

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS MATEMÁTICAS
Departamento de Estadística e Investigación Operativa I



TESIS DOCTORAL

**Análisis inferencial basado en medidas de Fi-divergencia para
modelos loglineales con muestreo Multinomial y
sobredispersión**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Juana María Alonso Revenga

Directores

**Leandro Pardo Llorente
Nirian Martín Apaolaza**

Madrid, 2018

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS MATEMÁTICAS

Departamento de Estadística e Investigación Operativa I



Análisis inferencial basado en medidas de Fi-divergencia
para modelos loglineales con muestreo Multinomial y
sobredispersión

Directores:

Dr. Leandro Pardo Llorente

Dra. Nirian Martín Apaolaza

Trabajo presentado para optar al título de Doctor por:
Juana María Alonso Revenga

Junio 2017

*A mi gran familia de Guadamur y de Córdoba,
en especial a mi marido Jesús y a mis hijos, Jesús, Ana y Lola.*

Agradecimientos

En primer lugar tengo que dar las gracias a todos los que me animaron a la realización de esta tesis, en especial a Lina por darme ese empujón inicial sin el cual no hubiera sido capaz de embarcarme en este largo camino.

Agradecer a mis directores de Tesis, los profesores Leandro Pardo y Nirian Martín, la confianza que depositaron en mi, su disponibilidad para el trabajo, y compartir conmigo toda su gran experiencia investigadora y su bagaje intelectual sin los cuales hubiera sido imposible llevar a cabo este trabajo.

Gracias también a mis compañeros de la Facultad de Estudios Estadísticos, por escuchar mis dudas, por su interés en el desarrollo de este largo trabajo y sus palabras de aliento en todo momento, en especial a Lina, Aida, Charo, José Luis, M^a Ángeles, Inés, Carmen, Teresa,...

Agradecer a Carmen, bibliotecaria de la Facultad, por su eficacia en la búsqueda de cualquier bibliografía, a todo el personal de la facultad que de una otra forma me ayudaron a desarrollar este trabajo.

Por último, dar las gracias a mi familia por aguantar la obsesión en la que en muchos momentos se convirtió la realización de este trabajo de investigación, por su paciencia conmigo durante estos años y por su apoyo incondicional en todo lo que hago. Muchas gracias.

Resumen

En los últimos años se han incrementado de forma importante los métodos estadísticos para analizar datos cualitativos. Quizá esto se ha debido en parte a la gran demanda, por parte de las Ciencias Biomédicas (particularmente en relación a estudios epidemiológicos), Sociales y del Comportamiento, de técnicas estadísticas específicas para el tratamiento de la gran cantidad de datos cualitativos de que disponían. El desarrollo de técnicas específicas para el tratamiento de datos cualitativos o categóricos ha permitido descartar, por innecesarios y muchas veces inapropiadas, muchas de las técnicas para variables continuas que se venían utilizando para este tipo de datos.

Por otra parte, una proporción sustancial de estudios epidemiológicos, particularmente estudios médicos, veterinarios, ensayos clínicos y de medidas repetidas para investigaciones clínicas, producen datos que, por su naturaleza, están agrupados en conglomerados o clusters. Estas agrupaciones generan observaciones altamente correladas dentro de cada uno de los grupos; así las técnicas estadísticas generales que implican muestreo aleatorio simple, pueden llevar en estos casos a conclusiones erróneas. Cuando se utiliza el muestreo por conglomerados hay dos fuentes de variación: la primera es la variación entre individuos de un mismo cluster y la segunda la variación entre clusters o conglomerados. El efecto del aumento de la variabilidad implica un aumento de los errores estándar de los estimadores, lo que da lugar a intervalos de confianza más anchos y errores de tipo I inflados en comparación con un estudio de igual tamaño muestral pero aleatorizado a los individuos. Luego, si se realiza este diseño de muestreo y se ignora su efecto utilizando las técnicas estadísticas habituales, se obtendrán intervalos falsamente estrechos y p-valores erróneamente pequeños que pueden llevar a tomar decisiones equivocadas en los estudios realizados.

Este problema ha sido abordado en la literatura estadística desde diferentes enfoques Choi y McHuhg (1989), Fellegi (1980), Holt et al. (1980), Rao y Scott (1981,1984), Bedrick (1983), Fay (1985), parten del diseño de muestreo concreto, para encontrar la distribución de los estadísticos necesarios para cada análisis.

Por otro lado, Altham (1976), Cohen (1976), Brier (1980), Fienberg (1979), Menéndez et al.(1995, 1996), enfocaron dicho estudio bajo la hipótesis de una distribución de probabilidad que modelizara este tipo de datos. La distribución Dirichlet-multinomial es históricamente la primera distribución disponible para modelizar datos multinomiales agrupados en clusters y que presentan sobredispersión debido a una correlación homogénea (Mosimann, 1962). Mas tarde, Cohen y Altham (1976) propusieron la distribución multinomial n -inflada y más recientemente, Morel and Nagaraaj (1993) propusieron la distribución multinomial agrupada aleatoriamente. La distribución binomial cero-inflada pertenece también a la familia de distribuciones que modelizan este tipo de datos.

En esta memoria, se presenta un enfoque diferente donde no se asume ninguna distribución subyacente concreta para los datos analizados, solamente se parte de la hipótesis de datos multinomiales con sobredispersión, que serán analizados mediante la utilización de un modelo loglineal. El objetivo es presentar nuevos estimadores basados en medidas de ϕ -divergencia, (Pardo, 2006) tanto para los parámetros propios del modelo loglineal como para los necesarios para medir el efecto del diseño del muestro por conglomerados.

En el Capítulo 1 se presentan las características de los modelos loglineales para datos multinomiales con datos agrupados en clusters, así como las distribuciones de probabilidad asociadas a este tipo de muestreo. Se introduce el concepto de sobredispersión en los datos con respecto a la distribución multinomial y su relación con el coeficiente de correlación intraccluster.

Puesto que no se supondrá ninguna distribución específica subyacente, sino simplemente hipótesis sobre los dos primeros momentos de la distribución que introducen la

medida de sobredispersión, en el Capítulo 2 se obtendrán los *estimadores de mínima cuasi ϕ -divergencia* ($MC\phi$) para los parámetros del modelo loglineal, así como su distribución asintótica. Los estimadores de máxima cuasi-verosimilitud ($EMCV$), introducidos por Wedderburn (1974), son un caso particular de los anteriores para modelos multinomiales con datos agrupados en clusters. Además, se presentarán nuevos estimadores del parámetro de sobredispersión para el caso en el que el tamaño de todos los clusters sea el mismo. Se finalizará este Capítulo con una aplicación práctica de los nuevos estimadores y con un estudio de simulación que compara el comportamiento de los nuevos estimadores, en cuanto a su sesgo y error cuadrático medio, con otros estimadores propuestos anteriormente.

En el Capítulo 3 se extienden los resultados presentados en el Capítulo 2 al caso en el que los tamaños muestrales de los clusters o conglomerados sean diferentes. Se define el parámetro de sobredispersión asociado al tamaño medio de cluster como una nueva medida del efecto del diseño en esta situación. Además, se presenta un estimador específico del parámetro de sobredispersión para el caso en el que el tamaño de los clusters es grande.

Una vez definidos los estimadores se analizará el procedimiento para encontrar la parametrización del modelo loglineal que mejor se ajuste a los datos. Esto llevará al problema de bondad de ajuste, de los datos bajo consideración, a un modelo loglineal específico. A tal fin, en el Capítulo 4, se introducirá una familia de estadísticos de contraste de bondad de ajuste basada en medidas de ϕ -divergencia, en la que el efecto del diseño se corregirá mediante los nuevos estimadores del parámetro de sobredispersión presentados en los Capítulos 2 y 3. Así, se abordará el problema cuando los tamaños de los clusters o conglomerados son iguales, y también para tamaños desiguales.

Por último, en el Capítulo 5 se estudiarán los modelos loglineales de tipo jerárquico con especial atención a su aplicación al estudio del grado de acuerdo entre evaluadores. Agresti (1988) presentó una secuencia de modelos jerárquicos para medir el grado de acuerdo, en este capítulo se estudian estos modelos cuando los individuos, examinados por los diferentes

evaluadores, están agrupados en clusters. Para todos los estadísticos presentados se realiza un estudio de simulación comparando con los estadísticos de bondad de ajuste en los que se corrige el efecto del diseño muestral mediante otros estimadores del parámetro de sobredispersión.

Para las aplicaciones numéricas y las simulaciones, desarrolladas en la memoria, se ha utilizado el software Matlab, así como SAS y Fortran con las rutinas IMSL, para la comprobación de los resultados obtenidos. Los programas desarrollados están disponibles en la website:

http://www.ucm.es/estadistica_io3/juana-maria-alonso-revenga-1

Abstract

Statistical methods to analyze qualitative data have increased significantly in recent years. Perhaps this is due in part to the high demand, by the Biomedical Sciences (particularly in relation to epidemiological studies), Social and Behaviour Sciences, of specific statistical techniques for the treatment of the large amount of qualitative data at disposal. The development of specific techniques for the study of qualitative or categorical data has allowed to discard, by unnecessary and often inappropriate, many of the techniques for continuous variables which were being used for this type of data.

On the other hand, a substantial proportion of epidemiological studies, particularly medical studies, veterinary, repeated measures for clinical research and clinical trials, produce data which, by their nature, are grouped in clusters. These groupings generate observations highly correlated within each of the groups; thus, the general statistical techniques, that involve simple random sampling, can lead to erroneous conclusions in these situations. When cluster sampling is used, there are two sources of variation: the first is the variation between individuals of a same cluster and the second, variation between clusters. The effect of the increase in the variability implies an increase of the standard errors of the estimators, which leads to wider confidence intervals and inflated type I errors in comparison with a study of equal sample size but randomized to individuals. As a result, there will be intervals falsely narrow and wrongly small p-values that can lead to wrong decisions in the carried out studies.

Correlated clustered multinomial frequency data have been dealt in the statistical literature for many years through two different approaches. Following Choi and McHugh (1989),

the design-based approach provides inference with respect to the sampling distribution of estimates over repetitions of the same design. The works of Fellegi (1980), Holt et al. (1980), Rao and Scott (1981,1984), Bedrick (1983), Landis et al (1984), Koch et al. (1975), Fay (1985), as well as references therein are good examples of this approach. On the other hand, Altham (1976), Cohen (1976), Brier (1980), Fienberg (1979), Menéndez et al. (1995, 1996) postulate a probability distribution to model the sample data. Dirichlet-multinomial is, historically, the first suitable distribution to modelize homogeneously correlated clustered overdispersed multinomial frequency with a fixed cluster size (see Mosimann, 1962). Later, Cohen (1976) and Altham (1976) proposed the n -inflated distribution and more recently, Morel and Nagaraj (1993) proposed the random-clumped distribution. The zero-inflated binomial distribution falls also inside this family of homogeneously correlated clustered overdispersed multinomial frequency data.

This report presents a different approach where no specific underlying distribution for the data analyzed is assumed, based on the sole knowledge of the vector mean and the variance-covariance matrix of the distribution, that will be analyzed using a log-linear model. The goal is to present new estimators based on ϕ -divergence measures (Pardo, 2006), for own log-linear model parameters and also for the required parameter to measure the effect of the design of cluster.

In Chapter 1 we show the features of loglinear modeling for overdispersed multinomial data, as well as the probability distributions associated with this kind of sampling. We shall introduce the concept of overdispersion in the data with respect to the multinomial distribution and its relationship with the intracluster correlation coefficient.

Since, we shall not suppose any specific underlying distribution, but simply hypotheses about the first two moments of the distribution, that introduce the overdispersion measure. In Chapter 2 we shall present a new family of estimators useful for log-linear modeling with overdispersion. These new estimators are the quasi-minimum ϕ -divergence estimators. One

member of these estimators family is the so-called quasi maximum likelihood estimator. Their corresponding asymptotic properties are also shown.

We shall propose, in Chapter 2, new estimators for the design effect, as well as for the intracluster correlation coefficient for equal cluster sizes. A practical application will be presented and a simulation study that compares the behavior of the new estimators proposed. These estimators exhibit by far a better behavior with regard to the mean square error (MSE) in comparison with the existing estimation method, which is fully non-parametric.

The assumption of equal cluster sizes is generalized to unequal cluster sizes in Chapter 3. In this setting, on one hand, a new family of consistent estimators of the design effect or intracluster correlation coefficient is provided and, on the other hand, a new estimator is proposed for the special case of large cluster sizes.

In Chapter 4, a family of statistics tests for goodness of fit based on ϕ -divergences will be introduced, in which the effect of the design will be corrected through the new overdispersion estimators. Thus, the problem will be addressed for equal cluster sizes and also for unequal cluster sizes.

Finally, in Chapter 5 we shall study hierarchical loglinear models, with special attention to its application for raters agreement. Agresti (1988) presented a sequence of hierarchical models to measure raters agreement. In this chapter, these models are studied when individuals, examined by different raters, are grouped in clusters. For all the statistics presented a simulation study is done comparing the goodness of fit statistics which corrects the effect of the sampling design with others overdispersion parameter estimators.

For numerical applications and simulations, developed in this memory, Matlab software, as well as SAS and Fortran with IMSL routines have been used . The programs developed are available on the website:

http://www.ucm.es/estadistica_io3/juana-maria-alonso-revenga-1

Índice de figuras

2.1	RECM y sesgo de los diferentes estimadores de ρ^2 para datos generados con la distribución Dirichlet multinomial.....	54
2.2	RECM y sesgo de los diferentes estimadores de ρ^2 para datos generados con la distribución multinomial n-inflada.....	55
2.3	RMSE y sesgo de los diferentes estimadores de ρ^2 para datos generados con la distribución multinomial agrupada aleatoriamente.....	56
2.4	RECM y sesgo de los estimadores de ρ^2 , no paramétrico y $MC\phi$ con $\lambda=2/3$, para datos generados con la distribución Dirichlet multinomial.....	57
2.5	RECM y sesgo de los estimadores de ρ^2 , no paramétrico y $MC\phi$ con $\lambda=2/3$, para datos generados con la distribución multinomial n-inflada.....	58
2.6	RMSE y sesgo de los estimadores de ρ^2 , no paramétrico y $MC\phi$ con $\lambda=2/3$, para datos generados con la distribución multinomial aleatoriamente agrupada	59
2.7	RECM y sesgo de los diferentes estimadores de ρ^2 , para datos generados con la distribución Dirichlet multinomial con $n =5$	61
2.8	RECM y sesgo de los estimadores de ρ^2 , para datos generados con la distribución multinomial n-inflada con $n =5$	62
2.9	RECM y sesgo de los estimadores de ρ^2 , para datos generados con la distribución multinomial aleatoriamente agrupada con $n =5$	63

2.10	RECM y sesgo de los diferentes estimadores de ρ^2 para datos generados con la distribución Dirichlet multinomial con $n = 10$	64
2.11	RMSE y sesgo de los diferentes estimadores de ρ^2 para datos generados con la distribución multinomial n-infl ada con $n = 10$	65
2.12	RECM y sesgo de los estimadores de ρ^2 , para datos generados con la distribución multinomial aleatoriamente agrupada con $n = 10$	66
2.13	RECM y sesgo de los diferentes estimadores de ρ^2 , para datos generados con la distribución Dirichlet multinomial con $n = 20$	67
2.14	RECM y sesgo de los diferentes estimadores de ρ^2 , para datos generados con la distribución multinomial n-infl ada con $n = 20$	68
2.15	RECM y sesgo de los estimadores de ρ^2 , para datos generados con la distribución multinomial aleatoriamente agrupada con $n = 20$	69
2.16	Distribución de las estimaciones para diferentes tamaños de cluster cuando $\rho^2 = 0,25$	70
2.17	Distribución de las estimaciones para diferentes tamaños de cluster cuando $\rho^2 = 0,75$	71
3.1	RECM y sesgo de los estimadores $CM\phi$ de ρ^2 para datos generados con la distribución Dirichlet multinomial.....	95
3.2	RECM y sesgo de los estimadores $CM\phi$ de ρ^2 para datos generados con la distribución multinomial n-infl ada.....	96
3.3	RECM y sesgo de los diferentes estimadores $CM\phi$ de ρ^2 para datos generados con la distribución multinomial agrupada aleatoriamente.....	97
3.4	RECM y sesgo para el estimador de ρ^2 de Brier.....	99
3.5	RECM y sesgo para el estimador de ρ^2 de Brier modificado.....	100

3.6 RECM y sesgo para el estimador $CM\phi$ de ρ^2	101
3.7 Histogramas para de los diferentes estimadores de ρ^2 , para el valor $\rho^2=0,25$	103
3.8 RECM y sesgo para los estimadores de ρ^2 , para locus D3. Distribución DM.....	104
3.9 RECM y sesgo para los estimadores de ρ^2 , para locus D3. Distribución RC.....	105
3.10 RECM y sesgo para los estimadores de ρ^2 , para locus D3. Distribución NI.....	106
3.11 RECM y sesgo para los estimadores de ρ^2 , para locus vWA. Distribución DM.....	107
3.12 RECM y sesgo para los estimadores de ρ^2 , para locus vWA. Distribución RC.....	108
3.13 RECM y sesgo para los estimadores de ρ^2 , para locus vWA. Distribución NI.....	109
3.14 RECM y sesgo para los estimadores de ρ^2 , para locus FGA. Distribución DM.....	110
3.15 RECM y sesgo para los estimadores de ρ^2 , para locus FGA. Distribución RC.....	111
3.16 RECM y sesgo para los estimadores de ρ^2 , para locus FGA. Distribución NI.....	112
3.17 RECM y sesgo para los estimadores de ρ^2 , para locus D8. Distribución DM.....	113
3.18 RECM y sesgo para los estimadores de ρ^2 , para locus D8. Distribución RC.....	114
3.19 RECM y sesgo para los estimadores de ρ^2 , para locus D8. Distribución NI.....	115
4.1 Probabilidad de error de Tipo I estimada para $T^{\phi_1}(\widehat{\theta}_{\phi_2}, \widetilde{\vartheta}_{\widehat{n}^*, N})$ con datos generados con la distribución Dirichlet Multinomial.....	146
4.2 Probabilidad de error de Tipo I estimada para $T^{\phi_1}(\widehat{\theta}_{\phi_2}, \widetilde{\vartheta}_{\widehat{n}^*, N})$ con datos generados con la distribución Multinomial n-inflada.....	147
4.3 Probabilidad de error de Tipo I estimada para $T^{\phi_1}(\widehat{\theta}_{\phi_2}, \widetilde{\vartheta}_{\widehat{n}^*, N})$ con datos generados con la distribución Agrupada Aleatoriamente.....	148

4.4	Probabilidad de error de Tipo I estimada para $T^{\phi_1}(\hat{\theta}_{\phi_2}, \hat{\vartheta}_{n^*,N})$, comparada por distribuciones.....	150
4.5	Probabilidad de error de Tipo I estimada para $T^{\phi_1}(\hat{\theta}_{\phi_2}, \hat{\vartheta}_{n^*,N})$ resumen comparativo.....	151
4.6	Probabilidad de error de Tipo II estimada para $T^{\phi_1}(\hat{\theta}_{\phi_2}, \hat{\vartheta}_{n^*,N})$ con datos generados con la distribución Dirichlet Multinomial.....	153
4.7	Probabilidad de error de Tipo II estimada para $T^{\phi_1}(\hat{\theta}_{\phi_2}, \hat{\vartheta}_{n^*,N})$ con datos generados con la distribución Multinomial n-inflada.....	154
4.8	Probabilidad de error de Tipo II estimada para $T^{\phi_1}(\hat{\theta}_{\phi_2}, \hat{\vartheta}_{n^*,N})$ con datos generados con la distribución Agrupada Aleatoriamente.....	155
5.1	Probabilidad de error de Tipo I estimada para $T_{\phi_1, \phi_2}^{(i)}$ con datos generados con la distribución Dirichlet Multinomial. (Modelo 2).....	180
5.2	Probabilidad de error de Tipo I estimada para $T_{\phi_1, \phi_2}^{(i)}$ con datos generados con la distribución Multinomial n-inflada, (Modelo 2).....	181
5.3	Probabilidad de error de Tipo I estimada para $T_{\phi_1, \phi_2}^{(i)}$ con datos generados con la distribución Agrupada Aleatoriamente, (Modelo 2).....	182
5.4	Probabilidad de error de Tipo I estimada para $T_{\phi_1, \phi_2}^{(i)}$ con datos generados con la distribución Dirichlet Multinomial. (Modelo 2*).....	183
5.5	Probabilidad de error de Tipo I estimada para $T_{\phi_1, \phi_2}^{(i)}$ con datos generados con la distribución Multinomial n-inflada, (Modelo 2*).....	184
5.6	Probabilidad de error de Tipo I estimada para $T_{\phi_1, \phi_2}^{(i)}$ con datos generados con la distribución agrupada aleatoriamente.....	185

Índice de Tablas

2.1	Número de parejas según Diagnóstico y Sexo.(Datos Cohen).....	44
2.2	Número de individuos según Edad, Diagnóstico y Sexo.....	48
2.3	Estimaciones de θ para diferentes valores de λ	49
2.4	Probabilidades estimadas para diferentes valores de λ	50
2.5	Individuos porcluster según Sexo,y Diagnóstico de Esquizofrenia.....	51
2.6	Número de individuos según Sexo y diagnóstico.....	52
2.7	Estimaciones de las probabilidades y de ρ^2 para diferentes valores de λ y no paramétricos	53
3.1	Satisfacción con la vivienda y con el vecindario (Datos de Brier, 1980).....	87
3.2	Número de vecinos según Grado de Satisfacción con la vivienda y con el vecindario.....	87
3.3	Estimaciones de θ para diferentes valores de λ	89
3.4	Estimaciones de las probabilidades estimadas para diferentes valores de λ y no paramétricos.....	89
3.5	Frecuencias de los alelos: 12, 13, 14, 15, 16, 17, 18, 19 (M=8) para el locus D3S1358 separados por subpoblaciones.....	90
3.6	Frecuencias de los alelos: 11, 13, 14, 15, 16, 17, 18, 19, 20, 21 (M=10) para el locus vWA, separados por subpoblaciones.....	90

3.7	Frecuencias de los alelos: 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 (M=13) para el locus FGA, separados por subpoblaciones.....	91
3.8	Frecuencias de los alelos: 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 (M=11) para el locus D8S1179, separados por subpoblaciones.....	91
3.9	Estimaciones de las probabilidades y de ρ^2 para los locus D3S1358, vWA, FGA y D8S1179.....	93
4.1	Estimaciones de las probabilidades y de ρ^2 para diferentes valores de λ y no paramétricos, (Datos Cohen).....	136
4.2	Valores del estadístico $T^{\phi_1}(\hat{\theta}_{\phi_2}, \vartheta_{n,N}^{\sim})$ y su p-valor.....	137
4.3	Valores del estadístico $T^{\phi_1}(\hat{\theta}_{\phi_2}, \vartheta_{n,N}^{\sim})$ y su p-valor.....	137
4.4	Valores del estadístico $T^{\phi_1}(\hat{\theta}_{\phi_2}, \vartheta_{n^*,N}^{\sim})$, y su p-valor, (Brier).....	143
4.5	Valores del estadístico $T^{\phi_1}(\hat{\theta}_{\phi_2}, \vartheta_{n^*,N}^{\sim})$, y su p-valor (semiparamétrico).....	143
5.1	Datos de Esclerosis. (Landis y Koch).....	172
5.2	Valores del estadístico $T_{\phi_1, \phi_2}^{(i)}$ y su p-valor.....	175
5.3	Valores del estadístico $T^{\phi_1}(\hat{\theta}_{\phi_2}, \vartheta_{n^*,N}^{\sim})$, y su p-valor.....	175
5.4	Valores del estadístico $T_{\phi_1, \phi_2}^{(i)}$ y su p-valor. (Modelo 2*).....	176
5.5	Valores del estadístico $T_{\phi_1, \phi_2}^{(i)}$ y su p-valor, contraste 2 (Modelo 2*).....	177

Índice

Agradecimientos	v
Resumen	vii
Abstract	xi
Índice de figuras	xv
Índice de tablas	xix

CAPÍTULOS

1. EL MODELO LOGLINEAL CON MUESTREO POR CONGLOMERADOS	1
1.1. Introducción	1
1.2. Formulación del modelo	2
1.2.1. Modelización de datos agrupados en clusters o conglomerados	2
1.2.2. El modelo loglineal	4
1.3. Distribuciones para muestreo complejo	6
1.4. Estimador no paramétrico. Distribución asintótica	15
2. EL ESTIMADOR DE MÍNIMA CUASI FI-DIVERGENCIA EN DATOS MULTINOMIALES POR CONGLOMERADOS	17
2.1. El estimador de mínima cuasi ϕ -divergencia en el modelo loglineal asociado a datos multinomiales por conglomerados	18
2.1.1. Propiedades asintóticas del estimador de mínima cuasi ϕ -divergencia	24

2.2.	Propiedades asintóticas del estimador de mínima cuasi ϕ -divergencia en modelos loglineales con muestreo por conglomerados.....	32
2.3.	Estimadores consistentes para ϑ_n y ρ^2	35
2.4.	Parametrización de un modelo loglineal para tablas de contingencia con muestreo por conglomerados	40
2.5.	Aplicación numérica	43
2.6.	Estudio de simulación	52
3.	NUEVOS ESTIMADORES PARA EL COEFICIENTE DE SOBREDISPERSIÓN CON CLUSTERS O CONGLOMERADOS DE TAMAÑOS DIFERENTES	73
3.1.	Notación y resultados básicos	73
3.2.	Estimadores consistentes de Brier para ϑ_{n^*} y ρ^2	77
3.3.	Nuevos estimadores no-paramétricos y semiparamétricos para ϑ_{n^*} y ρ^2	81
3.3.1.	Caso 1: Existen grupos de clusters con tamaños iguales.	81
3.3.2.	Caso 2: Clusters de tamaño suficientemente grande suficientemente grande	84
3.4.	Aplicaciones numéricas	86
3.4.1.	Datos de Brier sobre satisfacción de los vecinos con su vivienda y vecindario	86
3.4.2.	Estudio sobre datos de ADN para diferentes genotipos (Weir and Hill, 2002)	90
3.5.	Estudio de simulación	94
4.	BONDAD DE AJUSTE PARA MODELOS MULTINOMIALES CON SOBREDISPERSIÓN	117
4.1.	Estadísticos de ϕ -divergencia para bondad de ajuste con clusters de igual tamaño	118
4.1.1.	Distribución asintótica	120
4.2.	Aplicaciones numéricas.	135
4.3.	Estadísticos de ϕ -divergencia para bondad de ajuste con clusters de diferente tamaño	138
4.3.1.	Distribución asintótica	139
4.4.	Aplicaciones numéricas.	141
4.5.	Estudio de simulación.	144

5. SECUENCIAS ANIDADAS DE MODELOS LOGLINEALES JERARQUIZADOS PARA DATOS MULTINOMIALES CON SOBREDISPERSIÓN.....	157
5.1. Planteamiento del problema.....	157
5.2. Estadísticos de ϕ -divergencia para secuencias anidadas de modelos loglineales jerarquizados	158
5.3. Distribución asintótica del estadístico ϕ -divergencia con datos multinomiales con sobredispersión bajo la hipótesis nula	160
5.4. Distribución asintótica de los estadísticos ϕ -divergencia con datos multinomiales con sobredispersión bajo hipótesis alternativas contiguas	165
5.5. Aplicación de modelos loglineales jerárquicos para medir el grado de acuerdo entre evaluadores (rater agreement).....	168
5.5.1. Aplicación sobre categorías de tipo ordinal.....	171
5.6. Estudio de simulación.	177
CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN	187
BIBLIOGRAFÍA	189

CAPÍTULO 1

EL MODELO LOGLINEAL CON MUESTREO POR CONGLOMERADOS

1.1. Introducción

Los diseños de muestreo utilizados habitualmente en modelos loglineales están basados en las distribuciones Multinomial, Poisson o Producto-Multinomial (Agresti 2002). Estos diseños de muestreo suponen que las frecuencias de aparición en las tablas de contingencia están negativamente correladas (en el caso de muestreo multinomial), son independientes (el caso de muestreo de Poisson) ó correladas negativamente en estratos independientes (en el caso de muestreo producto-multinomial).

Sin embargo, los datos categóricos provienen, en muchas ocasiones, de muestras donde los datos están agrupados por familias (cluster) dentro de las cuales existe una dependencia que no se puede ignorar. En estos casos es necesario realizar el estudio en un contexto de muestreos complejos distintos a los anteriormente mencionados. Cohen (1976) y Altham (1976) presentaron modelos que describen la correlación entre individuos dentro de un mismo grupo. Sus modelos suponen grupos de igual tamaño. Más tarde, Brier (1980) propuso la utilización de la distribución Dirichlet-multinomial para modelizar este tipo de muestreos permitiendo la generalización a tamaños muestrales distintos entre grupos. En la mayoría de estas situaciones, la correlación intracluster tiende a ser positiva y esto da lugar a que la varianza de las variables indicadoras bajo el muestreo por clusters sea mayor que bajo muestreo aleatorio que da lugar a la distribución multinomial. Este concepto, se conoce como *efecto del diseño* o *sobredispersión* con respecto al muestreo multinomial. El

efecto del diseño representa el coeficiente por el que hay que multiplicar al tamaño muestral para conseguir los mismos niveles de error que con un muestreo aleatorio simple.

En este Capítulo, en la Sección 1.2, se presentan los modelos loglineales en un contexto diferente al habitual, en donde el muestreo utilizado da lugar a grupos de individuos (clusters o conglomerados) de forma que, dentro de los clusters los individuos presentan dependencia, mientras que los individuos de clusters diferentes son independientes. Además, los clusters son todos ellos homogéneos entre sí. Es decir, la Sección 1.2 está dedicada a la formulación del modelo objeto de estudio en esta memoria.

Se presentan, en la Sección 1.3, distintos modelos de probabilidad que permiten modelizar la situación anteriormente planteada. Así, se consideran la distribución *Multinomial n -inflada* propuesta por Cohen y Altham (1976), la *Dirichlet Multinomial*, propuesta por Brier (1980) y la *Multinomial Agrupada Aleatoriamente*, propuesta por Morel y Nagaraj (1993).

Se finalizará el Capítulo estableciendo la distribución asintótica del estimador no paramétrico del vector de probabilidades para este tipo de muestreo.

1.2. Formulación del modelo

1.2.1. Modelización de datos agrupados en clusters o conglomerados

Un cluster es un conjunto de individuos o unidades, como puede ser una familia, colegio, hospital, pacientes de un mismo médico, etc. Los datos agrupados mediante clusters pueden ser el resultado de un ensayo clínico aleatorizado dirigido a grupos de pacientes en lugar de a individuos, esto es lo que se llama muestreo aleatorio por clusters. Hay varias razones por las que los investigadores aplican este tipo de muestreo en lugar de aleatorizar por los individuos pero la fundamental es la sencillez y economía en la recogida de datos, siendo esta, en ocasiones, la única forma posible de recogida. Se presenta a continuación

un modelo para el estudio estadístico de datos así agrupados teniendo en cuenta el grado de dependencia que puede existir entre los individuos pertenecientes a un mismo cluster.

Se consideran N clusters y para cada uno de ellos se obtiene un número fijo, n , de observaciones $\mathbf{X}^{(\ell)} = (X_1^{(\ell)}, \dots, X_n^{(\ell)})^T$, $\ell = 1, 2, \dots, N$. Las variables $X_s^{(\ell)}$ toman valores en el espacio muestral \mathcal{X} . Por tanto, $\mathbf{X}^{(\ell)} = (X_1^{(\ell)}, \dots, X_n^{(\ell)})^T$ tomará valores en el espacio \mathcal{X}^n . Supóngase que se tiene un clasificador

$$\delta : \mathcal{X} \rightarrow \{1, \dots, M\},$$

de forma que cada valor $r \in \{1, \dots, M\}$ es asignado con una probabilidad determinada, $\Pr(\delta(X_s^{(\ell)}) = r) = p_r$, y dicha probabilidad puede depender de un parámetro desconocido, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^{M_0}$, es decir $p_r = p_r(\boldsymbol{\theta})$. Se denotará mediante $Y_r^{(\ell)}$ el número de observaciones procedentes del ℓ -ésimo cluster, $\ell = 1, \dots, N$, a las que el clasificador asigna el valor r de $\{1, \dots, M\}$. Por tanto,

$$Y_r^{(\ell)} = \sum_{s=1}^n I_{\{r\}}(\delta(X_s^{(\ell)})), \quad r = 1, \dots, M, \quad \ell = 1, \dots, N,$$

con la función indicadora $I_{\{r\}}(\delta(X_s^{(\ell)}))$ igual a 1 si $\delta(X_s^{(\ell)}) = r$ y 0 en otro caso. El vector M -dimensional de clasificación asociado al ℓ -ésimo cluster,

$$\mathbf{Y}^{(\ell)} = (Y_1^{(\ell)}, \dots, Y_M^{(\ell)})^T, \tag{1.1}$$

se llama también tabla de contingencia.

Se supondrá que los N vectores $\mathbf{X}^{(\ell)} = (X_1^{(\ell)}, \dots, X_n^{(\ell)})^T$ son independientes y como consecuencia los N vectores de clasificación, $\mathbf{Y}^{(\ell)} = (Y_1^{(\ell)}, \dots, Y_M^{(\ell)})^T$, también lo son. Obsérvese, sin embargo, que con el esquema desarrollado hasta ahora, fijado ℓ , si las variables

$X_1^{(\ell)}, \dots, X_n^{(\ell)}$ no son independientes entre si, tampoco lo serán las variables de clasificación $Y_1^{(\ell)}, \dots, Y_M^{(\ell)}$.

El esquema dado anteriormente se podría resumir de la siguiente forma: para cada cluster se tienen las muestras $\mathbf{X}^{(\ell)}, \mathbf{Y}^{(\ell)}$ que provienen de las poblaciones de vectores aleatorios $\mathbf{X} = (X_1, \dots, X_n)^T$ e $\mathbf{Y} = (Y_1, \dots, Y_M)^T$ con espacios muestrales respectivos \mathcal{X}^n y

$$\{\mathbf{y} = (y_1, \dots, y_M)^T \in \{0, 1, \dots, n\}^M : \sum_{r=1}^M y_r = n\}.$$

En el caso de que las componentes de \mathbf{X} sean independientes se tendrá que el vector \mathbf{Y} se distribuye según una Multinomial de parámetros

$$(n; p_1(\boldsymbol{\theta}), \dots, p_M(\boldsymbol{\theta})).$$

En lo que sigue, se denotará $m_r(\boldsymbol{\theta}) = np_r(\boldsymbol{\theta}), r = 1, \dots, M$ y $\mathbf{m}(\boldsymbol{\theta}) = (m_1(\boldsymbol{\theta}), \dots, m_M(\boldsymbol{\theta}))$.

1.2.2. El modelo loglineal

Se asumirá a partir de ahora que el vector de valores esperados, $\mathbf{m}(\boldsymbol{\theta})$, se puede justificar mediante un modelo loglineal de la forma,

$$\log \mathbf{m}(\boldsymbol{\theta}) = \mathbf{X}\boldsymbol{\theta}^*, \quad \boldsymbol{\theta}^* = (u, \boldsymbol{\theta}^T)^T, \quad \boldsymbol{\theta}^T = (\theta_1, \dots, \theta_{M_0})^T, \quad (1.2)$$

siendo \mathbf{X} una matriz de dimensión $M \times (M_0 + 1)$ con $rg(\mathbf{X}) = M_0 + 1$. Es decir, se supondrá que $\mathbf{m}(\boldsymbol{\theta})$ pertenece a la clase

$$\mathcal{C}(\mathbf{X}) = \{\mathbf{m}(\boldsymbol{\theta}) : \log \mathbf{m}(\boldsymbol{\theta}) = \mathbf{X}\boldsymbol{\theta}^* \text{ para algún } \boldsymbol{\theta}^* \in \mathbb{R}^{M_0+1}\}$$

de los modelos loglineales sobre \mathbf{X} , siempre que el vector $\mathbf{1}_M = (1, \dots, 1)^T$ pertenezca a $\mathcal{C}(\mathbf{X})$. Evidentemente $\mathcal{C}(\mathbf{X})$ es el espacio vectorial engendrado por las columnas de la matriz \mathbf{X} .

No es difícil establecer que, si $\mathbf{1}_M \in \mathcal{C}(\mathbf{X})$, entonces, sin pérdida de generalidad, la matriz \mathbf{X} se puede escribir de la forma

$$\mathbf{X} = \begin{pmatrix} 1 & w_{11} & \cdots & w_{1M_0} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & w_{M1} & \cdots & w_{MM_0} \end{pmatrix} = (\mathbf{1}_M, \mathbf{W}),$$

con

$$\mathbf{W} = \begin{pmatrix} w_{11} & \cdots & w_{1M_0} \\ \vdots & \vdots & \vdots \\ w_{M1} & \cdots & w_{MM_0} \end{pmatrix} = \begin{pmatrix} w_1^T \\ \vdots \\ w_M^T \end{pmatrix}.$$

A partir de lo visto anteriormente se verifica, por (1.2), que

$$\log m_r(\boldsymbol{\theta}) = u + \sum_{g=1}^{M_0} w_{rg} \theta_g, \quad r = 1, \dots, M. \quad (1.3)$$

Además, puesto que $m_r(\boldsymbol{\theta}) = np_r(\boldsymbol{\theta})$, $r = 1, \dots, M$, se tiene que

$$\sum_{r=1}^M m_r(\boldsymbol{\theta}) = n,$$

y teniendo en cuenta (1.3)

$$n = \exp\{u\} \sum_{r=1}^M \exp\left\{\sum_{g=1}^{M_0} w_{rg} \theta_g\right\},$$

luego

$$\exp\{u\} = \frac{n}{\sum_{r=1}^M \exp\left\{\sum_{g=1}^{M_0} w_{rg} \theta_g\right\}}.$$

Se puede expresar $\sum_{g=1}^{M_0} w_{rg} \theta_g = \mathbf{w}_r^T \boldsymbol{\theta}$, y así por (1.3) se tiene

$$m_r(\boldsymbol{\theta}) = \frac{n \exp\{\mathbf{w}_r^T \boldsymbol{\theta}\}}{\sum_{r=1}^M \exp\{\mathbf{w}_r^T \boldsymbol{\theta}\}}, \quad r = 1, \dots, M, \quad (1.4)$$

o lo que es lo mismo,

$$p_r(\boldsymbol{\theta}) = \frac{\exp\{\mathbf{w}_r^T \boldsymbol{\theta}\}}{\sum_{r=1}^M \exp\{\mathbf{w}_r^T \boldsymbol{\theta}\}}, \quad r = 1, \dots, M, \quad (1.5)$$

que en expresión matricial viene dado por

$$\mathbf{p}(\boldsymbol{\theta}) = \frac{\exp\{\mathbf{W}\boldsymbol{\theta}\}}{\mathbf{1}_M^T \exp\{\mathbf{W}\boldsymbol{\theta}\}}.$$

Se supondrá que $rg(\mathbf{W}) = M_0 \leq M - 1$ y que las columnas son linealmente independientes del vector $\mathbf{1}_M = (1, \dots, 1)^T$. La matriz \mathbf{W} es la matriz de diseño y se supondrá conocida, ya que caracteriza el modelo log-lineal.

1.3. Distribuciones para muestreo complejo

Hay muchas situaciones reales en las que no se puede suponer la independencia entre las observaciones, $\mathbf{X}^{(\ell)} = (X_1^{(\ell)}, \dots, X_n^{(\ell)})^T$, dentro de un mismo cluster y esto lleva a que la distribución del vector aleatorio $\mathbf{Y}^{(\ell)} = (Y_1^{(\ell)}, \dots, Y_M^{(\ell)})^T$ no es una multinomial. En la literatura se han propuesto varios modelos para caracterizar la dependencia entre las observaciones. A continuación se presentan algunas de las distribuciones que pueden modelizar esta situación: la distribución *Multinomial n-inflada* propuesta por Cohen y Altham (1976), la *Dirichlet Multinomial*, introducida por Brier (1980) y la *Multinomial Agrupada Aleatoriamente*, propuesta por Morel y Nagaraj (1993).

Cohen presentó su modelo para explicar la dependencia existente en datos demográficos entre miembros de una misma familia (cluster). En concreto, supuso que se partía de un diseño muestral en el que se tienen N clusters en los que se observan dos individuos ($n = 2$). Posteriormente, Altham generalizó este modelo para n observaciones por cada

cluster. La suposición esencial de este modelo es que la distribución del vector aleatorio $\mathbf{Y} = (Y_1, \dots, Y_M)^T$ es de la forma

$$\Pr(Y_1 = y_1, \dots, Y_M = y_M) = (1 - \rho^2) \frac{n!}{y_1! \dots y_M!} \prod_{r=1}^M p_r(\boldsymbol{\theta})^{y_r} + \rho^2 \sum_{r=1}^M I_{\{y_r\}}(n) p_r(\boldsymbol{\theta}), \quad (1.6)$$

para $r = 1, \dots, M$, siendo ρ^2 el coeficiente de correlación intracluster e $I_{\{y_r\}}(n)$ definido mediante

$$I_{\{y_r\}}(n) = \begin{cases} 1, & \text{si } y_r = n \\ 0, & \text{si } y_r \neq n \end{cases} \quad r = 1, \dots, M.$$

Esta distribución ha sido posteriormente denominada como *multinomial n-inflada*. Obsérvese que es una mixtura entre una multinomial y la distribución que asigna probabilidad p_r a los resultados en los que los individuos caen todos en la celda r ($y_r = n$), $r = 1, \dots, M$, es decir

$$\Pr(Y_1 = 0, \dots, Y_{r-1} = 0, Y_r = n, Y_{r+1} = 0, \dots, Y_M = 0) = (1 - \rho^2) p_r(\boldsymbol{\theta})^n + \rho^2 p_r(\boldsymbol{\theta}),$$

en los demás casos la probabilidad se calcula según la distribución Multinomial multiplicada por el coeficiente $(1 - \rho^2)$.

Para aclarar la interpretación de este modelo de probabilidad, se plantea el ejemplo presentado por Cohen en el que se estudió la dependencia entre padecer esquizofrenia y el sexo del paciente. Para realizar este estudio se disponía de una muestra que estaba formada por parejas de hermanos de diferentes familias. Parece razonable suponer que dentro de cada cluster (familia en este caso) los individuos no son independientes. Así pues, en este caso tenemos $M = 2 \times 2$ categorías $\{YM, YF, NM, NF\}$ que se obtienen de cruzar los valores de las variables: esquizofrenia (Y=yes, N=No) y sexo (M=male, F=female). Los valores posibles de $Y_r^{(\ell)}$, número de individuos que en la familia ℓ -ésima presentan la modalidad r ,

son $\{0, 1, 2\}$, para $r = 1, \dots, 4$ y $\ell = 1, \dots, N$, donde N es el número de familias muestreadas.

En cada par,

$$\Pr(Y_1 = y_1, \dots, Y_4 = y_4) = (1 - \rho^2) \frac{2!}{y_1! \times \dots \times y_4!} \prod_{r=1}^4 p_r(\boldsymbol{\theta})^{y_r} + \rho^2 \sum_{r=1}^4 I_{\{y_r\}}(2) p_r(\boldsymbol{\theta}),$$

para $r = 1, \dots, 4$, donde

$$I_{\{y_r\}}(2) = \begin{cases} 1 & \text{si } y_r = 2 \\ 0 & \text{si } y_r \neq 2 \end{cases}.$$

Por ejemplo,

$$\Pr(Y_1 = 1, Y_2 = 0, Y_3 = 0, Y_4 = 1) = (1 - \rho^2) 2p_1(\boldsymbol{\theta}) p_4(\boldsymbol{\theta}),$$

$$\Pr(Y_1 = 0, Y_2 = 2, Y_3 = 0, Y_4 = 0) = (1 - \rho^2) p_2(\boldsymbol{\theta})^2 + \rho^2 p_2(\boldsymbol{\theta}).$$

Obsérvese que si $\rho^2 = 0$, estamos en el caso de independencia y la probabilidad corresponde a una multinomial. Mientras que para $\rho^2 = 1$ estaríamos en el caso de total dependencia entre el sexo y la esquizofrenia dentro de cada familia. Es decir, la probabilidad conjunta se calcula como una mixtura entre una multinomial y los casos extremos, con un coeficiente $\rho^2 > 0$ que mide el grado de dependencia dentro de cada familia.

Brier propuso la distribución *Dirichlet multinomial*, cuya definición se basa en la hipótesis de que el vector de probabilidad en cada cluster $\mathbf{p} = (p_1, \dots, p_M)^T$ es una variable aleatoria con distribución de Dirichlet

$$\Pr(P_1 = p_1, \dots, P_M = p_M | \mathbf{p}(\boldsymbol{\theta}), c) = \frac{\Gamma(c)}{\prod_{r=1}^M \Gamma(cp_r(\boldsymbol{\theta}))} \prod_{r=1}^M p_r^{cp_r(\boldsymbol{\theta})-1},$$

cuyos parámetros son $\mathbf{p}(\boldsymbol{\theta})$ y $c = \rho^{-2}(1 - \rho^2) > 0$ y la distribución condicionada $(\mathbf{Y} | \mathbf{p}(\boldsymbol{\theta})) \sim \mathcal{M}(n, \mathbf{p}(\boldsymbol{\theta}))$. De forma que, la distribución de $\mathbf{Y} = (Y_1, \dots, Y_M)^T$ viene dada por la función de probabilidad

$$\Pr(Y_1 = y_1, \dots, Y_M = y_M) = \frac{n!}{y_1! \cdots y_M!} \frac{\Gamma(c)}{\Gamma(n+c)} \frac{\prod_{r=1}^M \Gamma(y_r + cp_r(\boldsymbol{\theta}))}{\prod_{r=1}^M \Gamma(cp_r(\boldsymbol{\theta}))}. \quad (1.7)$$

Esta distribución fue introducida por Mossiman (1962).

Morel y Nagaraj (1993) propusieron la distribución *Multinomial agrupada aleatoriamente* (random-clumped), generada a partir de la mixtura de distribuciones multinomiales, cuyo vector de probabilidad es una media ponderada entre $\mathbf{p}(\boldsymbol{\theta})$ y la probabilidad degenerada en la celda r -ésima

$$U^{(r)} \sim \mathcal{M}(n, (1 - \rho)\mathbf{p}(\boldsymbol{\theta}) + \rho\mathbf{e}_r), \quad r = 1, \dots, M,$$

donde $\rho \in [0, 1]$, y \mathbf{e}_r es el r -ésimo vector unitario de dimensión M (1 en la posición r y cero en el resto), siendo $p_r(\boldsymbol{\theta})$, $r = 1, \dots, M$, la ponderación de la mixtura, la función de masa es

$$\Pr(Y_1 = y_1, \dots, Y_M = y_M) = \sum_{r=1}^M p_r(\boldsymbol{\theta}) \Pr(U^{(r)} = \mathbf{y}). \quad (1.8)$$

Hall (2000) propuso la distribución *Binomial 0-inflada* generada a partir de la mixtura de una variable binomial de parámetros $(n, (\mathbf{p}_1(\boldsymbol{\theta}), \mathbf{p}_2(\boldsymbol{\theta}))^T)$ con otra binomial de parámetros $(n, (0, 1)^T)$. Se toma una ponderación genérica $(\omega, 1 - \omega)$ para ambas variables aleatorias.

