= \int_0^1 p^{\alpha-2} (1-p)^\beta dp \\
 &= \frac{\Gamma(\alpha-1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta)} \underbrace{\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha-1)\Gamma(\beta+1)} p^{\alpha-2} (1-p)^\beta dp}_{\sim \text{Beta}(\alpha-1, \beta+1)} \\
 &= \frac{\Gamma(\alpha-1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta)}.
 \end{aligned}$$

Por lo tanto,  $\frac{\partial}{\partial t^*} \frac{p^\alpha (1-p)^{\beta-1}}{1-(1-p)e^{t^*}} \Big|_{t^*=0}$  es integrable.

Luego,

$$\begin{aligned}
 E(y) &= \frac{\partial M_y(t^*)}{\partial t^*} \Big|_{t^*=0} \\
 &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \frac{\partial}{\partial t^*} \frac{p^\alpha (1-p)^{\beta-1}}{1-(1-p)e^{t^*}} \Big|_{t^*=0} dp \quad (\text{A.2}) \\
 &= \frac{\beta}{\alpha-1}.
 \end{aligned}$$

Análogamente, se puede mostrar que  $\frac{\partial^2}{\partial t^{*2}} \frac{p^\alpha (1-p)^{\beta-1}}{1-(1-p)e^{t^*}} \Big|_{t^*=0}$  es integrable, y por lo tanto que el segundo momento está dado por,

$$\begin{aligned}
E(y^2) &= \frac{\partial^2 M_y(t^*)}{\partial t^{*2}} \Big|_{t^*=0} \\
&= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \frac{\partial^2}{\partial t^{*2}} \frac{p^\alpha(1-p)^{\beta-1}}{1-(1-p)e^{t^*}} \Big|_{t^*=0} dp \\
&= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \frac{p^\alpha(1-p)^\beta e^{t^*} (1-(1-p)e^{t^*})^2 + 2p^\alpha(1-p)^{\beta+1}(1-(1-p)e^{t^*})e^{2t^*}}{(1-(1-p)e^{t^*})^4} \Big|_{t^*=0} dp \\
&= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \frac{p^{\alpha+2}(1-p)^\beta + 2p^{\alpha+1}(1-p)^{\beta+1}}{p^4} dp \\
&= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 (p^{\alpha-2}(1-p)^\beta + 2p^{\alpha-3}(1-p)^{\beta+1}) dp \\
&= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \left[ \frac{\Gamma(\alpha - 1)\Gamma(\beta + 1)}{\Gamma(\alpha + \beta)} \underbrace{\int_0^1 \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha - 1)\Gamma(\beta + 1)} p^{\alpha-2}(1-p)^\beta dp}_{\sim \text{Beta}(\alpha-1, \beta+1)} \right. \\
&\quad \left. + \frac{2\Gamma(\alpha - 2)\Gamma(\beta + 2)}{\Gamma(\alpha + \beta)} \underbrace{\int_0^1 \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha - 2)\Gamma(\beta + 2)} p^{\alpha-3}(1-p)^{\beta+1} dp}_{\sim \text{Beta}(\alpha-2, \beta+2)} \right] \\
&= \frac{\beta}{\alpha - 1} + \frac{2(\beta + 1)\beta}{(\alpha - 1)(\alpha - 2)}.
\end{aligned} \tag{A.3}$$

A partir de, (A.2) y (A.3), se obtiene la varianza de la distribución,

$$\begin{aligned}
Var(y) &= \mathbf{E}(y^2) - \mathbf{E}^2(y) \\
&= \frac{\beta}{\alpha - 1} + \frac{2(\beta + 1)\beta}{(\alpha - 1)(\alpha - 2)} - \frac{\beta^2}{(\alpha - 1)^2} \\
&= \frac{\alpha\beta(\alpha + \beta - 1)}{(\alpha - 1)^2(\alpha - 2)},
\end{aligned} \tag{A.4}$$

donde (A.2) y (A.4) coincide con (2.12).

## Función Generadora de Momentos de la distribución Waring Generalizada

En esta sección se demuestra la **Proposición 2** y a partir de ella se obtienen la media y varianza de la distribución Waring Generalizada.

**Dem. Proposición 2:** Por definición la Función Generadora de Momentos (FGM) para esta distribución, en que  $y \sim \text{WG}(\vartheta, \kappa, \rho)$ , estará dada por,

$$M_y(t^*) = \mathbf{E}(e^{t^*y}) = \sum_{y=0}^{\infty} e^{t^*y} P(y|\vartheta, \kappa, \rho),$$

donde  $P(y|\vartheta, \kappa, \rho)$  corresponde a la función de probabilidad de la distribución Waring Generalizada, resultante de la mezcla dada en (2.19), es decir,

$$P(y|\vartheta, \kappa, \rho) = \int_0^{\infty} \frac{\Gamma(y + \vartheta)}{y!\Gamma(\vartheta)} \left(\frac{1}{1+v}\right)^{\vartheta} \left(\frac{v}{1+v}\right)^y \frac{\Gamma(\kappa + \rho)}{\Gamma(\kappa)\Gamma(\rho)} v^{\kappa-1} (1+v)^{-(\kappa+\rho)} dv,$$

con lo cual, la FGM estará dada por,

$$M_y(t^*) = \sum_{y=0}^{\infty} e^{t^*y} \int_0^{\infty} \frac{\Gamma(y + \vartheta)}{y!\Gamma(\vartheta)} \left(\frac{1}{1+v}\right)^{\vartheta} \left(\frac{v}{1+v}\right)^y \frac{\Gamma(\kappa + \rho)}{\Gamma(\kappa)\Gamma(\rho)} v^{\kappa-1} (1+v)^{-(\kappa+\rho)} dv,$$

con

$$\begin{aligned} M_y(t^*) &= \lim_{n \rightarrow \infty} \sum_{y=0}^n e^{t^*y} \int_0^{\infty} \frac{\Gamma(y + \vartheta)}{y!\Gamma(\vartheta)} \left(\frac{1}{1+v}\right)^{\vartheta} \left(\frac{v}{1+v}\right)^y \frac{\Gamma(\kappa + \rho)}{\Gamma(\kappa)\Gamma(\rho)} v^{\kappa-1} (1+v)^{-(\kappa+\rho)} dv \\ &= \int_0^{\infty} \lim_{n \rightarrow \infty} \sum_{y=0}^n e^{t^*y} \frac{\Gamma(y + \vartheta)}{y!\Gamma(\vartheta)} \left(\frac{1}{1+v}\right)^{\vartheta} \left(\frac{v}{1+v}\right)^y \frac{\Gamma(\kappa + \rho)}{\Gamma(\kappa)\Gamma(\rho)} v^{\kappa-1} (1+v)^{-(\kappa+\rho)} dv \\ &= \int_0^{\infty} \lim_{n \rightarrow \infty} \sum_{y=0}^n M_1 dv, \end{aligned}$$

pues,  $M_1 = e^{t^*y} \frac{\Gamma(y + \vartheta)}{y!\Gamma(\vartheta)} \left(\frac{1}{1+v}\right)^{\vartheta} \left(\frac{v}{1+v}\right)^y \frac{\Gamma(\kappa + \rho)}{\Gamma(\kappa)\Gamma(\rho)} v^{\kappa-1} (1+v)^{-(\kappa+\rho)} \geq 0$ . Además,

definiendo la sucesión,  $A_n = \sum_{y=0}^n M_1$ , se tiene que,  $0 \leq A_n \leq A_{n+1}$ ; por lo tanto, por teorema de convergencia monótona,

$$M_y(t^*) = \int_0^{\infty} \sum_{y=0}^{\infty} e^{t^*y} \frac{\Gamma(y + \vartheta)}{y!\Gamma(\vartheta)} \left(\frac{1}{1+v}\right)^{\vartheta} \left(\frac{v}{1+v}\right)^y \frac{\Gamma(\kappa + \rho)}{\Gamma(\kappa)\Gamma(\rho)} v^{\kappa-1} (1+v)^{-(\kappa+\rho)} dv.$$