Para todos estos modelos se puede comprobar que la media y la varianza tienen, en todos ellos, la misma expresión:

$$\mathbf{E} \left[(Y_1, \dots, Y_M)^T \right] = n(p_1(\boldsymbol{\theta}), \dots, p_M(\boldsymbol{\theta}))^T \quad \mathbf{y} \quad \mathbf{Var} \left[(Y_1, \dots, Y_M)^T \right] = \vartheta_n n \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})}, \quad (1.9)$$

con

$$\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})} = \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})} - \mathbf{p}(\boldsymbol{\theta}) \mathbf{p}(\boldsymbol{\theta})^T,$$

siendo $\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}$ la matriz diagonal formada con el vector $\mathbf{p}(\boldsymbol{\theta})$ y

$$\vartheta_n = 1 + (n - 1)\rho^2 > 1. \quad (1.10)$$

Al término ϑ_n se le denomina *efecto del diseño* o *parámetro de sobredispersión*, ya que representa el factor por el que se multiplica la matrix de varianzas covarianzas $\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})}$ de la distribución multinomial $\mathcal{M}(n; (p_1(\boldsymbol{\theta}), \dots, p_M(\boldsymbol{\theta}))^T)$. Esta expresión será, a partir de ahora, la condición que se exigirá a la distribución de probabilidad, de forma que los resultados se desarrollarán, no para una distribución de probabilidad concreta, sino simplemente para todas aquellas que verifiquen la condición (1.9).

A continuación se demuestra la verificación de las expresiones (1.9) para el caso particular de la distribución binomial 0-inflada.

Teorema 1.1 *La distribución Binomial 0-inflada en la primera celda, es decir n-inflada en la segunda celda expresada como*

$$\left[(Y_1, Y_2)^T \mid V = v \right] = \begin{cases} \mathcal{M}\left(n, (p_1(\boldsymbol{\theta}), p_2(\boldsymbol{\theta}))^T\right) & \text{si } v = 1, \quad \text{con } \Pr(V = 1) = \omega \\ n\mathbf{e}_2 & \text{si } v = 0, \quad \text{con } \Pr(V = 0) = 1 - \omega \end{cases},$$

verifica que

$$\mathbf{E} \left[(Y_1, Y_2)^T \right] = n(\omega p_1(\boldsymbol{\theta}), 1 - \omega p_1(\boldsymbol{\theta}))^T \quad \text{y} \quad \mathbf{Var} \left[(Y_1, Y_2)^T \right] = \vartheta_n n \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})}. \quad (1.11)$$

Demostración Su vector de momentos de primer orden viene dado por

$$\begin{aligned}
E[\mathbf{Y}] &= E[E[\mathbf{Y}|V]] = E[\mathbf{Y}|V=1]\Pr(V=1) + E[n\mathbf{e}_2|V=0]\Pr(V=0) \\
&= n \begin{pmatrix} p_1(\boldsymbol{\theta}) \\ 1-p_1(\boldsymbol{\theta}) \end{pmatrix} \omega + \begin{pmatrix} 0 \\ n \end{pmatrix} (1-\omega) \\
&= n \begin{pmatrix} \omega p_1(\boldsymbol{\theta}) \\ 1-\omega p_1(\boldsymbol{\theta}) \end{pmatrix}.
\end{aligned}$$

Para encontrar la expresión de la matriz de los momentos de segundo orden, primero se calcula

$$\begin{aligned}
E[\text{Var}[\mathbf{Y}|V]] &= \text{Var}[\mathbf{Y}|V=1]\Pr(V=1) + \text{Var}[n\mathbf{e}_2|V=0]\Pr(V=0) \\
&= n\omega p_1(\boldsymbol{\theta})(1-p_1(\boldsymbol{\theta})) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},
\end{aligned}$$

y además

$$\begin{aligned}
\text{Var}[E[\mathbf{Y}|V]] &= E[E[\mathbf{Y}|V]E^T[\mathbf{Y}|V]] - E[E[\mathbf{Y}|V]]E^T[E[\mathbf{Y}|V]] \\
&= E[\mathbf{Y}|V=1]E^T[\mathbf{Y}|V=1]\omega + E[\mathbf{Y}|V=0]E^T[\mathbf{Y}|V=0](1-\omega) \\
&\quad - n^2 \begin{pmatrix} \omega p_1(\boldsymbol{\theta}) \\ 1-\omega p_1(\boldsymbol{\theta}) \end{pmatrix} \begin{pmatrix} \omega p_1(\boldsymbol{\theta}) & 1-\omega p_1(\boldsymbol{\theta}) \end{pmatrix} \\
&= n^2 \begin{pmatrix} p_1(\boldsymbol{\theta}) \\ 1-p_1(\boldsymbol{\theta}) \end{pmatrix} \begin{pmatrix} p_1(\boldsymbol{\theta}) & 1-p_1(\boldsymbol{\theta}) \end{pmatrix} \omega + \begin{pmatrix} 0 \\ n \end{pmatrix} \begin{pmatrix} 0 & n \end{pmatrix} (1-\omega) \\
&\quad - n^2 \begin{pmatrix} \omega p_1(\boldsymbol{\theta}) \\ 1-\omega p_1(\boldsymbol{\theta}) \end{pmatrix} \begin{pmatrix} \omega p_1(\boldsymbol{\theta}) & 1-\omega p_1(\boldsymbol{\theta}) \end{pmatrix}
\end{aligned}$$

luego

$$\begin{aligned}
\text{Var}[E[\mathbf{Y}|V]] &= n^2\omega \begin{pmatrix} p_1^2(\boldsymbol{\theta}) & p_1(\boldsymbol{\theta})(1-p_1(\boldsymbol{\theta})) \\ p_1(\boldsymbol{\theta})(1-p_1(\boldsymbol{\theta})) & (1-p_1(\boldsymbol{\theta}))^2 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & n^2 \end{pmatrix} (1-\omega) \\
&\quad - n^2 \begin{pmatrix} \omega^2 p_1^2(\boldsymbol{\theta}) & \omega p_1(\boldsymbol{\theta})(1-\omega p_1(\boldsymbol{\theta})) \\ \omega p_1(\boldsymbol{\theta})(1-\omega p_1(\boldsymbol{\theta})) & (1-\omega p_1(\boldsymbol{\theta}))^2 \end{pmatrix} \\
&= n^2(1-\omega)\omega p_1^2(\boldsymbol{\theta}) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},
\end{aligned}$$

y así

$$\begin{aligned}
\text{Var}[\mathbf{Y}] &= E[\text{Var}[\mathbf{Y}|V]] + \text{Var}[E[\mathbf{Y}|V]] \\
&= n\omega p_1(\boldsymbol{\theta})(1-p_1(\boldsymbol{\theta})) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} + n^2(1-\omega)\omega p_1^2(\boldsymbol{\theta}) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \\
&= (1+\rho^2(n-1))n\omega p_1(\boldsymbol{\theta})(1-\omega p_1(\boldsymbol{\theta})) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}
\end{aligned}$$

donde

$$\rho^2 = \frac{(1-\omega)p_1(\boldsymbol{\theta})}{1-\omega p_1(\boldsymbol{\theta})} \quad \text{para cualquier } \omega \in (0,1).$$

Teniendo en cuenta que

$$\begin{aligned}
\Sigma_{\mathbf{p}(\boldsymbol{\theta})} &= \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})} - \mathbf{p}(\boldsymbol{\theta})\mathbf{p}(\boldsymbol{\theta})^T \\
&= \begin{pmatrix} \omega p_1(\boldsymbol{\theta}) & 0 \\ 0 & 1-\omega p_1(\boldsymbol{\theta}) \end{pmatrix} - \begin{pmatrix} \omega p_1(\boldsymbol{\theta})(1-\omega p_1(\boldsymbol{\theta})) & \omega p_1(\boldsymbol{\theta})(1-\omega p_1(\boldsymbol{\theta})) \\ \omega p_1(\boldsymbol{\theta})(1-\omega p_1(\boldsymbol{\theta})) & \omega p_1(\boldsymbol{\theta})(1-\omega p_1(\boldsymbol{\theta})) \end{pmatrix} \\
&\quad \omega p_1(\boldsymbol{\theta})(1-\omega p_1(\boldsymbol{\theta})) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},
\end{aligned}$$

se tiene (1.11). ■

Este resultado coincide con el establecido por Morel y Neerchal (2012, pg 83). Se puede extender esta definición siendo

$$(\mathbf{Y} | V = v) = \begin{cases} \mathcal{M}(n, \mathbf{p}(\boldsymbol{\theta})) & \text{si } v = 1, \quad \text{con } \Pr(V = 1) = \omega \\ ne_M & \text{si } v = 0, \quad \text{con } \Pr(V = 0) = 1 - \omega \end{cases}$$

la distribución multinomial cero inflada en las $M - 1$ primeras celdas, es decir n-inflada en celda M -ésima. Para $M \geq 3$ no parece apropiado utilizar un coeficiente de correlación ρ^2 homogéneo en todas las categorías porque la correlación intracluster para cada una de las celdas parece ser heterogénea. La razón para esto es que para $M \geq 3$ no se puede expresar su matriz de varianzas-covarianzas como (1.9). Para el resto de las distribuciones puede consultarse Morel y Neerchal (2012).

En todos estos modelos es el coeficiente de correlación intracluster, ρ^2 , el que más fácilmente permite interpretar la naturaleza de la relación entre individuos de un mismo cluster. Suponiendo que la correlación es la misma para cada par de individuos h y s ($h, s = 1, \dots, n_\ell, \quad h \neq s$) de cualquier cluster $\ell = 1, \dots, N$, el coeficiente de correlación intracluster para la categoría r , ρ_r^2 , se define como

$$\begin{aligned} \rho_r^2 &= Cor \left[I_{\{r\}}(\delta(X_h^{(\ell)})), I_{\{r\}}(\delta(X_s^{(\ell)})) \right] \\ &= \frac{E \left[I_{\{r\}}(\delta(X_h^{(\ell)})), I_{\{r\}}(\delta(X_s^{(\ell)})) \right] - E \left[I_{\{r\}}(\delta(X_h^{(\ell)})) \right] E \left[I_{\{r\}}(\delta(X_s^{(\ell)})) \right]}{\sqrt{Var \left(I_{\{r\}}(\delta(X_h^{(\ell)})) \right) Var \left(I_{\{r\}}(\delta(X_s^{(\ell)})) \right)}}. \end{aligned}$$

Puesto que la correlación es homogénea en todas las posibles categorías $\rho_r^2 = \rho^2, \quad r = 1, \dots, M$, y si se tiene en cuenta que $E \left[I_{\{r\}}(\delta(X_h^{(\ell)})) \right] = p_r(\boldsymbol{\theta})$, entonces

$$\rho^2 = \frac{\Pr(\delta(X_h^{(\ell)}) = r, \delta(X_s^{(\ell)}) = r) - p_r^2(\boldsymbol{\theta})}{p_r(\boldsymbol{\theta})(1 - p_r(\boldsymbol{\theta}))}, \quad \ell = 1, \dots, N, \quad h, s = 1, \dots, n_\ell, \quad h \neq s,$$

luego

$$\text{Cov} \left[I_{\{r\}}(\delta(X_h^{(\ell)})), I_{\{r\}}(\delta(X_s^{(\ell)})) \right] = \begin{cases} \rho^2 p_r(\boldsymbol{\theta}) (1 - p_r(\boldsymbol{\theta})) & h \neq s \\ p_r(\boldsymbol{\theta}) (1 - p_r(\boldsymbol{\theta})) & h = s \end{cases}.$$

Teorema 1.2 *Supuesto que existe el mismo grado de sobredispersión en las M categorías bajo consideración; es decir $\rho_1^2 = \dots = \rho_M^2 = \rho^2$, $\vartheta_n^{(1)} = \dots = \vartheta_n^{(M)} = \vartheta_n$ se verifica que*

$$\text{Var} [Y_r^{(\ell)}] = n\vartheta_n p_r(\boldsymbol{\theta}) (1 - p_r(\boldsymbol{\theta})), \quad r = 1, \dots, M, \quad \ell = 1, \dots, N.$$

Demostración Se pueden considerar las variables latentes Z_r que modelizan la probabilidad de cada suceso de forma que $E[Z_r] = p_r(\boldsymbol{\theta})$, es decir,

$$(Y_r | Z_r = p_r(\boldsymbol{\theta})) \sim \text{Bin}(n, p_r(\boldsymbol{\theta})).$$

Como el soporte de la variable Z_r es $[0, 1]$, se sigue que $Z_r \geq Z_r^2$ y por tanto $E[Z_r] \geq E[Z_r^2]$

ó

$$\begin{aligned} E[Z_r] - E^2[Z_r] &\geq E[Z_r^2] - E^2[Z_r] \\ p_r(\boldsymbol{\theta}) (1 - p_r(\boldsymbol{\theta})) &\geq \text{Var}(Z_r). \end{aligned} \tag{1.12}$$

De (1.12), se deduce que existe $\rho_r^2 \in [0, 1]$ tal que $\text{Var}(Z_r) = \rho_r^2 p_r(\boldsymbol{\theta}) (1 - p_r(\boldsymbol{\theta}))$ y así

$$E[Y_r^{(\ell)}] = E[E[Y_r^{(\ell)} | Z_r]] = np_r(\boldsymbol{\theta}),$$

y

$$\begin{aligned}
\text{Var} [Y_r^{(\ell)}] &= E [\text{Var} [Y_r^{(\ell)} | Z_r]] + \text{Var} [E [Y_r^{(\ell)} | Z_r]] = E [nZ_r(1 - Z_r)] + \text{Var} [nZ_r] \\
&= n [E [Z_r] - \text{Var}(Z_r) - E^2 [Z_r]] + n^2 \rho_r^2 p_r(\boldsymbol{\theta}) (1 - p_r(\boldsymbol{\theta})) \\
&= np_r(\boldsymbol{\theta}) - n\rho_r^2 p_r(\boldsymbol{\theta}) (1 - p_r(\boldsymbol{\theta})) - np_r^2(\boldsymbol{\theta}) + n^2 \rho_r^2 p_r(\boldsymbol{\theta}) - n^2 \rho_r^2 p_r^2(\boldsymbol{\theta}) \\
&= n [p_r(\boldsymbol{\theta}) - \rho_r^2 p_r(\boldsymbol{\theta}) + \rho_r^2 p_r^2(\boldsymbol{\theta}) - p_r^2(\boldsymbol{\theta}) + n\rho_r^2 p_r(\boldsymbol{\theta}) - n\rho_r^2 p_r^2(\boldsymbol{\theta})] \\
&= n [p_r(\boldsymbol{\theta}) (1 + \rho_r^2(n - 1) - p_r^2(\boldsymbol{\theta}) (1 + \rho_r^2(n - 1)))] \\
&= n(1 + \rho_r^2(n - 1))p_r(\boldsymbol{\theta}) (1 - p_r(\boldsymbol{\theta})) = n\vartheta_n^{(r)} p_r(\boldsymbol{\theta}) (1 - p_r(\boldsymbol{\theta})).
\end{aligned}$$

Puesto que se considera el mismo grado de sobredispersión en las M categorías, se verifica que $\rho_1^2 = \dots = \rho_M^2 = \rho^2$, $\vartheta_n^{(1)} = \dots = \vartheta_n^{(M)} = \vartheta_n$, y así

$$\text{Var} [Y_r^{(\ell)}] = n\vartheta_n p_r(\boldsymbol{\theta}) (1 - p_r(\boldsymbol{\theta})) \quad r = 1, \dots, M, \quad \ell = 1, \dots, N,$$

que coincide con la diagonal de la matriz de varianzas-covarianzas (1.9). ■

1.4. Estimador no paramétrico. Distribución asintótica

A continuación se presenta la distribución asintótica del estimador no paramétrico para datos con sobredispersión. El estimador no paramétrico de $p_r(\boldsymbol{\theta})$, $r = 1, \dots, M$, basado en los N clusters viene dado por

$$\widehat{p}_r = \frac{1}{nN} \sum_{\ell=1}^N Y_r^{(\ell)}. \quad (1.13)$$

Se observa que dicho estimador se puede expresar en función de los estimadores de cada cluster $\widehat{p}_r^{(\ell)}$,

$$\widehat{p}_r = \frac{1}{N} \sum_{r=1}^N \frac{Y_r^{(\ell)}}{n} = \frac{1}{N} \sum_{\ell=1}^N \widehat{p}_r^{(\ell)}.$$

Aplicando la ley fuerte de los grandes números cuando el número de clusters crece hacia ∞ , se tiene

$$\widehat{\mathbf{p}} = (\widehat{p}_1, \dots, \widehat{p}_M)^T \xrightarrow[N \rightarrow \infty]{c.s.} \mathbf{p}(\boldsymbol{\theta}) = (p_1(\boldsymbol{\theta}), \dots, p_M(\boldsymbol{\theta}))^T.$$

Aplicando el Teorema Central del Límite cuando el número de clusters crece hacia ∞ , y bajo la suposición de que las componentes $(Y_1, \dots, Y_M)^T$ son independientes, es decir, en el caso de muestreo multinomial se tiene

$$\sqrt{N} (\hat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta})) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_M, \frac{1}{n} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})}).$$

Siendo $\boldsymbol{\theta}_0$ el valor verdadero y desconocido del parámetro, para el modelo con sobredispersión definido en (1.9), cuando el número de clusters crece hacia ∞ , para el estimador no paramétrico definido en (1.13), se verifica

$$\sqrt{N} (\hat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_M, \frac{\vartheta_n}{n} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}). \quad (1.14)$$

CAPÍTULO 2

EL ESTIMADOR DE MÍNIMA CUASI FI-DIVERGENCIA EN DATOS MULTINOMIALES POR CONGLOMERADOS

En este Capítulo se presenta un enfoque diferente para el análisis de datos multinomiales con sobredispersión mediante la utilización de un modelo loglineal donde no se supone ninguna distribución subyacente concreta para los datos analizados. El objetivo es presentar nuevos estimadores basados en medidas de ϕ -divergencia, (Pardo, 2006), tanto para los parámetros propios del modelo loglineal como para los necesarios para medir el efecto del diseño del muestro por conglomerados. Puesto que no se supondrá ninguna distribución específica subyacente, sino simplemente hipótesis sobre los dos primeros momentos de la distribución que introducen la medida de sobredispersión, se obtendrán los *estimadores de mínima cuasi ϕ -divergencia* ($EMC\phi$) para los parámetros del modelo loglineal, así como su distribución asintótica. Los estimadores de máxima cuasi-verosimilitud ($EMCV$), introducidos por Wedderburn (1974), son un caso particular de los anteriores para modelos multinomiales con datos agrupados en clusters. Además, se presentarán nuevos estimadores del parámetro de sobredispersión para el caso en el que el tamaño de todos los clusters sea el mismo. Se finalizará este Capítulo con una aplicación práctica de los nuevos estimadores y con un estudio de simulación que compara el comportamiento de los nuevos estimadores, en cuanto a su sesgo y error cuadrático medio, con otros estimadores propuestos anteriormente.

2.1. El estimador de mínima cuasi ϕ -divergencia en el modelo loglineal asociado a datos multinomiales por conglomerados

Si los vectores $\mathbf{Y}^{(\ell)}$, $\ell = 1, \dots, N$ definidos en (1.1) siguen una distribución multinomial de parámetros n y $\mathbf{p}(\theta)$, el vector

$$\mathbf{Y} = \sum_{\ell=1}^N \mathbf{Y}^{(\ell)}$$

sigue una distribución multinomial de parámetros nN y $\mathbf{p}(\theta)$. Así, el estimador de máxima verosimilitud de θ será el valor del espacio paramétrico que maximice la verosimilitud,

$$\Pr(Y_1 = y_1, \dots, Y_M = y_M) = \binom{nN}{y_1, \dots, y_M} p_1(\theta)^{y_1} \cdot \dots \cdot p_M(\theta)^{y_M},$$

o equivalentemente

$$\log \Pr(Y_1 = y_1, \dots, Y_M = y_M) = -nN d_{Kullback}(\hat{\mathbf{p}}, \mathbf{p}(\theta)) + K,$$

donde $d_{Kullback}(\hat{\mathbf{p}}, \mathbf{p}(\theta))$ es la divergencia de Kullback entre los vectores de probabilidad $\hat{\mathbf{p}}$ y $\mathbf{p}(\theta)$; es decir,

$$d_{Kullback}(\hat{\mathbf{p}}, \mathbf{p}(\theta)) = \sum_{r=1}^M \hat{p}_r \log \frac{\hat{p}_r}{p_r(\theta)}$$

y K una constante arbitraria que no depende de θ . Luego, el estimador de máxima verosimilitud para el modelo multinomial es el valor $\hat{\theta} = \hat{\theta}(\mathbf{Y})$ verificando

$$\hat{\theta}(\mathbf{Y}) = \arg \min_{\theta \in \Theta} d_{Kullback}(\hat{\mathbf{p}}, \mathbf{p}(\theta)). \quad (2.1)$$

Para modelos loglineales multinomiales, Cressie y Pardo (2000) introdujeron el concepto de estimador de mínima ϕ -divergencia en base a las medidas de ϕ -divergencia. La medida de ϕ -divergencia entre los vectores $\hat{\mathbf{p}}$ y $\mathbf{p}(\theta)$ viene dada por

$$d_\phi(\hat{\mathbf{p}}, \mathbf{p}(\theta)) = \sum_{r=1}^M p_r(\theta) \phi\left(\frac{\hat{p}_r}{p_r(\theta)}\right), \quad \text{con } \phi \in \Phi^*, \quad (2.2)$$

siendo Φ^* la clase de todas las funciones convexas $\phi(x)$, $x > 0$, tales que en $x = 1$, $\phi(1) = 0$, $\phi''(1) > 0$ y en $x = 0$, $0\phi(0/0) = 0$ y $0\phi(p/0) = \lim_{u \rightarrow \infty} p\phi(u)/u$. De esta forma, para cada $\phi \in \Phi^*$ que sea diferenciable en $x = 1$, la función

$$\psi(x) = \phi(x) - \phi'(1)(x - 1),$$

también pertenece a Φ^* . Para más detalles acerca de las medidas de ϕ -divergencia ver Pardo (2006).

A partir de (2.2) se define el estimador de mínima ϕ -divergencia para el modelo multinomial mediante

$$\hat{\boldsymbol{\theta}}(\mathbf{Y}) = \arg \min_{\boldsymbol{\theta} \in \Theta} d_{\phi}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})). \quad (2.3)$$

Obsérvese que si se considera la función

$$\phi(x) = x \log x - x + 1,$$

se obtiene la conocida divergencia de Kullback

$$d_{Kullback}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})) = \sum_{r=1}^M \hat{p}_r \log \frac{\hat{p}_r}{p_r(\boldsymbol{\theta})},$$

y como consecuencia el estimador de máxima verosimilitud es un elemento de la familia de estimadores de mínima ϕ -divergencia. Una familia de ϕ -divergencias que será importante a lo largo de la presente memoria será la familia de ϕ -divergencias introducida y estudiada por Cressie y Read (1984), en lo sucesivo familia de divergencias Cressie y Read, y que se obtiene a partir de las funciones

$$\phi_{\lambda}(x) = \begin{cases} \frac{1}{\lambda(1+\lambda)} [x^{\lambda+1} - x - \lambda(x-1)], & \lambda \notin \{-1, 0\} \\ \lim_{v \rightarrow \lambda} \frac{1}{v(1+v)} [x^{v+1} - x - v(x-1)], & \lambda \in \{-1, 0\} \end{cases},$$

con $\lambda \in \mathbb{R}$. Así, las diferentes medidas de ϕ -divergencia de esta familia dependiendo del valor del parámetro λ son las siguientes

$$d_{\phi_\lambda}(\widehat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})) = \begin{cases} \frac{1}{\lambda(\lambda+1)} \sum_{r=1}^M \left(\frac{\widehat{p}_r^{\lambda+1}}{p_r^{\lambda+1}(\boldsymbol{\theta})} - p_r(\boldsymbol{\theta}) \right), & \lambda \notin \{-1, 0\} \\ d_{Kullback}(\mathbf{p}(\boldsymbol{\theta}), \widehat{\mathbf{p}}), & \lambda = -1 \\ d_{Kullback}(\widehat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})), & \lambda = 0 \end{cases} \quad (2.4)$$

A partir de (2.4), el estimador de mínima divergencia de Cressie y Read, para el muestreo multinomial, se define mediante

$$\widehat{\boldsymbol{\theta}}_\lambda = \arg \min_{\boldsymbol{\theta} \in \Theta} d_{\phi_\lambda}(\widehat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})).$$

Obsérvese que para $\lambda = 0$ se obtiene el estimador de máxima verosimilitud para modelos multinomiales. Para $\lambda = 1$, $d_{\phi_1}(\widehat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta}))$, es el estimador de mínima ji-cuadrado.

Si $\rho^2 = 0$ la tabla de contingencia asociada al único cluster o conglomerado ($N = 1$) $\mathbf{Y} = \mathbf{Y}^{(1)}$, que sigue una distribución multinomial de parámetros n y $\mathbf{p}(\boldsymbol{\theta})$, es suficiente para aplicar la ley Débil de los Grandes Números o el Teorema Central del Límite sin más que hacer el número, n , de observaciones independientes lo suficientemente grande. Si $\rho^2 > 0$ entonces n debe estar fijo y N debe ser suficientemente grande. La definición que se dará seguidamente va a poner de manifiesto que la definición (2.3), válida en principio para $\rho^2 = 0$, va a ser útil también para el caso $\rho^2 > 0$. Ahora sólo se supondrá que

$$E[\mathbf{Y}^{(\ell)}] = n\mathbf{p}(\boldsymbol{\theta}), \quad y \quad Var[\mathbf{Y}^{(\ell)}] = n\vartheta_n \boldsymbol{\Sigma}_p(\boldsymbol{\theta}_0), \quad (2.5)$$

con $\vartheta_n = 1 + (n-1)\rho^2 \in (1, n]$; es decir, no se supondrá la existencia de una distribución subyacente por lo que el estimador de $\boldsymbol{\theta}$ obtenido será un estimador “de mínima cuasi ϕ -divergencia” ($MC\phi$).

Definición 2.1 Sea un modelo estadístico verificando (2.5). El estimador de mínima cuasi ϕ -divergencia (MC ϕ) para $\boldsymbol{\theta}$, se define mediante

$$\widehat{\boldsymbol{\theta}}_\phi = \widehat{\boldsymbol{\theta}}_\phi(\mathbf{Y}) = \arg \min_{\boldsymbol{\theta} \in \Theta} d_\phi(\widehat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})),$$

donde $d_\phi(\widehat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta}))$ es la ϕ -divergencia entre los vectores $\widehat{\mathbf{p}}$ y $\mathbf{p}(\boldsymbol{\theta})$.

El siguiente teorema da un procedimiento para la obtención del estimador de cuasi mínima ϕ -divergencia en términos de la resolución de un sistema de ecuaciones no lineales en $\boldsymbol{\theta}$.

Teorema 2.1 El estimador de mínima cuasi ϕ -divergencia, para el parámetro $\boldsymbol{\theta}$ se obtiene como resolución del sistema de ecuaciones, no lineales en $\boldsymbol{\theta}$, dado por

$$\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-1} \boldsymbol{\Psi}^\phi(\boldsymbol{\theta}) = \mathbf{0}_{M_0}. \quad (2.6)$$

siendo

$$\boldsymbol{\Psi}^\phi(\boldsymbol{\theta}) = \left(\Psi_1^\phi(\boldsymbol{\theta}), \dots, \Psi_M^\phi(\boldsymbol{\theta}) \right)^T,$$

con

$$\Psi_r^\phi(\boldsymbol{\theta}) = \widehat{p}_r \phi' \left(\frac{\widehat{p}_r}{p_r(\boldsymbol{\theta})} \right) - p_r(\boldsymbol{\theta}) \phi \left(\frac{\widehat{p}_r}{p_r(\boldsymbol{\theta})} \right), \quad r = 1, \dots, M. \quad (2.7)$$

Demostración La búsqueda del estimador de mínima cuasi ϕ -divergencia para el parámetro $\boldsymbol{\theta}$ llevará a la resolución del sistema de ecuaciones

$$\frac{\partial d_\phi(\widehat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta}))}{\partial \theta_g} = 0; \quad g = 1, \dots, M_0,$$

con las condiciones que establece el modelo loglineal

$$p_r(\boldsymbol{\theta}) = \frac{\exp\{\mathbf{w}_r^T \boldsymbol{\theta}\}}{\sum_{u=1}^M \exp\{\mathbf{w}_u^T \boldsymbol{\theta}\}}, \quad r = 1, \dots, M.$$

Al ser

$$d_\phi(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})) = \sum_{r=1}^M p_r(\boldsymbol{\theta}) \phi\left(\frac{\hat{p}_r}{p_r(\boldsymbol{\theta})}\right),$$

derivando la ϕ -divergencia con respecto θ_g , $g = 1, \dots, M_0$ e igualando a cero, se tiene

$$\sum_{r=1}^M \phi\left(\frac{\hat{p}_r}{p_r(\boldsymbol{\theta})}\right) \frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_g} - \sum_{r=1}^M p_r(\boldsymbol{\theta}) \phi'\left(\frac{\hat{p}_r}{p_r(\boldsymbol{\theta})}\right) \frac{\hat{p}_r}{p_r(\boldsymbol{\theta})^2} \frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_g} = 0,$$

o lo que es lo mismo

$$\sum_{r=1}^M \frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_g} \left[\phi\left(\frac{\hat{p}_r}{p_r(\boldsymbol{\theta})}\right) - \phi'\left(\frac{\hat{p}_r}{p_r(\boldsymbol{\theta})}\right) \frac{\hat{p}_r}{p_r(\boldsymbol{\theta})} \right] = 0.$$

Así

$$\frac{\partial d_\phi(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta}))}{\partial \theta_g} = - \sum_{r=1}^M \frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_g} \frac{\Psi_r^\phi(\boldsymbol{\theta})}{p_r(\boldsymbol{\theta})}, \quad g = 1, \dots, M_0,$$

donde $\Psi_r^\phi(\boldsymbol{\theta})$ se define en (2.7). Ahora bien, para el modelo loglineal se tiene

$$\begin{aligned} \frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_g} &= \frac{\exp\{\mathbf{w}_r^T \boldsymbol{\theta}\} w_{rg} \sum_{u=1}^M \exp\{\mathbf{w}_u^T \boldsymbol{\theta}\} - \exp\{\mathbf{w}_r^T \boldsymbol{\theta}\} \sum_{u=1}^M \exp\{\mathbf{w}_u^T \boldsymbol{\theta}\} w_{ug}}{\left(\sum_{u=1}^M \exp\{\mathbf{w}_u^T \boldsymbol{\theta}\}\right)^2} \quad (2.8) \\ &= \frac{\exp\{\mathbf{w}_r^T \boldsymbol{\theta}\} w_{rg}}{\sum_{u=1}^M \exp\{\mathbf{w}_u^T \boldsymbol{\theta}\}} - \frac{\exp\{\mathbf{w}_r^T \boldsymbol{\theta}\}}{\sum_{u=1}^M \exp\{\mathbf{w}_u^T \boldsymbol{\theta}\}} \sum_{u=1}^M \frac{\exp\{\mathbf{w}_u^T \boldsymbol{\theta}\} w_{ug}}{\sum_{u=1}^M \exp\{\mathbf{w}_u^T \boldsymbol{\theta}\}} \\ &= p_r(\boldsymbol{\theta}) w_{rg} - p_r(\boldsymbol{\theta}) \sum_{u=1}^M p_u(\boldsymbol{\theta}) w_{ug} \quad g = 1, \dots, M_0, \end{aligned}$$

con lo cual el sistema de ecuaciones que se deberá resolver para encontrar el estimador de cuasi mínima ϕ -divergencia en un modelo loglineal viene dado por,

$$- \sum_{r=1}^M \left(w_{rg} - \sum_{u=1}^M w_{ug} p_u(\boldsymbol{\theta}) \right) \Psi_r^\phi(\boldsymbol{\theta}) = 0 \quad g = 1, \dots, M_0. \quad (2.9)$$

Si se tiene en cuenta que $\Sigma_{p(\theta)} \mathbf{D}_{p(\theta)}^{-1} = \mathbf{I}_M - \mathbf{p}(\theta) \mathbf{1}_M^T$, se obtiene que el sistema anterior se puede escribir de forma matricial

$$\mathbf{W}^T \Sigma_{p(\theta)} \mathbf{D}_{p(\theta)}^{-1} \Psi^\phi(\theta) = \mathbf{0}_{M_0}.$$

■

Las ecuaciones (2.6) son funciones no lineales del parámetro desconocido θ . Para resolver este problema se utiliza el método de Newton-Raphson, de tal forma que en el paso $(t + 1)$ -ésimo el estimador $\hat{\theta}_\phi^{(t+1)}$, se obtiene a partir de $\hat{\theta}_\phi^{(t)}$ como

$$\hat{\theta}_\phi^{(t+1)} = \hat{\theta}_\phi^{(t)} - \hat{\theta}_\phi^{(t)} \mathbf{G}_\phi^{-1}(\hat{\theta}_\phi^{(t)}) \mathbf{W}^T \Sigma_{p(\hat{\theta}_\phi^{(t)})} \mathbf{D}_{p(\hat{\theta}_\phi^{(t)})}^{-1} \Psi^\phi(\hat{\theta}_\phi^{(t)}),$$

donde

$$\begin{aligned} \mathbf{G}_\phi(\theta) &= (G_{\phi,g,h}(\theta))_{g,h=1,\dots,M_0} = \mathbf{W}^T \Sigma_{p(\theta)} \mathbf{D}_{p(\theta)}^{-1} D_{\bar{\Psi}^\phi(\theta)} \mathbf{D}_{p(\theta)}^{-1} \Sigma_{p(\theta)} \mathbf{W} \quad (2.10) \\ &= \mathbf{W}^T (\mathbf{I}_M - \mathbf{p}(\theta) \mathbf{1}^T) D_{\bar{\Psi}^\phi(\theta)} (\mathbf{I}_M - \mathbf{p}(\theta) \mathbf{1}^T) \mathbf{W}, \end{aligned}$$

con

$$\begin{aligned} G_{\phi,g,h}(\theta) &= \frac{\partial^2}{\partial \theta_g \partial \theta_h} d_\phi(\hat{\mathbf{p}}, \mathbf{p}(\theta)) \\ &= \sum_{r=1}^M \phi'' \left(\frac{\hat{p}_r}{p_r(\theta)} \right) \left(\frac{\hat{p}_r}{p_r(\theta)} \right)^2 \frac{1}{p_r(\theta)} \frac{\partial p_r(\theta)}{\partial \theta_g} \frac{\partial p_r(\theta)}{\partial \theta_h} - \sum_{r=1}^M \frac{\partial^2 p_r(\theta)}{\partial \theta_j \partial \theta_g} \frac{\Psi_r^\phi(\theta)}{p_r(\theta)} \\ &= \sum_{r=1}^M \bar{\Psi}_r^\phi(\theta) \frac{1}{p_r^2(\theta)} \frac{\partial p_r(\theta)}{\partial \theta_g} \frac{\partial p_r(\theta)}{\partial \theta_h}, \end{aligned}$$

y

$$\begin{aligned} \bar{\Psi}^\phi(\theta) &= \left(\bar{\Psi}_1^\phi(\theta), \dots, \bar{\Psi}_M^\phi(\theta) \right)^T, \\ \bar{\Psi}_r^\phi(\theta) &= \phi'' \left(\frac{\hat{p}_r}{p_r(\theta)} \right) \frac{\hat{p}_r^2}{p_r(\theta)} - \Psi_r^\phi(\theta) \quad r = 1, \dots, M. \quad (2.11) \end{aligned}$$

Puesto que no se está suponiendo ninguna distribución específica subyacente sino simplemente una hipótesis sobre los dos primeros momentos de la distribución expresadas en (1.9), se puede decir que los estimadores aquí obtenidos son *estimadores de mínima cuasi ϕ -divergencia* (de aquí en adelante, $EMC\phi$), definidos por primera vez en Vos (1992). Llegados a este punto conviene mencionar que los estimadores de máxima cuasi-verosimilitud ($EMCV$), introducidos por Wedderburn (1974), son muy útiles para modelos multinomiales con datos agrupados en clusters, en particular para modelos loglineales multinomiales. Los $EMCVs$ se obtienen resolviendo el sistema de ecuaciones (2.6) con ϕ la divergencia de Kullback, es decir,

$$\Psi^\phi(\boldsymbol{\theta}) = \widehat{\boldsymbol{p}} - \boldsymbol{p}(\boldsymbol{\theta}).$$

Así, los $EMC\phi$ son generalizaciones de los $EMCVs$, para modelos loglineales con datos agrupados en clusters.

2.1.1. Propiedades asintóticas del estimador de mínima cuasi ϕ -divergencia

En esta sección se establecen las propiedades asintóticas del estimador de mínima cuasi ϕ -divergencia bajo las condiciones de regularidad de Birch que se detallan a continuación:

1. $\boldsymbol{\theta}_0$ es un punto interior de Θ .
2. $\pi_r = p_r(\boldsymbol{\theta}_0) > 0$ para $r = 1, \dots, M$. Así $\boldsymbol{\pi} = (\pi_1, \dots, \pi_M)^T$ es un punto interior de

$$\Delta_M = \{\boldsymbol{p} = (p_1, \dots, p_M)^T : p_r > 0, r = 1, \dots, M, \sum_{r=1}^M p_r = 1\}.$$

3. La función $\boldsymbol{p} : \Theta \rightarrow \Delta_M$ es totalmente diferenciable en $\boldsymbol{\theta}_0$, es decir existen las derivadas parciales de $p_r(\boldsymbol{\theta})$ con respecto a θ_g en $\boldsymbol{\theta}_0$ y $p_r(\boldsymbol{\theta})$ tiene una aproximación lineal en $\boldsymbol{\theta}_0$ dada por

$$p_r(\boldsymbol{\theta}) = p_r(\boldsymbol{\theta}_0) + \sum_{g=1}^{M_0} (\theta_g - \theta_{0g}) \frac{\partial p_r(\boldsymbol{\theta}_0)}{\partial \theta_g} + o(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|),$$

donde $o(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|)$ denota una función que verifica

$$\lim_{\boldsymbol{\theta} \rightarrow \boldsymbol{\theta}_0} \frac{o(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|)}{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|} = 0.$$

4. La matriz jacobiana

$$\mathbf{J}(\boldsymbol{\theta}) = \left. \frac{\partial \mathbf{p}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \left(\frac{\partial p_r(\boldsymbol{\theta}_0)}{\partial \theta_g} \right)_{\substack{r=1,\dots,M \\ g=1,\dots,M_0}}$$

es de rango completo, es decir $rg(\mathbf{J}(\boldsymbol{\theta})) = M_0$.

5. La función inversa $\mathbf{p}^{-1} : \Delta_M \longrightarrow \Theta$ es continua en $\mathbf{p}(\boldsymbol{\theta}_0) = \boldsymbol{\pi}$.

6. La función $\mathbf{p} : \Theta \longrightarrow \Delta_M$ es continua en cada punto $\boldsymbol{\theta} \in \Theta$.

Bajo estas hipótesis se establece la descomposición BAN (Best Asintotically Normal) del estimador de mínima cuasi ϕ -divergencia. Antes se enunciará el Teorema de la Función Implícita ya que se utilizará para su obtención:

“Sea

$$\mathbf{F}^T = (F_1, \dots, F_{M_0}) : \mathbb{R}^{M+M_0} \longrightarrow \mathbb{R}^{M_0},$$

una función continuamente diferenciable en un conjunto abierto $U \subset \mathbb{R}^{M+M_0}$ que contiene al punto

$$\left(\mathbf{x}^* = (x_1^*, \dots, x_M^*)^T, \mathbf{x}_0 = (x_1^0, \dots, x_{M_0}^0)^T \right),$$

para el cual $F(\mathbf{x}^*, \mathbf{x}_0) = \mathbf{0}_{M_0}$ y se supone que la matriz

$$\left(\frac{\partial F_i}{\partial x_j} \right)_{\substack{i=1,\dots,M_0 \\ j=M+1,\dots,M+M_0}}$$

es no singular en $(\mathbf{x}^*, \mathbf{x}_0)$. Entonces, existe un entorno M -dimensional U_0 de \mathbf{x}^* en \mathbb{R}^M y una única función diferenciable $\mathbf{g} : U_0 \longrightarrow \mathbb{R}^{M_0}$ de forma que $\mathbf{g}(\mathbf{x}^*) = \mathbf{x}_0$ y $F(\mathbf{x}, \mathbf{g}(\mathbf{x})) = \mathbf{0}_{M_0} \forall \mathbf{x} \in U_0$ ”.

Teorema 2.2 *Se considera un modelo estadístico verificando (2.5). El estimador de mínima cuasi ϕ -divergencia (EMC ϕ), $\widehat{\boldsymbol{\theta}}_\phi$, para el modelo loglineal,*

$$p_r(\boldsymbol{\theta}) = \frac{\exp\{\mathbf{w}_r^T \boldsymbol{\theta}\}}{\sum_{u=1}^M \exp\{\mathbf{w}_u^T \boldsymbol{\theta}\}} \quad r = 1, \dots, M.$$

verifica

$$\widehat{\boldsymbol{\theta}}_\phi = \boldsymbol{\theta}_0 + \mathbf{I}_F^{-1}(\boldsymbol{\theta}_0) \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} (\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) + o(\|\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)\|), \quad (2.12)$$

siendo $\mathbf{I}_F(\boldsymbol{\theta}_0)$ la matriz de información de Fisher asociada al modelo loglineal, que se supone no singular, y que viene dada por

$$\mathbf{I}_F(\boldsymbol{\theta}_0) = \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W},$$

con $\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} = \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)} - \mathbf{p}(\boldsymbol{\theta}_0) \mathbf{p}(\boldsymbol{\theta}_0)^T$.