Luego,

$$\begin{aligned} M_y(t^*) &= \int_0^\infty \frac{\Gamma(\kappa + \rho)}{\Gamma(\kappa)\Gamma(\rho)} v^{\kappa-1} (1+v)^{-(\kappa+\rho)} \left[ \sum_{y=0}^\infty e^{t^*y} \frac{\Gamma(y+\vartheta)}{y!\Gamma(\vartheta)} \left(\frac{1}{1+v}\right)^\vartheta \left(\frac{v}{1+v}\right)^y \right] dv \\ &= \int_0^\infty \frac{\Gamma(\kappa + \rho)}{\Gamma(\kappa)\Gamma(\rho)} v^{\kappa-1} (1+v)^{-(\kappa+\rho)} \left(\frac{1}{1+v - ve^{t^*}}\right)^\vartheta dv, \end{aligned}$$

pues,

$$M_{y|v}(t^*) = \left(\frac{1}{1+v - ve^{t^*}}\right)^\vartheta, \text{ ya que } y|v \sim \text{BN}(\vartheta, p), \text{ con } p = \frac{1}{1+v}.$$

Finalmente,

$$M_y(t^*) = \frac{\Gamma(\kappa + \rho)}{\Gamma(\kappa)\Gamma(\rho)} \int_0^\infty v^{\kappa-1} (1+v)^{-(\kappa+\rho)} \left(\frac{1}{1+v - ve^{t^*}}\right)^\vartheta dv. \quad \# \quad (\text{A.5})$$

A partir de (A.5) se puede obtener la media y varianza de la distribución Waring Generalizada.

Sea,

$$\begin{aligned} &\int_0^\infty \frac{\partial}{\partial t^*} v^{\kappa-1} (1+v)^{-(\kappa+\rho)} \left(\frac{1}{1+v - ve^{t^*}}\right)^\vartheta \Big|_{t^*=0} dv \\ &= \int_0^\infty v^{\kappa-1} (1+v)^{-(\kappa+\rho)} \vartheta \left(\frac{1}{1+v - ve^{t^*}}\right)^{\vartheta-1} \frac{\vartheta e^{t^*}}{(1+v - ve^{t^*})^2} \Big|_{t^*=0} dv \\ &= \vartheta \int_0^\infty v^\kappa (1+v)^{-(\kappa+\rho)} dv \\ &= \vartheta \frac{\Gamma(\kappa+1)\Gamma(\rho-1)}{\Gamma(\kappa+\rho)} \underbrace{\int_0^\infty \frac{\Gamma(\kappa+\rho)}{\Gamma(\kappa+1)\Gamma(\rho-1)} v^\kappa (1+v)^{-(\kappa+\rho)} dv}_{\sim \text{BetaII}(\kappa+1, \rho-1)} \\ &= \vartheta \frac{\Gamma(\kappa+1)\Gamma(\rho-1)}{\Gamma(\kappa+\rho)}. \end{aligned}$$

Por lo tanto,  $\frac{\partial}{\partial t^*} v^{\kappa-1} (1+v)^{-(\kappa+\rho)} \left(\frac{1}{1+v - ve^{t^*}}\right)^\vartheta \Big|_{t^*=0}$  es integrable.

Entonces,

$$\begin{aligned} E(y) &= \frac{\partial M_y(t^*)}{\partial t^*} \Big|_{t^*=0} \\ &= \frac{\Gamma(\kappa + \rho)}{\Gamma(\kappa)\Gamma(\rho)} \int_0^\infty \frac{\partial}{\partial t^*} v^{\kappa-1} (1+v)^{-(\kappa+\rho)} \left(\frac{1}{1+v - ve^{t^*}}\right)^\vartheta dv \\ &= \frac{\vartheta \kappa}{\rho - 1}. \end{aligned} \quad (\text{A.6})$$