Demostración Sea l^M el interior del cubo unitario M -dimensional y

$$\Delta_M = \{\mathbf{p} = (p_1, \dots, p_M) : p_r > 0, r = 1, \dots, M, \sum_{r=1}^M p_r = 1\}.$$

Es claro que $\Delta_M \subset l^M$. Sea U un entorno de $\boldsymbol{\theta}_0$ en el cual $p_r(\boldsymbol{\theta}) = \exp\{\mathbf{w}_r^T \boldsymbol{\theta}\} / \sum_{u=1}^M \exp\{\mathbf{w}_u^T \boldsymbol{\theta}\}$, con $r = 1, \dots, M$, tiene derivadas parciales continuas. Considérese la función

$$\mathbf{F}^T = (F_1, \dots, F_{M_0}) : l^M \times U \longrightarrow \mathbb{R}^{M_0},$$

de forma que

$$F_g(p_1, \dots, p_M; \theta_1, \dots, \theta_{M_0}) = \frac{\partial d_\phi(\mathbf{p}, \mathbf{p}(\boldsymbol{\theta}))}{\partial \theta_g}, \quad g = 1, \dots, M_0,$$

siendo

$$d_\phi(\mathbf{p}, \mathbf{p}(\boldsymbol{\theta})) = \sum_{r=1}^M p_r(\boldsymbol{\theta}) \phi\left(\frac{p_r}{p_r(\boldsymbol{\theta})}\right).$$

Para $\pi_r = p_r(\boldsymbol{\theta}_0)$, $r = 1, \dots, M$, se tiene que

$$F_g(\pi_1, \dots, \pi_M; \theta_{01}, \dots, \theta_{0M_0}) = 0, \quad g = 1, \dots, M_0,$$

ya que

$$\begin{aligned} \frac{\partial d_\phi(\mathbf{p}, \mathbf{p}(\boldsymbol{\theta}))}{\partial \theta_g} &= \frac{\partial}{\partial \theta_g} \sum_{r=1}^M p_r(\boldsymbol{\theta}) \phi\left(\frac{p_r}{p_r(\boldsymbol{\theta})}\right) \\ &= \sum_{r=1}^M \left[\frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_g} \phi\left(\frac{p_r}{p_r(\boldsymbol{\theta})}\right) - p_r(\boldsymbol{\theta}) \phi'\left(\frac{p_r}{p_r(\boldsymbol{\theta})}\right) \frac{p_r}{p_r^2(\boldsymbol{\theta})} \frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_g} \right] \\ &= \sum_{r=1}^M \frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_g} \left[\phi\left(\frac{p_r}{p_r(\boldsymbol{\theta})}\right) - \phi'\left(\frac{p_r}{p_r(\boldsymbol{\theta})}\right) \frac{p_r}{p_r(\boldsymbol{\theta})} \right] \quad g = 1, \dots, M_0. \end{aligned} \quad (2.13)$$

Por otra parte, para el modelo loglineal se tiene

$$\frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_g} = p_r(\boldsymbol{\theta}) w_{rg} - p_r(\boldsymbol{\theta}) \sum_{u=1}^M w_{ug} p_u(\boldsymbol{\theta}), \quad r = 1, \dots, M,$$

cuya sustitución en (2.13) da lugar a

$$\frac{\partial d_\phi(\mathbf{p}, \mathbf{p}(\boldsymbol{\theta}))}{\partial \theta_g} = \sum_{r=1}^M \left(p_r(\boldsymbol{\theta}) w_{rg} - p_r(\boldsymbol{\theta}) \sum_{u=1}^M w_{ug} p_u(\boldsymbol{\theta}) \right) \left[\phi\left(\frac{p_r}{p_r(\boldsymbol{\theta})}\right) - \phi'\left(\frac{p_r}{p_r(\boldsymbol{\theta})}\right) \frac{p_r}{p_r(\boldsymbol{\theta})} \right]$$

y entonces

$$F_g(\pi_1, \dots, \pi_M; \theta_{01}, \dots, \theta_{0M_0}) = \sum_{r=1}^M \left(p_r(\boldsymbol{\theta}_0) w_{rg} - p_r(\boldsymbol{\theta}_0) \sum_{u=1}^M w_{ug} p_u(\boldsymbol{\theta}_0) \right) [\phi(1) - \phi'(1)] = 0$$

$g = 1, \dots, M_0.$

Seguidamente se comprueba que la matriz

$$\frac{\partial \mathbf{F}}{\partial \boldsymbol{\theta}^T} = \left(\frac{\partial F_g}{\partial \theta_j} \right)_{\substack{g=1, \dots, M_0 \\ j=1, \dots, M_0}}$$

es no singular en $(\pi_1, \dots, \pi_M; \theta_{01}, \dots, \theta_{0M_0})$. En efecto, al ser

$$\begin{aligned} \frac{\partial}{\partial \theta_j} \left(\frac{\partial d_\phi(\mathbf{p}, \mathbf{p}(\boldsymbol{\theta}))}{\partial \theta_g} \right) &= - \sum_{r=1}^M \frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_j} \frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_g} \phi' \left(\frac{p_r}{p_r(\boldsymbol{\theta})} \right) \frac{p_r}{p_r^2(\boldsymbol{\theta})} \\ &+ \sum_{r=1}^M \frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_j} \frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_g} \phi'' \left(\frac{p_r}{p_r(\boldsymbol{\theta})} \right) \frac{p_r}{p_r^2(\boldsymbol{\theta})} \frac{p_r}{p_r(\boldsymbol{\theta})} \\ &+ \sum_{r=1}^M \frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_j} \frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_g} \phi' \left(\frac{p_r}{p_r(\boldsymbol{\theta})} \right) \frac{p_r}{p_r^2(\boldsymbol{\theta})} \\ &+ \sum_{r=1}^M \frac{\partial^2 p_r(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_g} \left[\phi \left(\frac{p_r}{p_r(\boldsymbol{\theta})} \right) - \phi' \left(\frac{p_r}{p_r(\boldsymbol{\theta})} \right) \frac{p_r}{p_r(\boldsymbol{\theta})} \right], \end{aligned}$$

se tiene que

$$\begin{aligned} \frac{\partial}{\partial \theta_j} \left(\frac{\partial d_\phi(\mathbf{p}, \mathbf{p}(\boldsymbol{\theta}))}{\partial \theta_g} \right) &= \sum_{r=1}^M \frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_j} \frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_g} \phi'' \left(\frac{p_r}{p_r(\boldsymbol{\theta})} \right) \frac{p_r}{p_r^2(\boldsymbol{\theta})} \frac{p_r}{p_r(\boldsymbol{\theta})} \\ &+ \sum_{r=1}^M \frac{\partial^2 p_r(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_g} \left[\phi \left(\frac{p_r}{p_r(\boldsymbol{\theta})} \right) - \phi' \left(\frac{p_r}{p_r(\boldsymbol{\theta})} \right) \frac{p_r}{p_r(\boldsymbol{\theta})} \right]. \end{aligned}$$

Luego, para $(\pi_1, \dots, \pi_M; \theta_{01}, \dots, \theta_{0M_0})$, con $\boldsymbol{\pi} = \mathbf{p}(\boldsymbol{\theta}_0)$, se obtiene

$$\begin{aligned} \frac{\partial \mathbf{F}}{\partial \boldsymbol{\theta}^T} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} &= \left(\frac{\partial F_g}{\partial \theta_j} \right) \Big|_{(\pi_1, \dots, \pi_M; \theta_{01}, \dots, \theta_{0M_0})} \\ &= \phi''(1) \left(\sum_{r=1}^M \frac{\partial p_r(\boldsymbol{\theta}_0)}{\partial \theta_j} \frac{\partial p_r(\boldsymbol{\theta}_0)}{\partial \theta_g} \frac{1}{p_r(\boldsymbol{\theta}_0)} \right)_{\substack{g=1, \dots, M_0 \\ j=1, \dots, M_0}}. \end{aligned}$$

Si se denota

$$\mathbf{A}(\boldsymbol{\theta}) = \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-1/2} \left(\frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_g} \right)_{\substack{r=1, \dots, M \\ g=1, \dots, M_0}} = \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-1/2} \mathbf{J}(\boldsymbol{\theta})$$

$$\mathbf{A}(\boldsymbol{\theta})^T \mathbf{A}(\boldsymbol{\theta}) = \left(\sum_{r=1}^M \frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_j} \frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_g} \frac{1}{p_r(\boldsymbol{\theta})} \right)_{\substack{g=1, \dots, M_0 \\ j=1, \dots, M_0}}$$

se puede expresar

$$\left. \frac{\partial \mathbf{F}}{\partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \phi''(1) \mathbf{A}(\boldsymbol{\theta}_0)^T \mathbf{A}(\boldsymbol{\theta}_0). \quad (2.14)$$

Se sabe que si \mathbf{B} es una matriz $p \times q$ con rango p y \mathbf{C} es $q \times s$ con rango q , entonces $\text{rg}(\mathbf{BC}) = p$. Luego, si $\mathbf{B} = \mathbf{J}(\boldsymbol{\theta}_0)^T$ y $\mathbf{C} = \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-1/2}$, son respectivamente matrices $M_0 \times M$ y $M \times M$, se sigue que $\mathbf{A}(\boldsymbol{\theta}_0)^T = \mathbf{BC}$ tiene rango M_0 por la condición 4 de Birch. Además,

$$\text{rg}(\mathbf{A}(\boldsymbol{\theta}_0)^T \mathbf{A}(\boldsymbol{\theta}_0)) = \text{rg}(\mathbf{A}(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0)^T) = \text{rg}(\mathbf{A}(\boldsymbol{\theta}_0)) = M_0.$$

Con lo que queda probado que $\left. \frac{\partial \mathbf{F}}{\partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ es no singular. Por otra parte, recuérdese que para el modelo loglineal

$$\begin{aligned} \mathbf{J}(\boldsymbol{\theta}) &= \left(\frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_g} \right)_{\substack{r=1, \dots, M \\ g=1, \dots, M_0}} = \left(p_r(\boldsymbol{\theta}) w_{rg} - p_r(\boldsymbol{\theta}) \sum_{u=1}^M p_u(\boldsymbol{\theta}) w_{ug} \right)_{\substack{r=1, \dots, M \\ g=1, \dots, M_0}} \\ &= (\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})} - \mathbf{p}(\boldsymbol{\theta}) \mathbf{p}(\boldsymbol{\theta})^T) \mathbf{W} = \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})} \mathbf{W}. \end{aligned}$$

Se comprueba entonces que

$$\begin{aligned} \mathbf{A}(\boldsymbol{\theta})^T \mathbf{A}(\boldsymbol{\theta}) &= \mathbf{W}^T \boldsymbol{\Sigma} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-1} \boldsymbol{\Sigma} \mathbf{W} = \mathbf{W}^T \boldsymbol{\Sigma} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-1} (\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})} - \mathbf{p}(\boldsymbol{\theta}) \mathbf{p}(\boldsymbol{\theta})^T) \mathbf{W} \\ &= \mathbf{W}^T \boldsymbol{\Sigma} (\mathbf{I}_{p(\boldsymbol{\theta})} - \mathbf{1}_M \mathbf{p}(\boldsymbol{\theta})^T) \mathbf{W} = \mathbf{W}^T (\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})} - \mathbf{p}(\boldsymbol{\theta}) \mathbf{p}(\boldsymbol{\theta})^T) (\mathbf{I}_{p(\boldsymbol{\theta})} - \mathbf{1}_M \mathbf{p}(\boldsymbol{\theta})^T) \mathbf{W} \\ &= \mathbf{W}^T (\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})} - \mathbf{p}(\boldsymbol{\theta}) \mathbf{p}(\boldsymbol{\theta})^T - \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})} \mathbf{1}_M \mathbf{p}(\boldsymbol{\theta})^T + \mathbf{p}(\boldsymbol{\theta}) \mathbf{p}(\boldsymbol{\theta})^T \mathbf{1}_M \mathbf{p}(\boldsymbol{\theta})^T) \mathbf{W} \\ &= \mathbf{W}^T (\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})} - \mathbf{p}(\boldsymbol{\theta}) \mathbf{p}(\boldsymbol{\theta})^T - \mathbf{p}(\boldsymbol{\theta}) \mathbf{p}(\boldsymbol{\theta})^T + \mathbf{p}(\boldsymbol{\theta}) \mathbf{p}(\boldsymbol{\theta})^T) \mathbf{W} = \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})} \mathbf{W}, \end{aligned} \quad (2.15)$$

y de esta forma se puede expresar (2.14) como

$$\left. \frac{\partial \mathbf{F}}{\partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \phi''(1) \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W} = \phi''(1) \mathbf{I}_F(\boldsymbol{\theta}_0), \quad (2.16)$$

donde $\Sigma_{p(\theta)}$ es la matriz de varianzas covarianzas de la distribución multinomial e $\mathbf{I}_F(\boldsymbol{\theta}_0)$ es la matriz de Información de Fisher de una observación para el modelo loglineal. Luego, se puede aplicar el Teorema de la Función Implícita. Entonces, existe un entorno M -dimensional U_0 de $\boldsymbol{\pi} = (\pi_1, \dots, \pi_M)^T$ en l^M y una única función diferenciable y continua $\tilde{\boldsymbol{\theta}} : \mathbb{R}^M \longrightarrow \mathbb{R}^{M_0}$, tal que $\tilde{\boldsymbol{\theta}}(\boldsymbol{\pi}) = \boldsymbol{\theta}_0$ y $\mathbf{F}(\mathbf{p}, \tilde{\boldsymbol{\theta}}(\mathbf{p})) = \mathbf{0}_{M_0} \quad \forall \mathbf{p} \in U_0$. Por la regla de la cadena

$$\frac{\partial \mathbf{F}(\mathbf{p}, \tilde{\boldsymbol{\theta}}(\mathbf{p}))}{\partial \mathbf{p}^T} + \frac{\partial \mathbf{F}(\mathbf{p}, \tilde{\boldsymbol{\theta}}(\mathbf{p}))}{\partial \tilde{\boldsymbol{\theta}}^T(\mathbf{p})} \frac{\partial \tilde{\boldsymbol{\theta}}(\mathbf{p})}{\partial \mathbf{p}^T} = \mathbf{0}_{M_0},$$

$\forall \mathbf{p} \in U_0$ y para $\mathbf{p} = \boldsymbol{\pi}$

$$\left. \frac{\partial \mathbf{F}}{\partial \mathbf{p}^T} \right|_{\mathbf{p}=\boldsymbol{\pi}} + \left. \frac{\partial \mathbf{F}}{\partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \left. \frac{\partial \boldsymbol{\theta}}{\partial \mathbf{p}^T} \right|_{\mathbf{p}=\boldsymbol{\pi}, \boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}_{M_0}.$$

Ahora bien, por una parte, se sabe que

$$\left. \frac{\partial \mathbf{F}}{\partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \phi''(1) \mathbf{A}(\boldsymbol{\theta}_0)^T \mathbf{A}(\boldsymbol{\theta}_0),$$

y por otra parte, posteriormente se demuestra que

$$\left. \frac{\partial \mathbf{F}}{\partial \mathbf{p}^T} \right|_{\mathbf{p}=\boldsymbol{\pi}} = -\phi''(1) \mathbf{A}(\boldsymbol{\theta}_0)^T \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2},$$

con lo cual

$$\left. \frac{\partial \tilde{\boldsymbol{\theta}}(\mathbf{p})}{\partial \mathbf{p}^T} \right|_{\mathbf{p}=\boldsymbol{\pi}, \boldsymbol{\theta}=\boldsymbol{\theta}_0} = - \left. \frac{\partial \mathbf{F}}{\partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}^{-1} \left. \frac{\partial \mathbf{F}}{\partial \mathbf{p}^T} \right|_{\mathbf{p}=\boldsymbol{\pi}} = (\mathbf{A}(\boldsymbol{\theta}_0)^T \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}(\boldsymbol{\theta}_0)^T \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2}.$$

Desarrollando por Taylor $\tilde{\boldsymbol{\theta}}(\mathbf{p})$ en un entorno de $\boldsymbol{\pi}$, se obtiene

$$\tilde{\boldsymbol{\theta}}(\mathbf{p}) = \tilde{\boldsymbol{\theta}}(\boldsymbol{\pi}) + \left. \frac{\partial \tilde{\boldsymbol{\theta}}(\mathbf{p})}{\partial \mathbf{p}^T} \right|_{\mathbf{p}=\boldsymbol{\pi}} (\mathbf{p} - \boldsymbol{\pi}) + o(\|\mathbf{p} - \boldsymbol{\pi}\|),$$

donde $o(\|\mathbf{p} - \boldsymbol{\pi}\|)$ denota una función que verifica

$$\lim_{\mathbf{p} \rightarrow \boldsymbol{\pi}} \frac{o(\|\mathbf{p} - \boldsymbol{\pi}\|)}{\|\mathbf{p} - \boldsymbol{\pi}\|} = 0,$$

y como $\tilde{\boldsymbol{\theta}}(\boldsymbol{\pi}) = \boldsymbol{\theta}_0$ se llega a

$$\tilde{\boldsymbol{\theta}}(\mathbf{p}) = \boldsymbol{\theta}_0 + \left(\mathbf{A}(\boldsymbol{\theta}_0)^T \mathbf{A}(\boldsymbol{\theta}_0) \right)^{-1} \mathbf{A}(\boldsymbol{\theta}_0)^T \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} (\mathbf{p} - \boldsymbol{\pi}) + o(\|\mathbf{p} - \boldsymbol{\pi}\|).$$

Ahora bien, $\hat{\mathbf{p}} \xrightarrow[n \rightarrow \infty]{c.s.} \boldsymbol{\pi}$, por lo tanto $\hat{\mathbf{p}} \in U_0$ y como consecuencia $\tilde{\boldsymbol{\theta}}(\mathbf{p})$ es solución única de las ecuaciones

$$\frac{\partial d_\phi(\hat{\mathbf{p}}, \tilde{\boldsymbol{\theta}}(\mathbf{p}))}{\partial \theta_g} = 0 \quad g = 1, \dots, M_0,$$

en un entorno de $\boldsymbol{\pi}$. Luego, $\tilde{\boldsymbol{\theta}}(\mathbf{p})$ es el estimador de mínima ϕ -divergencia, $\hat{\boldsymbol{\theta}}_\phi$, que, como consecuencia de lo anterior, verifica

$$\hat{\boldsymbol{\theta}}_\phi = \boldsymbol{\theta}_0 + \left(\mathbf{A}(\boldsymbol{\theta}_0)^T \mathbf{A}(\boldsymbol{\theta}_0) \right)^{-1} \mathbf{A}(\boldsymbol{\theta}_0)^T \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} (\hat{\mathbf{p}} - \boldsymbol{\pi}) + o(\|\hat{\mathbf{p}} - \boldsymbol{\pi}\|). \quad (2.17)$$

Expresándolo en función de la Información de Fisher y la matriz de diseño del modelo loglineal se tendría

$$\hat{\boldsymbol{\theta}}_\phi = \boldsymbol{\theta}_0 + \mathbf{I}_F^{-1}(\boldsymbol{\theta}_0) \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} (\hat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) + o(\|\hat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)\|).$$

Para finalizar la demostración del teorema queda por establecer que efectivamente

$$\left. \frac{\partial \mathbf{F}}{\partial \mathbf{p}^T} \right|_{\mathbf{p}=\boldsymbol{\pi}} = -\phi''(1) \mathbf{A}(\boldsymbol{\theta}_0)^T \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2}.$$

En efecto,

$$\begin{aligned} \frac{\partial}{\partial p_r} \left(\frac{\partial d_\phi(\mathbf{p}, \mathbf{p}(\boldsymbol{\theta}))}{\partial \theta_g} \right) &= \frac{\partial}{\partial p_r} \left(\sum_{u=1}^M \frac{\partial p_u(\boldsymbol{\theta})}{\partial \theta_g} \left[\phi \left(\frac{p_u}{p_u(\boldsymbol{\theta})} \right) - \phi' \left(\frac{p_u}{p_u(\boldsymbol{\theta})} \right) \frac{p_u}{p_u(\boldsymbol{\theta})} \right] \right) \\ &= - \left(\phi'' \left(\frac{p_r}{p_r(\boldsymbol{\theta})} \right) \frac{1}{p_r(\boldsymbol{\theta})} \frac{p_r}{p_r(\boldsymbol{\theta})} \right) \frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_g}. \end{aligned}$$

Para $(\pi_1, \dots, \pi_M; \theta_{01}, \dots, \theta_{0M_0})$ se tiene

$$\frac{\partial}{\partial p_r} \left(\frac{\partial d_\phi(\boldsymbol{\pi}, \mathbf{p}(\boldsymbol{\theta}_0))}{\partial \theta_g} \right) = -\phi''(1) \frac{1}{p_r(\boldsymbol{\theta}_0)} \left(p_r(\boldsymbol{\theta}_0) w_{rg} - p_r(\boldsymbol{\theta}_0) \sum_{u=1}^M w_{ug} p_u(\boldsymbol{\theta}_0) \right).$$

■

2.2. Propiedades asintóticas del estimador de mínima cuasi ϕ -divergencia en modelos loglineales con muestreo por conglomerados

El siguiente resultado establece la distribución asintótica del estimador de mínima cuasi ϕ -divergencia en un modelo loglineal con muestreo complejo.

Teorema 2.3 *Sea $\widehat{\boldsymbol{\theta}}_\phi$ el estimador de MC ϕ para un parámetro desconocido $\boldsymbol{\theta}$ en un modelo loglineal para un modelo con datos multinomiales agrupados en clusters y verificando 2.5, entonces se cumple*

i)

$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_\phi - \boldsymbol{\theta}_0) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_{M_0}, \frac{\vartheta_n}{n} (\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W})^{-1}), \quad (2.18)$$

ii)

$$\sqrt{N}(\mathbf{p}(\widehat{\boldsymbol{\theta}}_\phi) - \mathbf{p}(\boldsymbol{\theta}_0)) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_M, \frac{\vartheta_n}{n} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W} (\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W})^{-1} \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}), \quad (2.19)$$

donde ϑ_n esta definido en (1.10).

Demostración

i) Como se vio en los resultados asintóticos del Capítulo 1,

$$\sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_M, \frac{\vartheta_n}{n} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}).$$

Por otra parte, por (2.25) del teorema anterior, $\widehat{\boldsymbol{\theta}}_\phi$ tiene la siguiente expresión

$$\widehat{\boldsymbol{\theta}}_\phi = \boldsymbol{\theta}_0 + \mathbf{I}_F^{-1}(\boldsymbol{\theta}_0) \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} (\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) + o(\|\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)\|),$$

si se multiplica por \sqrt{N} se tiene

$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_\phi - \boldsymbol{\theta}_0) = \mathbf{I}_F^{-1}(\boldsymbol{\theta}_0) \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) + o(\mathbf{1}_{M_0}), \quad (2.20)$$

luego se verifica

$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_\phi - \boldsymbol{\theta}_0) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_{M_0}, \frac{\vartheta_n}{n} \boldsymbol{\Sigma}^*), \quad (2.21)$$

con

$$\boldsymbol{\Sigma}^* = \mathbf{I}_F^{-1}(\boldsymbol{\theta}_0) \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \frac{\vartheta_n}{n} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W} (\mathbf{I}_F^{-1}(\boldsymbol{\theta}_0))^T.$$

Para simplificar la expresión anterior se comprueba previamente que

$$\begin{aligned} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-1} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})} &= \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})} (\mathbf{I}_{\mathbf{p}(\boldsymbol{\theta})} - \mathbf{1}_M \mathbf{p}(\boldsymbol{\theta})^T) \\ &= \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})} - \mathbf{p}(\boldsymbol{\theta}) \mathbf{p}(\boldsymbol{\theta})^T - \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})} \mathbf{1}_M \mathbf{p}(\boldsymbol{\theta})^T + \mathbf{p}(\boldsymbol{\theta}) \mathbf{p}(\boldsymbol{\theta})^T \mathbf{1}_M \mathbf{p}(\boldsymbol{\theta})^T \\ &= \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})}, \end{aligned}$$

luego,

$$\begin{aligned}
\Sigma^* &= \frac{\vartheta_n}{n} \mathbf{I}_F^{-1}(\boldsymbol{\theta}_0) \mathbf{W}^T \Sigma_{p(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \Sigma_{p(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \Sigma_{p(\boldsymbol{\theta}_0)} \mathbf{W} (\mathbf{I}_F^{-1}(\boldsymbol{\theta}_0))^T \\
&= \frac{\vartheta_n}{n} \mathbf{I}_F^{-1}(\boldsymbol{\theta}_0) \mathbf{W}^T \Sigma_{p(\boldsymbol{\theta}_0)} \mathbf{W} (\mathbf{I}_F^{-1}(\boldsymbol{\theta}_0))^T \\
&= \frac{\vartheta_n}{n} (\mathbf{W}^T \Sigma_{p(\boldsymbol{\theta}_0)} \mathbf{W})^{-1} \mathbf{W}^T \Sigma_{p(\boldsymbol{\theta}_0)} \mathbf{W} (\mathbf{W}^T \Sigma_{p(\boldsymbol{\theta}_0)} \mathbf{W})^{-1} \\
&= \frac{\vartheta_n}{n} (\mathbf{W}^T \Sigma_{p(\boldsymbol{\theta}_0)} \mathbf{W})^{-1}.
\end{aligned}$$

Finalmente se tiene (2.18).

ii) Puesto que la función $\mathbf{p} : \Theta \longrightarrow \Delta_M$ es totalmente diferenciable en $\boldsymbol{\theta}_0$, es decir existen las derivadas parciales de $p_r(\boldsymbol{\theta})$ con respecto a θ_g en $\boldsymbol{\theta}_0$ y $p_r(\boldsymbol{\theta})$ tiene una aproximación lineal en $\boldsymbol{\theta}_0$ dada por

$$p_r(\boldsymbol{\theta}) = p_r(\boldsymbol{\theta}_0) + \sum_{g=1}^{M_0} (\theta_g - \theta_{0g}) \frac{\partial p_r(\boldsymbol{\theta}_0)}{\partial \theta_g} + o(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|),$$

y puesto que

$$\mathbf{J}(\boldsymbol{\theta}) = \left(\frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_g} \right)_{\substack{r=1, \dots, M \\ g=1, \dots, M_0}} = \Sigma_{p(\boldsymbol{\theta})} \mathbf{W},$$

para $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_\phi$ se obtiene

$$\sqrt{N}(\mathbf{p}(\widehat{\boldsymbol{\theta}}_\phi) - \mathbf{p}(\boldsymbol{\theta}_0)) = \Sigma_{p(\boldsymbol{\theta}_0)} \mathbf{W} \sqrt{N}(\widehat{\boldsymbol{\theta}}_\phi - \boldsymbol{\theta}_0) + o(\mathbf{1}_M), \quad (2.22)$$

y por i)

$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_\phi - \boldsymbol{\theta}_0) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_{M_0}, \frac{\vartheta_n}{n} (\mathbf{W}^T \Sigma_{p(\boldsymbol{\theta}_0)} \mathbf{W})^{-1}), \quad (2.23)$$

se tiene

$$\sqrt{N}(\mathbf{p}(\widehat{\boldsymbol{\theta}}_\phi) - \mathbf{p}(\boldsymbol{\theta}_0)) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_{M_0}, \frac{\vartheta_n}{n} \Sigma_{p(\boldsymbol{\theta}_0)} \mathbf{W} (\mathbf{W}^T \Sigma_{p(\boldsymbol{\theta}_0)} \mathbf{W})^{-1} \mathbf{W}^T \Sigma_{p(\boldsymbol{\theta}_0)}).$$

■

2.3. Estimadores consistentes para ϑ_n y ρ^2

En el muestreo multinomial por conglomerados, una finalidad importante será estimar el efecto del diseño, ϑ_n , o el coeficiente de correlación intracluster ρ^2 . Recuerdese que, $\vartheta_n = 1 + (n - 1)\rho^2$, y en consecuencia,

$$\frac{\vartheta_n}{n} = \frac{1}{n} + \frac{(n-1)}{n}\rho^2$$

es una función creciente de ρ^2 , de forma que

$$\rho^2 = \frac{k-1}{n-1} \Rightarrow \frac{\vartheta_n}{n} = \frac{k}{n}, \quad k \in \{1, \dots, n\},$$

y

$$\frac{k-2}{n-1} < \rho^2 < \frac{k-1}{n-1} \Rightarrow \frac{k-1}{n} < \frac{\vartheta_n}{n} < \frac{k}{n}, \quad k \in \{2, \dots, n\}.$$

Por otro lado, si el tamaño del cluster fuera grande

$$\frac{\vartheta_n}{n} \simeq \rho^2. \tag{2.24}$$

Se consideran N tablas de contingencia en un único vector NM -dimensional,

$$\tilde{\mathbf{Y}} = ((\mathbf{Y}^{(1)})^T, \dots, (\mathbf{Y}^{(N)})^T)^T,$$

donde

$$\mathbf{Y}^{(\ell)} = (Y_1^{(\ell)}, \dots, Y_M^{(\ell)})^T,$$

y se define su correspondiente vector de probabilidades, $\tilde{\mathbf{p}}$, como

$$\tilde{\mathbf{p}} = \frac{1}{nN} \tilde{\mathbf{Y}}.$$

Por otra parte, el estimador no paramétrico del vector de probabilidades suponiendo homogeneidad entre clusters viene dado por

$$\tilde{\mathbf{p}}^* = \left(\frac{1}{N} \hat{\mathbf{p}}^T, \dots, \frac{1}{N} \hat{\mathbf{p}}^T \right)^T, \quad (2.25)$$

donde $\hat{\mathbf{p}}$ se definió en (1.13). Brier (1980) propuso un estimador de ϑ_n basado en comparar la discrepancia entre $\tilde{\mathbf{p}}$ y $\tilde{\mathbf{p}}^*$ en el siguiente sentido

$$X^2(\tilde{\mathbf{Y}}) = \sum_{\ell=1}^N \left(\mathbf{Y}^{(\ell)} - n\hat{\mathbf{p}} \right)^T \frac{1}{n} \mathbf{D}_{\hat{\mathbf{p}}}^{-1} \left(\mathbf{Y}^{(\ell)} - n\hat{\mathbf{p}} \right) = n \sum_{\ell=1}^N \sum_{r=1}^M \frac{(\hat{p}_r^{(\ell)} - \hat{p}_r)^2}{\hat{p}_r}. \quad (2.26)$$

La forma de este estimador recuerda al estadístico ji-cuadrado para el contraste de homogeneidad entre clusters.

El siguiente teorema establece estimadores consistentes de ϑ_n y ρ^2 utilizando la misma expresión propuesta por Brier (1980), por esta razón se hará referencia a ellos como estimadores de Brier. Sin embargo, estos estimadores son válidos, para todas las distribuciones reseñadas en (2.5), como por ejemplo la distribución Dirichlet-Multinomial dada en (1.7), la multinomial n-inflada (1.6), y la multinomial agrupada aleatoriamente dada en (1.8) y la binomial 0-inflada.

Teorema 2.4 *Para $X^2(\tilde{\mathbf{Y}})$ dado en (2.26), cuando N tiende a infinito, se tiene*

i)

$$\frac{X^2(\tilde{\mathbf{Y}})}{(N-1)(M-1)} \xrightarrow[N \rightarrow \infty]{\mathcal{P}} \vartheta_n, \quad (2.27)$$

ii)

$$\frac{\frac{X^2(\tilde{\mathbf{Y}})}{(N-1)(M-1)} - 1}{n-1} \xrightarrow[N \rightarrow \infty]{\mathcal{P}} \rho^2. \quad (2.28)$$

Demostración Sea

$$\mathbf{S}_Y = \frac{1}{N-1} \left(\mathbf{Y}^{(\ell)} - n\hat{\mathbf{p}} \right) \left(\mathbf{Y}^{(\ell)} - n\hat{\mathbf{p}} \right)^T,$$

la matriz de cuasi varianzas covarianzas de la muestra aleatoria simple $\mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(N)}$ y sea

$$\bar{\mathbf{S}}_Y = \text{diag} \{ \mathbf{S}_Y \} = \begin{pmatrix} S_{Y_1}^2 & & \\ & \ddots & \\ & & S_{Y_M}^2 \end{pmatrix},$$

con

$$S_{Y_r}^2 = \sum_{\ell=1}^N (Y^{(\ell,r)} - n\hat{p}_r)^2 / (N-1).$$

Cada elemento de la diagonal de $\bar{\mathbf{S}}_Y$ es un estimador consistente de cada elemento de la diagonal de $\vartheta_n n \Sigma_{p(\theta)}$, ya que

$$E [\bar{\mathbf{S}}_Y] = \text{diag} \{ E [\mathbf{S}_Y] \} = \text{diag} \left\{ \text{Var} \left[\mathbf{Y}^{(\ell)} \right] \right\} = \text{diag} \{ \vartheta_n n \Sigma_{p(\theta)} \}$$

y

$$S_{Y_r}^2 \xrightarrow[N \rightarrow \infty]{\mathcal{P}} \vartheta_n n p_r(\boldsymbol{\theta})(1 - p_r(\boldsymbol{\theta})), \quad r = 1, \dots, M, \quad (2.29)$$

o en forma matricial

$$\bar{\mathbf{S}}_Y \xrightarrow[N \rightarrow \infty]{\mathcal{P}} \text{diag} \{ \vartheta_n n \Sigma_{p(\theta)} \}.$$

Se tiene que

$$\text{tr}(\bar{\mathbf{S}}_Y) = \sum_{r=1}^M S_{Y_r}^2 = \text{tr}(\mathbf{S}_Y) = \frac{1}{N-1} \sum_{\ell=1}^N \left(\mathbf{Y}^{(\ell)} - n\hat{\mathbf{p}} \right)^T \left(\mathbf{Y}^{(\ell)} - n\hat{\mathbf{p}} \right) \quad (2.30)$$

es un estimador consistente para $\text{tr} \{ \vartheta_n n \Sigma_{p(\theta)} \} = \vartheta_n n \sum_{r=1}^M p_r(\boldsymbol{\theta})(1 - p_r(\boldsymbol{\theta}))$. Basado en la expresión de, $X^2(\tilde{\mathbf{Y}})$, y siguiendo los mismos pasos que para establecer (2.30), se tiene que

$$\text{tr}(\bar{\mathbf{S}}_Y \frac{1}{n} \mathbf{D}_{\mathbf{p}(\theta)}^{-1}) = \text{tr}(\mathbf{D}_{\mathbf{p}(\theta)}^{-1/2} \bar{\mathbf{S}}_Y \frac{1}{n} \mathbf{D}_{\mathbf{p}(\theta)}^{-1/2}) = \frac{1}{N-1} \sum_{\ell=1}^N \left(\mathbf{Y}^{(\ell)} - n\hat{\mathbf{p}} \right)^T \frac{1}{n} \mathbf{D}_{\mathbf{p}(\theta)}^{-1} \left(\mathbf{Y}^{(\ell)} - n\hat{\mathbf{p}} \right)$$

y

$$\begin{aligned}
E \left[\text{tr}(\overline{\mathbf{S}}_{\mathbf{Y}} \frac{1}{n} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-1}) \right] &= \text{tr} E \left[\overline{\mathbf{S}}_{\mathbf{Y}} \frac{1}{n} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-1} \right] = \text{tr} \left(E \left[\overline{\mathbf{S}}_{\mathbf{Y}} \right] \frac{1}{n} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-1} \right) = \text{tr} \left(\vartheta_n n \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})} \frac{1}{n} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-1} \right) \\
&= \vartheta_n \text{tr} \left(\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-1} \right) = \vartheta_n \text{tr} \left((\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})} - \mathbf{p}(\boldsymbol{\theta}) \mathbf{p}(\boldsymbol{\theta})^T) \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-1} \right) \\
&= \vartheta_n \left[\text{tr}(\mathbf{I}_M) - \text{tr}(\mathbf{p}(\boldsymbol{\theta}) \mathbf{1}^T) \right] = \vartheta_n (M - 1).
\end{aligned}$$

Así,

$$\begin{aligned}
E \left[\frac{1}{M-1} \text{tr}(\overline{\mathbf{S}}_{\mathbf{Y}} \frac{1}{n} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-1}) \right] &= E \left[\frac{1}{(M-1)(N-1)} \sum_{\ell=1}^N \left(\mathbf{Y}^{(\ell)} - n \widehat{\mathbf{p}} \right)^T \frac{1}{n} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-1} \left(\mathbf{Y}^{(\ell)} - n \widehat{\mathbf{p}} \right) \right] \\
&= \vartheta_n,
\end{aligned}$$

y teniendo en cuenta que $\widehat{\mathbf{p}}$ es un estimador consistente de $\mathbf{p}(\boldsymbol{\theta})$ cuando $N \rightarrow \infty$, así como (2.29), se tiene

$$\begin{aligned}
\frac{1}{M-1} \text{tr}(\overline{\mathbf{S}}_{\mathbf{Y}} \frac{1}{n} \mathbf{D}_{\widehat{\mathbf{p}}}^{-1}) &= \frac{1}{(M-1)(N-1)} \sum_{\ell=1}^N \left(\mathbf{Y}^{(\ell)} - n \widehat{\mathbf{p}} \right)^T \frac{1}{n} \mathbf{D}_{\widehat{\mathbf{p}}}^{-1} \left(\mathbf{Y}^{(\ell)} - n \widehat{\mathbf{p}} \right) \\
&= \frac{X^2(\widetilde{\mathbf{Y}})}{(M-1)(N-1)}
\end{aligned}$$

tiende en probabilidad a ϑ_n . En otras palabras,

$$\frac{X^2(\widetilde{\mathbf{Y}})}{(M-1)(N-1)} = \frac{1}{(M-1)n} \sum_{r=1}^M \frac{1}{\widehat{p}_r} S_{Y_r}^2 \xrightarrow[N \rightarrow \infty]{\mathcal{P}} \frac{1}{(M-1)n} \vartheta_n n \sum_{r=1}^M \frac{p_r(\boldsymbol{\theta})}{p_r(\boldsymbol{\theta})} (1 - p_r(\boldsymbol{\theta})) = \vartheta_n.$$

De la relación entre ρ_n^2 y ϑ_n , expresada en $\vartheta_n = 1 + (n-1)\rho^2$, se deduce la segunda parte de este teorema. ■

La demostración de este teorema es completamente diferente a la dada por Brier (1980). Este resultado es la base para la definición de los estimadores consistentes tanto para el efecto del diseño, como para el coeficiente de correlación intra-cluster.

Definición 2.2 El estimador consistente de Brier del efecto del diseño, ϑ_n , es

$$\tilde{\vartheta}_{n,N}(\tilde{\mathbf{Y}}) = \frac{X^2(\tilde{\mathbf{Y}})}{(M-1)(N-1)},$$

donde $X^2(\tilde{\mathbf{Y}})$ se define en (2.26). De forma análoga, el estimador consistente de Brier del coeficiente de correlación intracluster, ρ^2 , es

$$\rho_{n,N}^2(\tilde{\mathbf{Y}}) = \frac{\tilde{\vartheta}_{n,N}(\tilde{\mathbf{Y}}) - 1}{n-1}.$$

El estimador del efecto del diseño, $\tilde{\vartheta}_{n,N}(\tilde{\mathbf{Y}})$, así como el estimador del coeficiente de correlación intracluster $\rho_{n,N}^2(\tilde{\mathbf{Y}})$, son totalmente no paramétricos. Utilizando la demostración del Teorema 2 es posible dar una definición similar a la anterior reemplazando la frecuencia relativa del denominador de $X^2(\tilde{\mathbf{Y}})$ por el estimador de la probabilidad de acuerdo a un modelo loglineal. Estos estimadores representan las versiones semiparamétricas.

Definición 2.3 La extensión semiparamétrica, $(EMC\phi)$, del estimador consistente de Brier de ϑ_n , es

$$\tilde{\vartheta}_{n,N}(\tilde{\mathbf{Y}}, \hat{\boldsymbol{\theta}}_\phi) = \frac{X^2(\tilde{\mathbf{Y}}, \hat{\boldsymbol{\theta}}_\phi)}{(M-1)(N-1)},$$

donde

$$X^2(\tilde{\mathbf{Y}}, \hat{\boldsymbol{\theta}}_\phi) = \sum_{\ell=1}^N \left(\mathbf{Y}^{(\ell)} - n\hat{\mathbf{p}} \right)^T \frac{1}{n} \mathbf{D}_{\mathbf{p}(\hat{\boldsymbol{\theta}}_\phi)}^{-1} \left(\mathbf{Y}^{(\ell)} - n\hat{\mathbf{p}} \right) = n \sum_{\ell=1}^N \sum_{r=1}^M \frac{(\hat{p}_r^{(\ell)} - \hat{p}_r)^2}{p_r(\hat{\boldsymbol{\theta}}_\phi)}. \quad (2.31)$$

De forma similar, la extensión semiparamétrica del estimador consistente de Brier de ρ^2 , es

$$\rho_{n,N}^2(\tilde{\mathbf{Y}}, \hat{\boldsymbol{\theta}}_\phi) = \frac{\tilde{\vartheta}_{n,N}(\tilde{\mathbf{Y}}, \hat{\boldsymbol{\theta}}_\phi) - 1}{n-1}. \quad (2.32)$$

2.4. Parametrización de un modelo loglineal para tablas de contingencia con muestreo por conglomerados

Antes de aplicar los resultados anteriores a los casos prácticos que se presentan, es necesario detallar la parametrización de un modelo loglineal para tablas de contingencia con muestreo por clusters. En concreto, si se considera una variable bidimensional (X_1, X_2) , con categorías I y J respectivamente, la muestra del ℓ -ésimo cluster será

$$\mathbf{X}^{(\ell)} = ((X_{1,1}^{(\ell)}, X_{2,1}^{(\ell)}), \dots, (X_{1,n}^{(\ell)}, X_{2,n}^{(\ell)}))^T, \quad \ell = 1, \dots, N,$$

$$(X_{1,s}^{(\ell)}, X_{2,s}^{(\ell)}) \in \mathcal{X} = \{1, \dots, I\} \times \{1, \dots, J\}, \quad s = 1, \dots, n,$$

y el vector de probabilidad

$$\mathbf{p}(\boldsymbol{\theta}) = (p_{11}(\boldsymbol{\theta}), p_{12}(\boldsymbol{\theta}), \dots, p_{IJ}(\boldsymbol{\theta}))^T,$$

$$p_{ij}(\boldsymbol{\theta}) = \Pr(X_1 = i, X_2 = j), \quad i = 1, \dots, I, \quad j = 1, \dots, J.$$

Nótese que el vector de probabilidades se ha ordenado lexicográficamente para identificarlo con el vector $\mathbf{p}(\boldsymbol{\theta}) = (p_1(\boldsymbol{\theta}), \dots, p_M(\boldsymbol{\theta}))^T$, introducido en el Capítulo 1, tomando $M = I \times J$. Para esta situación, un posible modelo es el de independencia de X_1 y X_2

$$p_{ij}(\boldsymbol{\theta}) = p_{i\bullet}(\boldsymbol{\theta}) p_{\bullet j}(\boldsymbol{\theta}), \quad i = 1, \dots, I, \quad j = 1, \dots, J,$$

donde $p_{i\bullet}(\boldsymbol{\theta}) = \sum_{j=1}^J p_{ij}(\boldsymbol{\theta})$, $p_{\bullet j}(\boldsymbol{\theta}) = \sum_{i=1}^I p_{ij}(\boldsymbol{\theta})$, o equivalentemente

$$\log p_{ij}(\boldsymbol{\theta}) = u + \theta_{1(i)} + \theta_{2(j)}, \quad i = 1, \dots, I, \quad j = 1, \dots, J,$$

con

$$\sum_{i=1}^I \theta_{1(i)} = \sum_{j=1}^J \theta_{2(j)} = 0.$$

Para este problema los contadores definidos en (1.1) para cada cluster $\ell = 1, \dots, N$, pueden denotarse utilizando la notación de doble índice

$$Y_{ij}^{(\ell)} = \sum_{s=1}^n I_{\{(i,j)\}}(X_{1,s}^{(\ell)}, X_{2,s}^{(\ell)}), \quad i = 1, \dots, I, \quad j = 1, \dots, J.$$

En esta situación, se tiene una tabla de contingencia bidimensional con I filas y J columnas para cada cluster, donde

$$\mathbf{Y}^{(\ell)} = (Y_{11}^{(\ell)}, Y_{12}^{(\ell)}, \dots, Y_{IJ}^{(\ell)})^T,$$

son los valores de las celdas que cuentan el número de apariciones de los valores de las variables X_1 y X_2 respectivamente. Si se considera una variable tridimensional (X_1, X_2, X_3) con categorías I , J y K respectivamente, la muestra del cluster ℓ -ésimo es

$$\begin{aligned} \mathbf{X}^{(\ell)} &= ((X_{1,1}^{(\ell)}, X_{2,1}^{(\ell)}, X_{3,1}^{(\ell)}), \dots, (X_{1,n}^{(\ell)}, X_{2,n}^{(\ell)}, X_{3,n}^{(\ell)}))^T, \quad \ell = 1, \dots, N, \\ (X_{1,s}^{(\ell)}, X_{2,s}^{(\ell)}, X_{3,s}^{(\ell)}) &\in \mathcal{X} = \{1, \dots, I\} \times \{1, \dots, J\} \times \{1, \dots, K\}, \quad s = 1, \dots, n, \end{aligned}$$

y el vector de probabilidad de las $M = I \times J \times K$ categorías con índice único $\mathbf{p}(\boldsymbol{\theta}) = (p_1(\boldsymbol{\theta}), \dots, p_M(\boldsymbol{\theta}))^T$ está en correspondencia con el que sigue el orden lexicográfico,

$$\begin{aligned} \mathbf{p}(\boldsymbol{\theta}) &= (p_{111}(\boldsymbol{\theta}), p_{112}(\boldsymbol{\theta}), \dots, p_{IJK}(\boldsymbol{\theta}))^T, \\ p_{ijk}(\boldsymbol{\theta}) &= \Pr(X_1 = i, X_2 = j, X_3 = k), \quad i = 1, \dots, I, \quad j = 1, \dots, J, \quad k = 1, \dots, K. \end{aligned}$$

En esta situación, un posible problema será contrastar si las variables categóricas X_1, X_2 y X_3 son independientes, es decir

$$p_{ijk}(\boldsymbol{\theta}) = p_{i\bullet\bullet}(\boldsymbol{\theta}) p_{\bullet j\bullet}(\boldsymbol{\theta}) p_{\bullet\bullet k}(\boldsymbol{\theta}), \quad \forall i = 1, \dots, I, \quad j = 1, \dots, J, \quad k = 1, \dots, K,$$

donde

$$p_{i\bullet\bullet}(\boldsymbol{\theta}) = \sum_{j=1}^J \sum_{k=1}^K p_{ijk}(\boldsymbol{\theta}), p_{\bullet j\bullet}(\boldsymbol{\theta}) = \sum_{i=1}^I \sum_{k=1}^K p_{ijk}(\boldsymbol{\theta}), p_{\bullet\bullet k}(\boldsymbol{\theta}) = \sum_{i=1}^I \sum_{j=1}^J p_{ijk}(\boldsymbol{\theta}),$$

o equivalentemente

$$\log p_{ijk}(\boldsymbol{\theta}) = u + \theta_{1(i)} + \theta_{2(j)} + \theta_{3(k)}, \quad i = 1, \dots, I, \quad j = 1, \dots, J, \quad k = 1, \dots, K,$$

con

$$\sum_{i=1}^I \theta_{1(i)} = \sum_{j=1}^J \theta_{2(j)} = \sum_{k=1}^K \theta_{3(k)} = 0.$$

En este caso, los valores muestrales de la variable contador para el cluster ℓ -ésimo se pueden denotar por

$$Y_{ijk}^{(\ell)} = \sum_{s=1}^n I_{\{(i,j,k)\}}(X_{1,s}^{(\ell)}, X_{2,s}^{(\ell)}, X_{3,s}^{(\ell)}), \quad i = 1, \dots, I, \quad j = 1, \dots, J, \quad k = 1, \dots, K.$$

Ahora, se tiene una tabla de contingencia tridimensional, $I \times J \times K$, para cada cluster,

$$\mathbf{Y}^{(\ell)} = (Y_{111}^{(\ell)}, Y_{112}^{(\ell)}, \dots, Y_{IJK}^{(\ell)})^T,$$

donde cada celda contiene la frecuencia conjunta de las tres variables $X_{1,s}^{(\ell)}, X_{2,s}^{(\ell)}, X_{3,s}^{(\ell)}$, respectivamente. Un posible modelo es el de independencia de la variable X_1 con respecto a X_2 y X_3 ,

$$p_{ijk}(\boldsymbol{\theta}) = p_{\bullet jk}(\boldsymbol{\theta}) p_{i\bullet\bullet}(\boldsymbol{\theta}) \quad i = 1, \dots, I, \quad j = 1, \dots, J, \quad k = 1, \dots, K,$$

donde $p_{\bullet jk}(\boldsymbol{\theta}) = \sum_{i=1}^I p_{ijk}(\boldsymbol{\theta})$, o equivalentemente

$$\log p_{ijk}(\boldsymbol{\theta}) = u + \theta_{1(i)} + \theta_{2(j)} + \theta_{3(k)} + \theta_{23(jk)},$$

con

$$\begin{aligned}\sum_{i=1}^I \theta_{1(i)} &= \sum_{j=1}^J \theta_{2(j)} = \sum_{k=1}^K \theta_{3(k)} = 0, \\ \sum_{j=1}^J \theta_{23(jk)} &= \sum_{k=1}^K \theta_{23(jk)} = 0.\end{aligned}$$

También podría ser de interés el modelo de independencia de las variables X_1 y X_3 dado X_2 ,

$$p_{ijk}(\boldsymbol{\theta}) = \frac{p_{ij\bullet}(\boldsymbol{\theta})}{p_{\bullet j\bullet}(\boldsymbol{\theta})} p_{\bullet jk}(\boldsymbol{\theta}), \quad i = 1, \dots, I, \quad j = 1, \dots, J, \quad k = 1, \dots, K,$$

o bien

$$\log p_{ijk}(\boldsymbol{\theta}) = u + \theta_{1(i)} + \theta_{2(j)} + \theta_{3(k)} + \theta_{12(ij)} + \theta_{23(jk)},$$

con

$$\begin{aligned}\sum_{i=1}^I \theta_{1(i)} &= \sum_{j=1}^J \theta_{2(j)} = \sum_{k=1}^K \theta_{3(k)} = 0, \\ \sum_{i=1}^I \theta_{12(ij)} &= \sum_{j=1}^J \theta_{12(ij)} = 0, \\ \sum_{j=1}^J \theta_{23(jk)} &= \sum_{k=1}^K \theta_{23(jk)} = 0.\end{aligned}$$

2.5. Aplicación numérica

En esta Sección se analiza un ejemplo que ya ha sido estudiado por Cohen (1976) y Altham (1976) para poner de manifiesto la teoría desarrollada en los apartados anteriores. El estudio trata sobre la relación entre el sexo y el diagnóstico de Esquizofrenia en parejas de hermanos. Se dispone de una muestra formada por $N = 71$ parejas de hermanos en donde se desea determinar si el diagnóstico de Esquizofrenia (X_3) es independiente del sexo (X_2). Para determinar el grado de dependencia existente dentro de los clusters, en este caso las parejas de hermanos, se utilizan los datos resumidos en la Tabla 2.1 en donde se recogen

las características del hermano mayor y el hermano menor. Donde, N =No esquizofrenia, S =Sí Esquizofrenia, H =Hombre y M =Mujer.

Tabla 2.1. Número de parejas según Diagnóstico y Sexo.

Hermano Mayor	Hermano Menor			
	NH	NM	SH	SM
NH	2	4	1	1
NM	3	15	3	8
SH	1	3	13	5
SM	1	1	4	6

Si se calcula el coeficiente de correlación de Pearson se tiene $\hat{\rho} = 0.3557$ con un p -valor= 0.0025 y al calcular el contraste χ^2 de independencia se obtiene el valor del estadístico=25.0840, con un p -valor= 0.0025, lo que lleva a rechazar la independencia entre hermanos cuando se estudia las variables diagnóstico y sexo conjuntamente.

Para ver cómo influye esta dependencia intra-cluster en las estimaciones de la probabilidad se va a ajustar el modelo loglineal con interacciones de orden dos entre las variables, pero sin tener en cuenta esta dependencia entre hermanos (miembros de un mismo cluster). Los estimadores de los parámetros se obtendrán haciendo uso de la familia de ϕ -divergencias de Cressie-Read. La Tabla 2.2 muestra los 142 individuos anteriormente descritos por parejas, clasificados ahora por las tres variables: Edad, con valores Mayor y Menor, Sexo y Diagnóstico anteriormente descritas.

Tabla 2.2. Número de individuos según Edad, Diagnóstico y Sexo.

Edad	Sexo	Diagnóstico	
		N	S
Mayor	H	8	22
	M	29	12
Menor	H	7	21
	M	23	20

Se considera el siguiente modelo loglineal, $p_j(\boldsymbol{\theta}) = \exp\{\mathbf{w}_j^T \boldsymbol{\theta}\} / \sum_{r=1}^M \exp\{\mathbf{w}_r^T \boldsymbol{\theta}\}$ con $j = 1, \dots, M$, que en este caso, se puede expresar lexicográficamente como

$$p_{ijk}(\boldsymbol{\theta}) = \exp\{u + \theta_{1(i)} + \theta_{2(j)} + \theta_{3(k)} + \theta_{12(ij)} + \theta_{13(ik)} + \theta_{23(jk)}\} \quad \text{con } i, j, k = 1, 2, \quad (2.33)$$

donde, el efecto de la variable Edad viene representado por $\theta_{1(i)}$, Sexo por $\theta_{2(j)}$ y el Diagnóstico de la enfermedad por $\theta_{3(k)}$ y sus respectivas interacciones dos a dos $\theta_{12(ij)}$, $\theta_{13(ik)}$ y $\theta_{23(jk)}$. Estos parámetros verifican las siguientes condiciones

$$\sum_{i=1}^2 \theta_{1(i)} = \sum_{j=1}^2 \theta_{2(j)} = \sum_{k=1}^2 \theta_{3(k)} = 0 \quad (2.34)$$

y

$$\sum_{i=1}^2 \theta_{12(ij)} = \sum_{j=1}^2 \theta_{12(ij)} = 0,$$

$$\sum_{i=1}^2 \theta_{13(ik)} = \sum_{k=1}^2 \theta_{13(ik)} = 0,$$

$$\sum_{j=1}^2 \theta_{23(jk)} = \sum_{k=1}^2 \theta_{23(jk)} = 0.$$

Para estimar los parámetros del modelo loglineal se considerará la familia de ϕ -divergencias, definidas en (2.4). Es decir, tomando

$$\begin{aligned} \phi_\lambda(x) &= \frac{1}{\lambda(1+\lambda)} [x^{\lambda+1} - x - \lambda(x-1)], & \lambda \notin \{-1, 0\} \\ \phi'_\lambda(x) &= \frac{1}{\lambda(1+\lambda)} [(1+\lambda)x^\lambda - 1 - \lambda], & \lambda \notin \{-1, 0\}, \\ \phi''_\lambda(x) &= x^{\lambda-1}, & \lambda \notin \{-1, 0\}, \end{aligned}$$

la función $\boldsymbol{\Psi}^{\phi_\lambda}(\boldsymbol{\theta}) = (\Psi_1^{\phi_\lambda}(\boldsymbol{\theta}), \dots, \Psi_M^{\phi_\lambda}(\boldsymbol{\theta}))$, definida en (2.7) es

$$\Psi_r^{\phi_\lambda}(\boldsymbol{\theta}) = \widehat{p}_r \phi'_\lambda\left(\frac{\widehat{p}_r}{p_r(\boldsymbol{\theta})}\right) - p_r(\boldsymbol{\theta}) \phi_\lambda\left(\frac{\widehat{p}_r}{p_r(\boldsymbol{\theta})}\right), \quad r = 1, \dots, M,$$

con $\lambda \in \mathbb{R} - \{-1, 0\}$, que tiene la expresión

$$\begin{aligned}\Psi_r^{\phi_\lambda}(\boldsymbol{\theta}) &= \frac{1}{\lambda(1+\lambda)} \left[\widehat{p}_r \left((1+\lambda) \frac{\widehat{p}_r^\lambda}{p_r(\boldsymbol{\theta})^\lambda} - (1+\lambda) \right) - p_r(\boldsymbol{\theta}) \left(\frac{\widehat{p}_r^{\lambda+1}}{p_r(\boldsymbol{\theta})^{\lambda+1}} - \frac{\widehat{p}_r}{p_r(\boldsymbol{\theta})} - \lambda \frac{\widehat{p}_r}{p_r(\boldsymbol{\theta})} + \lambda \right) \right] \\ &= \frac{1}{\lambda(1+\lambda)} \left[\frac{\widehat{p}_r^{\lambda+1}}{p_r^\lambda(\boldsymbol{\theta})} + \lambda \frac{\widehat{p}_r^{\lambda+1}}{p_r^\lambda(\boldsymbol{\theta})} - (1+\lambda)\widehat{p}_r - \frac{\widehat{p}_r^{\lambda+1}}{p_r^\lambda(\boldsymbol{\theta})} + (1+\lambda)\widehat{p}_r - \lambda p_r(\boldsymbol{\theta}) \right].\end{aligned}$$

Luego

$$\Psi_r^{\phi_\lambda}(\boldsymbol{\theta}) = \frac{1}{1+\lambda} \left(\frac{\widehat{p}_r^{\lambda+1}}{p_r^\lambda(\boldsymbol{\theta})} - p_r(\boldsymbol{\theta}) \right), \quad r = 1, \dots, M,$$

y si se expresa matricialmente

$$\boldsymbol{\Psi}^{\phi_\lambda}(\boldsymbol{\theta}) = \frac{1}{1+\lambda} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-\lambda} (\widehat{\mathbf{p}}^{\lambda+1} - \mathbf{p}^{\lambda+1}(\boldsymbol{\theta})).$$

Puesto que, el estimador de $MC\phi$ de $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}_{\phi_\lambda}$, tal y como se vio en (2.6) se obtiene resolviendo el sistema de ecuaciones

$$\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-1} \boldsymbol{\Psi}^{\phi_\lambda}(\boldsymbol{\theta}) = \mathbf{0}_{M_0},$$

sustituyendo la expresión de $\boldsymbol{\Psi}^{\phi_\lambda}(\boldsymbol{\theta})$ se tiene

$$\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-(\lambda+1)} (\widehat{\mathbf{p}}^{\lambda+1} - \mathbf{p}^{\lambda+1}(\boldsymbol{\theta})) = \mathbf{0}_{M_0},$$

teniendo en cuenta que $\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-1} = \mathbf{I}_M - \mathbf{p}(\boldsymbol{\theta}) \mathbf{1}^T$,

$$\begin{aligned}\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-(\lambda+1)} (\widehat{\mathbf{p}}^{\lambda+1} - \mathbf{p}^{\lambda+1}(\boldsymbol{\theta})) &= (\mathbf{I}_M - \mathbf{p}(\boldsymbol{\theta}) \mathbf{1}_M^T) \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-\lambda} (\widehat{\mathbf{p}}^{\lambda+1} - \mathbf{p}^{\lambda+1}(\boldsymbol{\theta})) \\ &= \begin{bmatrix} \frac{1-p_1(\boldsymbol{\theta})}{p_1^\lambda(\boldsymbol{\theta})} & -\frac{p_1(\boldsymbol{\theta})}{p_2^\lambda(\boldsymbol{\theta})} & \cdots & -\frac{p_1(\boldsymbol{\theta})}{p_M^\lambda(\boldsymbol{\theta})} \\ -\frac{p_2(\boldsymbol{\theta})}{p_1^\lambda(\boldsymbol{\theta})} & \frac{1-p_2(\boldsymbol{\theta})}{p_2^\lambda(\boldsymbol{\theta})} & \cdots & -\frac{p_2(\boldsymbol{\theta})}{p_M^\lambda(\boldsymbol{\theta})} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{p_M(\boldsymbol{\theta})}{p_1^\lambda(\boldsymbol{\theta})} & -\frac{p_M(\boldsymbol{\theta})}{p_2^\lambda(\boldsymbol{\theta})} & \cdots & \frac{1-p_M(\boldsymbol{\theta})}{p_M^\lambda(\boldsymbol{\theta})} \end{bmatrix} \begin{bmatrix} \widehat{p}_1^\lambda - p_1^\lambda(\boldsymbol{\theta}) \\ \widehat{p}_2^\lambda - p_2^\lambda(\boldsymbol{\theta}) \\ \vdots \\ \widehat{p}_M^\lambda - p_M^\lambda(\boldsymbol{\theta}) \end{bmatrix}.\end{aligned}$$

Por otra parte, como se definió en (2.11)

$$\bar{\Psi}^\phi(\boldsymbol{\theta}) = \left(\bar{\Psi}_1^\phi(\boldsymbol{\theta}), \dots, \bar{\Psi}_M^\phi(\boldsymbol{\theta}) \right),$$

siendo

$$\bar{\Psi}_r^\phi(\boldsymbol{\theta}) = \phi''\left(\frac{\hat{p}_r}{p_r(\boldsymbol{\theta})}\right) \frac{\hat{p}_r^2}{p_r(\boldsymbol{\theta})} - \Psi_r^\phi(\boldsymbol{\theta}) \quad r = 1, \dots, M.$$

En el caso de la familia de Cressie-Read se tiene

$$\begin{aligned} \bar{\Psi}_r^{\phi_\lambda}(\boldsymbol{\theta}) &= \phi''\left(\frac{\hat{p}_r}{p_r(\boldsymbol{\theta})}\right) \frac{\hat{p}_r^2}{p_r(\boldsymbol{\theta})} - \Psi_r^\phi(\boldsymbol{\theta}) \\ &= \frac{\hat{p}_r^{\lambda+1}}{p_r^\lambda(\boldsymbol{\theta})} - \frac{1}{1+\lambda} \left(\frac{\hat{p}_r^{\lambda+1}}{p_r^\lambda(\boldsymbol{\theta})} - p_r(\boldsymbol{\theta}) \right) \\ &= \frac{1}{1+\lambda} \left(p_r(\boldsymbol{\theta}) + \lambda \frac{\hat{p}_r^{\lambda+1}}{p_r^\lambda(\boldsymbol{\theta})} \right) \quad r = 1, \dots, M, \end{aligned}$$

cuya expresión matricial es

$$\bar{\Psi}^{\phi_\lambda}(\boldsymbol{\theta}) = \frac{1}{1+\lambda} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})} \left(\mathbf{I}_M + \lambda \mathbf{D}_{\hat{\mathbf{p}}}^{\lambda+1} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-(\lambda+1)} \right) \mathbf{1}_M.$$

Por último, la expresión de $\mathbf{G}_\phi(\boldsymbol{\theta})$ definida en general en (2.10) como

$$\mathbf{G}_\phi(\boldsymbol{\theta}) = \mathbf{W}^T (\mathbf{I}_M - \mathbf{p}(\boldsymbol{\theta}) \mathbf{1}^T) \mathbf{D}_{\bar{\Psi}^\phi(\boldsymbol{\theta})} (\mathbf{I}_M - \mathbf{p}(\boldsymbol{\theta}) \mathbf{1}^T) \mathbf{W},$$

en el caso particular de la familia de Cressie-Read, $\mathbf{G}_{\phi_\lambda}(\boldsymbol{\theta})$, sustituyendo la expresión de $\bar{\Psi}^\phi(\boldsymbol{\theta})$, se tiene la siguiente expresión

$$\begin{aligned} \mathbf{G}_{\phi_\lambda}(\boldsymbol{\theta}) &= \frac{1}{1+\lambda} \mathbf{W}^T (\mathbf{I}_M - \mathbf{p}(\boldsymbol{\theta}) \mathbf{1}^T) \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})} \left(\mathbf{I}_M + \lambda \mathbf{D}_{\hat{\mathbf{p}}}^{\lambda+1} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-(\lambda+1)} \right) (\mathbf{I}_M - \mathbf{1} \mathbf{p}^T(\boldsymbol{\theta})) \mathbf{W} \\ &= \frac{1}{1+\lambda} \left(\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})} \mathbf{W} + \lambda \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})} \mathbf{D}_{\hat{\mathbf{p}}}^{\lambda+1} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta})}^{-(\lambda+2)} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})} \mathbf{W} \right). \end{aligned}$$

Luego, el algoritmo de Newton-Raphson para obtener el de $\boldsymbol{\theta}$, en el paso $(t + 1)$ -ésimo el estimador $\widehat{\boldsymbol{\theta}}_{\phi}^{(t+1)}$, se obtiene a partir de $\widehat{\boldsymbol{\theta}}_{\phi}^{(t)}$ como

$$\widehat{\boldsymbol{\theta}}_{\phi}^{(t+1)} = \widehat{\boldsymbol{\theta}}_{\phi}^{(t)} - \widehat{\boldsymbol{\theta}}_{\phi}^{(t)} \mathbf{G}_{\phi}^{-1}(\widehat{\boldsymbol{\theta}}_{\phi}^{(t)}) \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi}^{(t)})} \mathbf{D}_{\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi}^{(t)})}^{-1} \boldsymbol{\Psi}^{\phi}(\widehat{\boldsymbol{\theta}}_{\phi}^{(t)}),$$

en donde sustituyendo la expresión de $\mathbf{G}_{\phi_{\lambda}}(\boldsymbol{\theta})$, se tiene

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_{\phi_{\lambda}}^{(t+1)} &= \widehat{\boldsymbol{\theta}}_{\phi_{\lambda}}^{(t)} - \left(\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_{\lambda}}^{(t)})} \mathbf{W} + \lambda \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_{\lambda}}^{(t)})} \mathbf{D}_{\widehat{\mathbf{p}}}^{\lambda+1} \mathbf{D}_{\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_{\lambda}}^{(t)})}^{-(\lambda+2)} \boldsymbol{\Sigma}_{\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_{\lambda}}^{(t)})} \mathbf{W} \right)^{-1} \\ &\times \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_{\lambda}}^{(t)})} \mathbf{D}_{\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_{\lambda}}^{(t)})}^{-(\lambda+1)} (\widehat{\mathbf{p}}^{\lambda+1} - \mathbf{p}^{\lambda+1}(\widehat{\boldsymbol{\theta}}_{\phi_{\lambda}}^{(t)})). \end{aligned}$$

Para el modelo loglineal definido en (2.33) se obtiene el estimador $\widehat{\boldsymbol{\theta}}_{\phi_{\lambda}}$, para los valores de $\lambda = \{-1/2, 0, 2/3, 1, 2\}$, utilizando rutinas de optimización no lineal con restricciones de Matlab, para minimizar las correspondientes ϕ -divergencias de Cressie-Read.

Tabla 2.3. Estimaciones de $\boldsymbol{\theta}$ para diferentes valores de λ .

λ	$-\frac{1}{2}$	0	$\frac{2}{3}$	1	2
$\widehat{\theta}_{1(1)}$	0.0212	0.0209	0.0206	0.0207	0.0206
$\widehat{\theta}_{2(1)}$	-0.2399	-0.2393	-0.2372	-0.2364	-0.2341
$\widehat{\theta}_{12(11)}$	0.0758	0.0750	0.0745	0.0746	0.0741
$\widehat{\theta}_{3(1)}$	-0.1458	-0.1456	-0.1435	-0.1425	-0.1402
$\widehat{\theta}_{13(11)}$	0.1265	0.1256	0.1253	0.1254	0.1250
$\widehat{\theta}_{23(11)}$	-0.3957	-0.3946	-0.3922	-0.3910	-0.3881

El resto de los estimadores de los parámetros se obtienen a partir de las condiciones (2.34) y la constante

$$\exp(\widehat{u}) = \frac{1}{\sum_{j=1}^M \exp \sum_{g=1}^{M_0} w_{jg} \widehat{\theta}_g} = \frac{1}{8.82}.$$

En la Tabla 2.4 se recogen las probabilidades estimadas con el modelo loglineal planteado, utilizando diferentes valores de λ y en la última columna se recoge la estimación no paramétrica.

Tabla 2.4. Probabilidades estimadas para diferentes valores de λ .

λ	$-\frac{1}{2}$	0	$\frac{2}{3}$	1	2	No paramétrico
$\widehat{p}_{111}(\boldsymbol{\theta})$	0.06410	0.06414	0.06458	0.06483	0.06536	0.05633
$\widehat{p}_{112}(\boldsymbol{\theta})$	0.14701	0.14698	0.14675	0.14665	0.14662	0.14662
$\widehat{p}_{121}(\boldsymbol{\theta})$	0.19643	0.19614	0.19592	0.19588	0.19562	0.20422
$\widehat{p}_{122}(\boldsymbol{\theta})$	0.09250	0.09273	0.09274	0.09272	0.09279	0.08450
$\widehat{p}_{211}(\boldsymbol{\theta})$	0.04099	0.04118	0.04155	0.04169	0.04211	0.04929
$\widehat{p}_{212}(\boldsymbol{\theta})$	0.15595	0.15597	0.15586	0.15574	0.15556	0.16197
$\widehat{p}_{221}(\boldsymbol{\theta})$	0.17011	0.17001	0.16987	0.16977	0.16953	0.16197
$\widehat{p}_{222}(\boldsymbol{\theta})$	0.1329	0.13284	0.13272	0.13272	0.1326	0.14085

Obsérvese que si se analizan las variables Sexo y Diagnóstico de la enfermedad sin distinguir la Edad (hermano mayor (M) o menor(m)), los datos se presentan en $N = 71$ clusters, que son parejas de hermanos, $n = 2$, en donde podría existir una relación de dependencia entre hermanos que es necesario estimar mediante el coeficiente de correlación intracluster. La Tabla 2.5 muestra los 142 individuos clasificados ahora por las dos variables: Sexo y Diagnóstico, anteriormente descritas. Las categorías de ambas variables serán identificadas a partir de ahora con números ordinales (1 y 2) para las variables, $X_1^{(\ell)}$ (Sexo) y $X_2^{(\ell)}$ (Diagnóstico de Esquizofrenia), por orden. Por ejemplo, (H, S) se asocia con $(X_1^{(\ell)}, X_2^{(\ell)}) = (1, 2)$. Se plantea el modelo loglineal para la tabla de contingencia $I \times J$ con $I = J = 2$. En la Tabla 2.5 se presentan los valores de la variable $\mathbf{Y}^{(\ell)} = (Y_{11}^{(\ell)}, Y_{12}^{(\ell)}, \dots, Y_{22}^{(\ell)})^T$. Cada uno de estos vectores se corresponde a una fila, (ℓ) de las 71, que contiene los valores de la tabla de contingencia ℓ -ésima para las $M = 4$ categorías, en orden lexicográfico de la variable bidimensional $(X_1^{(\ell)}, X_2^{(\ell)})$.

Se plantea el modelo loglineal no saturado con hipótesis de independencia

$$p_{ij}(\boldsymbol{\theta}) = \exp\{u + \theta_{1(i)} + \theta_{2(j)}\}, \quad \text{con } i, j = 1, 2,$$

Tabla 2.5. Individuos por cluster según Sexo y Diagnóstico de Esquizofrenia .

ℓ	(H, N) $Y_{11}^{(\ell)}$	(H, S) $Y_{12}^{(\ell)}$	(M, N) $Y_{21}^{(\ell)}$	(M, S) $Y_{22}^{(\ell)}$	ℓ	(H, N) $Y_{11}^{(\ell)}$	(H, S) $Y_{12}^{(\ell)}$	(M, N) $Y_{21}^{(\ell)}$	(M, S) $Y_{22}^{(\ell)}$	ℓ	(H, N) $Y_{11}^{(\ell)}$	(H, S) $Y_{12}^{(\ell)}$	(M, N) $Y_{21}^{(\ell)}$	(M, S) $Y_{22}^{(\ell)}$
1	0	2	0	0	25	0	1	1	0	49	2	0	0	0
2	0	2	0	0	26	0	1	1	0	50	1	0	1	0
3	0	2	0	0	27	0	1	1	0	51	1	0	1	0
4	0	2	0	0	28	0	1	0	1	52	1	0	1	0
5	0	2	0	0	29	0	1	0	1	53	1	0	1	0
6	0	2	0	0	30	0	1	0	1	54	0	1	1	0
7	0	2	0	0	31	0	1	0	1	55	0	1	1	0
8	0	2	0	0	32	0	1	0	1	56	0	1	1	0
9	0	2	0	0	33	0	0	0	2	57	0	0	1	1
10	0	2	0	0	34	0	0	0	2	58	0	0	1	1
11	0	2	0	0	35	0	0	0	2	59	0	0	1	1
12	0	2	0	0	36	0	0	0	2	60	0	0	1	1
13	0	2	0	0	37	0	0	0	2	61	0	0	1	1
14	0	1	0	1	38	1	0	0	1	62	0	0	1	1
15	0	1	0	1	39	0	0	1	1	63	0	0	1	1
16	0	1	0	1	40	1	1	0	0	64	0	0	1	1
17	0	1	0	1	41	1	0	0	1	65	0	0	2	0
18	1	1	0	0	42	2	0	0	0	66	0	0	2	0
19	0	0	2	0	43	0	0	1	1	67	0	0	2	0
20	0	0	2	0	44	0	0	1	1	68	0	0	2	0
21	0	0	2	0	45	1	0	1	0	69	0	0	2	0
22	0	0	2	0	46	1	0	1	0	70	0	0	2	0
23	0	0	2	0	47	1	0	1	0	71	0	0	2	0
24	0	0	2	0	48	0	0	2	0					

donde el efecto de la variable Sexo viene representado por $\theta_{1(i)}$, y el diagnóstico de la enfermedad por $\theta_{2(k)}$ con la condición

$$\sum_{i=1}^2 \theta_{1(i)} = \sum_{j=1}^2 \theta_{2(j)} = 0. \quad (2.35)$$

Agrupando todos los individuos en una tabla de contingencia sin distinguir los clusters a los que pertenecen, se tiene la Tabla 2.6.

Tabla 2.6. Número de individuos según Sexo y Diagnóstico .

		<i>Diagnóstico</i>	
<i>Sexo</i>	<i>N</i>	<i>S</i>	
<i>H</i>	15	43	
<i>M</i>	52	32	

Con el modelo así planteado, los dos parámetros de interés, $\beta = (\rho^2, \mathbf{p}(\theta))^T$, son estimados por

$$\widehat{\beta}_{\widehat{n}^*, N, \lambda} = (\widehat{\rho}_{n, N, \lambda}^2, \mathbf{p}(\widehat{\theta}_{\phi_\lambda}))^T = \left(\frac{\widetilde{\vartheta}_{n, N, \lambda} - 1}{n - 1}, \frac{\exp\{\mathbf{W}\widehat{\theta}_{\phi_\lambda}\}}{\mathbf{1}_M^T \exp\{\mathbf{W}\widehat{\theta}_{\phi_\lambda}\}} \right)^T,$$

donde $\widetilde{\vartheta}_{n, N}$ se calcula según (2.32), y $\widehat{\theta}_{\phi_\lambda}$ utilizando la familia de ϕ -divergencias de Cressie y Read definidas en (2.4). Las estimaciones obtenidas aparecen en la Tabla 2.7, donde se observa que las estimaciones del coeficiente de correlación son muy similares, la estimación no paramétrica es la mayor de todas ellas y, para las obtenidas a partir de los estimadores de $MC\phi$, cuando $\lambda = -1/2$ parece ser que es claramente menor a las demás estimaciones. En cuanto a las probabilidades estimadas, sí existe bastante diferencia entre las estimaciones no paramétricas y las obtenidas mediante el modelo loglineal, esto es debido a que el modelo utilizado supone la independencia entre las dos variables.

Tabla 2.7. Estimaciones de las probabilidades y ρ para diferentes valores de λ y no paramétricas.

λ	$-\frac{1}{2}$	0	$\frac{2}{3}$	1	2	No paramétrico
$\rho_{n,N}^2(\tilde{\mathbf{Y}}, \hat{\theta}_\phi)$	0.2555	0.2758	0.2778	0.2823	0.2886	0.2926
$\hat{p}_{11}(\boldsymbol{\theta})$	0.2374	0.2475	0.2376	0.2283	0.2209	0.1056
$\hat{p}_{12}(\boldsymbol{\theta})$	0.2564	0.2953	0.3004	0.3054	0.3103	0.3662
$\hat{p}_{21}(\boldsymbol{\theta})$	0.2424	0.2122	0.2063	0.2007	0.1954	0.3028
$\hat{p}_{22}(\boldsymbol{\theta})$	0.2638	0.2450	0.2557	0.2657	0.2734	0.2253

2.6. Estudio de simulación

Para analizar el comportamiento de los nuevos estimadores del coeficiente de correlación intracluster, a continuación se realizará un estudio de simulación Monte Carlo basado en los datos del ejemplo anterior. En dicho estudio se generará $R = 10000$ repeticiones de muestras con $N = 71$ parejas ($n = 2$) para diferentes valores del coeficiente de correlación intra-cluster y se comparará la raíz del error cuadrático medio (RECM) cometido con las estimaciones realizadas con cada uno de los estimadores definidos. Es decir,

$$RECM = \sqrt{\frac{\sum_{r=1}^R (\tilde{\rho}_{n,N,\lambda}^2 - \rho_0^2)^2}{R}}$$

así como los sesgos medios

$$Sg = \frac{\sum_{r=1}^R \tilde{\rho}_{n,N,\lambda,r}^2 - \rho_0^2}{R}$$

de dichos estimadores. Puesto que, el espacio paramétrico de ρ^2 es $[0, 1]$, los estimadores serán truncados a valores dentro de dicho intervalo. Para generar las muestras se utilizarán las tres distribuciones que se introdujeron en el Capítulo 1 para modelizar la dependencia dentro de los clusters: Dirichlet-multinomial (DM), la multinomial n-inflada (NI) y la distribución multinomial agrupada aleatoriamente (RC).

La raíz del error cuadrático medio calculada con las muestras simuladas con las tres distribuciones, tiene un comportamiento similar con las tres, como se puede ver en los gráficos de las Figuras 2.1, 2.2 y 2.3. Se podría destacar que, la distribución multinomial

n-inflada es la que da lugar a un peor comportamiento de todos los estimadores, ya que presenta una mayor *RECM* en comparación con las otras dos distribuciones. En ellos se puede apreciar que el estimador con mayor sesgo medio es el no paramétrico (Brier) y, de todos los estimadores de $MC\phi$, el que tiene menor sesgo medio es el calculado para $\lambda = 2$. Se observa además que para los valores de ρ^2 entre 0.2 y 0.6, el sesgo es menor, mientras que la raíz del error cuadrático medio (*RECM*), para estos valores del coeficiente de correlación, es mayor, esto es porque las estimaciones de ρ^2 están centradas en el verdadero valor del parámetro y, al calcular el sesgo medio se contrarrestan las estimaciones mayores y menores. Puesto que, todos los estimadores de $MC\phi$ tienen una *RECM* prácticamente idéntico, en las figuras 2.4, 2.5 y 2.6, hemos elegido el estimador de $MC\phi$ que se obtiene con $\lambda = 2/3$, por ser uno de los que menor sesgo tiene, para compararlo con el estimador no paramétrico. En estas figuras se aprecia mejor la reducción, tanto en la *RECM* como especialmente en el sesgo, que ofrece el estimador de $MC\phi$ con respecto a la estimación no paramétrica, sea cual sea la distribución utilizada.

Figura 2.1. RECM y sesgo de los diferentes estimadores de ρ^2 , para datos generados con la distribución Dirichlet multinomial .

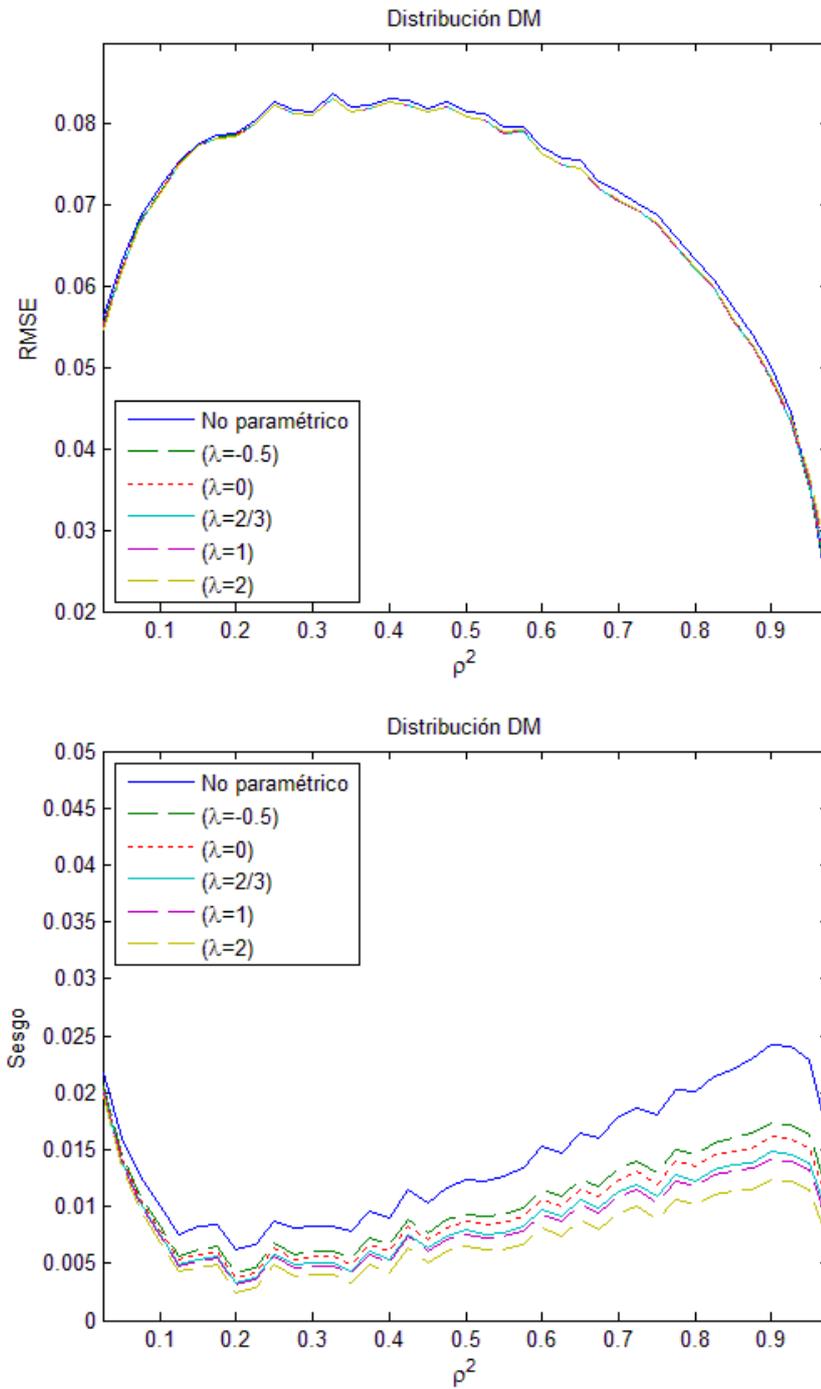


Figura 2.2. RECM y sesgo de los diferentes estimadores de ρ^2 , para datos generados con la distribución multinomial n-inflada .

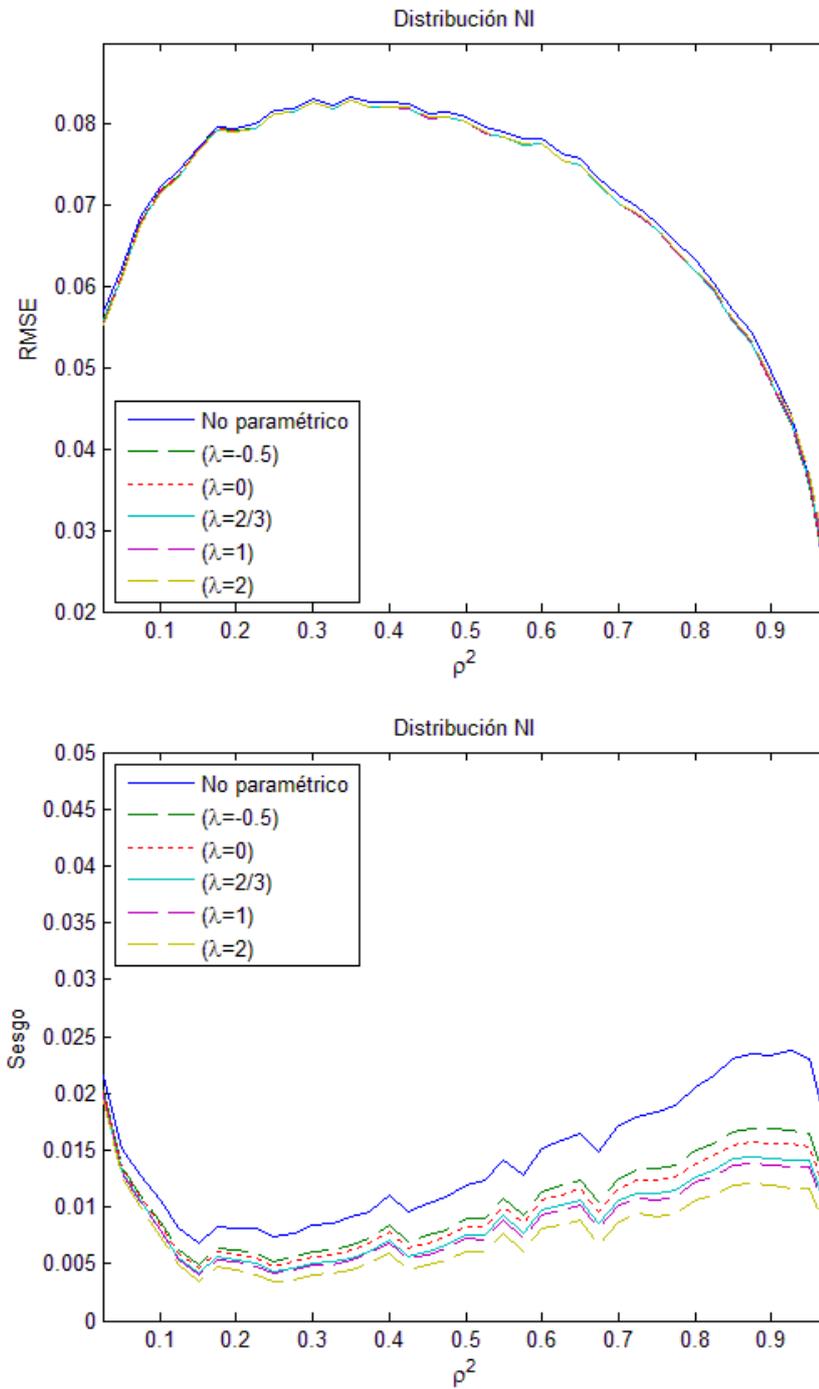


Figura 2.3. RECM y sesgo de los diferentes estimadores de ρ^2 , para datos generados con la distribución multinomial agrupada aleatoriamente .

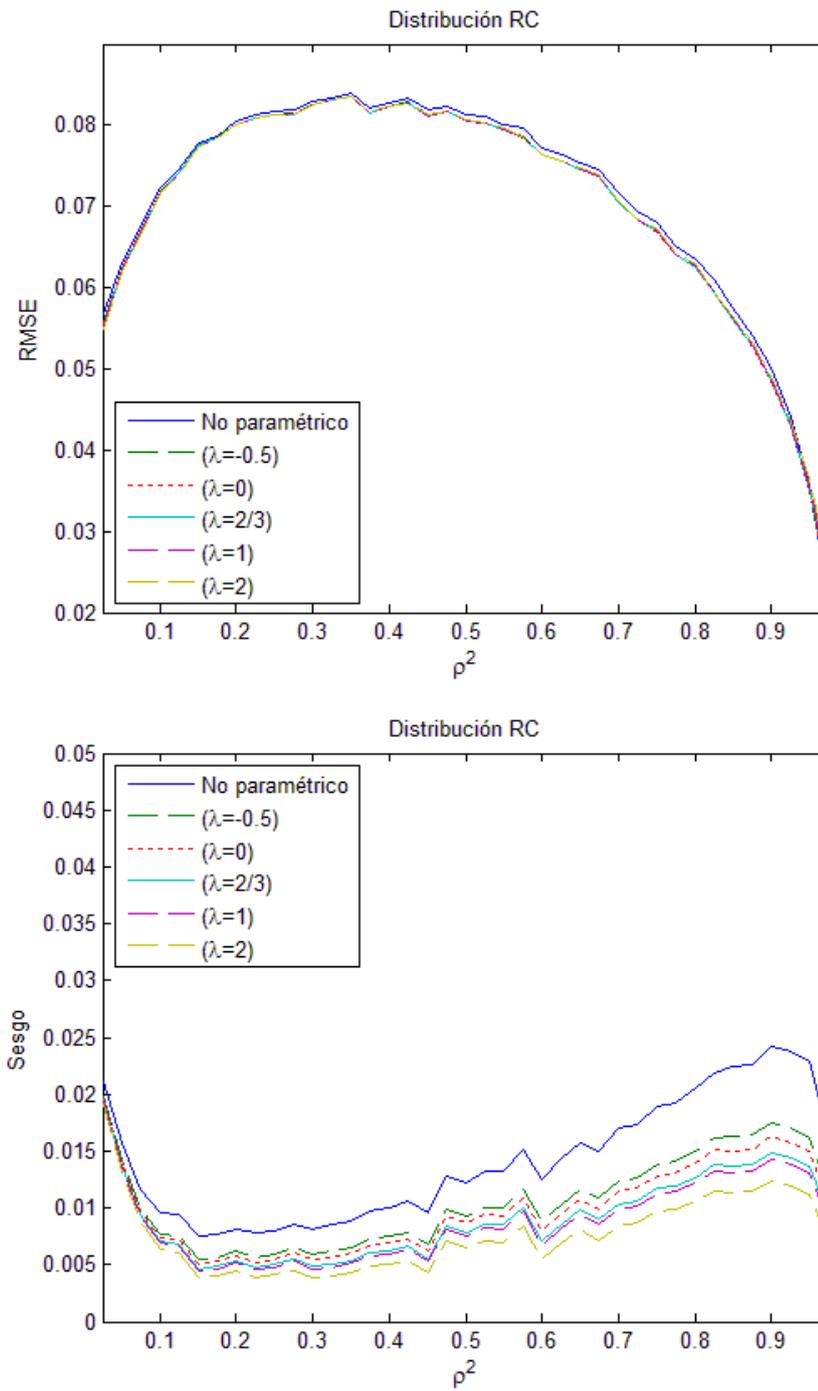


Figura 2.4. RECM y sesgo de los estimadores de ρ^2 , no paramétrico y de $MC\phi$ con $\lambda = 2/3$, para datos generados con la distribución Dirichlet multinomial.

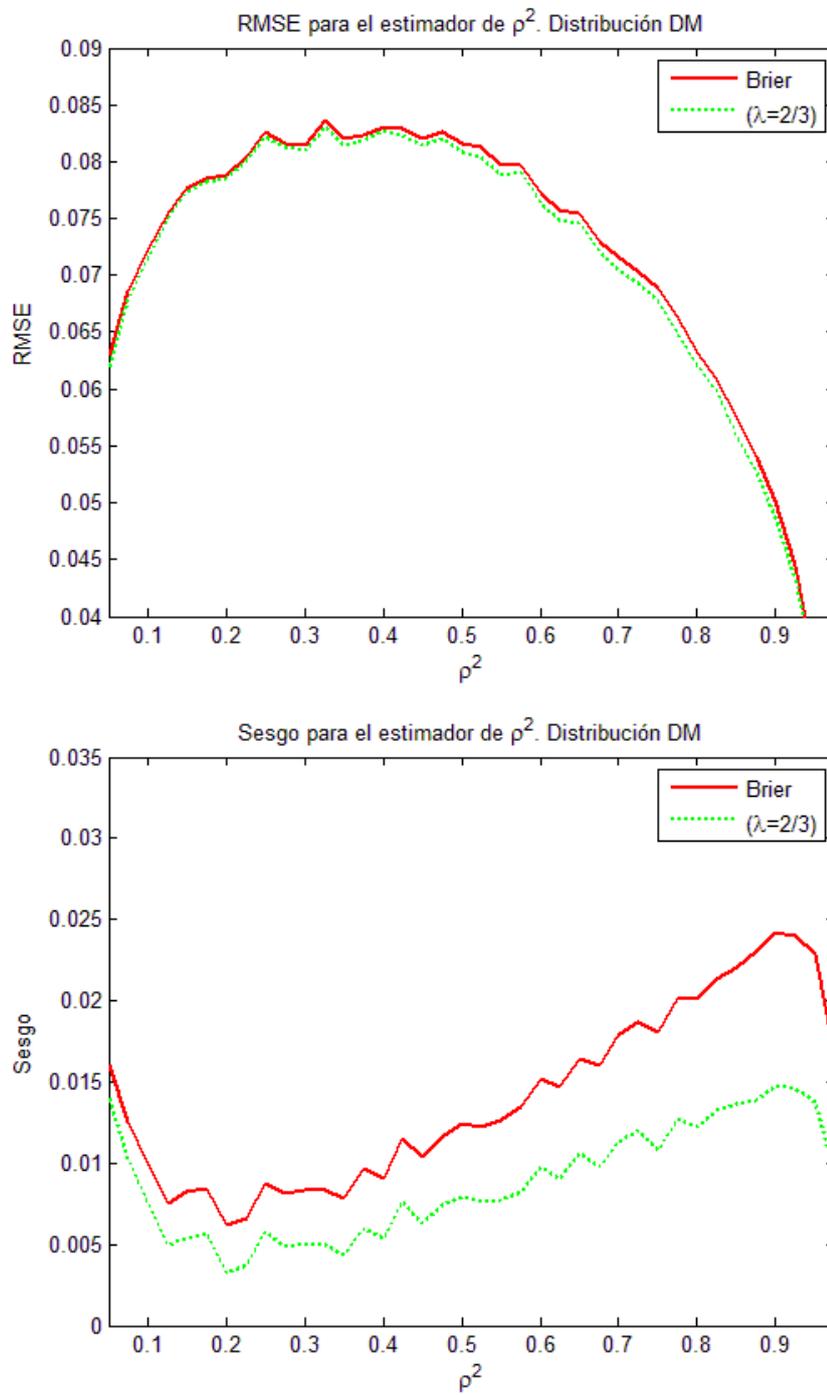


Figura 2.5. RECM y sesgo para los estimadores de ρ^2 , no paramétrico y de $MC\phi$ con $\lambda = 2/3$, para datos generados con la distribución multinomial n-inflada.

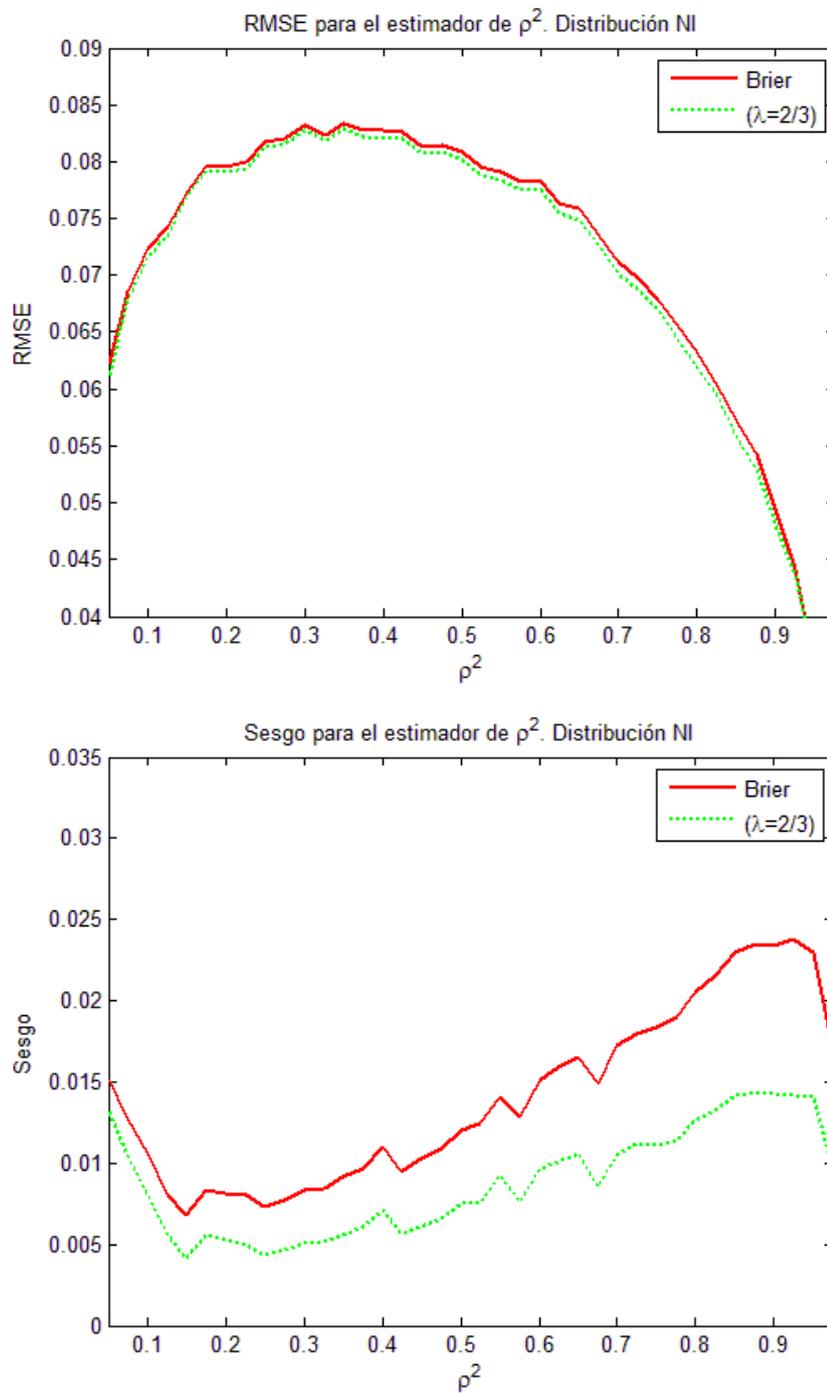
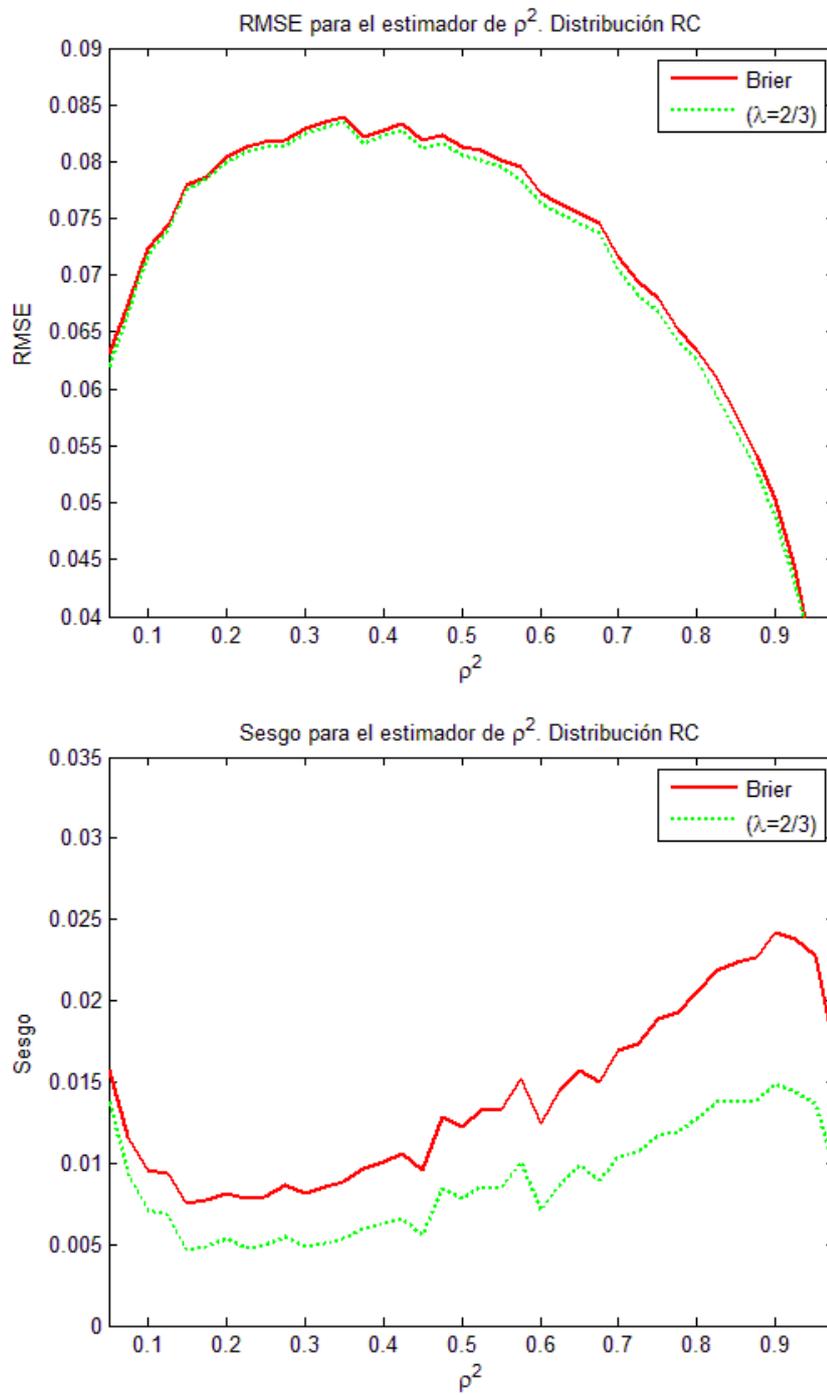


Figura 2.6. RECM y sesgo para los estimadores de ρ^2 , no paramétrico y de $MC\phi$ con $\lambda = 2/3$, para datos generados con la distribución agrupada aleatoriamente.