De la misma manera se puede mostrar que  $\frac{\partial^2}{\partial t^{*2}} v^{\kappa-1}(1+v)^{-(\kappa+\rho)} \left( \frac{1}{1+v-ve^{t^*}} \right)^\vartheta \Big|_{t^*=0}$  es integrable, y por lo tanto que el segundo momento está dado por,

$$\begin{aligned}
E(y^2) &= \frac{\partial^2 M_y(t)}{\partial t^{*2}} \Big|_{t^*=0} \\
&= \int_0^\infty \frac{\Gamma(\kappa+\rho)}{\Gamma(\kappa)\Gamma(\rho)} v^{\kappa-1}(1+v)^{-(\kappa+\rho)} \left[ \vartheta(\vartheta-1) \left( \frac{1}{1+v-ve^{t^*}} \right)^{\vartheta-2} \frac{v^2 e^{2t^*}}{(1+v-ve^{t^*})^4} \right. \\
&\quad \left. + \vartheta \left( \frac{1}{1+v-ve^{t^*}} \right)^{\vartheta-1} \left[ \frac{ve^{t^*}(1+v-ve^{t^*})^2 + 2(1+v-ve^{t^*})v^2 e^{2t^*}}{(1+v-ve^{t^*})^4} \right] \right] \Big|_{t^*=0} dv \\
&= \int_0^\infty \frac{\Gamma(\kappa+\rho)}{\Gamma(\kappa)\Gamma(\rho)} v^{\kappa-1}(1+v)^{-(\kappa+\rho)} (\vartheta^2 v^2 + \vartheta v + \vartheta v^2) dv \\
&= (\vartheta^2 + \vartheta) \frac{\Gamma(\kappa+\rho)}{\Gamma(\kappa)\Gamma(\rho)} \frac{\Gamma(\kappa+2)\Gamma(\rho-2)}{\Gamma(\kappa+\rho)} \underbrace{\int_0^\infty \frac{\Gamma(\kappa+\rho)}{\Gamma(\kappa+2)\Gamma(\rho-2)} v^{\kappa+1}(1+v)^{-(\kappa+\rho)} dv}_{\sim \text{BetaII}(\kappa+2, \rho-2)} \\
&\quad + \vartheta \frac{\Gamma(\kappa+\rho)}{\Gamma(\kappa)\Gamma(\rho)} \frac{\Gamma(\kappa+1)\Gamma(\rho-1)}{\Gamma(\kappa+\rho)} \underbrace{\int_0^\infty \frac{\Gamma(\kappa+\rho)}{\Gamma(\kappa+1)\Gamma(\rho-1)} v^\kappa (1+v)^{-(\kappa+\rho)} dv}_{\sim \text{BetaII}(\kappa+1, \rho-1)} \\
&= \frac{(\vartheta^2 + \vartheta)(\kappa+1)\kappa}{(\rho-1)(\rho-2)} + \frac{\vartheta\kappa}{\rho-1}.
\end{aligned} \tag{A.7}$$

Utilizando (A.6) y (A.7), se obtiene la varianza de la distribución,

$$\begin{aligned}
Var(y) &= \mathbf{E}(y^2) - \mathbf{E}^2(y) \\
&= \frac{(\vartheta^2 + \vartheta)(\kappa+1)\kappa}{(\rho-1)(\rho-2)} + \frac{\vartheta\kappa}{\rho-1} - \frac{\vartheta^2 \kappa^2}{(\rho-1)^2} \\
&= \frac{\vartheta\kappa}{\rho-1} + \frac{\vartheta\kappa(\kappa+1)}{(\rho-1)(\rho-2)} + \frac{\vartheta\kappa(\kappa+\rho-1)}{(\rho-1)^2(\rho-2)},
\end{aligned} \tag{A.8}$$

donde (A.6) y (A.8) coincide con (2.20) y (2.21), respectivamente.

## Apéndice B

# Función score y matriz hessiana: Modelo Waring

En esta sección se obtiene la función score y la matriz de información de Fisher para el vector de parámetros  $\theta = (\beta^T, \phi)^T$  del modelo Waring.