Para estudiar la influencia del tamaño del cluster sobre el comportamiento de los estimadores se ha realizado la misma simulación anterior pero con tamaños de cluster $n = 5$, $n = 10$ y $n = 20$. Los resultados obtenidos se muestran en las figuras 2.7, a 2.15, en donde se comprueba que la *RECM* disminuye a medida que aumenta el tamaño del cluster. De nuevo se observa que, las estimaciones generadas utilizando la distribución multinomial n-inflada tienen un menor descenso en la *RECM* así como un peor comportamiento del sesgo medio. El sesgo medio va creciendo a medida que aumenta la correlación intra-cluster y es para estos valores en los que se observa mayor diferencia entre las estimaciones generadas con el estimador no paramétrico (Brier) y los *EMC* ϕ s.

Si se analizan específicamente las muestras generadas con baja correlación, en concreto para $\rho^2 = 0.25$, en la figura 2.16, y muestras generadas con alta correlación, $\rho^2 = 0.75$, en la figura 2.17, se observa que, en el caso de menor correlación, todos los estimadores y, para todas las distribuciones, están centrados, mientras que en el caso de mayor correlación intra-cluster ($\rho^2 = 0.75$), los estimadores tienen un sesgo medio positivo. Se comprueba además que la reducción en la dispersión de las estimaciones al aumentar el tamaño del cluster es mayor en las distribuciones multinomial agrupada aleatoriamente y Dirichlet multinomial, que en la n-inflada.

Figura 2.7. RECM y sesgo de los diferentes estimadores de ρ^2 para datos generados con la distribución Dirichlet multinomial con $n=5$.

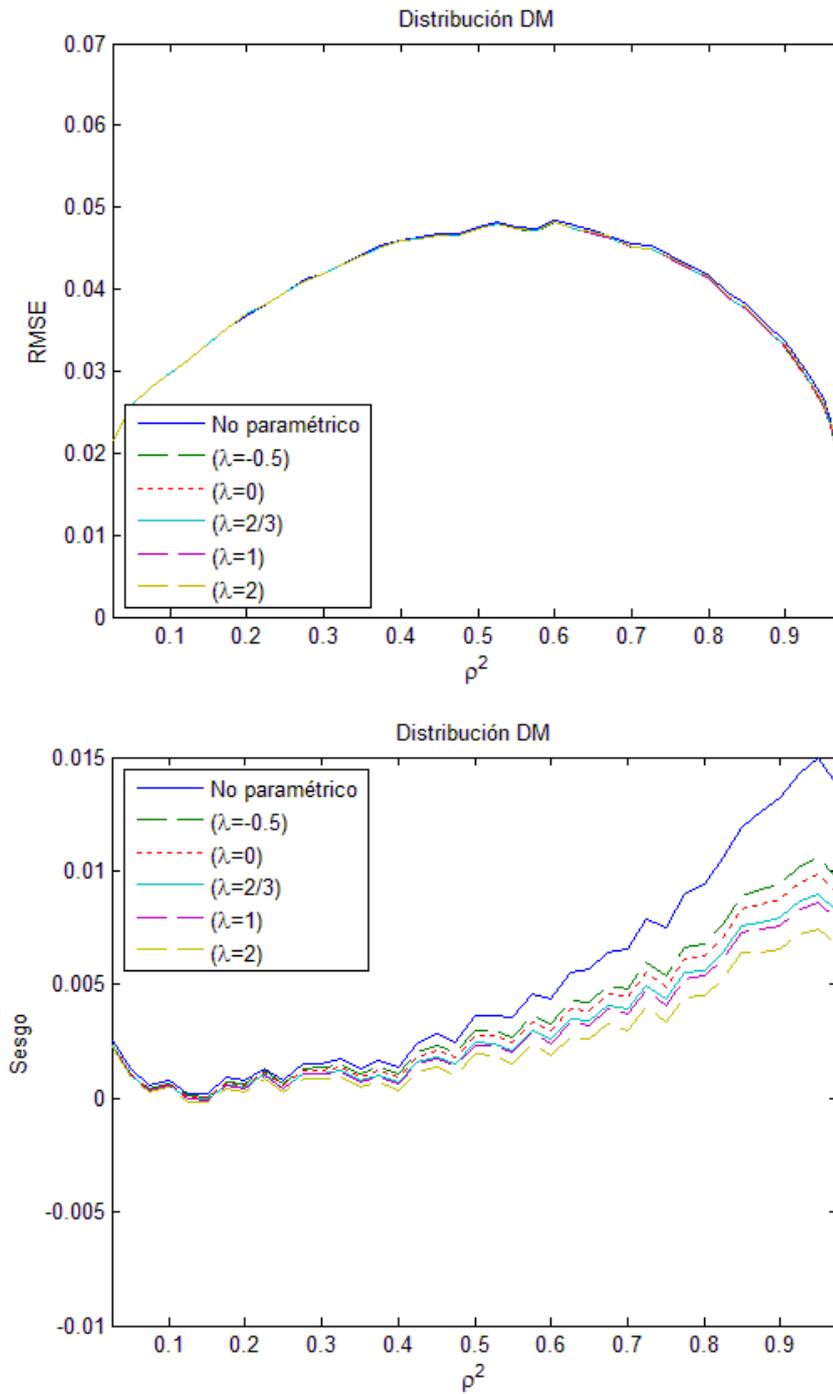


Figura 2.8. RECM y sesgo de los diferentes estimadores de ρ^2 para datos generados con la distribución multinomial n-inflada con $n=5$.

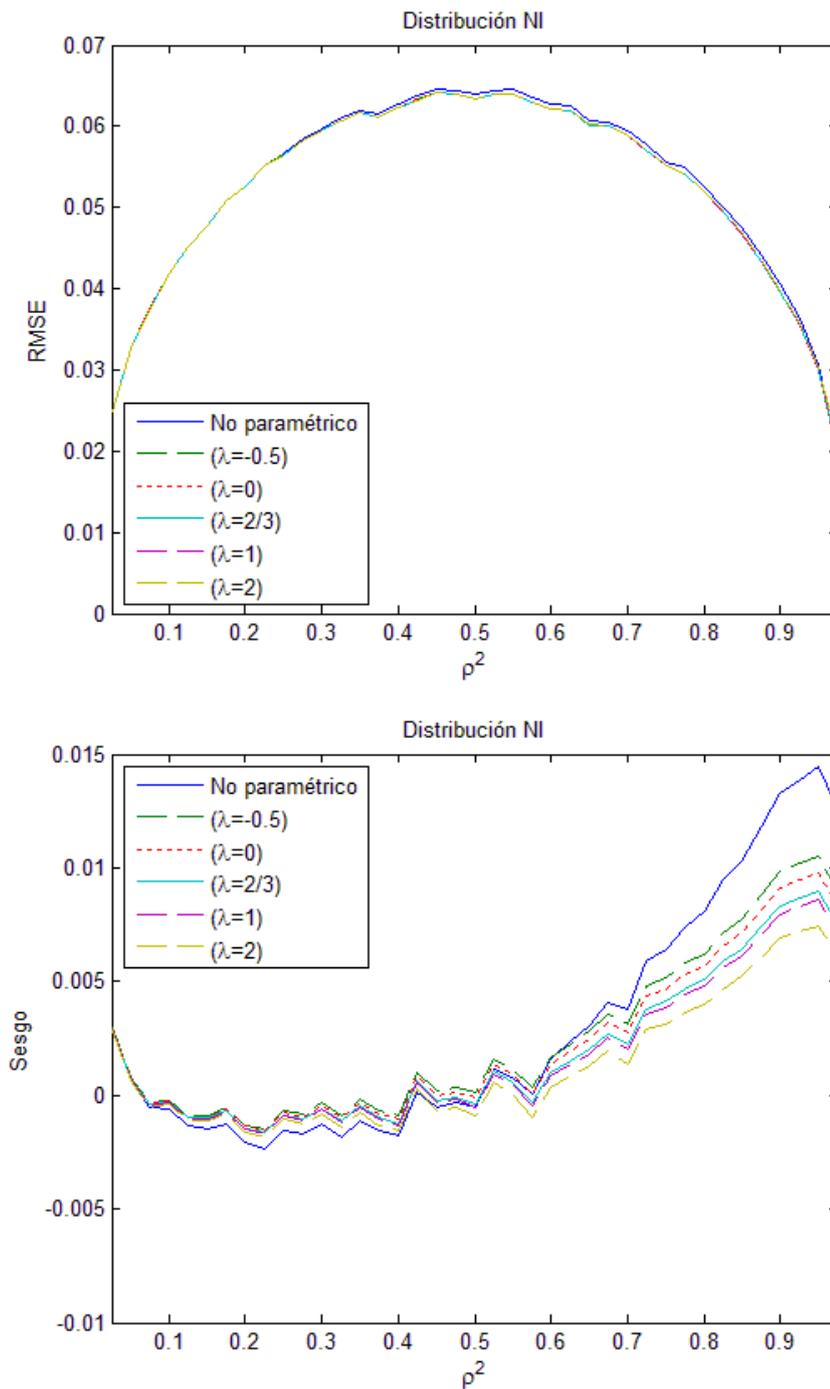


Figura 2.9. RECM y sesgo de los diferentes estimadores de ρ^2 para datos generados con la distribución multinomial agrupada aleatoriamente con $n=5$.

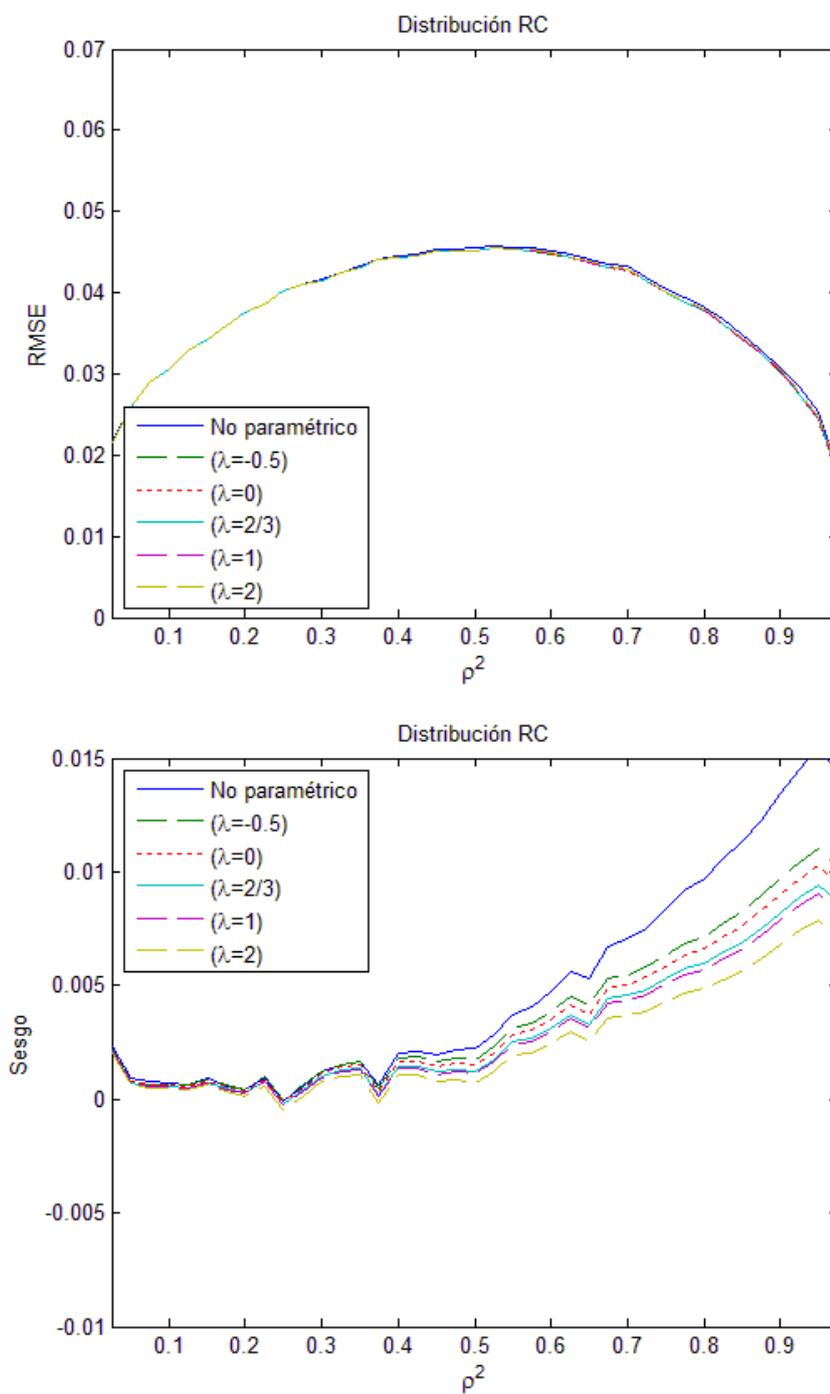


Figura 2.10. RECM y sesgo para los diferentes estimadores de ρ^2 con para datos generados con la distribución Dirichlet multinomial $n=10$.

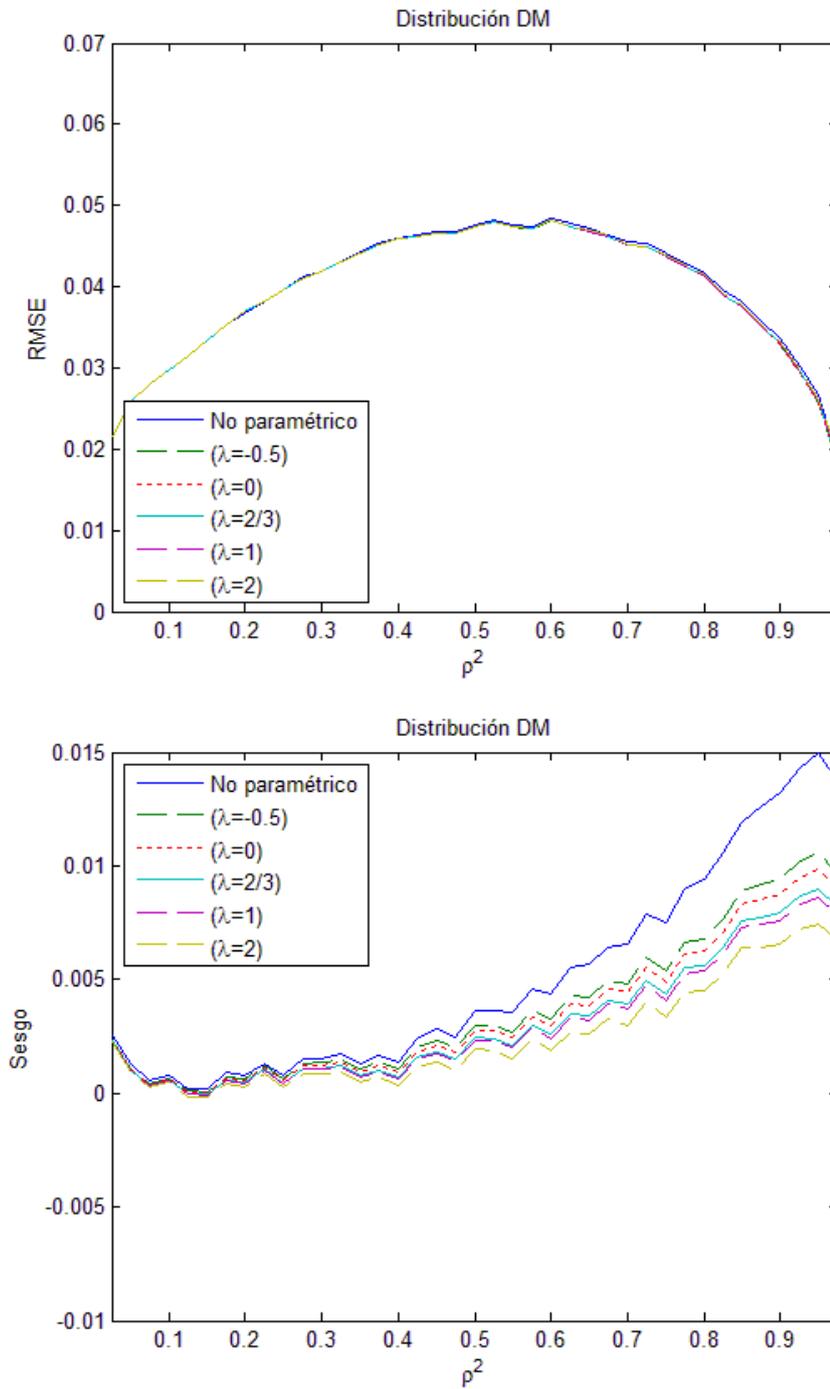


Figura 2.11. RECM y sesgo para los diferentes estimadores de ρ^2 para datos generados con la distribución multinomial n-inflada con $n=10$.

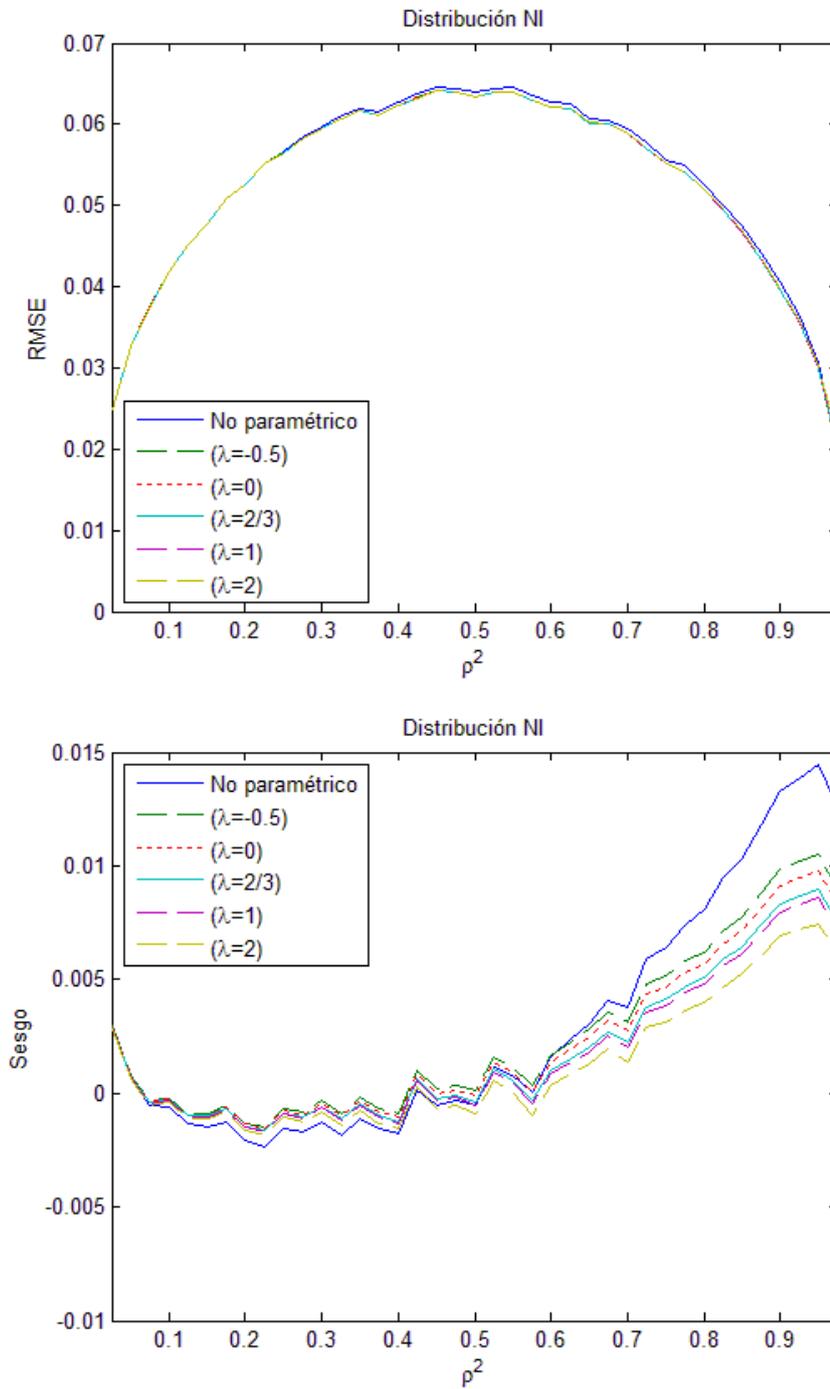


Figura 2.12. RECM y sesgo para los diferentes estimadores de ρ^2 para datos generados con la distribución multinomial agrupada aleatoriamente con $n=10$.

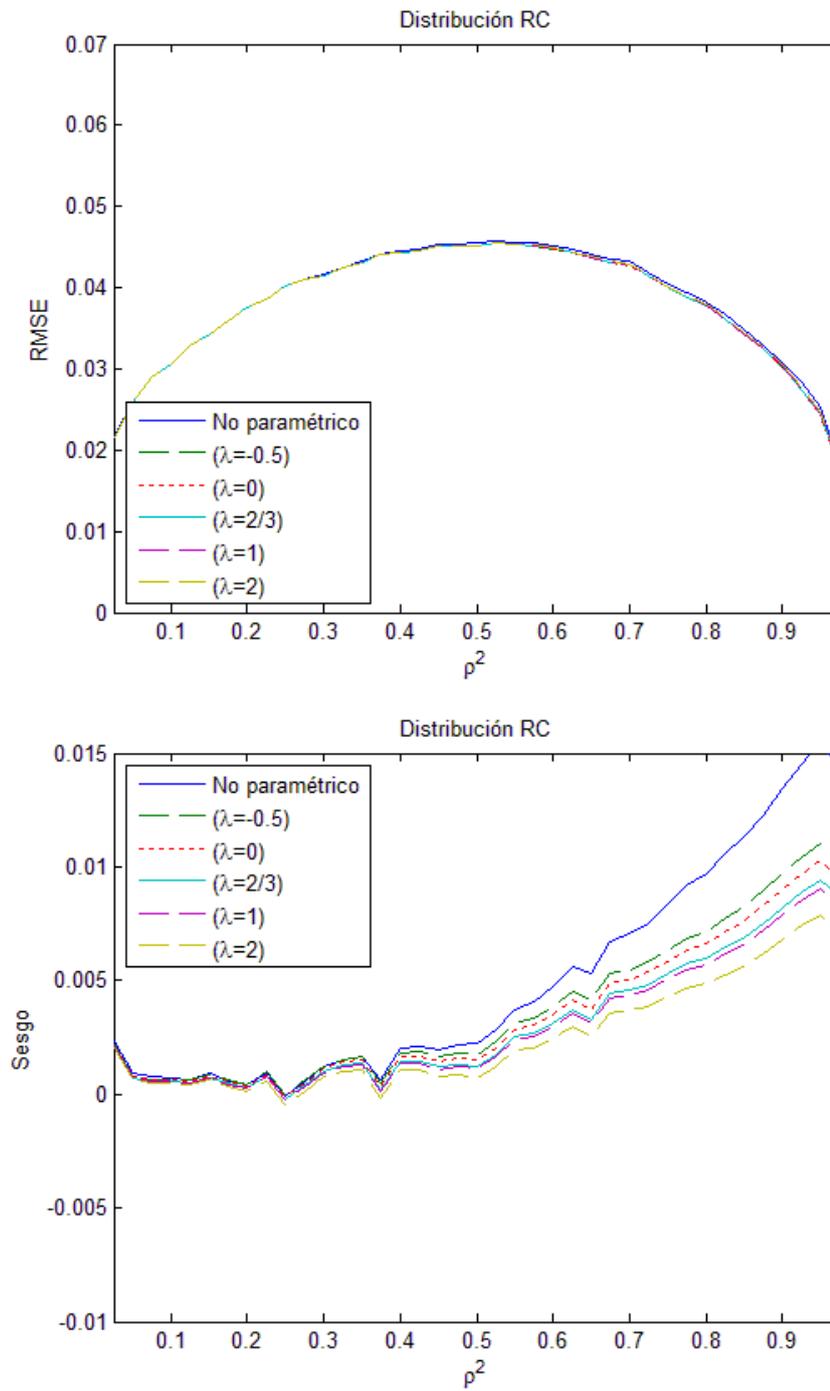


Figura 2.13. RECM y sesgo para los diferentes estimadores de ρ^2 para datos generados con la distribución Dirichlet multinomial con $n=20$.

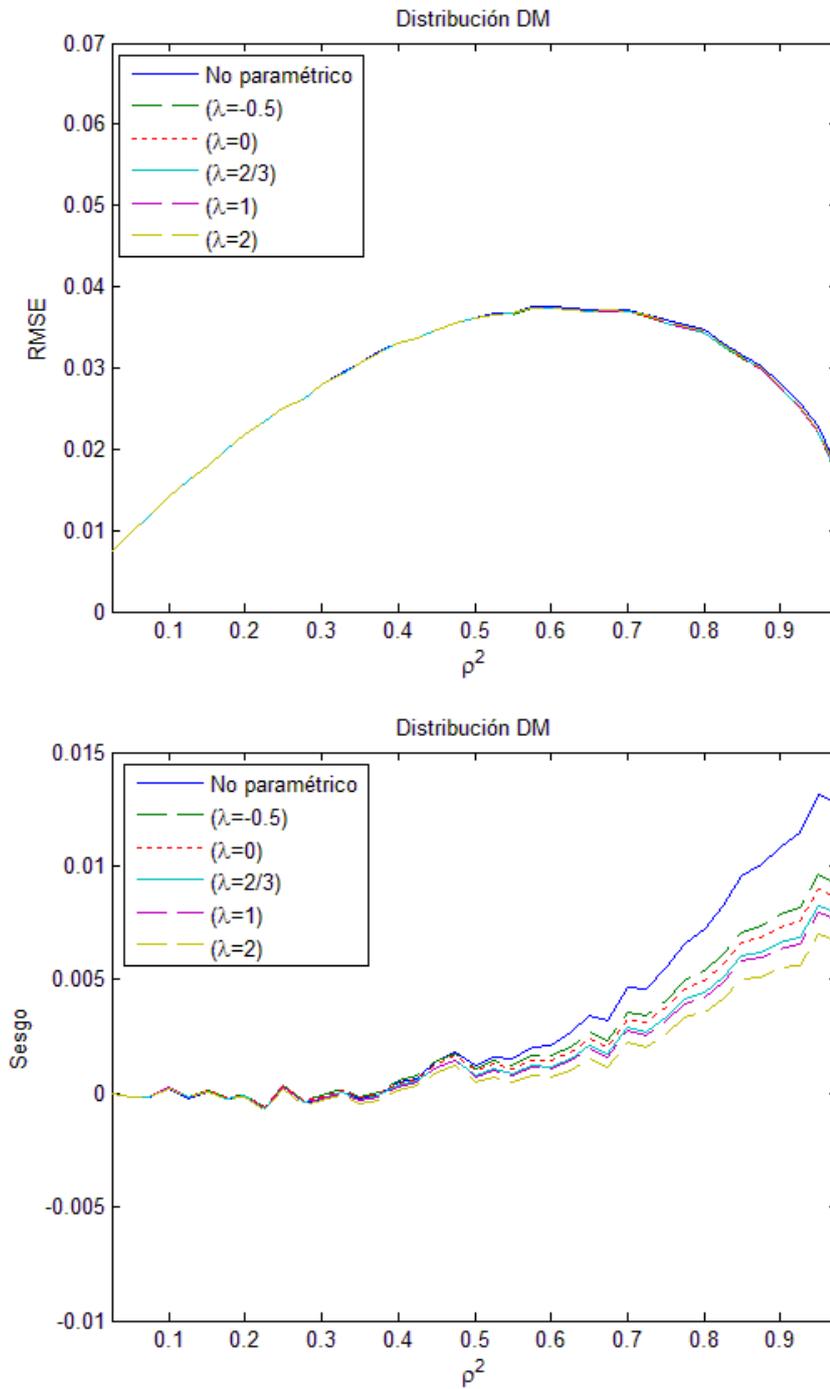


Figura 2.14. RECM y sesgo para los diferentes estimadores de ρ^2 para datos generados con la distribución multinomial n-inflada con $n=20$.

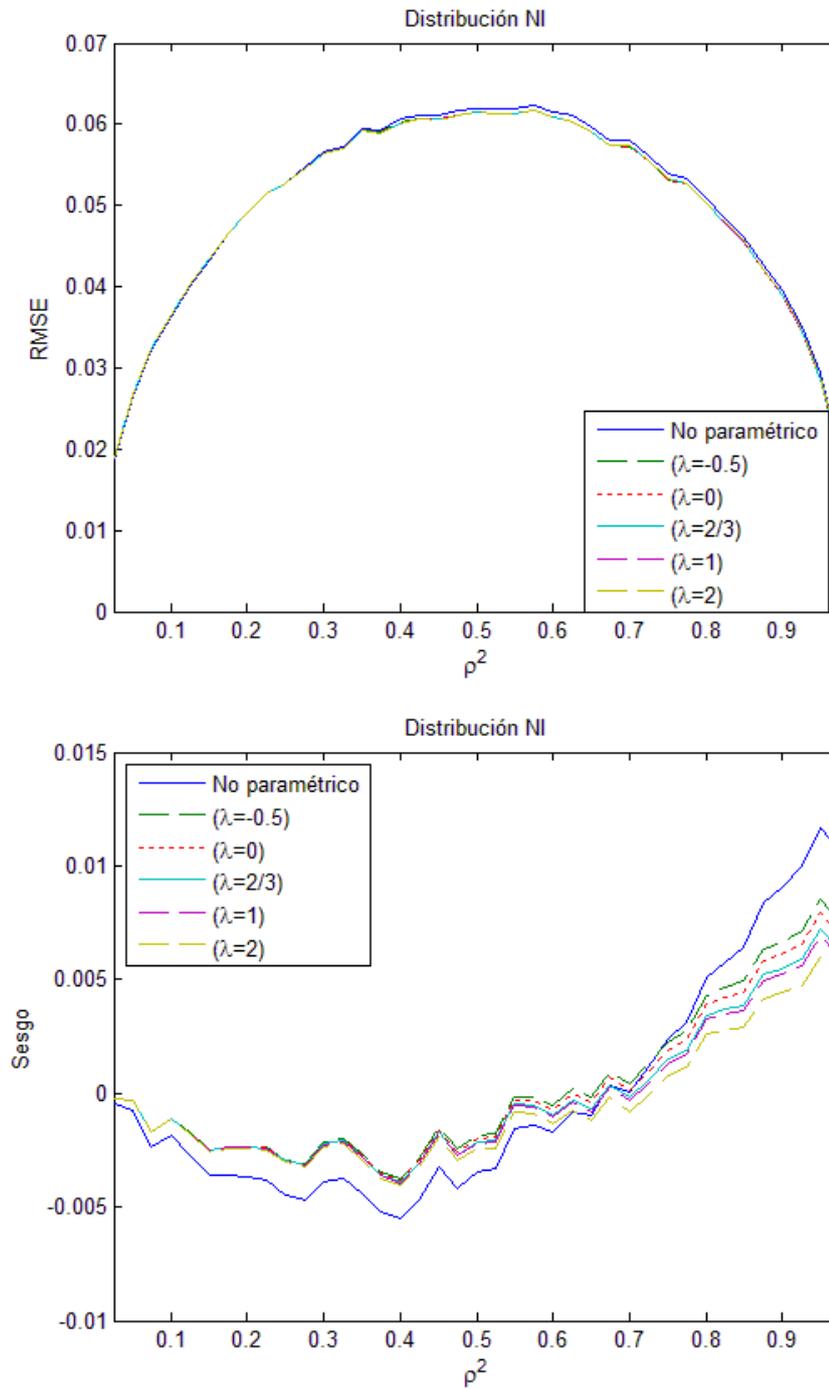


Figura 2.15. RECM y sesgo para los diferentes estimadores de ρ^2 para datos generados con la distribución multinomial agrupada aleatoriamente con $n=20$.

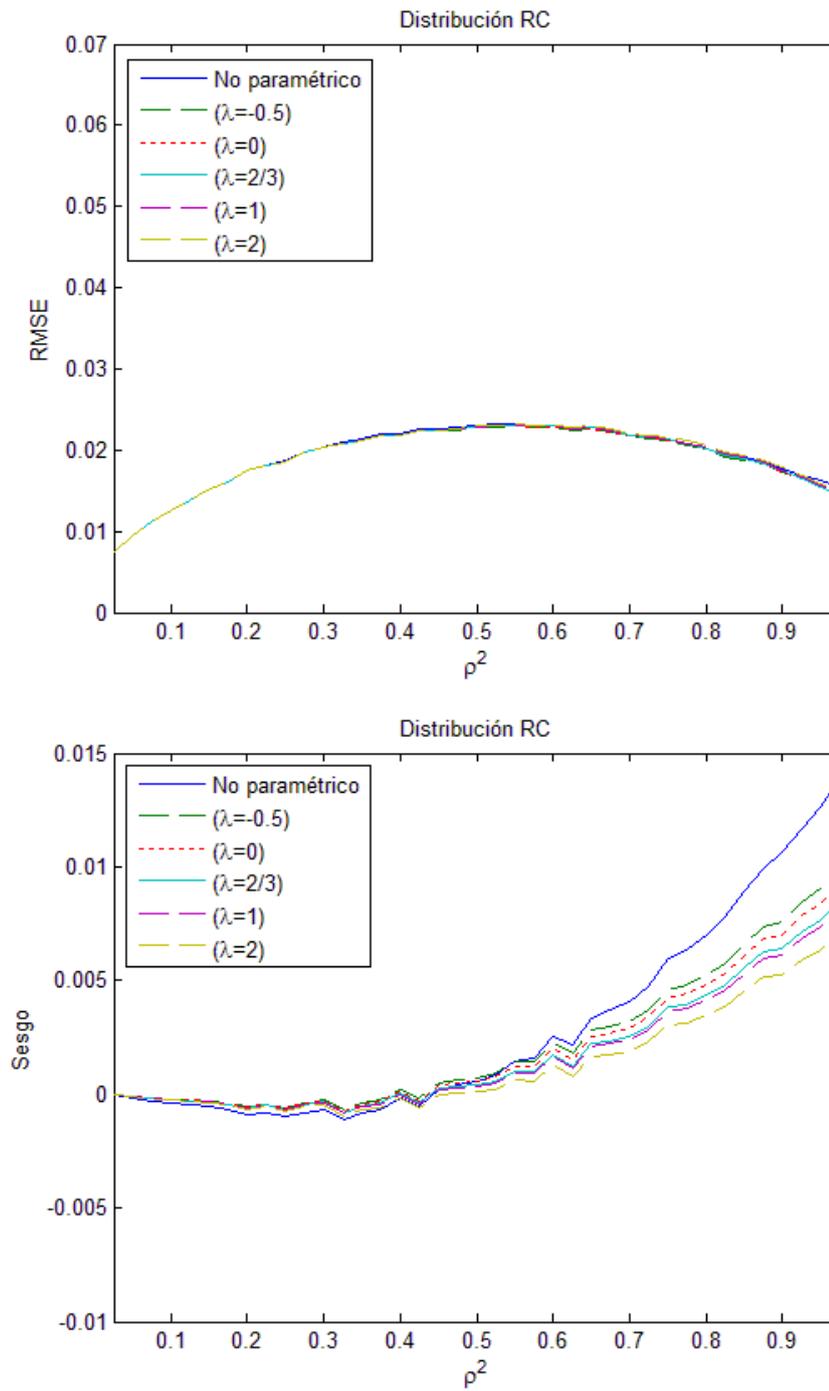


Figura 2.16. Distribución de las estimaciones para diferentes tamaños de cluster, $n=5$, $n=10$, y $n=20$ cuando $\rho^2=0.25$.

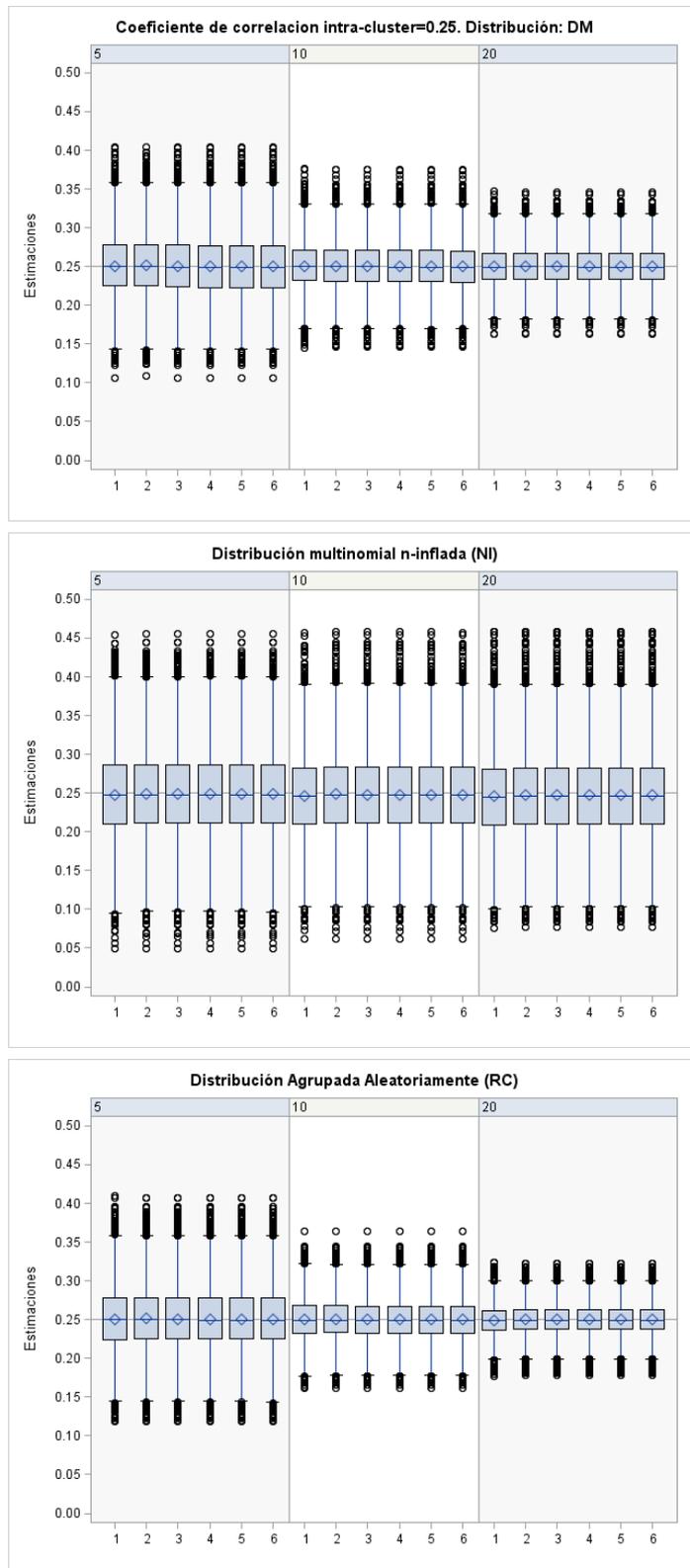
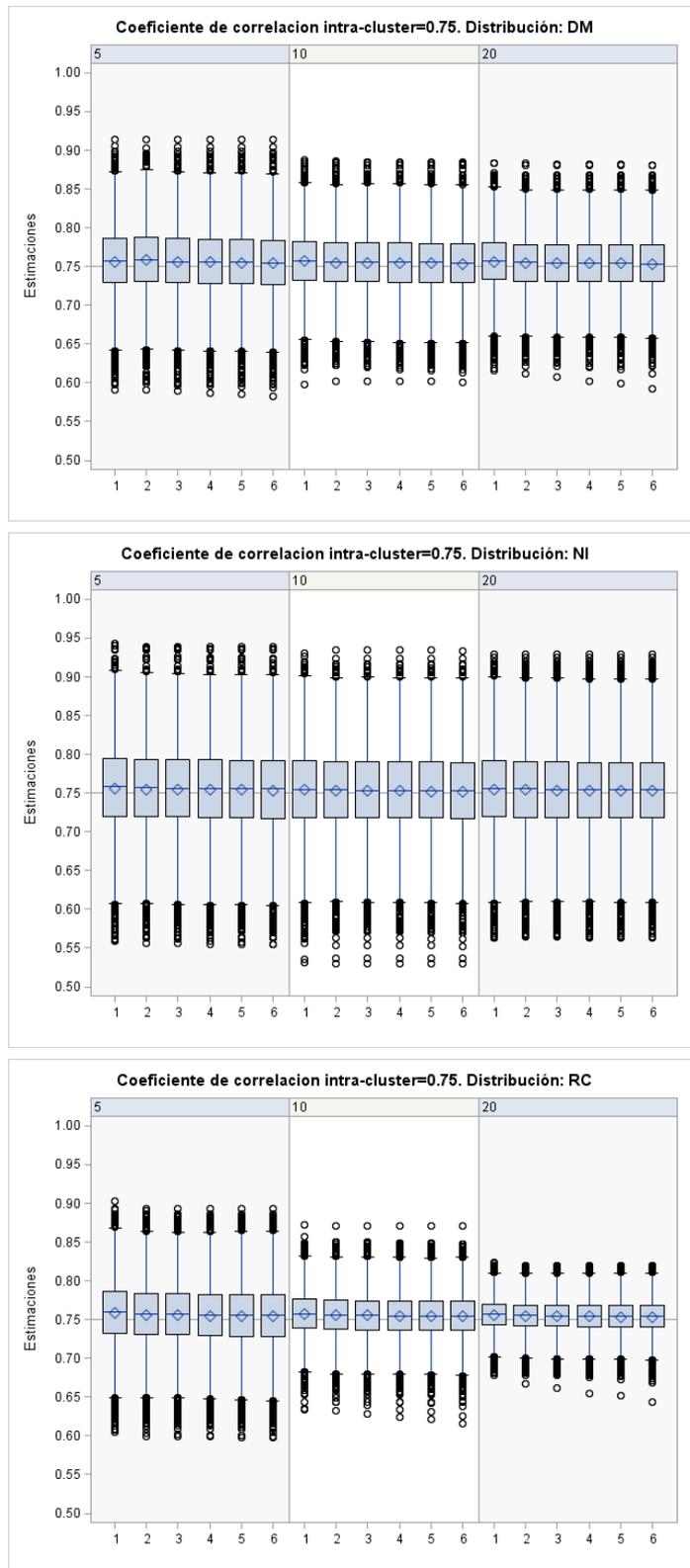


Figura 2.17. Distribución de las estimaciones para diferentes tamaños de cluster, $n=5$, $n=10$, y $n=20$, cuando $\rho^2=0.75$.



CAPÍTULO 3

NUEVOS ESTIMADORES PARA EL COEFICIENTE DE SOBREDISPERSIÓN CON CLUSTERS O CONGLOMERADOS DE TAMAÑOS DIFERENTES

La hipótesis de clusters de igual tamaño, bajo la cual se desarrolló el Capítulo 2, no siempre es posible encontrar en la práctica, especialmente cuando el muestreo no ha sido previamente planificado. Por esta razón, en este Capítulo se presentan nuevos estimadores semiparamétricos para el efecto del diseño ó coeficiente de sobredispersión, y para el coeficiente de correlación intracluster, cuando los tamaños de los clusters son diferentes. Además, se presenta un caso especial para el caso en el que el tamaño de los clusters es grande. Estos estimadores se calculan sobre un caso práctico con datos de identificación de ADN. Para finalizar se analiza el comportamiento de los nuevos estimadores del coeficiente de correlación intracluster en el caso de tamaños desiguales, a través de dos estudios de simulación Monte Carlo.

3.1. Notación y resultados básicos

Sin pérdida de generalidad, se consideran G grupos de clusters de forma que todas las tablas de contingencia, $\mathbf{Y}^{(g,\ell)} = (Y_1^{(g,\ell)}, \dots, Y_M^{(g,\ell)})^T$, $\ell = 1, \dots, N_g$ del mismo grupo de clusters tienen el mismo tamaño muestral n_g , $g = 1, \dots, G$, y $N = \sum_{g=1}^G N_g$. Se supone que al menos existe un índice g tal que $n_g > 1$. Si se sustituye la hipótesis $N \rightarrow \infty$ por $N_g \rightarrow \infty$ para cada uno de los grupos de clusters, entonces todos los resultados establecidos en el capítulo anterior se cumplen de forma separada para cada uno de los grupos de clusters ya que en estos sí se cumple la condición de tamaños de cluster iguales.

Los estimadores no paramétricos de $\mathbf{p}(\boldsymbol{\theta})$, calculados en función de los valores observados en cada uno de los N_g grupos de clusters, vienen dados por

$$\widehat{\mathbf{p}}^{(g)} = \frac{1}{N_g n_g} \sum_{\ell=1}^{N_g} \mathbf{Y}^{(g,\ell)},$$

es decir, $\widehat{\mathbf{p}}^{(g)} = (\widehat{p}_1^{(g)}, \dots, \widehat{p}_M^{(g)})^T$, con $\widehat{p}_r^{(g)} = \frac{1}{N_g n_g} \sum_{\ell=1}^{N_g} Y_r^{(g,\ell)}$, $r = 1, \dots, M$. Estos estimadores se pueden también expresar como la media de los estimadores no paramétricos de $\mathbf{p}(\boldsymbol{\theta})$, para cada uno de los clusters con índice “ ℓ ”, $\widehat{\mathbf{p}}^{(g,\ell)} = (\widehat{p}_1^{(g,\ell)}, \dots, \widehat{p}_M^{(g,\ell)})^T = \frac{1}{n_g} \mathbf{Y}^{(g,\ell)}$, $\ell = 1, \dots, N_g$, de la siguiente forma

$$\begin{aligned} \widehat{\mathbf{p}}^{(g)} &= \frac{1}{N_g} \sum_{\ell=1}^{N_g} \widehat{\mathbf{p}}^{(g,\ell)}, \\ \widehat{p}_r^{(g,\ell)} &= \frac{1}{n_g} Y_r^{(g,\ell)}. \end{aligned}$$

Por otro lado, el estimador no paramétrico de $\mathbf{p}(\boldsymbol{\theta})$, basado en el total de los G grupos de N_1, \dots, N_G clusters con tamaños muestrales n_1, \dots, n_G se puede expresar en función de los anteriores como

$$\widehat{\mathbf{p}} = \frac{\sum_{g=1}^G \sum_{\ell=1}^{N_g} \mathbf{Y}^{(g,\ell)}}{\sum_{g=1}^G N_g n_g} = \frac{\sum_{g=1}^G N_g n_g \frac{1}{N_g} \sum_{\ell=1}^{N_g} \frac{1}{n_g} \mathbf{Y}^{(g,\ell)}}{\sum_{g=1}^G N_g n_g} = \sum_{g=1}^G w_g \widehat{\mathbf{p}}^{(g)}, \quad (3.1)$$

donde

$$w_g = \frac{N_g n_g}{\sum_{h=1}^G N_h n_h} > 0, \quad g = 1, \dots, G, \quad (3.2)$$

con $\sum_{g=1}^G w_g = 1$.