### Función score

A partir de la ecuación (2.15) se obtiene que,

$$\frac{\partial l(\theta)}{\partial \beta_j} = \frac{\partial l(\beta, \phi)}{\partial \beta_j} = \sum_{i=1}^n \frac{l_i(\mu_i, \phi)}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j}, \quad \forall j. \quad (\text{B.1})$$

Note que  $\frac{\partial \mu_i}{\partial \eta_i} = \frac{1}{g'(\mu_i)}$ , luego, derivando la ecuación (2.16) con respecto a  $\mu_i$ .

$$\frac{l_i(\mu_i, \phi)}{\partial \mu_i} = \phi_2(a_i - b_i), \quad (\text{B.2})$$

donde  $a_i = \psi(\phi_1 + \mu_i \phi_2) - \psi(y_i + \phi_1 + \mu_i \phi_2 + 1)$ ,  $b_i = \psi(\mu_i \phi_2) - \psi(y_i + \mu_i \phi_2)$  y  $S_i = \frac{1}{g'(\mu_i)}$  y  $\psi(\cdot)$  corresponde a la función digamma dada por,  $\psi(z) = \frac{\partial \log \Gamma(z)}{\partial z}$ ,  $z > 0$ .

Sustituyendo la ecuación (B.2) en (B.1), se obtiene,

$$\frac{\partial l(\theta)}{\partial \beta_j} = \frac{\partial l(\beta, \phi)}{\partial \beta_j} = \phi_2 \sum_{i=1}^n (a_i - b_i) \frac{1}{g'(\mu_i)} x_{ij}. \quad (\text{B.3})$$

La ecuación (B.3) escrita en forma matricial estará dada por,

$$\mathbf{U}_\beta = \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \phi_2 \mathbf{X}^T \mathbf{J}_2 (\mathbf{a} - \mathbf{b}), \quad (\text{B.4})$$

donde,  $\mathbf{J}_2 = \text{diag} \left( \frac{1}{g'(\mu_1)}, \dots, \frac{1}{g'(\mu_n)} \right)$ . Note que la ecuación (B.4) coincide con (4.1).

Análogamente, se obtiene la función score para  $\phi$ , obteniéndose lo expresado en (4.2).

## Matrix hessiana

Luego, a partir de (B.1), se puede obtener la segunda derivada de  $l(\boldsymbol{\theta})$  con respecto a  $\boldsymbol{\beta}$ ,

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \beta_k \partial \beta_j} &= \sum_{i=1}^n \frac{\partial}{\partial \mu_i} \left( \frac{\partial l_i(\mu_i, \phi)}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \right) \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} x_{ij} \\ &= \sum_{i=1}^n \left( \frac{\partial^2 l_i(\mu_i, \phi)}{\partial \mu_i^2} \frac{\partial \mu_i}{\partial \eta_i} + \frac{\partial l_i(\mu_i, \phi)}{\partial \mu_i} \frac{\partial}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \right) \frac{\partial \mu_i}{\partial \eta_i} x_{ik} x_{ij}. \end{aligned}$$

Como,  $\mathbf{E} \left( \frac{\partial l(\boldsymbol{\theta})}{\partial \mu_i} \right) = 0$ , se tiene que,

$$\begin{aligned} \mathbf{E} \left( -\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \beta_k \partial \beta_j} \right) &= -\sum_{i=1}^n \mathbf{E} \left( \frac{\partial^2 l_i(\mu_i, \phi)}{\partial \mu_i^2} \right) \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2 x_{ik} x_{ij} \\ &= -\phi_2^2 \sum_{i=1}^n \mathbf{E}(d_i) \frac{1}{g'(\mu_i)^2} x_{ik} x_{ij}, \end{aligned}$$

donde,  $d_i = a_i^* - b_i^*$ , con  $a_i^* = \psi'(\phi_1 + \mu_i \phi_2) - \psi'(y_i + \phi_1 + \mu_i \phi_2 + 1)$ ,

$b_i = \psi'(\mu_i \phi_2) - \psi'(y_i + \mu_i \phi_2)$  y  $\psi'(\cdot)$  corresponde a la función trigamma. En forma matricial,

$$K_{\beta\beta} = -\phi_2^2 \mathbf{X}^T \mathbf{M} \mathbf{X},$$

donde,  $m_i = \mathbf{E}(d_i) \frac{1}{g'(\mu_i)^2}$ .