Teorema 3.1 Para el estimador no paramétrico de $\mathbf{p}(\boldsymbol{\theta})$, dado en (3.1) se verifica

$$\sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_M, \frac{\vartheta_{n^*}}{n} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}), \quad (3.3)$$

siendo $\vartheta_{n^*} = 1 + \rho^2 (n^* - 1) \in (1, n^*]$.

Demostración Aplicando el Teorema Central del Límite, para el g -ésimo grupo, se sigue que

$$\sqrt{N_g}(\widehat{\mathbf{p}}^{(g)} - \mathbf{p}(\boldsymbol{\theta}_0)) \xrightarrow[N_g \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_M, \frac{\vartheta_{n_g}}{n_g} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}). \quad (3.4)$$

Multiplicando por $\sqrt{N_g n_g} / \sum_{h=1}^G N_h n_h$,

$$w_g(\widehat{\mathbf{p}}^{(g)} - \mathbf{p}(\boldsymbol{\theta}_0)) \xrightarrow[N_g \rightarrow \infty]{\mathcal{L}} \mathcal{N}\left(\mathbf{0}_M, \frac{N_g n_g \vartheta_{n_g}}{(\sum_{h=1}^G N_h n_h)^2} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}\right),$$

con w_g definido en (3.2). Sumando desde $g = 1$ a G y teniendo en cuenta la independencia entre los clusters

$$\sum_{g=1}^G w_g(\widehat{\mathbf{p}}^{(g)} - \mathbf{p}(\boldsymbol{\theta}_0)) = (\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) \xrightarrow[N_g \rightarrow \infty, g=1, \dots, G]{\mathcal{L}} \mathcal{N}\left(\mathbf{0}_M, \frac{\sum_{g=1}^G N_g n_g \vartheta_{n_g}}{(\sum_{h=1}^G N_h n_h)^2} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}\right).$$

Finalmente, si se multiplica por $\sum_{h=1}^G N_h n_h / \sqrt{\sum_{g=1}^G N_g n_g \vartheta_{n_g}}$ se obtiene

$$\frac{\sum_{h=1}^G N_h n_h}{\sqrt{\sum_{g=1}^G N_g n_g \vartheta_{n_g}}} (\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) \xrightarrow[N_1, \dots, N_G \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_M, \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}). \quad (3.5)$$

Si además, se supone que existe una sucesión $\{N_h^*\}_{h=1}^G$, tal que

$$\frac{N_h}{N} \xrightarrow[N \rightarrow \infty]{P} N_h^* \in (0, 1],$$

Si dividimos numerador y denominador de (3.5) por N , se obtiene que

$$\frac{\sqrt{N} \sqrt{\sum_{h=1}^G \frac{N_h}{N} n_h}}{\sqrt{\sum_{g=1}^G \left(\frac{N_g}{N} n_g \vartheta_{n_g} \right) \left(\sum_{h=1}^G \frac{N_h}{N} n_h \right)^{-1}}} (\hat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) \xrightarrow[N_1, \dots, N_G \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_M, \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}).$$

Si se llama

$$\bar{n} = \sum_{g=1}^G N_g^* n_g, \quad (3.6)$$

y

$$\vartheta_{n^*} = \sum_{g=1}^G w_g^* \vartheta_{n_g}, \quad (3.7)$$

con

$$w_g^* = \frac{N_g^* n_g}{\sum_{h=1}^G N_h^* n_h} > 0, \quad g = 1, \dots, G,$$

de forma que los pesos verifican

$$w_g \xrightarrow[N \rightarrow \infty]{P} w_g^*, \quad \sum_{g=1}^G w_g^* = 1, \quad (3.8)$$

y se llega al resultado deseado. ■

Definición 3.1 *Se define el parámetro de sobredispersión ó efecto del diseño cuando el tamaño del cluster es el tamaño medio, como*

$$\vartheta_{n^*} = 1 + \rho^2 (n^* - 1) \in (1, n^*], \quad (3.9)$$

donde

$$n^* = \sum_{g=1}^G w_g^* n_g. \quad (3.10)$$

En particular, si $\vartheta_{n^*} = 1$ ($\rho^2 = 0$ o $n_1 = \dots = n_G = 1$) se estaría en el caso de muestreo multinomial.

Observación 3.1 *Obsérvese que $\bar{n} = \sum_{g=1}^G N_g^* n_g$, es el tamaño muestral medio de un cluster calculado considerando como peso para cada grupo de clusters la proporción de clusters de cada tamaño y $n^* = \sum_{g=1}^G w_g^* n_g$ es el tamaño muestral medio de un cluster calculado considerando como peso para cada grupo de clusters la proporción de observaciones en cada grupo de clusters.*

Es importante mencionar que Brier (1980) propuso un estimador para el parámetro desconocido ϑ_{n^*} , dado en (3.9), bajo la hipótesis, más fuerte, de que la distribución de las variables $\mathbf{Y}^{(g,\ell)}$ es la Dirichlet-multinomial (ver 1.7). Por esta razón, podría ser útil un nuevo estimador consistente de ϑ_{n^*} para definir nuevos estadísticos de contraste para medir la bondad de ajuste en modelos loglineales con datos multinomiales agrupados en clusters bajo sobredispersión sin la hipótesis de ninguna distribución subyacente en particular. En el siguiente capítulo se introducirán estos estadísticos de contraste que requerirían unas hipótesis más débiles en comparación con las utilizadas en el artículo de Brier.

3.2. Estimadores consistentes de Brier para ϑ_{n^*} y ρ^2

En esta sección se definen los estimadores consistentes no paramétricos para el efecto del diseño y para el coeficiente de correlación intracluster cuando los tamaños de cluster son desiguales. Además, en el caso en que un modelo loglineal haya sido validado, se define también el estimador semiparamétrico basado en medidas de divergencia.

Definición 3.2 *Los estimadores no paramétricos del efecto del diseño y el coeficiente de correlación intra-cluster en el muestreo por conglomerados con tamaños desiguales son respectivamente*

$$\tilde{\vartheta}_{\hat{n}^*, N} = \sum_{g=1}^G w_g \tilde{\vartheta}_{n_g}(\tilde{\mathbf{Y}}_g), \quad (3.11)$$

$$\tilde{\rho}_{\hat{n}^*, N}^2 = \frac{\tilde{\vartheta}_{\hat{n}^*, N} - 1}{\hat{n}^* - 1}, \quad (3.12)$$

donde

$$\hat{n}^* = \sum_{g=1}^G w_g n_g,$$

es un estimador consistente de n^* dado en (3.10) y

$$\begin{aligned} \tilde{\mathbf{Y}}_g &= ((\mathbf{Y}^{(g,1)})^T, \dots, (\mathbf{Y}^{(g,N_g)})^T)^T, \\ \tilde{\vartheta}_{n_g}(\tilde{\mathbf{Y}}_g) &= \frac{X^2(\tilde{\mathbf{Y}}_g)}{(N_g - 1)(M - 1)}, \\ X^2(\tilde{\mathbf{Y}}_g) &= n_g \sum_{\ell=1}^{N_g} \sum_{r=1}^M \frac{(\hat{p}_r^{(\ell,g)} - \hat{p}_r^{(g)})^2}{\hat{p}_r^{(g)}}, \end{aligned}$$

con $g = 1, \dots, G$.

Teorema 3.2 *Los estimadores definidos en (3.11) y (3.12), son estimadores consistentes, es decir*

$$\tilde{\vartheta}_{\hat{n}^*, N} \xrightarrow[N \rightarrow \infty]{P} \vartheta_{n^*}, \quad \tilde{\rho}_{\hat{n}^*, N}^2 \xrightarrow[N \rightarrow \infty]{P} \rho^2.$$

Demostración Para un tamaño de cluster específico n_g , aplicando el Teorema 2.4 se tiene que

$$\tilde{\vartheta}_{n_g}(\tilde{\mathbf{Y}}_g) = \frac{X^2(\tilde{\mathbf{Y}}_g)}{(N_g - 1)(M - 1)} \xrightarrow[N_g \rightarrow \infty]{P} \vartheta_{n_g},$$

luego, teniendo en cuenta (3.8) se sigue que

$$\tilde{\vartheta}_{\hat{n}^*, N} = \sum_{g=1}^G w_g \tilde{\vartheta}_{n_g}(\tilde{\mathbf{Y}}_g) \xrightarrow[N_1, \dots, N_G \rightarrow \infty]{P} \sum_{g=1}^G w_g^* \vartheta_{n_g} = \vartheta_{n^*}.$$

Por otra parte,

$$\hat{n}^* = \sum_{g=1}^G w_g n_g \xrightarrow[N_1, \dots, N_G \rightarrow \infty]{P} \sum_{g=1}^G w_g^* n_g = n^*,$$

luego teniendo en cuenta la definición (3.9)

$$\tilde{\rho}_{\hat{n}^*, N}^2 = \frac{\tilde{\vartheta}_{\hat{n}^*, N} - 1}{\hat{n}^* - 1} \xrightarrow[N_1, \dots, N_G \rightarrow \infty]{P} \frac{\vartheta_{n^*} - 1}{n^* - 1} = \rho^2.$$

■

Además $\tilde{\vartheta}_{n_g, \hat{n}^*, N} = 1 + \tilde{\rho}_{\hat{n}^*, N}^2(n_g - 1)$ es un estimador consistente alternativo para ϑ_{n_g} , $g = 1, \dots, G$, pero que requiere el cálculo previo de los estimadores de ϑ_{n^*} y ρ^2 , es decir, (3.11) y (3.12) respectivamente.

Con el nuevo parámetro, ϑ_{n^*} , definido para medir la sobredispersión en caso de tamaños desiguales, en el caso de que un modelo loglineal haya sido validado, es posible obtener el estimador de mínima ϕ -divergencia de $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_\phi = \hat{\boldsymbol{\theta}}_\phi(\mathbf{Y})$, cuyo comportamiento asintótico se da en el siguiente teorema.

Teorema 3.3 *Para un modelo loglineal con muestreo por conglomerados y tamaños desiguales se tiene*

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_\phi - \boldsymbol{\theta}_0) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_{M_0}, \frac{\vartheta_{n^*}}{\bar{n}} (\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W})^{-1}) \quad (3.13)$$

y

$$\sqrt{N}(\mathbf{p}(\hat{\boldsymbol{\theta}}_\phi) - \mathbf{p}(\boldsymbol{\theta}_0)) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_M, \frac{\vartheta_{n^*}}{\bar{n}} \mathbf{W} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} (\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W})^{-1} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W}^T). \quad (3.14)$$

Demostración En el Teorema 3.1 se demostró que

$$\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_M, \frac{\vartheta_{n^*}}{\bar{n}} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}).$$

Por otra parte, en el Teorema 2.3 se estableció que, en un modelo loglineal, el estimador de $MC\phi$, $\widehat{\boldsymbol{\theta}}_\phi$, admite la siguiente expresión,

$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_\phi - \boldsymbol{\theta}_0) = \mathbf{I}_F^{-1}(\boldsymbol{\theta}_0) \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) + o(\mathbf{1}_{M_0}),$$

luego

$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_\phi - \boldsymbol{\theta}_0) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_{M_0}, \boldsymbol{\Sigma}^*),$$

con

$$\begin{aligned} \boldsymbol{\Sigma}^* &= \mathbf{I}_F^{-1}(\boldsymbol{\theta}_0) \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \frac{\vartheta_{n^*}}{\bar{n}} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W} (\mathbf{I}_F^{-1}(\boldsymbol{\theta}_0))^T \\ &= \frac{\vartheta_{n^*}}{n} (\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W})^{-1}. \end{aligned}$$

Así, se tiene (3.13).

Puesto que la función $\mathbf{p} : \Theta \rightarrow \Delta_M$ es totalmente diferenciable en $\boldsymbol{\theta}_0$, es decir existen las derivadas parciales de $p_r(\boldsymbol{\theta})$ con respecto a θ_g en $\boldsymbol{\theta}_0$ y $p_r(\boldsymbol{\theta})$ tiene una aproximación lineal en $\boldsymbol{\theta}_0$ dada por

$$p_r(\boldsymbol{\theta}) = p_r(\boldsymbol{\theta}_0) + \sum_{g=1}^{M_0} (\theta_g - \theta_{0g}) \frac{\partial p_r(\boldsymbol{\theta}_0)}{\partial \theta_g} + o(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|),$$

y puesto que

$$\mathbf{J}(\boldsymbol{\theta}) = \left(\frac{\partial p_r(\boldsymbol{\theta})}{\partial \theta_g} \right)_{\substack{r=1, \dots, M \\ g=1, \dots, M_0}} = \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})} \mathbf{W},$$

para $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_\phi$ se obtiene

$$\sqrt{N}(\mathbf{p}(\widehat{\boldsymbol{\theta}}_\phi) - \mathbf{p}(\boldsymbol{\theta}_0)) = \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W} \sqrt{N}(\widehat{\boldsymbol{\theta}}_\phi - \boldsymbol{\theta}_0) + o(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|),$$

luego, utilizando la primera parte de este teorema, resultado (3.13), se tiene que

$$\sqrt{N} \left(\mathbf{p}(\hat{\boldsymbol{\theta}}_\phi) - \mathbf{p}(\boldsymbol{\theta}_0) \right) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_{M_0}, \frac{\vartheta_{n^*}}{n} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W} (\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W})^{-1} \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}).$$

Se obtiene por tanto (3.14). ■

3.3. Nuevos estimadores no-paramétricos y semiparamétricos para ϑ_{n^*} y ρ^2

En esta sección se definen nuevos estimadores consistentes para el parámetro de sobre-dispersión y el coeficiente de correlación intracluster. Se presentan dos casos dependiendo de si existen grupos de cluster del mismo tamaño. En el primer caso, se considera la situación en la que existen grupos de cluster, se calculan los estimadores para cada grupo de cluster del mismo tamaño, y el estimador global se define como una media ponderada de dichos estimadores. En el segundo caso se considera la situación en la que los clusters son todos de tamaño diferente. Esto ocurre habitualmente cuando los tamaños de los clusters son grandes

3.3.1. Caso 1: Existen grupos de clusters con tamaños iguales.

Sea $\tilde{\mathbf{Y}} = (\tilde{\mathbf{Y}}_1, \dots, \tilde{\mathbf{Y}}_G)^T$, el vector que contiene la muestra completa de observaciones, de forma que la dimensión de $\tilde{\mathbf{Y}}_g$ es $N_g > 1$. Es posible proponer un nuevo estimador consistente semiparamétrico de ϑ_{n^*} , con mayor tasa de convergencia que el introducido por Brier, utilizando la información de la muestra completa ($\tilde{\mathbf{Y}}$) para ponderar las diferencias entre las frecuencias estimadas en cada grupo de clusters de igual tamaño.

Definición 3.3 *Se denominan estimadores modificados de Brier de ϑ_{n^*} y ρ^2 en el muestreo por conglomerados con tamaños desiguales respectivamente a*

$$\tilde{\vartheta}_{\hat{n}^*, N, \bullet} = \sum_{g=1}^G w_g \tilde{\vartheta}_{n_g}(\tilde{\mathbf{Y}}_g, \tilde{\mathbf{Y}}), \quad (3.15)$$

$$\tilde{\rho}_{\hat{n}^*, N, \bullet}^2 = \frac{\tilde{\vartheta}_{\hat{n}^*, N, \bullet} - 1}{\hat{n}^* - 1}, \quad (3.16)$$

donde

$$\begin{aligned} \tilde{\vartheta}_{n_g}(\tilde{\mathbf{Y}}_g, \tilde{\mathbf{Y}}) &= \frac{X^2(\tilde{\mathbf{Y}}_g, \tilde{\mathbf{Y}})}{(N_g - 1)(M - 1)}, \\ X^2(\tilde{\mathbf{Y}}_g, \tilde{\mathbf{Y}}) &= n_g \sum_{\ell=1}^{N_g} \sum_{r=1}^M \frac{(\hat{p}_r^{(\ell, g)} - \hat{p}_r^{(g)})^2}{\hat{p}_r} = n_g \sum_{r=1}^M \frac{1}{\hat{p}_r} \sum_{\ell=1}^{N_g} (\hat{p}_r^{(\ell, g)} - \hat{p}_r^{(g)})^2, \end{aligned}$$

en lugar de $X^2(\tilde{\mathbf{Y}}_g)$ y $\tilde{\vartheta}_{n_g}(\tilde{\mathbf{Y}}_g)$ respectivamente, $g = 1, \dots, G$.

Teorema 3.4 *Los estimadores definidos en (3.15) y (3.16), son estimadores consistentes, es decir*

$$\tilde{\vartheta}_{\hat{n}^*, N, \bullet} \xrightarrow[N \rightarrow \infty]{P} \vartheta_{n^*}, \quad \tilde{\rho}_{\hat{n}^*, N, \bullet}^2 \xrightarrow[N \rightarrow \infty]{P} \rho^2.$$

Demostración Para un tamaño de cluster específico n_g , aplicando el Teorema 2.4 se tiene que

$$\tilde{\vartheta}_{n_g}(\tilde{\mathbf{Y}}_g, \tilde{\mathbf{Y}}) = \frac{X^2(\tilde{\mathbf{Y}}_g, \tilde{\mathbf{Y}})}{(N_g - 1)(M - 1)} \xrightarrow[N_g \rightarrow \infty]{P} \vartheta_{n_g},$$

luego, teniendo en cuenta (3.8) se sigue que

$$\tilde{\vartheta}_{\hat{n}^*, N, \bullet} = \sum_{g=1}^G w_g \tilde{\vartheta}_{n_g}(\tilde{\mathbf{Y}}_g, \tilde{\mathbf{Y}}) \xrightarrow[N_1, \dots, N_G \rightarrow \infty]{P} \sum_{g=1}^G w_g^* \vartheta_{n_g} = \vartheta_{n^*}.$$

Por otra parte, como

$$\hat{n}^* \xrightarrow[N \rightarrow \infty]{P} n^*,$$

se tiene

$$\tilde{\rho}_{\hat{n}^*, N, \bullet}^2 = \frac{\tilde{\vartheta}_{\hat{n}^*, N, \bullet} - 1}{\hat{n}^* - 1} \xrightarrow[N_1, \dots, N_G \rightarrow \infty]{P} \frac{\vartheta_{n^*} - 1}{n^* - 1} = \rho^2.$$

■

Además, si los datos se ajustan a un modelo loglineal, se puede proponer un nuevo estimador consistente semiparamétrico de ϑ_{n_g} incluso con una mayor tasa de convergencia.

Definición 3.4 *Se denominan estimadores basados en ϕ -divergencias de ϑ_{n^*} y ρ^2 en el muestreo por conglomerados con tamaños desiguales cuando un modelo loglineal ha sido validado a*

$$\tilde{\vartheta}_{\hat{n}^*, N, \lambda} = \sum_{g=1}^G w_g \tilde{\vartheta}_{n_g}(\tilde{\mathbf{Y}}_g, \hat{\boldsymbol{\theta}}_{\phi_\lambda}), \quad (3.17)$$

$$\tilde{\rho}_{\hat{n}^*, N, \lambda}^2 = \frac{\tilde{\vartheta}_{\hat{n}^*, N, \lambda} - 1}{\hat{n}^* - 1}, \quad (3.18)$$

respectivamente, donde

$$\begin{aligned} \tilde{\vartheta}_{n_g}(\tilde{\mathbf{Y}}_g, \hat{\boldsymbol{\theta}}_{\phi_\lambda}) &= \frac{X^2(\tilde{\mathbf{Y}}_g, \hat{\boldsymbol{\theta}}_{\phi_\lambda})}{(N_g - 1)(M - 1)}, \\ X^2(\tilde{\mathbf{Y}}_g, \hat{\boldsymbol{\theta}}_{\phi_\lambda}) &= n_g \sum_{\ell=1}^{N_g} \sum_{r=1}^M \frac{(\hat{p}_r^{(\ell, g)} - \hat{p}_r^{(g)})^2}{p_r(\hat{\boldsymbol{\theta}}_{\phi_\lambda})} = n_g \sum_{r=1}^M \frac{1}{p_r(\hat{\boldsymbol{\theta}}_{\phi_\lambda})} \sum_{\ell=1}^{N_g} (\hat{p}_r^{(\ell, g)} - \hat{p}_r^{(g)})^2, \end{aligned}$$

con $g = 1, \dots, G$.

Teorema 3.5 *Los estimadores definidos en (3.17) y (3.18), son estimadores consistentes, es decir*

$$\tilde{\vartheta}_{\hat{n}^*, N, \lambda} \xrightarrow[N \rightarrow \infty]{P} \vartheta_{n^*}, \quad \tilde{\rho}_{\hat{n}^*, N, \lambda}^2 \xrightarrow[N \rightarrow \infty]{P} \rho^2.$$

Demostración La demostración de este teorema es similar a la del teorema anterior sin más que sustituir $\tilde{\vartheta}_{n_g}(\tilde{\mathbf{Y}}_g, \tilde{\mathbf{Y}})$ y $X^2(\tilde{\mathbf{Y}}_g, \tilde{\mathbf{Y}})$ por $\tilde{\vartheta}_{n_g}(\tilde{\mathbf{Y}}_g, \hat{\boldsymbol{\theta}}_{\phi_\lambda})$ y $X^2(\tilde{\mathbf{Y}}_g, \hat{\boldsymbol{\theta}}_{\phi_\lambda})$, respectivamente. ■

En lo que sigue, se abreviará notación con $\tilde{\vartheta}_{n_g}$, $\tilde{\vartheta}_{n_g, \bullet}$, $\tilde{\vartheta}_{n_g, \lambda}$, denotando las tres versiones de $\tilde{\vartheta}_{n_g}(\tilde{\mathbf{Y}}_g)$, $\tilde{\vartheta}_{n_g}(\tilde{\mathbf{Y}}_g, \tilde{\mathbf{Y}})$, $\tilde{\vartheta}_{n_g}(\tilde{\mathbf{Y}}_g, \hat{\boldsymbol{\theta}}_{\phi_\lambda})$ respectivamente. A partir de cada uno de ellos se calcularán los correspondientes estimadores del efecto del diseño para el conjunto de los N

clusters, denotando por $\tilde{\vartheta}_{\hat{n}^*,N}$ al estimador de Brier, $\tilde{\vartheta}_{\hat{n}^*,N,\bullet}$, al estimador que se llamará Brier modificado y $\tilde{\vartheta}_{\hat{n}^*,N,\lambda}$, al estimador semiparamétrico y sus respectivos estimadores del coeficiente de correlación intracluster $\tilde{\rho}_{\hat{n}^*,N}^2$, $\tilde{\rho}_{\hat{n}^*,N,\bullet}^2$ y $\tilde{\rho}_{\hat{n}^*,N,\lambda}^2$.

3.3.2. Caso 2: Clusters de tamaño suficientemente grande

En el caso en el que el número de individuos en todos los clusters ($n_g \quad g = 1, \dots, G$) es grande, sin pérdida de generalidad, se puede suponer $N_g = 1$ y $G = N$. En esta situación se definirán nuevos estimadores del coeficiente de correlación intra-cluster y a partir de ellos se puede encontrar el estimador correspondiente del efecto del diseño mediante la relación $\tilde{\vartheta}_{\hat{n}^*,N,\lambda} = 1 + (\hat{n}^* - 1) \tilde{\rho}_{\hat{n}^*,N,\lambda}^2$.

Definición 3.5 *Se define el estimador modificado de Brier de ρ^2 en el muestreo por conglomerados con tamaños desiguales cuando $n_g, N \rightarrow \infty$ como*

$$\begin{aligned} \hat{\rho}^2 &= \frac{1}{(N-1)(M-1)} \sum_{\ell=1}^N \left(\hat{\mathbf{p}}^{(\ell)} - \frac{1}{N} \sum_{\ell=1}^N \hat{\mathbf{p}}^{(\ell)} \right)^T \mathbf{D}_{\hat{\mathbf{p}}}^{-1} \left(\hat{\mathbf{p}}^{(\ell)} - \frac{1}{N} \sum_{\ell=1}^N \hat{\mathbf{p}}^{(\ell)} \right) \\ &= \frac{1}{(N-1)(M-1)} \sum_{r=1}^M \frac{1}{\hat{p}_r} \sum_{\ell=1}^N \left(\hat{p}_r^{(\ell)} - \frac{1}{N} \sum_{\ell=1}^N \hat{p}_r^{(\ell)} \right)^2, \end{aligned} \quad (3.19)$$

y cuando un modelo loglineal haya sido validado se define el estimador semiparamétrico basado en ϕ -divergencias como

$$\begin{aligned} \hat{\rho}^2(\hat{\boldsymbol{\theta}}_{\phi_\lambda}) &= \frac{1}{(N-1)(M-1)} \sum_{\ell=1}^N \left(\hat{\mathbf{p}}^{(\ell)} - \frac{1}{N} \sum_{\ell=1}^N \hat{\mathbf{p}}^{(\ell)} \right)^T \mathbf{D}_{\mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_\lambda})}^{-1} \left(\hat{\mathbf{p}}^{(\ell)} - \frac{1}{N} \sum_{\ell=1}^N \hat{\mathbf{p}}^{(\ell)} \right) \\ &= \frac{1}{(N-1)(M-1)} \sum_{r=1}^M \frac{1}{p_r(\hat{\boldsymbol{\theta}}_{\phi_\lambda})} \sum_{\ell=1}^N \left(\hat{p}_r^{(\ell)} - \frac{1}{N} \sum_{\ell=1}^N \hat{p}_r^{(\ell)} \right)^2. \end{aligned} \quad (3.20)$$

Teorema 3.6 *Los estimadores definidos en (3.19) y (3.20), son estimadores consistentes de ρ^2 , es decir*

$$\hat{\rho}^2 \xrightarrow[n_g, N \rightarrow \infty]{P} \rho^2, \quad \hat{\rho}^2(\hat{\boldsymbol{\theta}}_{\phi_\lambda}) \xrightarrow[n_g, N \rightarrow \infty]{P} \rho^2. \quad (3.21)$$

Demostración Tal y como se vio en la expresión (2.24) del Capítulo 2 se verifica que

$$\lim_{n_g \rightarrow \infty} \frac{\vartheta_{n_g}}{n_g} = \rho^2,$$

luego para el estimador de la probabilidad en cada cluster se cumple

$$\widehat{\mathbf{p}}^{(g)} \xrightarrow[n_g \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{p}(\boldsymbol{\theta}_0), \rho^2 \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}),$$

lo que significa que $\widehat{\mathbf{p}}^{(1)}, \dots, \widehat{\mathbf{p}}^{(G)}$ son variables aleatorias M -dimensionales independientes e idénticamente distribuidas. Teniendo en cuenta argumentos similares a los utilizados para demostrar (3.3) se obtienen estimadores consistentes para ρ^2 cuando $n_g, N \rightarrow \infty$. Aplicando el Teorema Central del Límite, para el g -ésimo grupo, se sigue que

$$(\widehat{\mathbf{p}}^{(g)} - \mathbf{p}(\boldsymbol{\theta}_0)) \xrightarrow[n_g \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_M, \rho^2 \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}). \quad (3.22)$$

Multiplicando por $w_g = n_g / \sum_{h=1}^G n_h$,

$$w_g (\widehat{\mathbf{p}}^{(g)} - \mathbf{p}(\boldsymbol{\theta}_0)) \xrightarrow[n_g \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_M, w_g^2 \rho^2 \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}),$$

sumando desde $g = 1$ a G y teniendo en cuenta la independencia entre los clusters

$$\sum_{g=1}^G w_g (\widehat{\mathbf{p}}^{(g)} - \mathbf{p}(\boldsymbol{\theta}_0)) = (\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) \xrightarrow[n_g \rightarrow \infty, g=1, \dots, G]{\mathcal{L}} \mathcal{N}\left(\mathbf{0}_M, \sum_{g=1}^G w_g^2 \rho^2 \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}\right).$$

dividiendo en el numerador y denominador de (3.5) por N , se obtiene (3.21). ■

3.4. Aplicaciones numéricas

3.4.1. Datos de Brier sobre satisfacción de los vecinos con su vivienda y vecindario

Brier estudió datos de 20 vecindarios en Montevideo (Minnesota, EEUU), para analizar la relación entre la satisfacción con el vecindario y el grado de satisfacción con la vivienda particular. Para ello, obtuvo una muestra formada por $N = 20$ clusters (vecindarios), de forma que en $N_1 = 18$ de ellos se observan $n_1 = 5$ casas y en $N_2 = 2$ $n_2 = 3$ casas. Se distingue así dos grupos de vecindarios muestreados, $g = 1$ ó $g = 2$, dependiendo de que el número de casas seleccionado en cada vecindario (tamaño del cluster) sea $n_1 = 5$ o $n_2 = 3$, respectivamente. Para el vecindario ℓ -ésimo ($\ell = 1, \dots, N_g$) del g -ésimo tamaño de cluster, en el s -ésimo hogar seleccionado ($s = 1, \dots, n_g$), se le preguntaba al vecino por su grado de satisfacción con el vecindario ($X_{1s}^{(g,\ell)}$) y con su vivienda en particular ($X_{2s}^{(g,\ell)}$). Para las dos cuestiones, las respuestas posibles se codificaron en Insatisfecho (US), Satisfecho (S) o Muy Satisfecho (VS). Las anteriores categorías se identificarán a partir de ahora con números ordinales (1, 2, y 3) para las variables, $X_1^{(g,\ell)}$ y $X_2^{(g,\ell)}$, por orden lexicográfico: por ejemplo, (US, S) se asocia con $(X_1^{(g,\ell)}, X_2^{(g,\ell)}) = (1, 2)$ y aparece en segunda posición.

Bajo la hipótesis de que la clasificación de las familias de acuerdo a la satisfacción con la vivienda propia (X_2) es independiente de su clasificación de acuerdo al nivel de satisfacción con el vecindario (X_1), se plantea el modelo loglineal para la tabla de contingencia $I \times J$ con $I = J = 3$. En la Tabla 3.1 se presentan los valores $\mathbf{Y}^{(g,\ell)} = (Y_{11}^{(g,\ell)}, Y_{12}^{(g,\ell)}, \dots, Y_{33}^{(g,\ell)})^T$, cada uno de estos vectores se corresponde a la fila, (ℓ) de un cluster g para $g = 1, 2$ y $\ell = 1, \dots, N_g$.

En la última fila de la Tabla 3.1 aparecen los datos totales de la tabla de contingencia que recoge la distribución de los 96 hogares muestreados, que se expresa en la Tabla 3.2 como tabla de doble entrada donde $Y_{ij} = \sum_{\ell=1}^N Y_{ij}^{(\ell)}$.

	(US, US)	(US, S)	(US, VS)	(S, US)	(S, S)	(S, VS)	(VS, US)	(VS, S)	(VS, VS)	
(g, ℓ)	$Y_{11}^{(g, \ell)}$	$Y_{12}^{(g, \ell)}$	$Y_{13}^{(g, \ell)}$	$Y_{21}^{(g, \ell)}$	$Y_{22}^{(g, \ell)}$	$Y_{23}^{(g, \ell)}$	$Y_{31}^{(g, \ell)}$	$Y_{32}^{(g, \ell)}$	$Y_{33}^{(g, \ell)}$	n_g
(1, 1)	1	0	0	2	2	0	0	0	0	5
(1, 2)	1	0	0	2	2	0	0	0	0	5
(1, 3)	0	2	0	0	2	0	0	1	0	5
(1, 4)	0	1	0	2	1	0	1	0	0	5
(1, 5)	0	0	0	0	4	0	0	1	0	5
(1, 6)	1	0	0	3	1	0	0	0	0	5
(1, 7)	3	0	0	0	1	0	0	1	0	5
(1, 8)	1	0	0	1	3	0	0	0	0	5
(1, 9)	3	0	0	0	0	0	1	0	1	5
(1, 10)	0	1	0	0	3	1	0	0	0	5
(1, 11)	1	1	0	0	2	0	1	0	0	5
(1, 12)	0	1	0	4	0	0	0	0	0	5
(1, 13)	0	0	0	4	1	0	0	0	0	5
(1, 14)	0	0	0	1	2	0	0	0	2	5
(1, 15)	2	0	0	2	1	0	0	0	0	5
(1, 16)	0	0	0	1	1	1	0	2	0	5
(1, 17)	2	0	0	2	1	0	0	0	0	5
(1, 18)	2	0	0	2	0	0	1	0	0	5
(2, 1)	1	0	0	1	1	0	0	0	0	3
(2, 2)	0	0	0	1	0	1	0	0	1	3
Total	18	6	0	28	28	3	4	5	4	$n = 96$

Tabla 3.1. Satisfacción con la vivienda y el vecindario en Montevideo (Brier, 1980).

Tabla 3.2. Número de vecinos según Grado de Satisfacción Comunitario e individual.

	X_2		
X_1	US	S	VS
US	18	6	0
S	28	28	3
VS	4	5	4

Para estos datos, se considera el siguiente modelo loglineal en el caso de independencia entre las dos variables estudiadas

$$p_{ij}(\boldsymbol{\theta}) = \exp\{u + \theta_{1(i)} + \theta_{2(j)}\}, \quad \text{con } i, j = 1, 3, \quad (3.23)$$

donde el efecto de la variable satisfacción con el vecindario (X_1) viene representado por $\theta_{1(i)}$, y con su vivienda en particular (X_2), mediante $\theta_{2(j)}$. La matriz de diseño y el vector de parámetros desconocidos son

$$\mathbf{W} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 0 & 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 0 & -1 & 1 & 0 & -1 & 1 & 0 & -1 \\ 0 & 1 & -1 & 0 & 1 & -1 & 0 & 1 & -1 \end{pmatrix}^T \quad \text{y } \boldsymbol{\theta} = (\theta_{1(1)}, \theta_{1(2)}, \theta_{2(1)}, \theta_{2(2)})^T.$$

Bajo el modelo loglineal de independencia los parámetros de interés,

$$\boldsymbol{\beta} = \begin{pmatrix} \rho^2 \\ \mathbf{p}(\boldsymbol{\theta}) \end{pmatrix},$$

se estiman por

$$\widehat{\boldsymbol{\beta}}_{\widehat{n}^*, N, \lambda} = \begin{pmatrix} \widehat{\rho}_{\widehat{n}^*, N, \lambda}^2 \\ \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_\lambda}) \end{pmatrix} = \begin{pmatrix} \frac{\widetilde{\vartheta}_{\widehat{n}^*, N, \lambda} - 1}{\widehat{n}^* - 1} \\ \frac{\exp\{\mathbf{W}\widehat{\boldsymbol{\theta}}_{\phi_\lambda}\}}{\mathbf{1}_M^T \exp\{\mathbf{W}\widehat{\boldsymbol{\theta}}_{\phi_\lambda}\}} \end{pmatrix},$$

donde $\widetilde{\vartheta}_{\widehat{n}^*, N}$ se calcula según (3.15), y $\widehat{\boldsymbol{\theta}}_{\phi_\lambda}$ utilizando la familia de ϕ -divergencias de Cressie y Read definidas en (2.4). Para este modelo, los estimadores obtenidos para los parámetros aparecen en la Tabla 3.3.

A partir de estas estimaciones se calculan las probabilidades estimadas con el modelo loglineal planteado y su error estándar, así como los estimadores del coeficiente de correlación intracluster y en la última columna, de la Tabla 3.4, se recoge la estimación no

Tabla 3.3. Estimadores para diferentes valores de λ .

λ	$-\frac{1}{2}$	0	$\frac{2}{3}$	1	2
$\widehat{\theta}_{1(1)}$	-0.1038	-0.0950	-0.0879	-0.0831	-0.0641
$\widehat{\theta}_{1(2)}$	0.8816	0.8040	0.7385	0.7160	0.6758
$\widehat{\theta}_{2(1)}$	0.8880	0.7381	0.6328	0.5978	0.5221
$\widehat{\theta}_{2(2)}$	0.6474	0.4896	0.3852	0.3527	0.2961

paramétrica donde el estimador del coeficiente de correlación intracluster es el de Brier ($\widetilde{\rho}_{n^*,N}^2$) y entre paréntesis el que hemos llamado Brier modificado ($\widetilde{\rho}_{n^*,N,\bullet}^2$).

Tabla 3.4. Probabilidades estimadas para diferentes valores de λ .

λ	$-\frac{1}{2}$	0	$\frac{2}{3}$	1	2	<i>No paramétrico</i>
$\widetilde{\rho}_{n^*,N,\lambda}^2$	0.3109	0.1545	0.0872	0.0712	0.0477	0.0199(0.0172)
$\widehat{p}_1(\boldsymbol{\theta})$	0.1273 (0.0387)	0.1302 (0.03332)	0.1315 (0.0303)	0.1318 (0.0295)	0.1321 (0.0283)	0.1875 (0.0398)
$\widehat{p}_2(\boldsymbol{\theta})$	0.1001 (0.0322)	0.1016 (0.0276)	0.1026 (0.0253)	0.1032 (0.0248)	0.1054 (0.0241)	0.0625 (0.0247)
$\widehat{p}_3(\boldsymbol{\theta})$	0.0112 (0.0082)	0.0182 (0.0093)	0.0252 (0.0104)	0.0280 (0.0108)	0.0346 (0.0119)	0 0
$\widehat{p}_4(\boldsymbol{\theta})$	0.3411 (0.0616)	0.3201 (0.0511)	0.3005 (0.0456)	0.2932 (0.0440)	0.2770 (0.0414)	0.2917 (0.0464)
$\widehat{p}_5(\boldsymbol{\theta})$	0.2682 (0.0564)	0.2497 (0.0464)	0.2346 (0.0411)	0.2295 (0.0397)	0.2209 (0.0375)	0.2917 (0.0464)
$\widehat{p}_6(\boldsymbol{\theta})$	0.0302 (0.0207)	0.0448 (0.0210)	0.0577 (0.0214)	0.0623 (0.0216)	0.0725 (0.0222)	0.0312 (0.0178)
$\widehat{p}_7(\boldsymbol{\theta})$	0.0649 (0.0278)	0.0705 (0.0245)	0.0749 (0.0229)	0.0761 (0.0224)	0.0764 (0.0215)	0.0417 (0.0204)
$\widehat{p}_8(\boldsymbol{\theta})$	0.0510 (0.0225)	0.0550 (0.0198)	0.0585 (0.0322)	0.0595 (0.0322)	0.0610 (0.0322)	0.0521 (0.0227)
$\widehat{p}_9(\boldsymbol{\theta})$	0.0057 (0.0045)	0.0099 (0.0055)	0.0144 (0.0066)	0.0162 (0.0070)	0.0200 (0.0078)	0.0417 (0.0204)

3.4.2. Estudio sobre datos de ADN para diferentes genotipos (Weir and Hill, 2002)

En la FBI Laboratory Division Publication, se publicaron por Budowle and Moretti (1999) [12], los ficheros de datos de diferentes genotipos para seis subpoblaciones de Estados Unidos. Se recogieron las frecuencias de aparición para los 13 alelos de uso más común en el sistema de genética forense CODIS (Combined DNA Index System): D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820, CSF1PO, TPOX, THO1 y D16S539. Para los primeros cuatro locus, las frecuencias de sus alelos en los seis clusters, Americanos Africanos (1), Caucasianos (2), Hispanos (3), Bahameños (4), Jamaicanos (5), y Trinitarios (6), se resumen en las siguientes Tablas 3.5-3.8.

ℓ	$Y_1^{(\ell)}$	$Y_2^{(\ell)}$	$Y_3^{(\ell)}$	$Y_4^{(\ell)}$	$Y_5^{(\ell)}$	$Y_6^{(\ell)}$	$Y_7^{(\ell)}$	$Y_8^{(\ell)}$	n_ℓ
1	1	5	37	86	99	62	19	2	311
2	0	0	53	85	85	79	63	2	367
3	0	1	28	150	100	49	33	6	367
4	0	0	21	88	96	59	19	1	284
5	2	5	19	95	81	64	15	2	283
6	0	0	8	48	42	32	17	0	147

Tabla 3.5. Frecuencias de los alelos: 12, 13, 14, 15, 16, 17, 18, 19 (M=8) para el locus D3S1358 separados por subpoblaciones.

ℓ	$Y_1^{(\ell)}$	$Y_2^{(\ell)}$	$Y_3^{(\ell)}$	$Y_4^{(\ell)}$	$Y_5^{(\ell)}$	$Y_6^{(\ell)}$	$Y_7^{(\ell)}$	$Y_8^{(\ell)}$	$Y_9^{(\ell)}$	$Y_{10}^{(\ell)}$	n_ℓ
1	0	2	21	76	84	60	37	22	9	0	311
2	0	1	35	41	78	97	79	32	4	0	367
3	1	19	25	127	89	73	28	5	0	0	367
4	3	8	16	43	74	59	51	23	7	0	284
5	1	1	19	62	81	53	42	15	7	2	283
6	1	1	13	18	44	39	21	7	3	0	147

Tabla 3.6. Frecuencias de los alelos: 11, 13, 14, 15, 16, 17, 18, 19, 20, 21 (M=10) para el locus vWA, separados por subpoblaciones.

ℓ	$Y_1^{(\ell)}$	$Y_2^{(\ell)}$	$Y_3^{(\ell)}$	$Y_4^{(\ell)}$	$Y_5^{(\ell)}$	$Y_6^{(\ell)}$	$Y_7^{(\ell)}$	$Y_8^{(\ell)}$	$Y_9^{(\ell)}$	$Y_{10}^{(\ell)}$	$Y_{11}^{(\ell)}$	$Y_{12}^{(\ell)}$	$Y_{13}^{(\ell)}$	n_ℓ
1	3	16	25	38	74	36	59	33	12	7	6	1	1	311
2	12	19	54	65	68	58	54	26	7	4	0	0	0	367
3	1	30	27	45	67	52	44	55	32	13	1	0	0	367
4	0	17	22	31	42	51	60	30	10	16	2	2	0	283
5	1	19	15	18	61	61	42	32	9	16	6	3	0	283
6	2	8	14	15	25	23	30	17	6	3	2	1	1	147

Tabla 3.7. Frecuencias de los alelos para el locus FGA separado por subpoblaciones: 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 (M=13).

ℓ	$Y_1^{(\ell)}$	$Y_2^{(\ell)}$	$Y_3^{(\ell)}$	$Y_4^{(\ell)}$	$Y_5^{(\ell)}$	$Y_6^{(\ell)}$	$Y_7^{(\ell)}$	$Y_8^{(\ell)}$	$Y_9^{(\ell)}$	$Y_{10}^{(\ell)}$	$Y_{11}^{(\ell)}$	n_ℓ
1	1	2	6	12	32	72	104	65	14	3	0	311
2	7	4	38	19	53	127	75	40	3	1	0	367
3	1	1	34	24	41	117	90	46	10	3	0	367
4	0	1	6	16	33	54	93	55	19	7	0	284
5	0	2	3	11	32	60	89	59	25	1	1	283
6	1	0	7	11	22	35	36	26	9	0	0	147

Tabla 3.8. Frecuencias de los alelos para el locus D8S1179 separados por subpoblaciones: 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 (M=11).

Weir y Hill (2002) propusieron la estimación de ρ^2 a partir del primer y segundo momento de la distribución de frecuencias de los alelos, basándose en el método de los momentos. La expresión del estimador viene dada por

$$\bar{\rho}^2 = \frac{\sum_{r=1}^M (MSP_r - MSG_r)}{\sum_{r=1}^M (MSP_r + (\bar{\eta} - 1)MSG_r)}, \quad (3.24)$$

donde

$$\begin{aligned} \bar{\eta} &= \frac{1}{(N-1)(\sum_{\ell=1}^N n_{\ell})} \left(\left(\sum_{\ell=1}^N n_{\ell} \right)^2 - \sum_{\ell=1}^N n_{\ell}^2 \right), \\ MSP_r &= \frac{1}{N-1} \sum_{\ell=1}^N n_{\ell} (\hat{p}_r^{(\ell)} - \hat{p}_r)^2, \\ MSG_r &= \frac{1}{\sum_{\ell=1}^N n_{\ell} - N} \sum_{\ell=1}^N n_{\ell} \hat{p}_r^{(\ell)} (1 - \hat{p}_r^{(\ell)}). \end{aligned}$$

En la Tabla 3.9 se muestran las estimaciones de \mathbf{p} y ρ^2 para los locus D3S1358, vWA, FGA y D8S1179. Se observa que las estimaciones del coeficiente de correlación intracluster calculadas por los dos métodos, el nuevo propuesto en esta memoria, $\hat{\rho}^2$, y el propuesto por Weir Hill, $\bar{\rho}^2$, son similares. Estos dos estimadores se compararán en la siguiente sección mediante un estudio de simulación.

Tabla 3.9. Estimaciones de \mathbf{p} y ρ^2 para locus D3S1358, vWA, FGA y D8S1179.

locus	D3S1358	vWA	FGA	D8S1179
$\widehat{\rho}^2$	0.00398	0.0041	0.0035	0.00398
$\overline{\rho}^2$	0.0116	0.0107	0.0051	0.0138
\widehat{p}_1	0.0017	0.0034	0.0108	0.0057
\widehat{p}_2	0.0062	0.0181	0.0620	0.0057
\widehat{p}_3	0.0944	0.0733	0.0892	0.0534
\widehat{p}_4	0.3138	0.2086	0.1205	0.0529
\widehat{p}_5	0.2859	0.2558	0.1916	0.1211
\widehat{p}_6	0.1961	0.2166	0.1597	0.2643
\widehat{p}_7	0.0944	0.1467	0.1643	0.2769
\widehat{p}_8	0.0074	0.0591	0.1097	0.1654
\widehat{p}_9	—	0.0170	0.0432	0.0459
\widehat{p}_{10}	—	0.0011	0.0335	0.0085
\widehat{p}_{11}	—	—	0.0097	0.0006
\widehat{p}_{12}	—	—	0.00398	0.01819
\widehat{p}_{13}	—	—	0.00113	—

3.5. Estudio de simulación

Para analizar el comportamiento de los nuevos estimadores del coeficiente de correlación intracluster en el caso de tamaños desiguales, a continuación se realizan dos estudios de simulación Monte Carlo, uno para el caso de tamaños desiguales y con $N_g > 1, g = 1, \dots, G$, y el siguiente en el que todos los clusters se suponen de diferente tamaño y además estos son grandes (n_g suficientemente grande $N_g \geq 1$).

Caso 1: $N_g > 1, g = 1, \dots, G$. A partir de los datos del estudio de satisfacción con la vivienda, se generan $R = 10,000$ repeticiones de muestras con $G = 3$ grupos con $N_1 = 18, N_2 = 2, N_3 = 5$, clusters en cada grupo, en los que los tamaños son $n_1 = 5, n_2 = 3, n_3 = 7$, respectivamente, utilizando valores del coeficiente de correlación intracluster de 0 a 0.95. Puesto que, el espacio paramétrico de ρ^2 es $[0, 1]$, los estimadores serán truncados a valores dentro de dicho intervalo. Para generar las muestras se utilizarán las tres distribuciones: Dirichlet-multinomial (DM), la distribución Agrupada Aleatoriamente (RC) y la Multinomial n-inflada (NI). El modelo loglineal utilizado se basa en la hipótesis de independencia con valores del parámetro $\boldsymbol{\theta} = (\theta_{1(1)}, \theta_{1(2)}, \theta_{2(1)}, \theta_{2(2)})^T = (0, 1, 0, 2, 0, 4, 0, 3)^T$. Los resultados de estas simulaciones aparecen reflejados en las gráficas que se comentarán a continuación.

Primero, en las Figuras 3.1, 3.2 y 3.3 se compara el comportamiento de los estimadores semiparamétricos, $EMC\phi_s$, estudiando la $RECM$ y el sesgo de $\tilde{\rho}_{n^*, N, \lambda}^2$ para $\lambda \in \{-0.5, 0, \frac{2}{3}, 1, 2\}$, para decidir cual de ellos es el mejor. Se observa que, salvo para el estimador calculado con $\lambda = -0.5$, el comportamiento es muy similar en cuanto a su $RECM$. El estimador de máxima cuasiverosimilitud ($\lambda = 0$) tiene mayor $RECM$ que el resto para valores del coeficiente de correlación intracluster hasta 0.8, mientras que, para valores superiores a 0.8 tiene menor $RECM$. Con respecto al comportamiento del sesgo, se confirma el mal comportamiento del $EMC\phi$ calculado con $\lambda = -0.5$, puesto que presenta el mayor sesgo medio, siendo este casi siempre positivo. Para los estimadores obtenidos con $\lambda \in \{\frac{2}{3}, 1, 2\}$ el sesgo medio siempre es negativo, mientras que para el estimador de máx-

ima verosimilitud el sesgo medio es casi siempre positivo, pero con valores muy próximos a cero. Se comprueba que, sea cual sea la distribución utilizada para generar los datos, el estimador que presenta menor *RECM* se obtiene con $\lambda = \frac{2}{3}$, debido a que el sesgo es menor.

Figura 3.1. RECM y sesgo para los EMC ϕ de ρ^2 para datos generados con la Dirichlet multinomial.

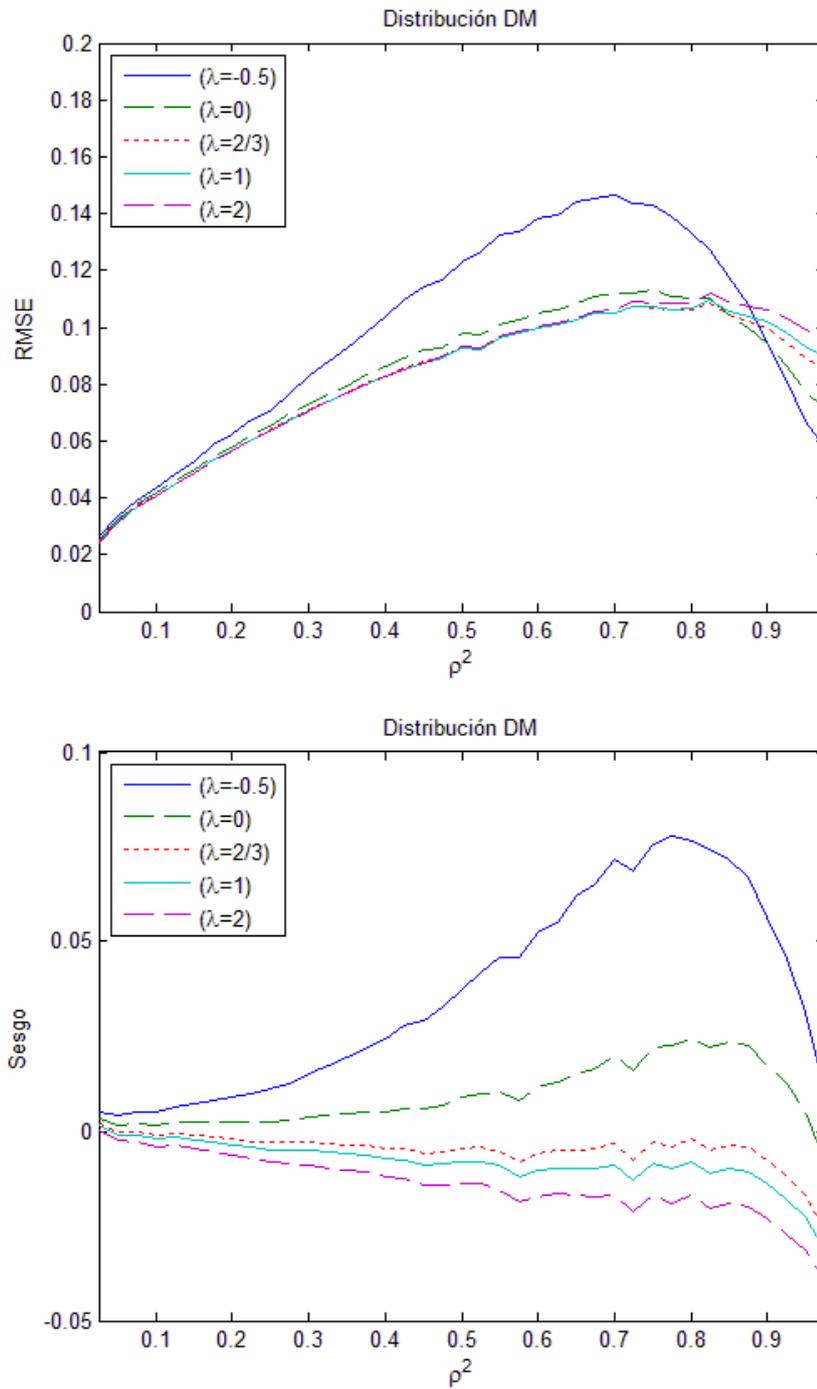


Figura 3.2. RECM y sesgo para los EMC ϕ de ρ^2 para datos generados con la distribución multinomial n-inflada.

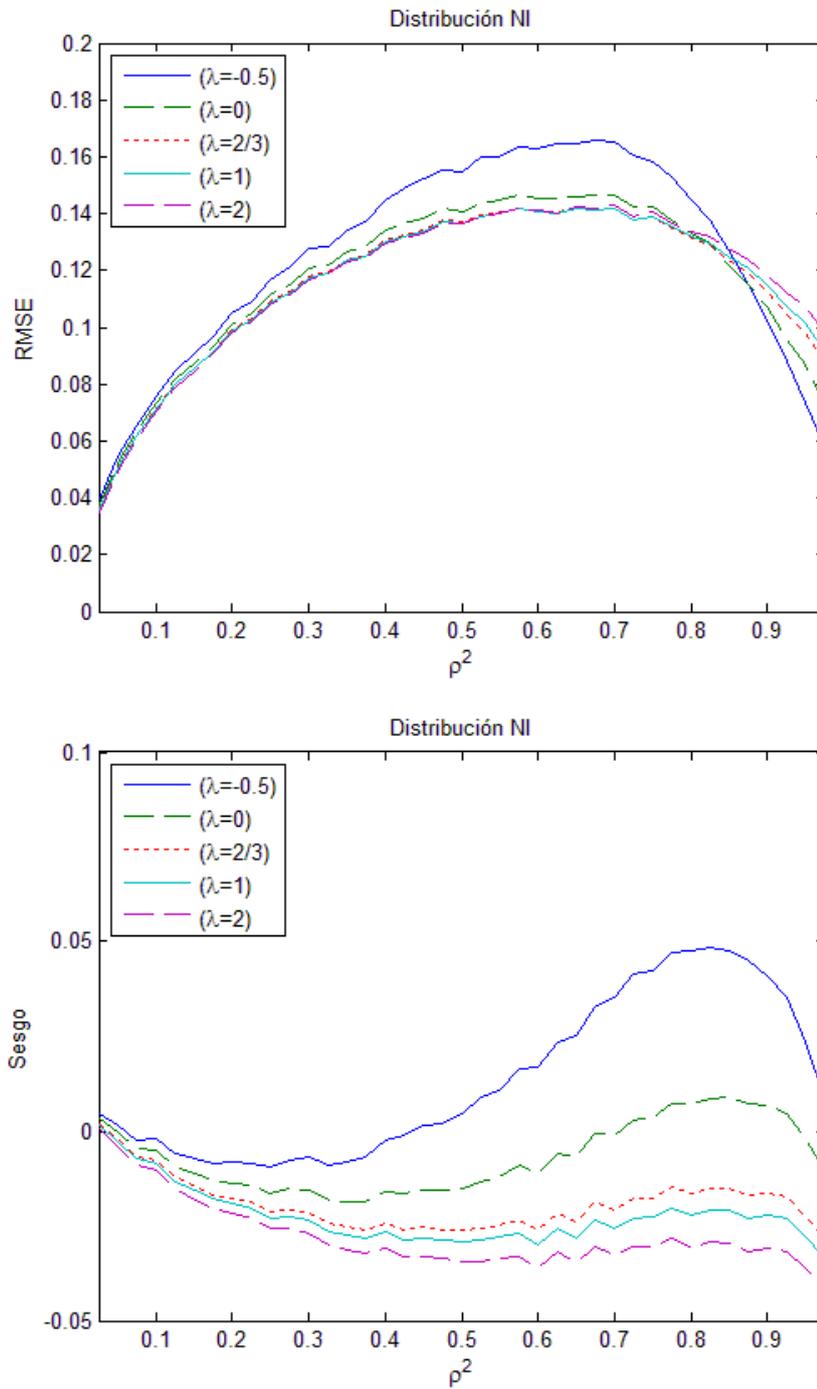
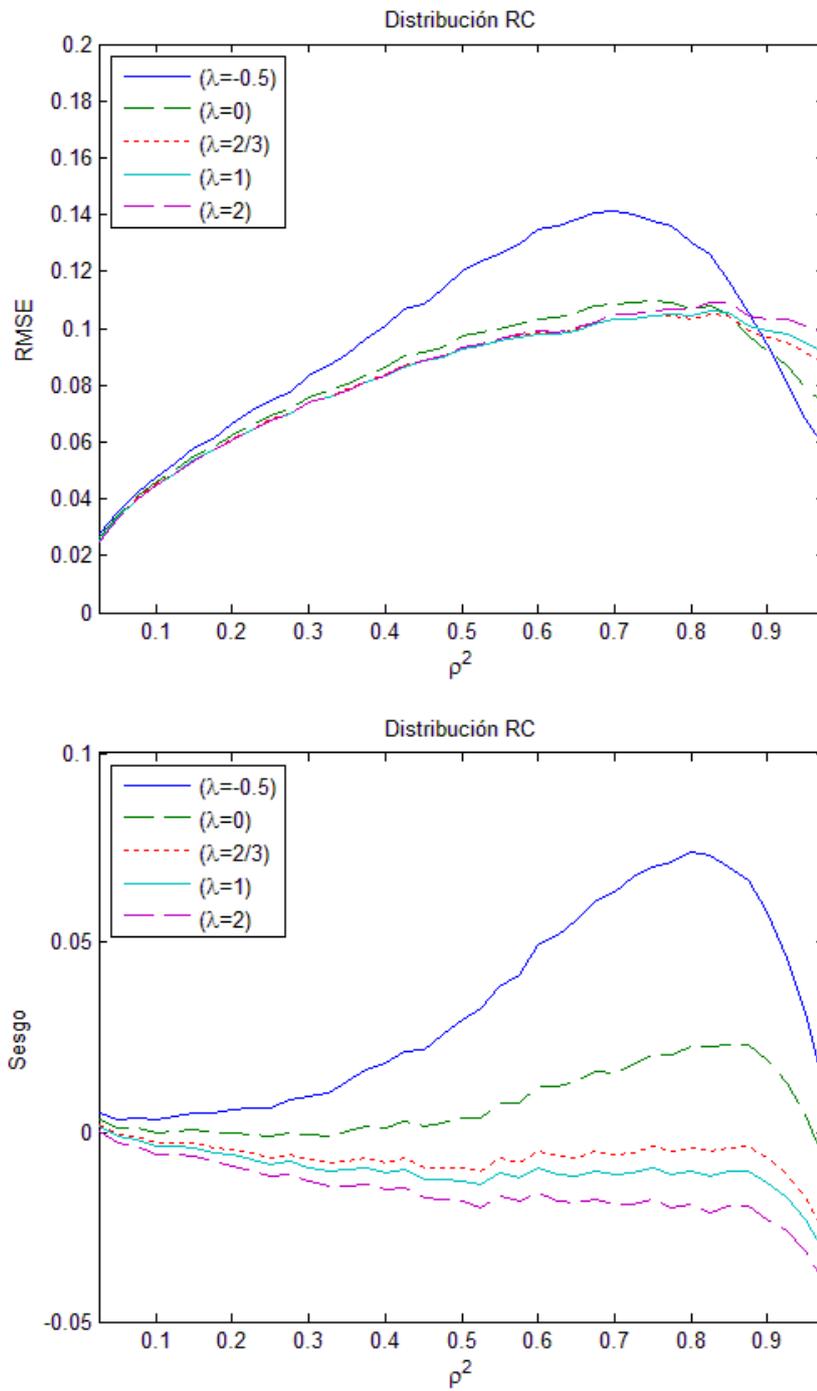


Figura 3.3. RECM y sesgo para los $EMC\phi$ de ρ^2 para datos generados con la distribución multinomial aleatoriamente agrupada.



Para estudiar el comportamiento de los tres estimadores cuando cambia la distribución subyacente en los datos, a continuación se presentan la *RECM* y sesgo medio de cada estimador con datos generados con las diferentes distribuciones en una misma gráfica. En la Figura 3.4 el estimador de Brier $\tilde{\rho}_{\hat{n}^*,N}^2$, en la Figura 3.5 $\tilde{\rho}_{\hat{n}^*,N,\bullet}^2$, el estimador de Brier modificado y en la Figura 3.6 $\tilde{\rho}_{\hat{n}^*,N,\lambda}^2$, el estimador de *MC* ϕ calculado con $\lambda = \frac{2}{3}$. Se comprueba que, sea cual sea la distribución, el *RECM* del estimador de Brier se dispara a medida que aumenta el coeficiente de correlación, mientras que, tanto para el estimador de Brier Modificado como para el *EMC* ϕ , la *RECM* disminuye a partir de valores de ρ^2 mayores que 0.5. Para este valor en concreto del coeficiente de correlación hemos realizado un estudio particularizado para entender mejor el comportamiento de estos estimadores. Para ello, se han representado los histogramas de las estimaciones realizadas con los tres estimadores para cada una de las distribuciones, (Figura 3.7), y se observa que el comportamiento en las tres distribuciones es prácticamente el mismo, puesto que los estimadores han sido contruidos sin considerar ninguna distribución subyacente. El estimador de Brier, $\tilde{\rho}_{\hat{n}^*,N}^2$, es claramente sesgado, dando lugar a sesgos negativos, mientras que con el estimador de Brier modificado, BM, $\tilde{\rho}_{\hat{n}^*,N,\bullet}^2$, se corrige bastante este sesgo y el *EMC* ϕ es prácticamente insesgado.

Figura 3.4. RECM y sesgo para el estimador de ρ^2 de Brier.

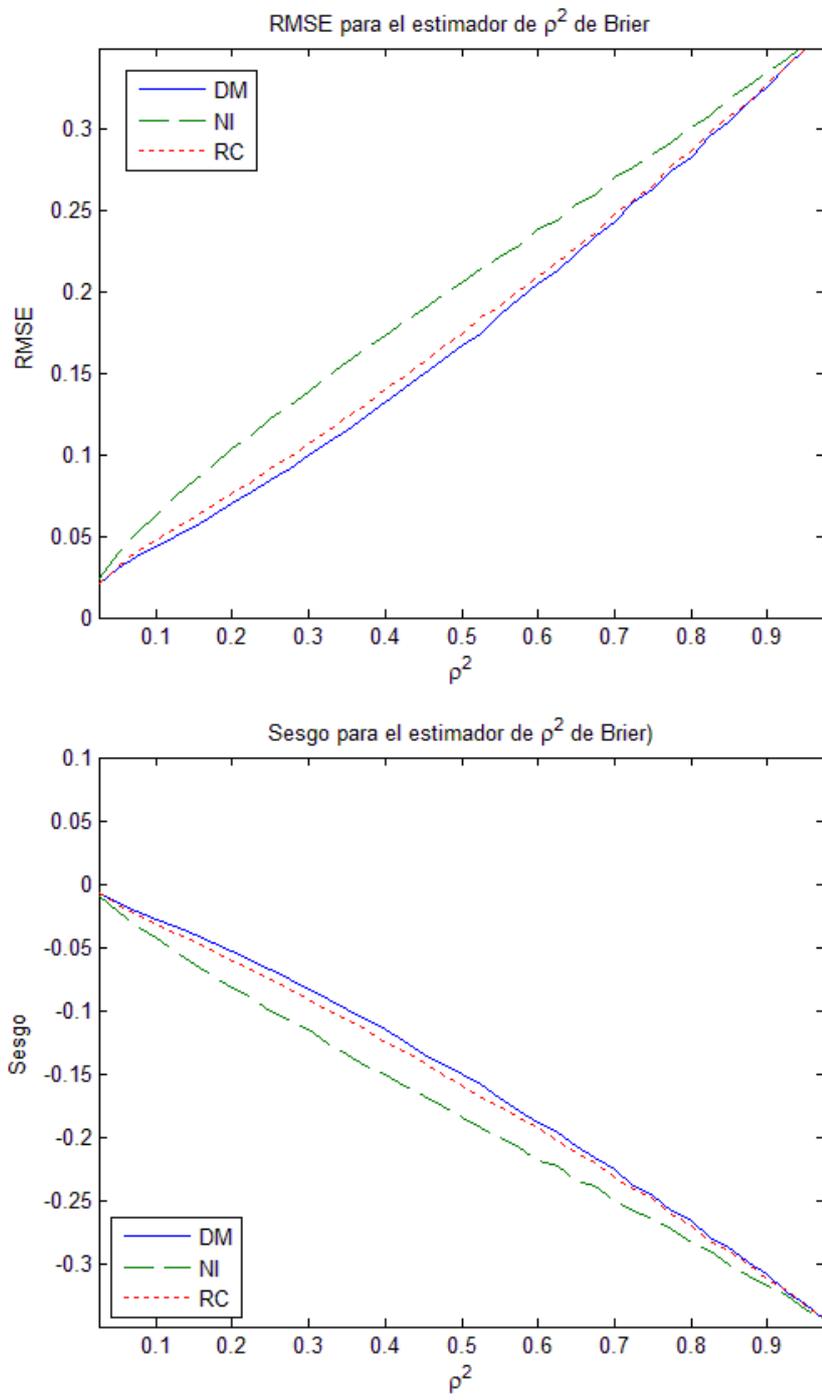


Figura 3.5. RECM y sesgo para el estimador de ρ^2 de Brier modificado.

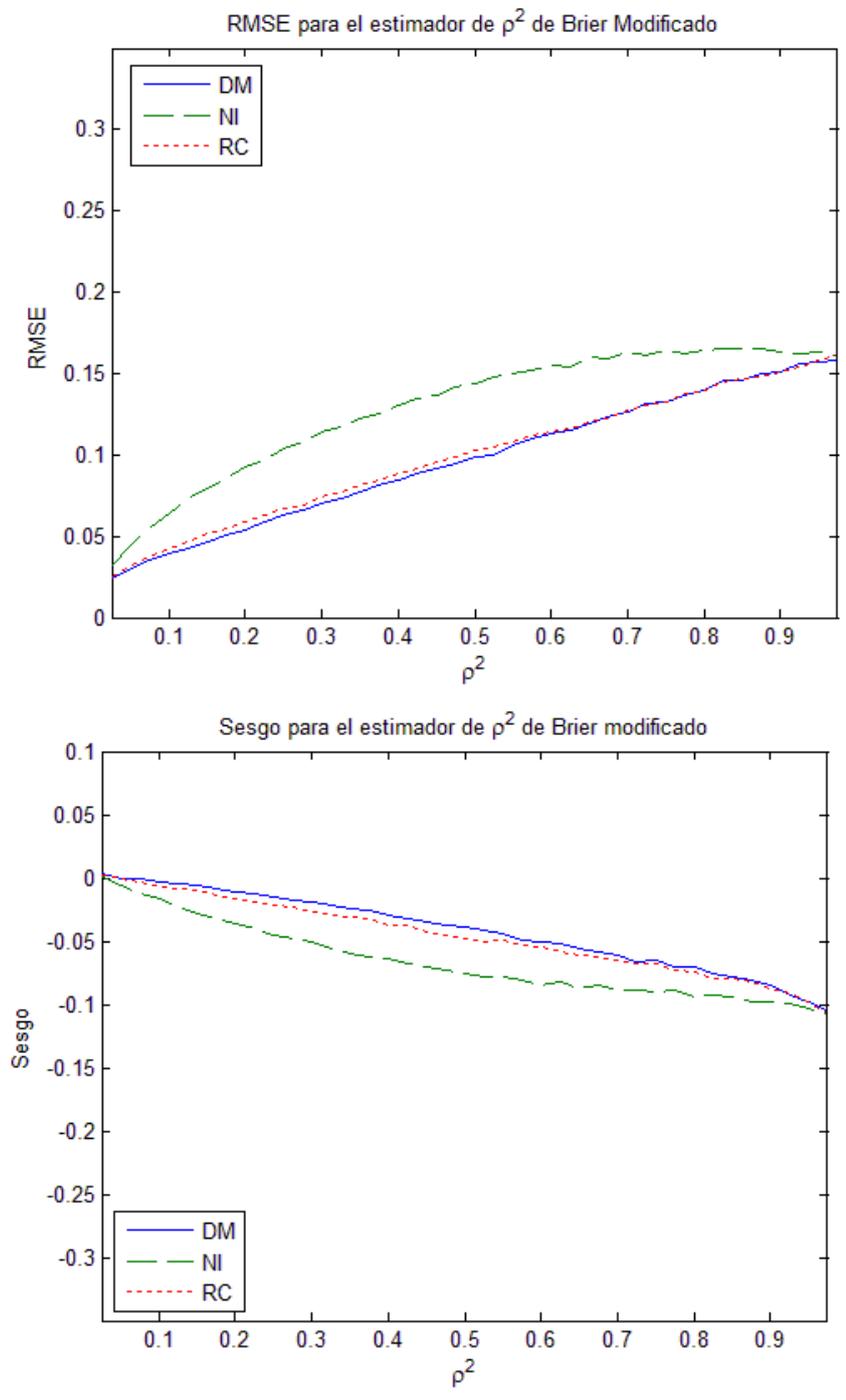


Figura 3.6. RECM y sesgo para el estimador $EMC\phi$ de ρ^2 .

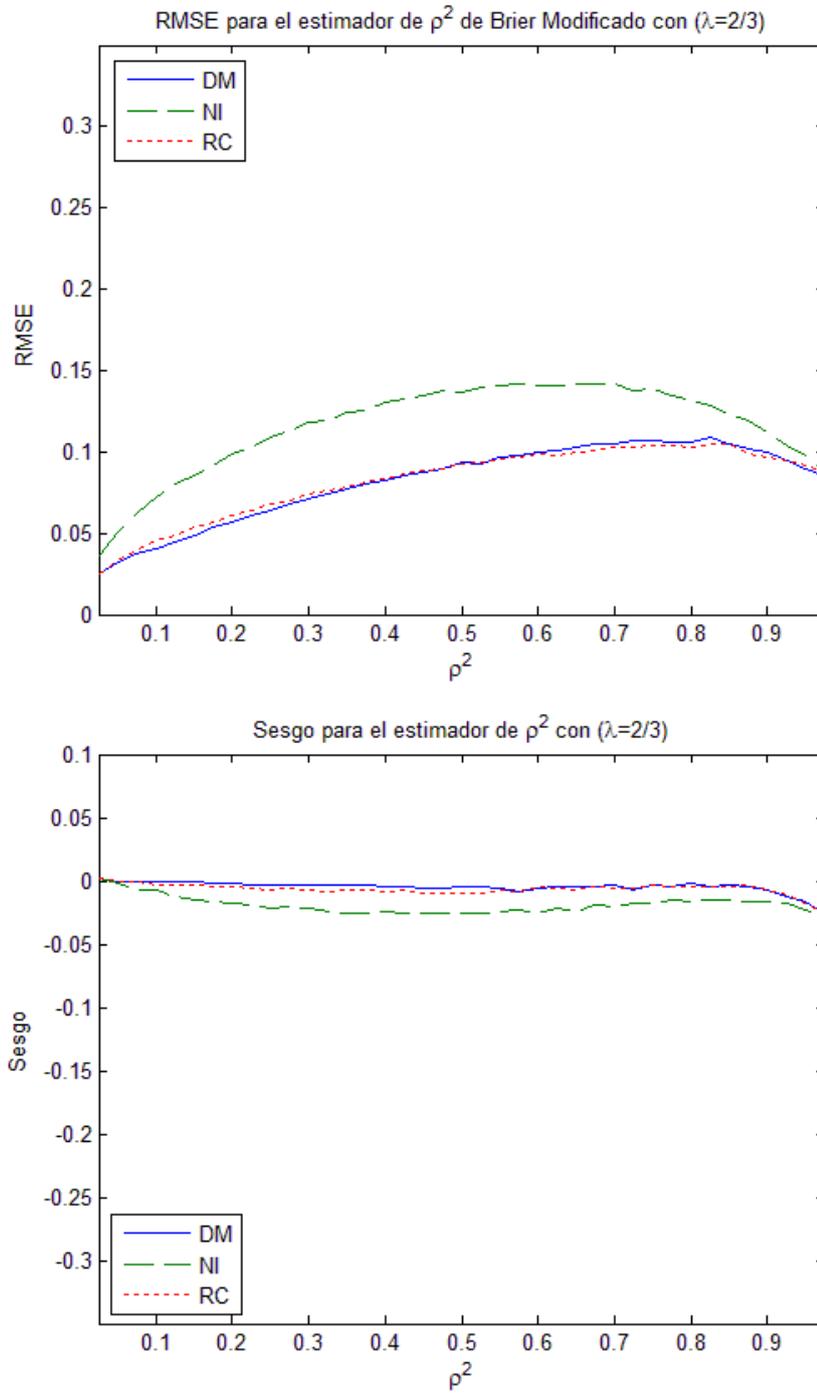
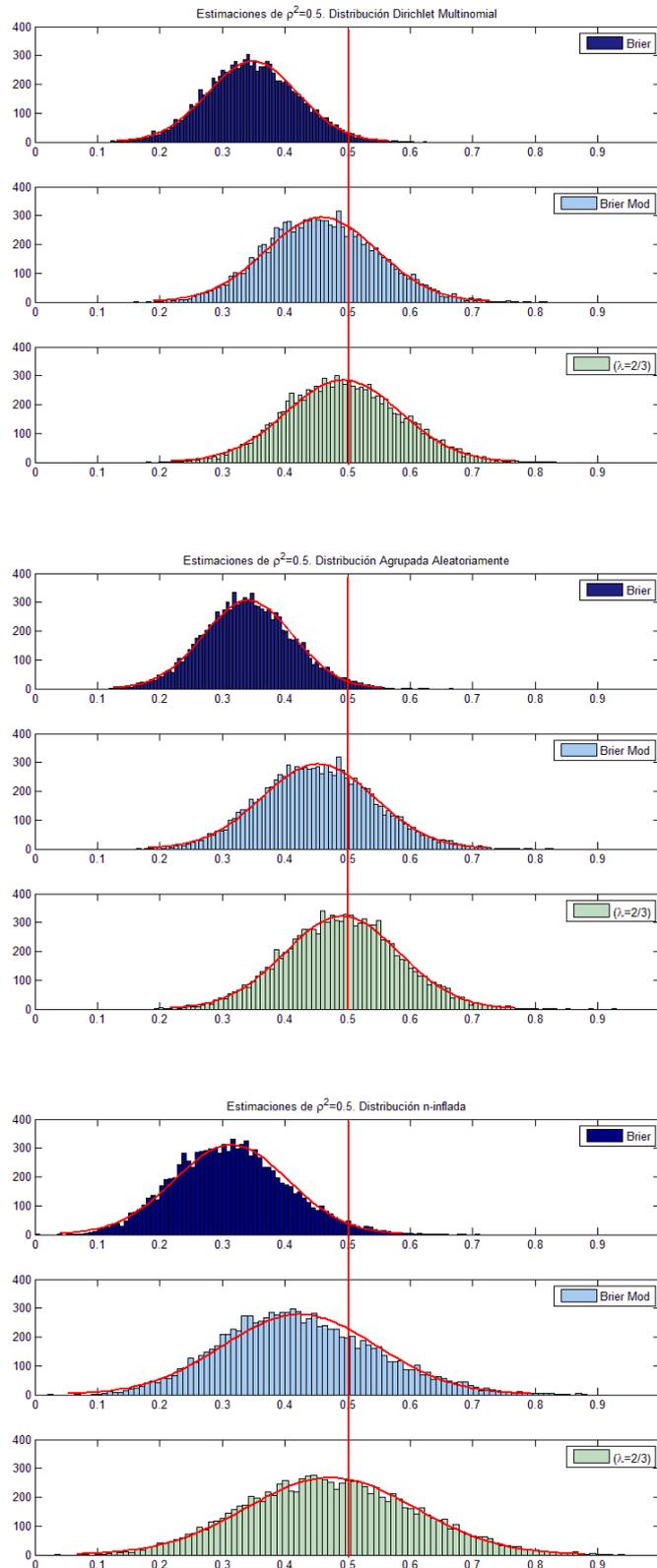


Figura 3.7. Histogramas para los diferentes estimadores de ρ^2 para el valor $\rho^2=0.5$.



Caso 2: n_g suficientemente grande $N_g \geq 1, g = 1, \dots, G$

Para analizar el comportamiento de los estimadores del coeficiente de correlación intra-cluster definidos para este caso, se toman como punto de partida los datos del estudio de ADN del FBI, dado que en este ejemplo los tamaños de los cluster (n_g) son grandes y todos ellos diferentes (311, 367, 367,284,283,147). Las distribuciones de probabilidad que aparecen en la Tabla 3.9 para los locus D3S1358, vWA, FGA y D8S1179, son las utilizadas para simular los diferentes conjuntos de datos, en este caso no se utilizará un modelo loglineal. Se han realizado simulaciones, con las que se ha calculado la RECM y el sesgo medio del nuevo estimador propuesto, $\hat{\rho}^2$ en (3.19) y del estimador propuesto por Weir y Hill ($\bar{\rho}^2$) comparándolos en las Figuras 3.8, a 3.19 correspondientes a los locus D3S1358, vWA, FGA y D8S1179. Debido a que esta clase de datos presentan valores del coeficiente de correlación, ρ^2 , pequeños se estudiará el comportamiento de los estimadores para valores de $\rho^2 \in (0,0,1)$. El sesgo del nuevo estimador propuesto, $\hat{\rho}^2$, tiende a ser más grande que el sesgo del estimador de Weir y Hill $\bar{\rho}^2$ y además negativo. Sin embargo, la RECM de $\hat{\rho}^2$ tiende a ser mas pequeño para los datos generados con la distribución n-inflada, mientras que para las otras dos distribuciones tiende a ser mayor. Seguramente mejoraría su comportamiento si fuera validado por un modelo loglineal.

Figura 3.8. RECM y sesgo para los estimadores de ρ^2 para el locus D3. Distribución Dirichlet Multinomial (DM).

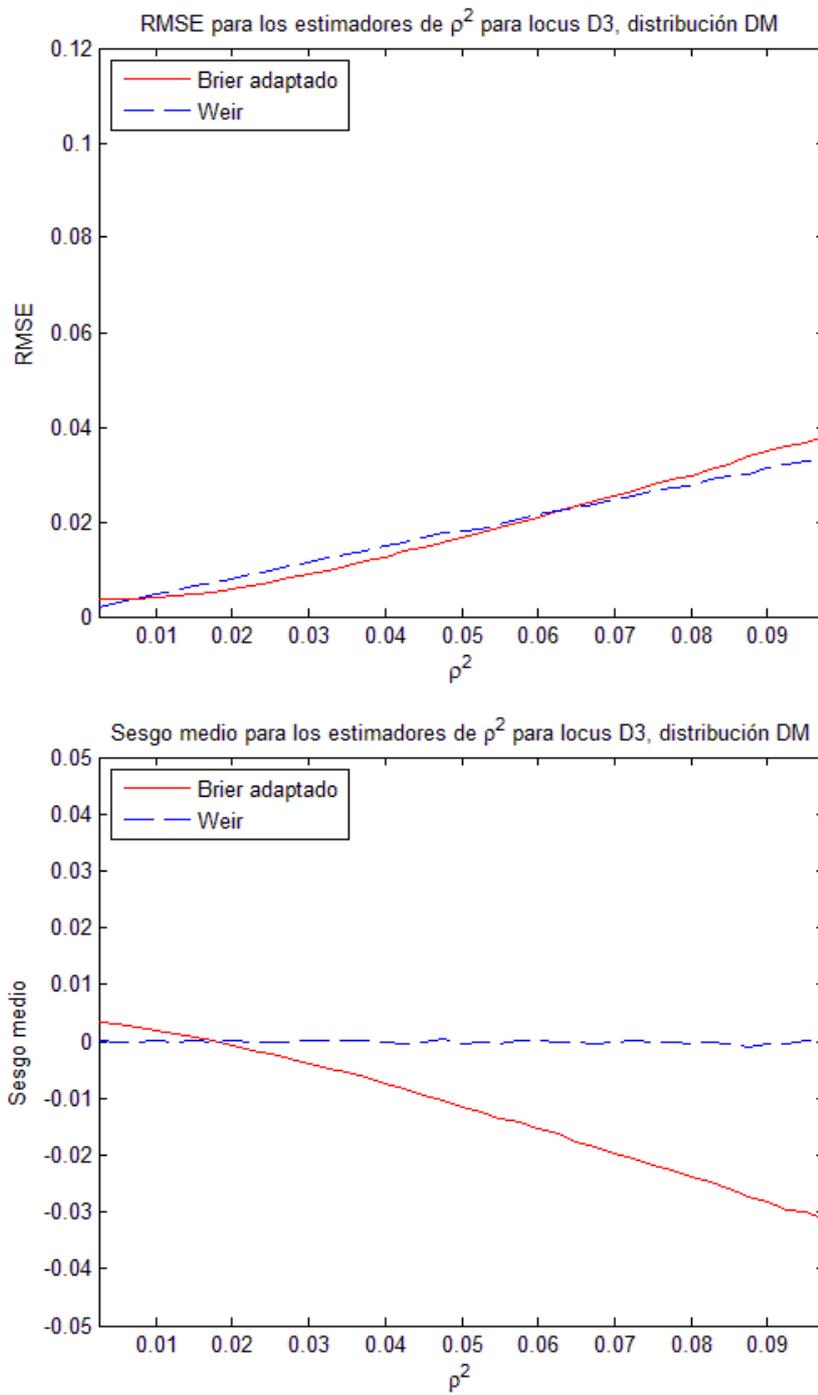


Figura 3.9. RECM y sesgo para los estimadores de ρ^2 para el locus D3. Distribución Agrupada Aleatoriamente (RC).

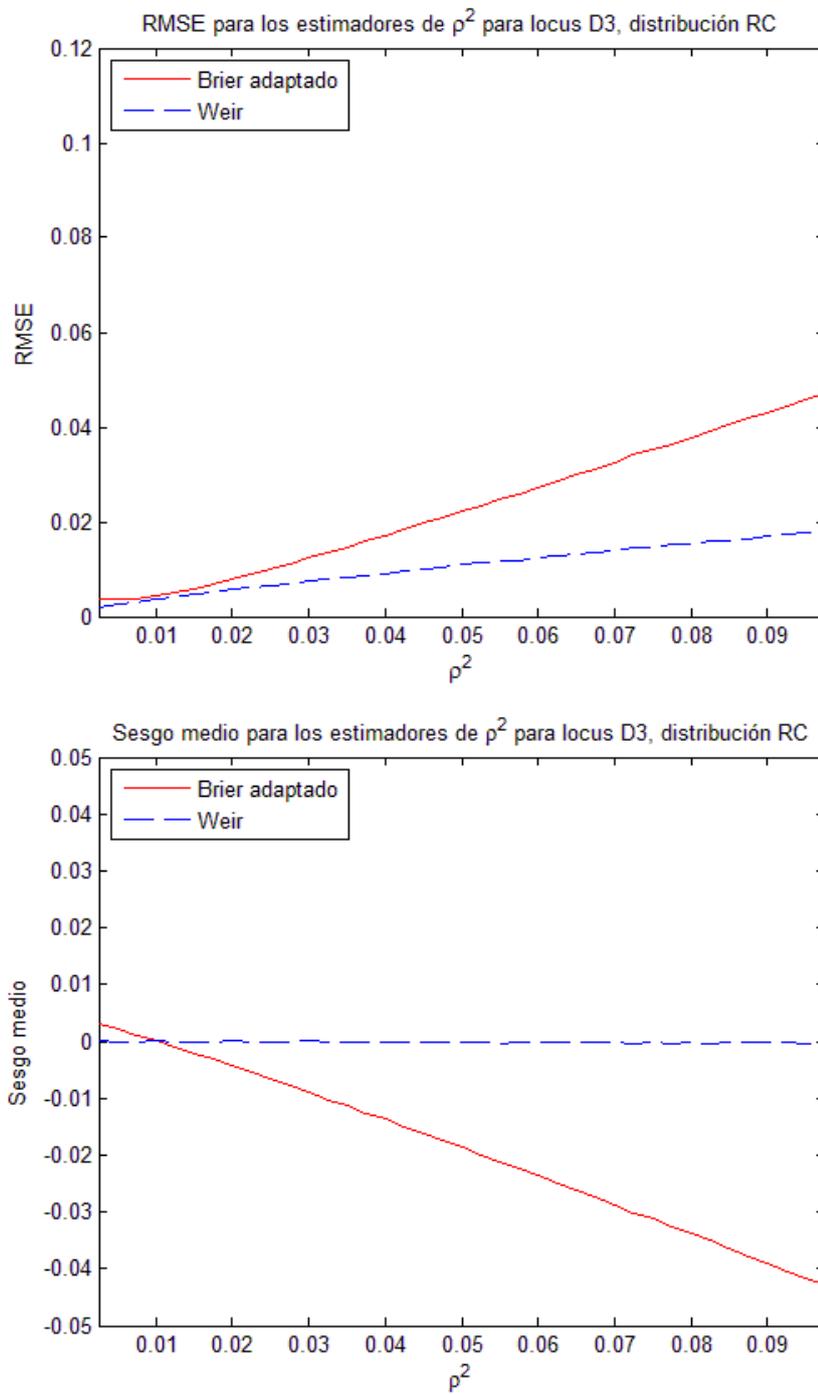


Figura 3.10. RECM y sesgo para los estimadores de ρ^2 para el locus D3. Distribución multinomial n-inflada (NI).

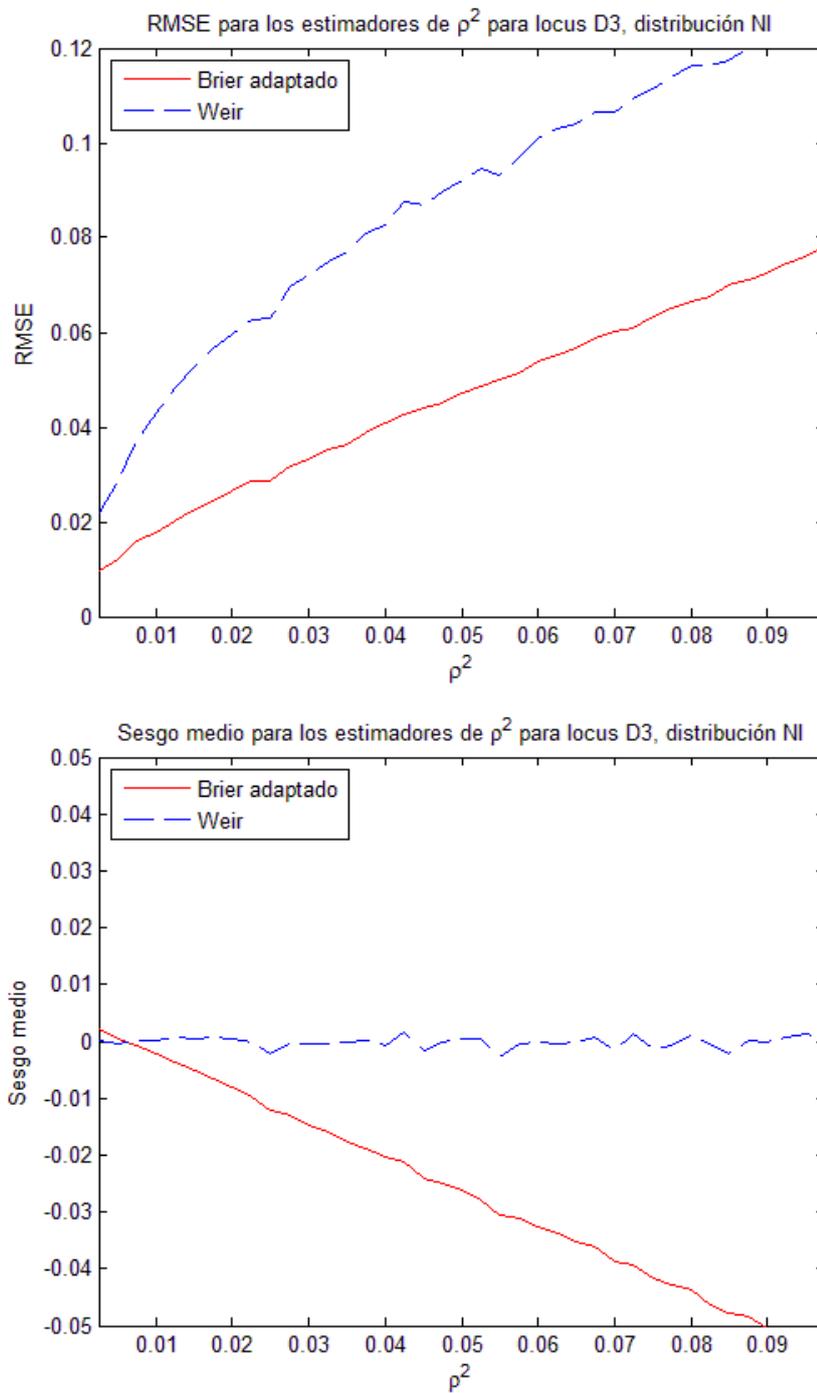


Figura 3.11. RECM y sesgo para los estimadores de ρ^2 para el locus vWA. Distribución Dirichlet multinomial (DM).

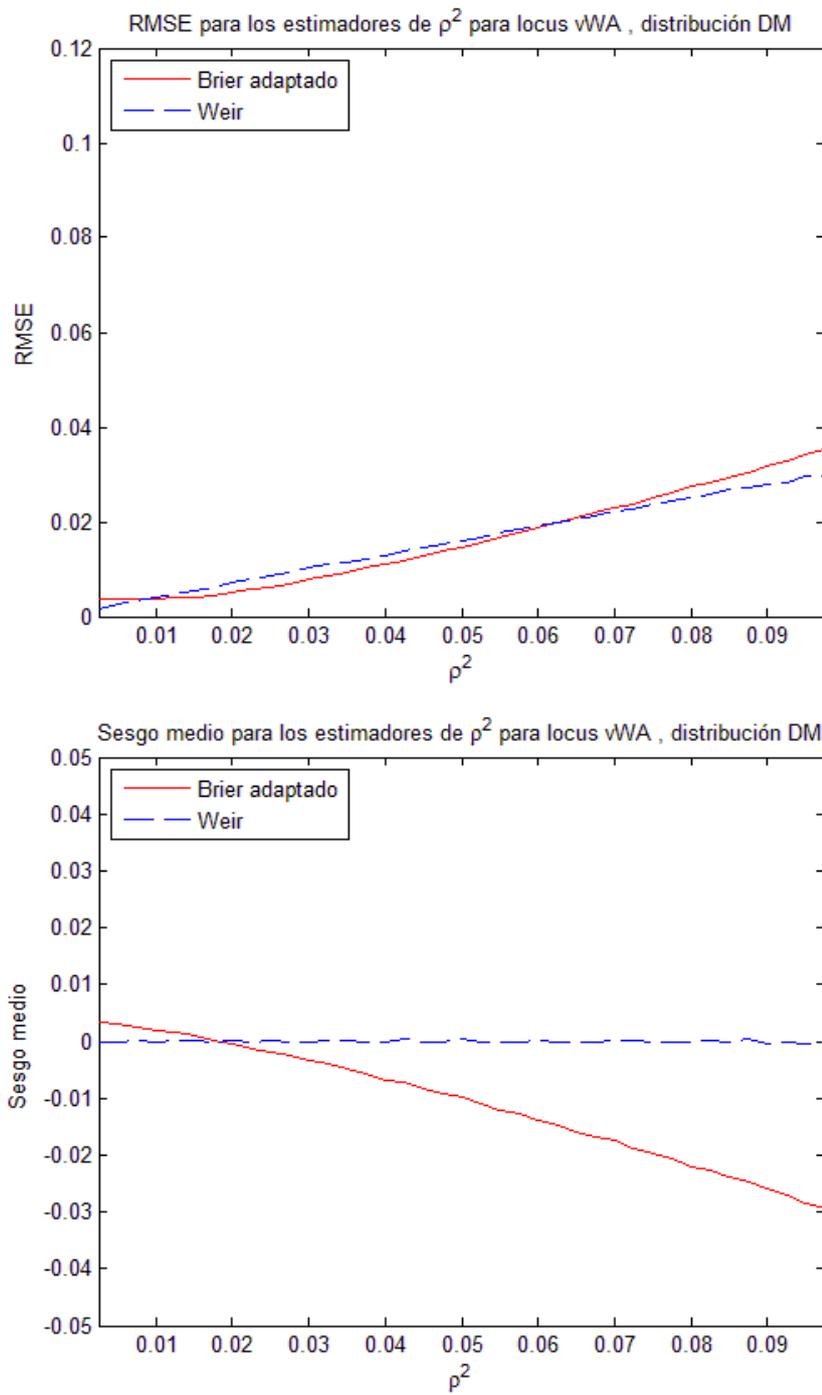


Figura 3.12. RECM y sesgo para los estimadores de ρ^2 para el locus vWA. Distribución Agrupada Aleatoriamente (RC).

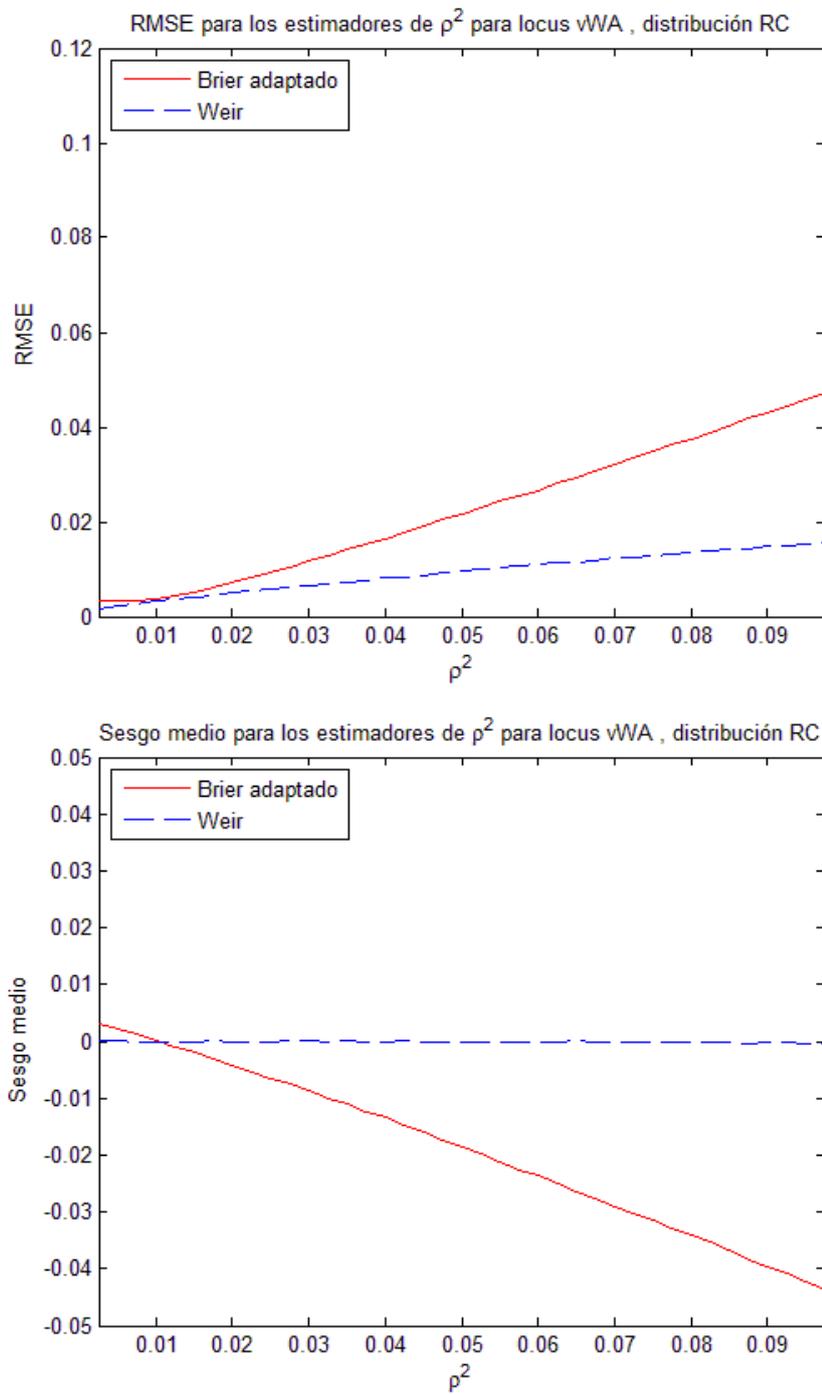


Figura 3.13. RECM y sesgo para los estimadores de ρ^2 para el locus vWA. Distribución multinomial n-inflada (NI).