A partir de (B.4) se obtiene la segunda derivada de la log-verosimilitud con respecto a  $\phi$ , la que puede escribirse como,

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \phi \partial \beta_j} = \frac{-2}{(\phi - 1)^2} \sum_{i=1}^n (a_i - b_i) \frac{\mu_i}{g'(\mu_i)} x_{ij} - 2\phi^2 \sum_{i=1}^n \left( a_i^* \frac{\mu_i + 1}{(\phi - 1)^2} - b_i^* \frac{\mu_i}{(\phi - 1)^2} \right) \frac{1}{g'(\mu_i)} x_{ij},$$

Luego,

$$K_{\phi\beta} = \mathbf{E} \left( -\frac{\partial^2 l(\boldsymbol{\beta}, \phi)}{\partial \phi \partial \boldsymbol{\beta}} \right)$$

Finalmente, la segunda derivada de (4.2) con respecto a  $\phi$  está dada por,

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \phi \partial \phi} = \sum_{i=1}^n \left( -\frac{1}{\phi^2} + \frac{1}{(\phi - 1)^2} + 4 \frac{(\mu_i + 1)^2}{(\phi - 1)^4} a_i^* + 4 \frac{\mu_i + 1}{(\phi - 1)^3} a_i - 4 \frac{\mu_i^2}{(\phi - 1)^4} b_i^* - 4 \frac{\mu_i}{(\phi - 1)^3} b_i \right).$$

Por lo tanto,

$$K_{\phi\phi} = \mathbf{E} \left( -\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \phi \partial \phi} \right).$$

Obteniéndose los elementos de la matriz de Información de Fisher esperada para  $\boldsymbol{\theta}$  dada en la ecuación (4.3)

# Referencias

- Agresti, A. (1996). *Categorical Data Analysis, Second Edition*. New York: John Wiley and Sons.
- Ajiferuke, I. (1991). A probabilistic model for the distribution of authorships. *Journal of the American Society for Information Science*, 42(4):279–289.
- Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723.
- Ariza-López, F. and Rodríguez-Avi, J. (2015). Estimating the count of completeness errors in geographic data sets by means of a generalized waring regression model. *International Journal of Geographical Information Science*, 29(8):1394–1418.
- Atkinson, A. C. (1985). *Plots, transformations, and regression: an introduction to graphical methods of diagnostic regression analysis*. Clarendon Press Oxford.
- Baird, D. D. and Wilcox, A. J. (1985). Cigarette smoking associated with delayed conception. *Jama*, 253(20):2979–2983.
- Barreto-Souza, W. and Simas, A. B. (2015). General mixed Poisson regression models with varying dispersion. *Statistics and Computing*, 26(6):1–18.
- Bartle, R. G. and Sherbert, D. R. (2000). *Introduction to Real Analysis*. New York: John Wiley and Sons.

- Burrell, Q. L. (1990). Using the gamma-Poisson model to predict library circulations. *Journal of the American Society for Information Science and Technology*, 41:164–170.
- Cameron, A. C. and Trivedi, P. K. (1986). Econometric models based on count data. Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics*, 1(1):29–53.
- Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbury Pacific Grove, CA.
- Chen, F., Zhu, H.-T., and Lee, S.-Y. (2009). Perturbation selection and local influence analysis for nonlinear structural equation model. *Psychometrika*, 74(3):493–516.
- Chen, J. J. and Ahn, H. (1996). Fitting Mixed Poisson Regression Models Using Quasi-Likelihood Methods. *Biometrical journal*, 38(1):81–96.
- Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1):15–18.
- Cook, R. D. (1986). Assessment of Local Influence. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(2):133–169.
- Crouchley, R. and Dassios, A. (1998). Interpreting the Beta Geometric in Comparative Fecundability Studies. *Biometrics*, 54(1):161–167.
- De Bastiani, F., de Aquino Cysneiros, A. H. M., Uribe-Opazo, M. A., and Galea, M. (2015). Influence diagnostics in elliptical spatial linear models. *TEST*, 24(2):322–340.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38.

- Escobar, L. A. and Villa, E. R. (2006). Using Moment Generating Functions to Derive Mixture Distributions. *American statistician*, 60(1):75–80.
- Everitt, B. and Hand, D. (1981). *Finite Mixture Distributions*. Chapman and Hall.
- Galea, M. (2013). Comparación de Métodos de Medición en Presencia de un Gold Estándar. *ESTADÍSTICA*, 65(185):83–99.
- Garay, A. M., Hashimoto, E. M., Ortega, E. M., and Lachos, V. H. (2011). On estimation and influence diagnostics for zero-inflated negative binomial regression models. *Computational Statistics & Data Analysis*, 55(3):1304–1318.
- Giménez, P. and Galea, M. (2013). Influence measures on corrected score estimators in functional heteroscedastic measurement error models. *Journal of Multivariate Analysis*, 114:1–15.
- Handcock, M. S. and Jones, J. H. (2004). Likelihood-based inference for stochastic models of sexual network formation. *Theoretical population biology*, 65(4):413–422.
- Handcock, M. S., Jones, J. H., and Morris, M. (2003). On Sexual contacts and epidemic thresholds, models and inference for sexual partnership distributions. *Seattle, WA: University of Washington Center for Statistics and the Social Sciences*, 35(1).
- Harris, T., Hilbe, J. M., Hardin, J. W., et al. (2014). Modeling count data with generalized distributions. *Stata Journal*, 14(3):562–579.
- Henry, L. (1957). Fécondité et famille. *Population*, 12(3):413–444.
- Hideto, F. (2015). Modelos semiparamétricos com resposta binomial negativa. *Dissertação de Tesis: Mestre em Ciências Universidade de Sao Paulo, Brasil*.
- Hilbe, J. M. (2011). *Negative Binomial Regression*. Cambridge: Cambridge University Press.

- Irwin, J. O. (1968). The generalized waring distribution applied to accident theory. *Journal of the Royal Statistical Society. Series A (General)*, pages 205–225.
- Irwin, J. O. (1975). The generalized waring distribution. part i. *Journal of the Royal Statistical Society. Series A (General)*, pages 18–31.
- Johnson, N. L., Kemp, A. W., and Kotz, S. (2005). *Univariate Discrete Distributions*. New York: John Wiley and Sons.
- Johnson, N. L., Kotz, S., and Balakrishnan, N. (1970). Distributions in statistics: Continuous Univariate Distributions. *New York: John Wiley and Sons*.
- Jowaheer, B. C. S. V. (2003). On familial longitudinal Poisson mixed models with gamma random effects. *Journal of Multivariate Analysis*, 87:398–412.
- Karlis, D. (2001). A general EM approach for maximum likelihood estimation in mixed Poisson regression models. *Statistical Modelling*, 1(4):305–318.
- Karlis, D. (2005). EM algorithm for mixed Poisson and other discrete distributions. *Astin bulletin*, 35(01):3–24.
- Lawless, J. F. (1987). Negative binomial and mixed poisson regression. *Canadian Journal of Statistics*, 15(3):209–225.
- Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 619–678.
- Levene, M., Fenner, T., Loizou, G., and Wheeldon, R. (2002). A stochastic model for the evolution of the web. *Computer Networks*, 39(3):277–287.
- Li, C.-S. (2010). Semiparametric negative binomial regression models. *Communications in Statistics—Simulation and Computation*®, 39(3):475–486.



- Lord, D. (2006). Modeling motor vehicle crashes using Poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. *Accident Analysis and Prevention*, 38:751–766.
- McLachlan, G. and Basford, K. (1987). *Mixture Models: Inference and Applications to Clustering*. New York: Dekker.
- McLachlan, G. J. and Krishnan, T. (1997). *The EM algorithm and extensions*. New York: John Wiley and Sons.
- McNeney, B. and Petkau, J. (1994). Overdispersed Poisson regression models for studies of air pollution and human health. *Canadian Journal of Statistics*, 22(4):421–440.
- Miller, A. J. (1961). A queueing model for road traffic flow. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 64–90.
- Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3):370–384.
- Newbold, E. M. (1927). Practical Applications of the Statistics of Repeated Events Particularly to Industrial Accidents. *Journal of the Royal Statistical Society*, pages 487–547.
- Paul, S. R. (2005). Testing goodness of fit of the geometric distribution: an application to human fecundability data. *Journal of Modern Applied Statistical Methods*, 4(2):8.
- Peng, Y., Lord, D., and Zou, Y. (2014). Applying the Generalized Waring model for

- investigating sources of variance in motor vehicle crash analysis. *Accident Analysis & Prevention*, 73:20–26.
- Pielou, E. (1962). Runs of one species with respect to another in transects through plant populations. *Biometrics*, 18(4):579–593.
- Poon, W.-Y. and Poon, Y. S. (1999). Conformal normal curvature and assessment of local influence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(1):51–61.
- Pregibon, D. (1981). Logistic regression diagnostics. *The Annals of Statistics*, pages 705–724.
- Ridout, M. S. and Morgan, B. J. (1991). Modelling digit preference in fecundability studies. *Biometrics*, pages 1423–1433.
- Rodríguez-Avi, J., Conde-Sánchez, A., Sáez-Castillo, A., and Olmo-Jiménez, M. (2007). A new generalization of the Waring distribution. *Computational Statistics & Data Analysis*, 51(12):6138–6150.
- Rodríguez-Avi, J., Conde-Sánchez, A., Sáez-Castillo, A., Olmo-Jiménez, M., and Martínez-Rodríguez, A. (2009). A generalized Waring regression model for count data. *Computational Statistics & Data Analysis*, 53(10):3717–3725.
- Rodríguez-Avi, J. and Olmo-Jiménez, M. J. (2015). A regression model for overdispersed data without too many zeros. *Statistical Papers*, pages 1–25.
- Sapatinas, T. (1995). Identifiability of mixtures of power-series distributions and related characterizations. *Annals of the Institute of Statistical Mathematics*, 47(3):447–459.

- Singh, B., Pudir, P., and Maheshwari, S. (2014). Parameter estimation of beta-geometric model with application to human fecundability data. *arXiv preprint arXiv:1405.6392*.
- Svetliza, Carolina F.; Paula, G. A. (2003). Diagnostics in Nonlinear Negative Binomial Models. *Communication in Statistics- Theory and Methods*, 32(6):1227–1250.
- Tesitelova, M. (1967). On the role of nouns in the lexical statistics. *Prague Stud. Math. Linguist*, 2:121–131.
- Titterington, D., Smith, A., and Makov, U. (1985). *Statistical analysis of finite mixture distributions*. New York: John Wiley and Sons.
- Van de Ven, R. and Weber, N. (1995). Log-linear models for mean and dispersion in mixed Poisson regression models. *Australian Journal of Statistics*, 37:205–205.
- Weinberg, C. R. and Gladen, B. C. (1986). The beta-geometric distribution applied to comparative fecundability studies. *Biometrics*, pages 547–560.
- Xekalaki, E. (1983a). Infinite divisibility, completeness and regression properties of the univariate generalized Waring distribution. *Annals of the Institute of Statistical Mathematics*, 35(1):279–289.
- Xekalaki, E. (1983b). The univariate generalized Waring distribution in relation to accident theory: proneness, spells or contagion? *Biometrics*, pages 887–895.
- Xekalaki, E. (1984). The bivariate generalized Waring distribution and its application to accident theory. *Journal of the Royal Statistical Society. Series A (General)*, pages 488–498.
- Xie, F.-C. and Wei, B.-C. (2008). Influence analysis for Poisson inverse Gaussian regression models based on the EM algorithm. *Metrika*, 67(1):49–62.

- Zhao, Y. and Lee, A. H. (1998). Theory and methods: Influence diagnostics for simultaneous equations models. *Australian and New Zealand Journal of Statistics*, 40(3):345–358.
- Zhu, H., Ibrahim, J. G., Lee, S., and Zhang, H. (2007). Perturbation selection and influence measures in local influence analysis. *The Annals of Statistics*, 35(6):2565–2588.
- Zhu, H., Lee, S.-Y., Wei, B.-C., and Zhou, J. (2001). Case-deletion measures for models with incomplete data. *Biometrika*, 88(3):727–737.
- Zhu, H.-T. and Lee, S.-Y. (2001). Local influence for incomplete data models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(1):111–126.