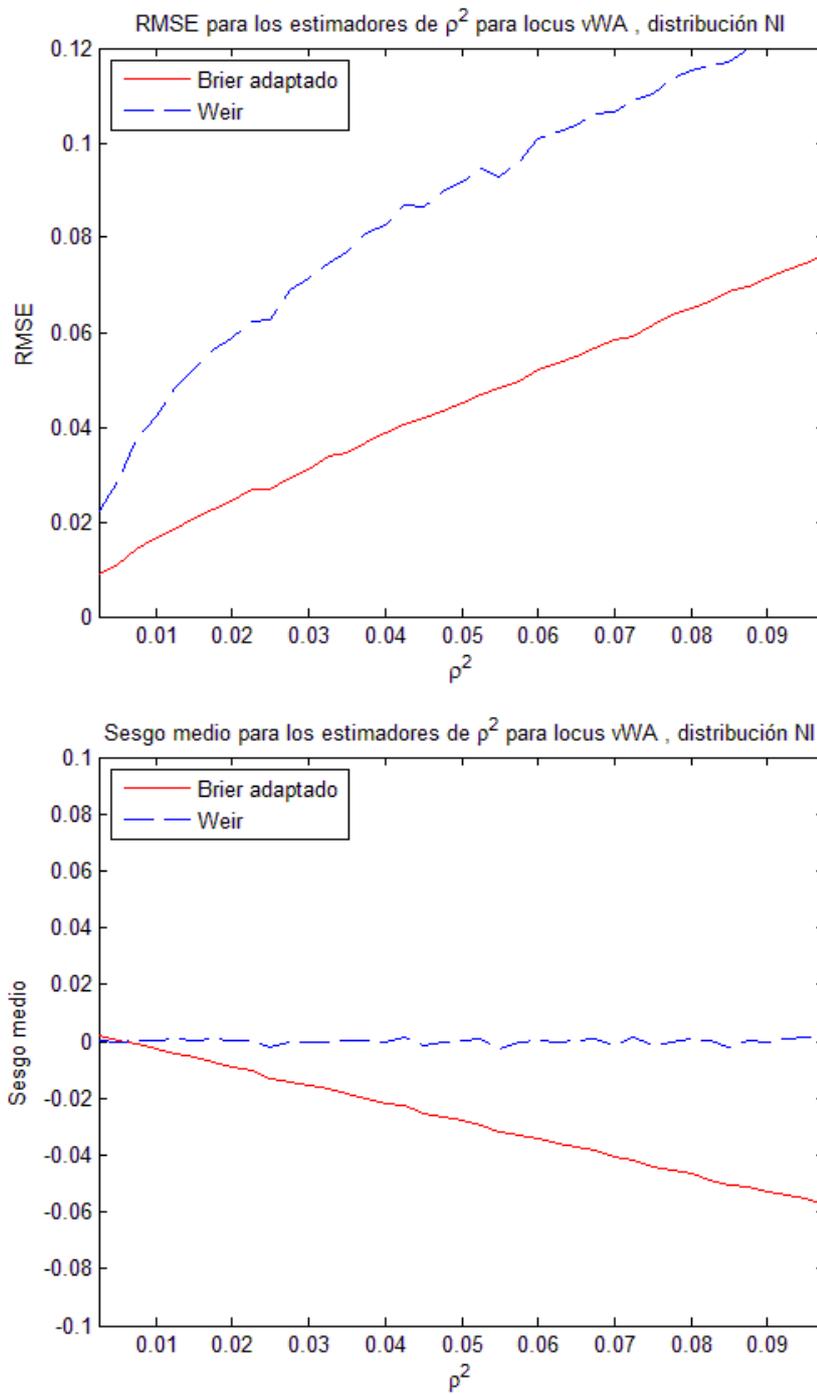


Figura 3.14. RECM y sesgo para los estimadores de ρ^2 para el locus FGA. Distribución Dirichlet Multinomial (DM) .

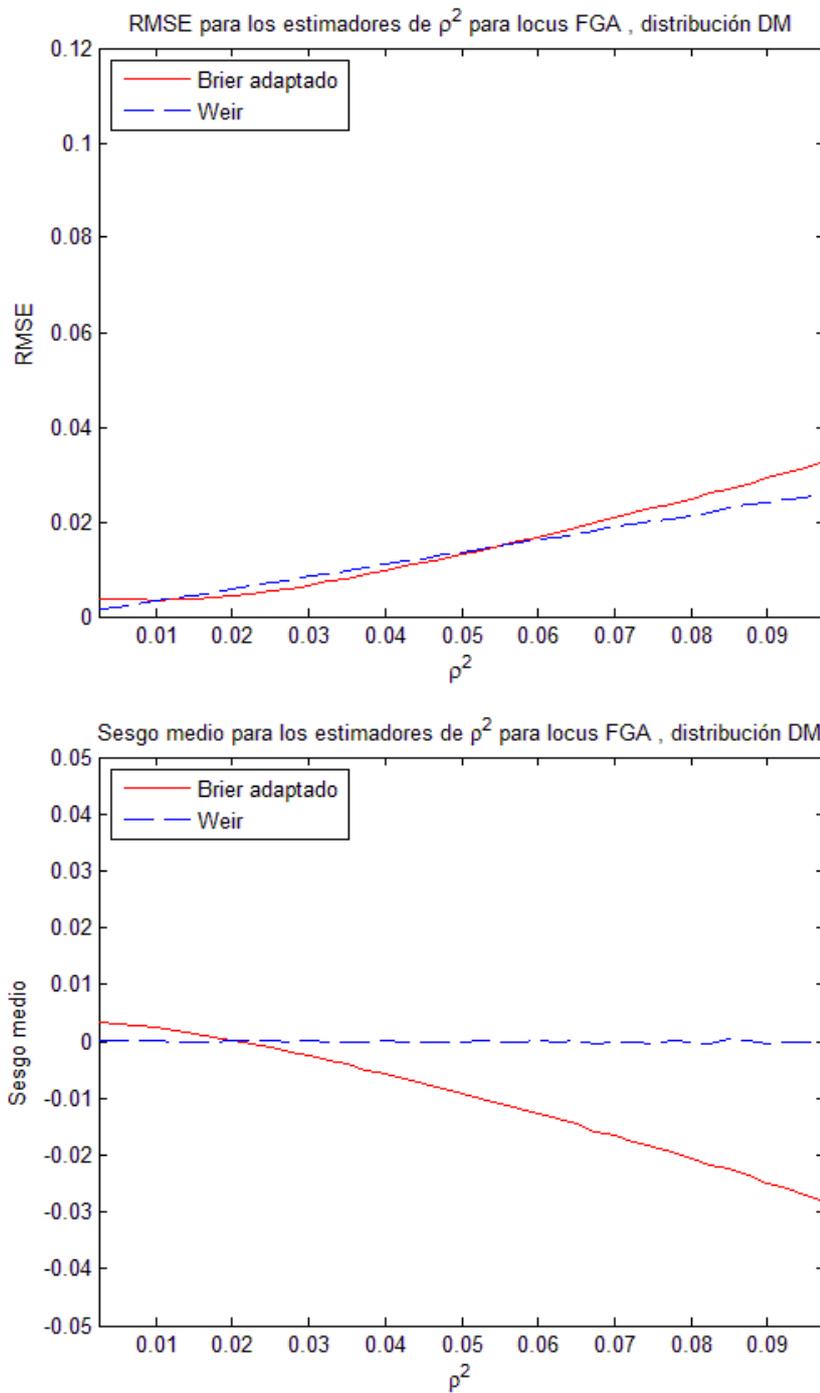


Figura 3.15. RECM y sesgo para los estimadores de ρ^2 para el locus FGA. Distribución Agrupada Aleatoriamente (RC).

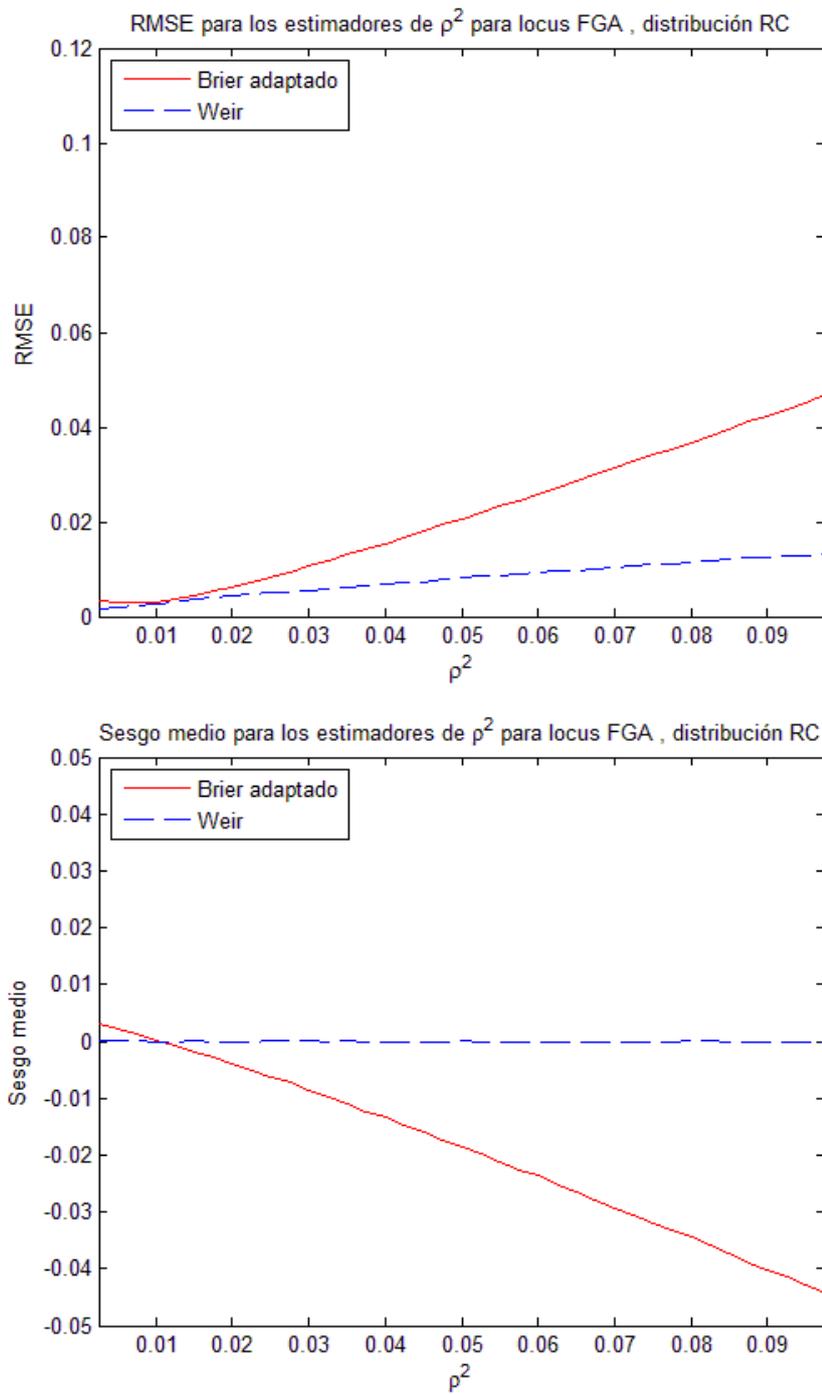


Figura 3.16. RECM y sesgo para los estimadores de ρ^2 para el locus FGA. Distribución multinomial n-inflada (NI).

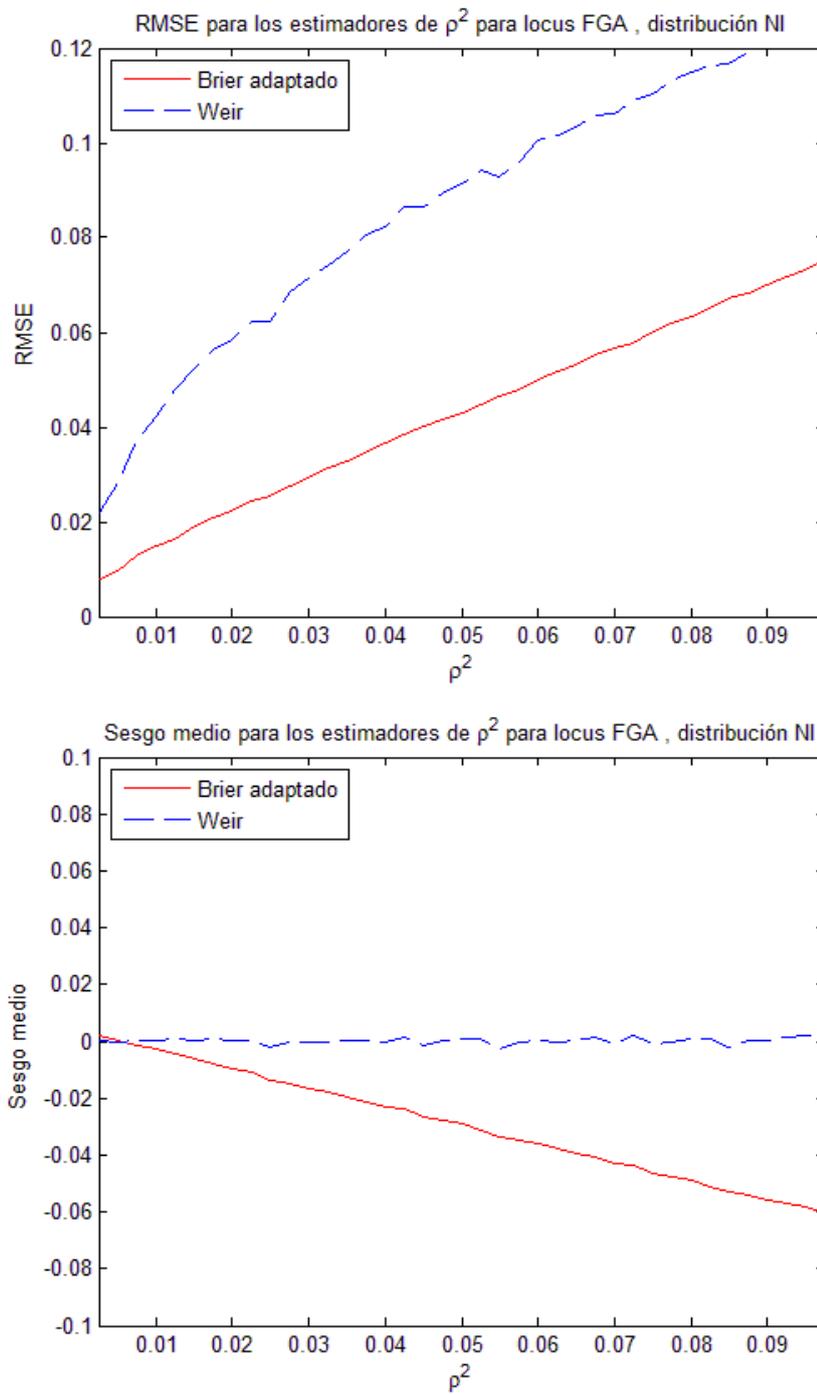


Figura 3.17. RECM y sesgo para los estimadores de ρ^2 para el locus D8. Distribución Dirichlet Multinomial (DM) 2.

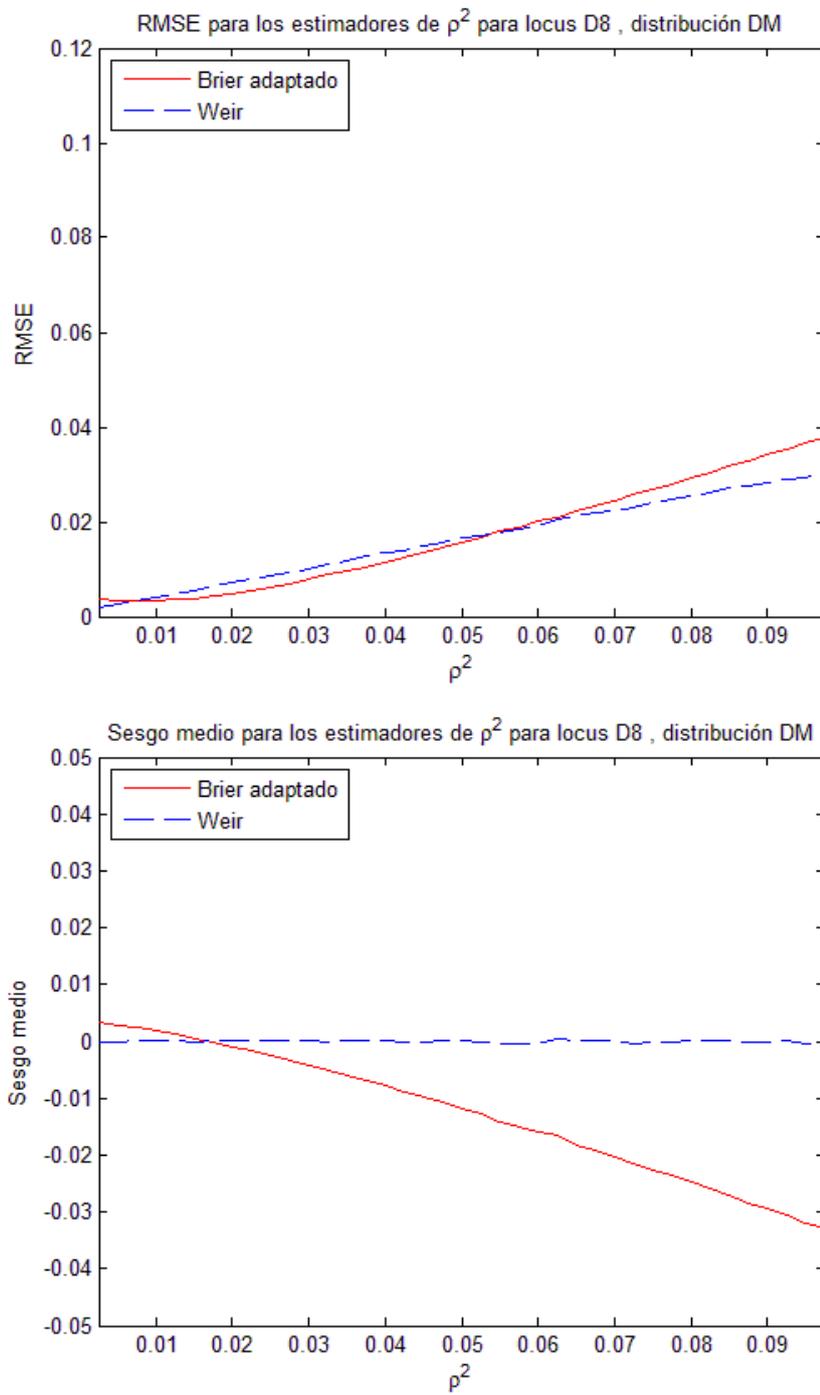


Figura 3.18. RECM y sesgo para los estimadores de ρ^2 para el locus D8. Distribución Agrupada Aleatoriamente (RC) .

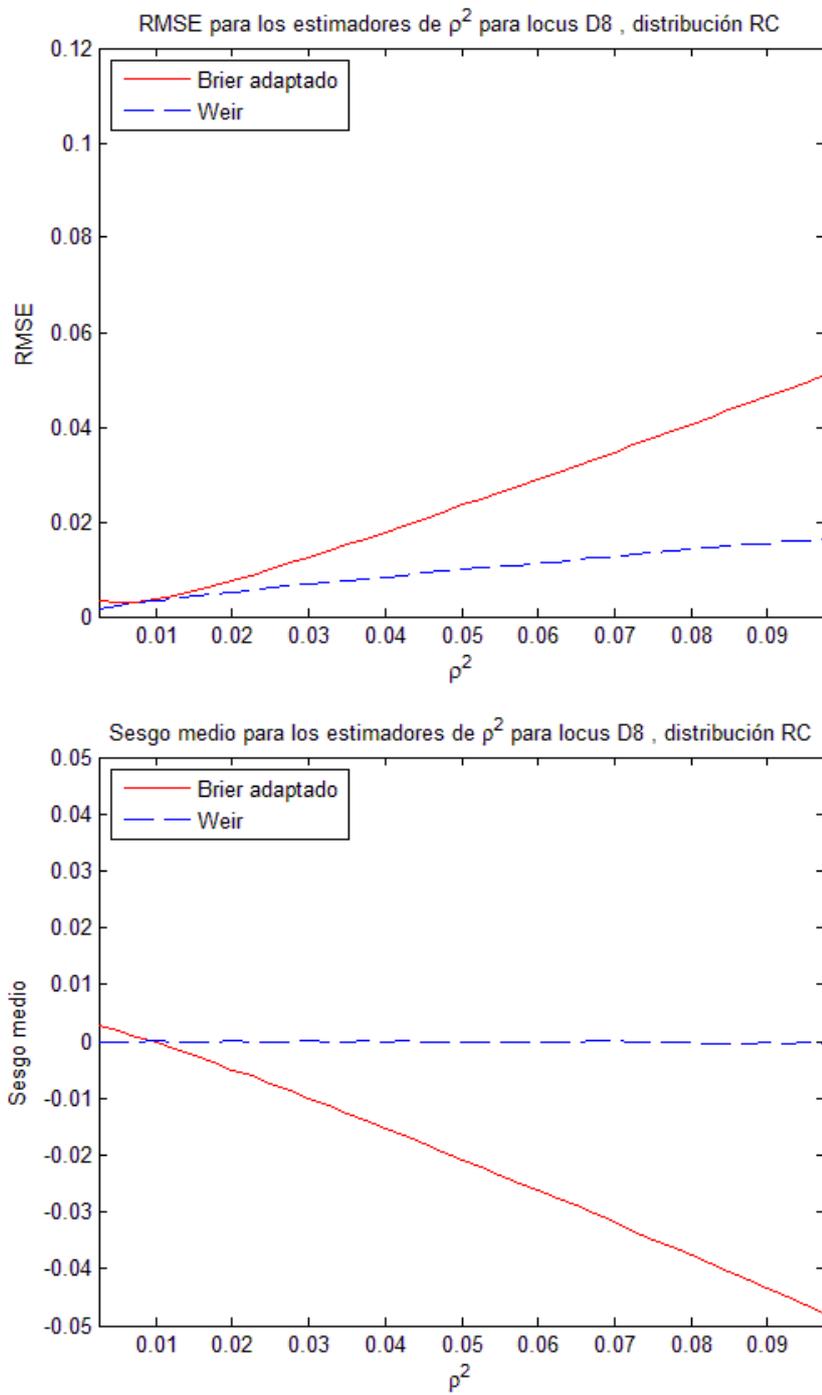
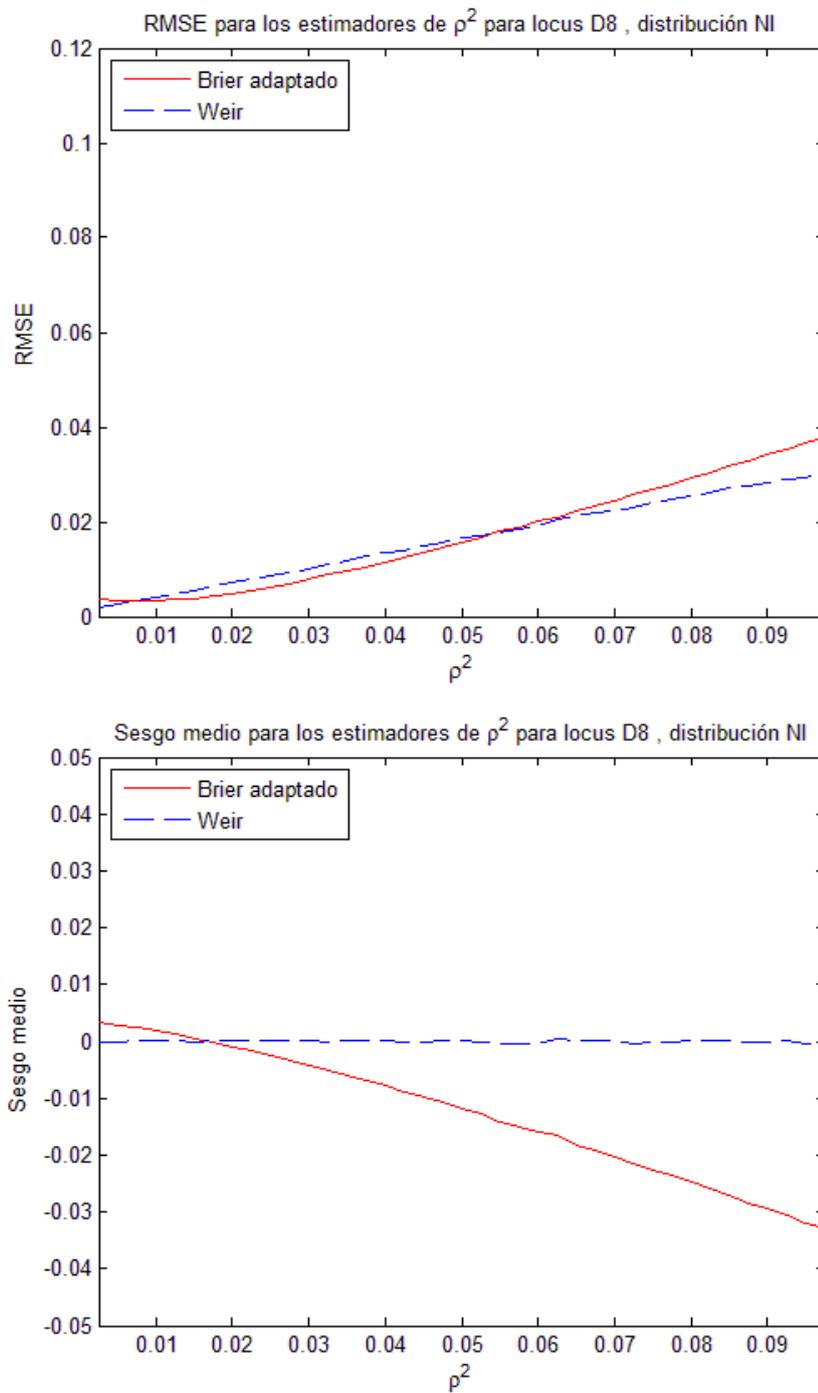


Figura 3.19. RECM y sesgo para los estimadores de ρ^2 para el locus D8. Distribución multinomial n-inflada (NI) .



CAPÍTULO 4

BONDAD DE AJUSTE PARA MODELOS MULTINOMIALES CON SOBREDISPERSIÓN

En los capítulos anteriores se introdujo y estudió el comportamiento del estimador de $MC\phi$ para el parámetro desconocido $\boldsymbol{\theta}$ en modelos multinomiales que verifican (1.9), es decir

$$\mathbf{E} \left[(Y_1, \dots, Y_M)^T \right] = n (p_1(\boldsymbol{\theta}), \dots, p_M(\boldsymbol{\theta}))^T \quad y \quad \mathbf{Var} \left[(Y_1, \dots, Y_M)^T \right] = \vartheta_n n \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})},$$

suponiendo que el vector $\mathbf{p}(\boldsymbol{\theta})$ se puede modelizar como un modelo loglineal. Además, se estableció un estimador consistente del parámetro de sobredispersión ϑ_n . Como se verá a lo largo del capítulo, tener un estimador consistente del efecto del diseño ϑ_n será fundamental para el desarrollo de contrastes de bondad de ajuste.

En este Capítulo se analizará el procedimiento para encontrar la parametrización del modelo loglineal que mejor se ajuste a los datos a analizar, de forma que queden claramente determinadas las dependencias e independencias existentes entre las variables que se están estudiando. Esto llevará al problema de bondad de ajuste, de los datos bajo consideración, a un modelo loglineal, o dicho de otra forma, a contrastar si los datos proceden de un modelo loglineal dado de acuerdo con la parametrización dada en (1.5), frente al modelo saturado. A tal fin, se introducirá una familia de estadísticos de contraste basada en medidas de divergencia en la que los parámetros desconocidos se estimarán a través del estimador de mínima cuasi ϕ -divergencia. Así, en la Sección 4.2 se abordará el problema cuando los tamaños de los clusters o conglomerados son iguales, presentando, en la Sección 4.3, una

aplicación numérica de esta situación. El problema con tamaños de cluster desiguales se analiza en la Sección 4.4 y se presenta una aplicación numérica de estos nuevos estadísticos. Por último, finalizamos el Capítulo con un estudio de simulación en donde se estudia el comportamiento de los nuevos estadísticos.

Es importante resaltar que con el planteamiento anterior no es necesario suponer una distribución para los datos que se están analizando; es decir, no es necesario hacer la hipótesis de si los datos proceden de una multinomial n -inflada, multinomial agrupada aleatoriamente, Dirichlet multinomial, etc...

4.1. Estadísticos de ϕ -divergencia para bondad de ajuste con clusters de igual tamaño

Dado el modelo loglineal planteado en (1.5) con una determinada matriz de diseño \mathbf{W} ,

$$\mathbf{p}(\boldsymbol{\theta}) = \frac{\exp\{\mathbf{W}\boldsymbol{\theta}\}}{\mathbf{1}_M^T \exp\{\mathbf{W}\boldsymbol{\theta}\}},$$

y asociado a una tabla de contingencia con muestreo por clusters o conglomerados, el objetivo es contrastar

$$H_0 : \mathbf{p} = \mathbf{p}(\boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^{M_0} \tag{4.1}$$

frente a

$$H_1 : \mathbf{p} \neq \mathbf{p}(\boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^{M_0}.$$

Bajo H_0 , $\mathbf{Y}^{(\ell)}$, $\ell = 1, \dots, N$, son vectores aleatorios independientes cuyos dos primeros momentos tienen la expresión dada en (1.9) donde $\mathbf{p}(\boldsymbol{\theta})$ sigue un modelo loglineal asociado a la matriz de diseño \mathbf{W} . Obsérvese que, para el problema que se abordará en este capítulo, no será necesario conocer si los vectores $\mathbf{Y}^{(\ell)}$ siguen una distribución concreta.

Para fijar ideas, si se consideran los datos de la Sección 3.4.1, se puede estar interesado en analizar si existe independencia entre la satisfacción con el vecindario y el grado de satisfacción con la vivienda particular. Es decir, se trata de contrastar

$$H_0 : \mathbf{p}(\boldsymbol{\theta}) = \exp \{u + \theta_{1(i)} + \theta_{2(j)}\}, \quad i, j = 1, 2, 3,$$

frente

$$H_1 : \mathbf{p}(\boldsymbol{\theta}) \neq \exp \{u + \theta_{1(i)} + \theta_{2(j)}\}, \quad i, j = 1, 2, 3.$$

Para abordar el problema, Brier, consideró los estadísticos de Pearson y del test de cociente de verosimilitudes, dados respectivamente por

$$X^2 = KN \sum_{r=1}^M \frac{(\hat{p}_r - p_r(\hat{\boldsymbol{\theta}}))^2}{p_r(\hat{\boldsymbol{\theta}})}$$

$$G^2 = 2KN \sum_{r=1}^M \hat{p}_r \log \frac{\hat{p}_r}{p_r(\hat{\boldsymbol{\theta}})},$$

bajo la suposición de que $\mathbf{Y}^{(\ell)}$, $\ell = 1, \dots, N$, es una muestra aleatoria simple procedente de un población Dirichlet multinomial, siendo $\hat{\boldsymbol{\theta}}$ el estimador de máxima verosimilitud de $\boldsymbol{\theta}$ y

$$K = \frac{1+c}{n+c}, \quad \text{siendo} \quad c = \frac{1-\rho^2}{\rho^2}.$$

A lo largo de todo el capítulo, se considerará $\hat{\boldsymbol{\theta}}_\phi$, el estimador de $MC\phi$, en lugar del estimador de máxima verosimilitud, y además se utilizará la siguiente familia de estadísticos de contraste

$$T^{\phi_1}(\mathbf{Y}, \hat{\boldsymbol{\theta}}_{\phi_2}) = K \frac{2N}{\phi_1''(1)} d_{\phi_1} \left(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}) \right), \quad (4.2)$$

basada en las medidas de ϕ -divergencia, que contendrá como casos particulares a los estadísticos ji-cuadrado X^2 , obtenido cuando $\phi_1(x) = \frac{1}{2}(x-1)$ correspondiente al valor $\lambda = 1$,

y $\phi_2(x) = x \log x - x + 1$ correspondiente al valor $\lambda = 0$, y al cociente de verosimilitudes G^2 , que se obtiene cuando $\phi_1(x) = \phi_2(x) = x \log x - x + 1$ correspondiente al valor $\lambda = 0$. Es de resaltar que, como no se supondrá ninguna distribución subyacente para los vectores $\mathbf{Y}^{(\ell)}$, la constante K de Brier será reemplazada por un estimador consistente de $\frac{n}{\vartheta_n}$.

Notesé que en (4.2) hay dos medidas de ϕ -divergencia, la asociada a ϕ_1 y la asociada a ϕ_2 . Esta última permitirá construir el estimador de $MC\phi$, $\hat{\boldsymbol{\theta}}_{\phi_2}$, mientras que la primera permitirá construir la familia de estadísticos de contraste.

4.1.1. Distribución asintótica

Bajo la suposición de que los clusters o conglomerados tienen el mismo tamaño, en este apartado se obtendrá la distribución asintótica de la familia de estadísticos de contraste dada en (4.2). El resultado esencial se presenta en el siguiente teorema.

Teorema 4.1 *La distribución asintótica de la diferencia entre el estimador no paramétrico y el de $MC\phi$ de $\mathbf{p}(\boldsymbol{\theta})$, cuando se tienen N clusters de tamaño n , es*

$$\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2})) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_M, \frac{\vartheta_n}{n} (\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} - \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W} (\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W})^{-1} \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)})).$$

donde $\boldsymbol{\theta}_0$ es el verdadero valor del parámetro $\boldsymbol{\theta}$.

Demostración En el Teorema 2.3 del Capítulo 2 se obtuvieron las distribuciones asintóticas de los estimadores de $MC\phi$, en el desarrollo de esta demostración se llegó a las expresiones (2.22),

$$\sqrt{N}(\mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}) - \mathbf{p}(\boldsymbol{\theta}_0)) = \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W} \sqrt{N}(\hat{\boldsymbol{\theta}}_{\phi_2} - \boldsymbol{\theta}_0) + o_p(\mathbf{1}_M),$$

y (2.25),

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_{\phi_2} - \boldsymbol{\theta}_0) = (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \sqrt{N}(\hat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) + o_p(\mathbf{1}_{M_0}),$$

donde

$$\mathbf{A}(\boldsymbol{\theta}_0) = \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W}. \quad (4.3)$$

Sustituyendo la expresión (2.22) de $\sqrt{N}(\hat{\boldsymbol{\theta}}_{\phi_2} - \boldsymbol{\theta}_0)$ en (2.25) se obtiene la siguiente relación

$$\sqrt{N}(\mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}) - \mathbf{p}(\boldsymbol{\theta}_0)) = \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \sqrt{N}(\hat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) + o_p(\mathbf{1}_M),$$

y sustrayendo a $\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0))$ ambos lados de la igualdad, se tiene

$$\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2})) = (\mathbf{I}_M - \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0)) \sqrt{N}(\hat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) - o_p(\mathbf{1}_M).$$

Por otro lado, en el Capítulo 1 se estableció (1.14),

$$\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_M, \frac{\partial_n}{n} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}),$$

luego $\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}))$ tiene una distribución asintótica Normal de media nula y si se sustituye $\mathbf{A}(\boldsymbol{\theta}_0)$ por su expresión (4.3) en función de \mathbf{W} y $\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}$ y se tiene en cuenta que

$$\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} = \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}, \quad (4.4)$$

y

$$\begin{aligned} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{A}(\boldsymbol{\theta}_0) &= \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W} \\ &= \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W} \\ &= \mathbf{A}(\boldsymbol{\theta}_0), \end{aligned} \quad (4.5)$$

se puede obtener su matriz de varianzas-covarianzas igual a

$$\begin{aligned}
& (\mathbf{I}_M - \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0)) \frac{\vartheta_n}{n} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \\
& \times (\mathbf{I}_M - \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0)) \\
& = \frac{\vartheta_n}{n} \left(\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} - \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \right. \\
& - \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \\
& + \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \\
& = \frac{\vartheta_n}{n} \left(\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} - \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W} (\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W})^{-1} \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \right. \\
& - \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W} (\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W})^{-1} \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \\
& + \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W} (\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W})^{-1} \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W} (\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W})^{-1} \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \left. \right) \\
& = \frac{\vartheta_n}{n} \left(\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} - \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W} (\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W})^{-1} \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \right. \\
& - \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W} (\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W})^{-1} \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \\
& + \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W} (\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W})^{-1} \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \left. \right) \\
& = \frac{\vartheta_n}{n} \left(\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} - \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W} (\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W})^{-1} \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \right).
\end{aligned}$$

■

Corolario 4.1 *El estadístico semiparamétrico de contraste de bondad de ajuste, para datos con sobredispersión agrupados en clusters, tiene la siguiente distribución asintótica*

$$\frac{X^2(\mathbf{Y}, \widehat{\boldsymbol{\theta}}_\phi)}{\widehat{\vartheta}_{n,N}} \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \chi_{M-M_0-1}^2,$$

donde

$$X^2(\mathbf{Y}, \widehat{\boldsymbol{\theta}}_\phi) = nN (\widehat{\mathbf{p}} - \mathbf{p}(\widehat{\boldsymbol{\theta}}_\phi))^T \mathbf{D}_{\mathbf{p}(\widehat{\boldsymbol{\theta}}_\phi)}^{-1} (\widehat{\mathbf{p}} - \mathbf{p}(\widehat{\boldsymbol{\theta}}_\phi)) = nN \sum_{r=1}^M \frac{(\widehat{p}_r - p_r(\widehat{\boldsymbol{\theta}}_\phi))^2}{p_r(\widehat{\boldsymbol{\theta}}_\phi)}. \quad (4.6)$$

Demostración

$$\frac{X^2(\mathbf{Y}, \hat{\boldsymbol{\theta}}_\phi)}{\tilde{\vartheta}_{n,N}} = \mathbf{Q}^T \mathbf{Q}, \quad (4.7)$$

donde

$$\mathbf{Q} = \sqrt{\frac{n}{\tilde{\vartheta}_{n,N}}} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_\phi)}^{-1/2} \sqrt{N} \left(\hat{\mathbf{p}} - \mathbf{p}(\hat{\boldsymbol{\theta}}_\phi) \right).$$

La distribución asintóticas de \mathbf{Q} se obtiene del Teorema 4.1 y de la aplicación del teorema de Slutsky

$$\mathbf{Q} \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_M, \mathbf{V}(\boldsymbol{\theta}_0)),$$

donde

$$\begin{aligned} \mathbf{V}(\boldsymbol{\theta}_0) &= \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_\phi)}^{-1} \left(\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} - \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W} (\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W})^{-1} \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \right) \\ &= \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_\phi)}^{-1} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} - \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_\phi)}^{-1/2} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W} (\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W})^{-1} \mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_\phi)}^{-1/2} \\ &= \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} - \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0). \end{aligned} \quad (4.8)$$

Finalmente, para hallar la distribución asintótica de la forma cuadrática dada en (4.7), se tendrá en cuenta que, si la matriz $\mathbf{V}(\boldsymbol{\theta}_0)$ es idempotente y simétrica, entonces la distribución de la forma cuadrática $\mathbf{Q}^T \mathbf{Q}$ es asintóticamente una ji-cuadrado con tantos grados de libertad como el rango o la traza de la matriz $\mathbf{V}(\boldsymbol{\theta}_0)$. Se mostrará que $\mathbf{V}(\boldsymbol{\theta}_0)$ es idempotente, ya que la simetría es directa,

$$\begin{aligned} \mathbf{V}(\boldsymbol{\theta}_0) \mathbf{V}(\boldsymbol{\theta}_0) &= \left(\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} - \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \right) \\ &\quad \times \left(\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} - \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \right) \end{aligned}$$

Utilizando de nuevo (4.4) y (4.5), se tiene que

$$\begin{aligned}
\mathbf{V}(\boldsymbol{\theta}_0) \mathbf{V}(\boldsymbol{\theta}_0) &= \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} - \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \\
&\quad - \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \\
&\quad + \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \\
&= \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} - \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \\
&= \mathbf{V}(\boldsymbol{\theta}_0).
\end{aligned}$$

Luego la matriz $\mathbf{V}(\boldsymbol{\theta}_0)$ es idempotente, por lo tanto su rango coincide con su traza y para calcular ésta basta tener en cuenta que las propiedades de la traza de un producto de matrices

$$\begin{aligned}
\text{rg}(\mathbf{V}(\boldsymbol{\theta}_0)) &= \text{tr}(\mathbf{V}(\boldsymbol{\theta}_0)) \\
&= \text{tr} \left(\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} - \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \right) \\
&= \text{tr}(\mathbf{I}_M - \mathbf{p}(\boldsymbol{\theta}_0)) - \text{tr}(\mathbf{I}_{M_0}) \\
&= M - 1 - M_0.
\end{aligned}$$

■

Observación 4.1 *Es importante aclarar las diferencias entre los estadísticos dados en (2.31) y (4.6) cuya expresión es similar pero, sin embargo su interpretación es muy diferente. Mientras que $X^2(\mathbf{Y}, \hat{\boldsymbol{\theta}}_\phi)$ es parte del estadístico de bondad de ajuste, en el Capítulo 2 se definió $X^2(\tilde{\mathbf{Y}}, \hat{\boldsymbol{\theta}}_\phi)$ a partir de la traza de la matriz de varianzas-covarianzas de $\tilde{\mathbf{Y}}$, para construir el estimador del efecto del diseño. Obsérvese que*

$$X^2(\tilde{\mathbf{Y}}, \hat{\boldsymbol{\theta}}_\phi) = \sum_{\ell=1}^N \left(\mathbf{Y}^{(\ell)} - n\hat{\mathbf{p}} \right)^T \frac{1}{n} \mathbf{D}_{\mathbf{p}(\hat{\boldsymbol{\theta}}_\phi)}^{-1} \left(\mathbf{Y}^{(\ell)} - n\hat{\mathbf{p}} \right) = n \sum_{\ell=1}^N \sum_{r=1}^M \frac{(\hat{p}_r^{(\ell)} - \hat{p}_r)^2}{p_r(\hat{\boldsymbol{\theta}}_\phi)}$$

se calcula a partir de los vectores $\mathbf{Y}^{(\ell)}$ de cada uno de los clusters, mientras que en

$$X^2(\mathbf{Y}, \hat{\boldsymbol{\theta}}_\phi) = nN(\hat{\mathbf{p}} - \mathbf{p}(\hat{\boldsymbol{\theta}}_\phi))^T \mathbf{D}_{\mathbf{p}(\hat{\boldsymbol{\theta}}_\phi)}^{-1} (\hat{\mathbf{p}} - \mathbf{p}(\hat{\boldsymbol{\theta}}_\phi)) = nN \sum_{r=1}^M \frac{(\hat{p}_r - p_r(\hat{\boldsymbol{\theta}}_\phi))^2}{p_r(\hat{\boldsymbol{\theta}}_\phi)}$$

las frecuencias se calculan sin distinguir los clusters.

Si se tiene en cuenta los resultados anteriores y se considera la familia de estadísticos de contraste basados en medidas de ϕ -divergencia introducida en (4.2) se llega, en el siguiente Teorema, a la distribución asintótica de dichos estadísticos.

Teorema 4.2 *El estadístico de contraste semiparamétrico para la bondad de ajuste en modelos loglineales con datos agrupados en clusters de igual tamaño y sobredispersión, bajo la hipótesis nula H_0 dada en (4.1), es decir, $\mathbf{p}(\boldsymbol{\theta})$ verifica el modelo loglineal asociado a la matriz de diseño \mathbf{W} , y se cumplen además las condiciones de regularidad de Birch establecidas en la Sección 2.1.1, verifica que*

$$\frac{T^{\phi_1}(\mathbf{Y}, \hat{\boldsymbol{\theta}}_{\phi_2})}{\tilde{\vartheta}_{n,N}} \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \chi_{M-M_0-1}^2 \quad (4.9)$$

donde

$$T^{\phi_1}(\mathbf{Y}, \hat{\boldsymbol{\theta}}_{\phi_2}) = \frac{2Nn}{\phi_1''(1)} d_{\phi_1}(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2})).$$

Demostración Primeramente, se expresará $d_{\phi_1}(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}))$ utilizando el desarrollo en serie de Taylor de la función ϕ_1 -divergencia en el punto $(\mathbf{p}(\boldsymbol{\theta}_0), \mathbf{p}(\boldsymbol{\theta}_0))$, para lo cual se requieren las derivadas parciales de primer y segundo orden. Al ser

$$d_{\phi_1}(\mathbf{p}, \mathbf{q}) = \sum_{j=1}^M q_j \phi_1\left(\frac{p_j}{q_j}\right),$$

se tiene

$$\frac{\partial d_{\phi_1}(\mathbf{p}, \mathbf{q})}{\partial p_j} = q_j \phi_1'\left(\frac{p_j}{q_j}\right) \frac{1}{q_j} = \phi_1'\left(\frac{p_j}{q_j}\right),$$

y

$$\frac{\partial d_{\phi_1}(\mathbf{p}, \mathbf{q})}{\partial q_j} = \phi_1\left(\frac{p_j}{q_j}\right) - \frac{p_j}{q_j} \phi_1'\left(\frac{p_j}{q_j}\right).$$

Luego, en el punto $(\mathbf{p}, \mathbf{q}) = (\mathbf{p}(\boldsymbol{\theta}_0), \mathbf{p}(\boldsymbol{\theta}_0))$, al ser $p_j = q_j = p_j(\boldsymbol{\theta}_0)$ y $\phi_1(1) = \phi_1'(1) = 0$, las derivadas primeras del desarrollo en serie de Taylor se anulan. Respecto a las segundas derivadas

$$\begin{aligned} \frac{\partial^2 d_{\phi_1}(\mathbf{p}, \mathbf{q})}{\partial p_i \partial p_j} &= \begin{cases} \phi_1''\left(\frac{p_j}{q_j}\right) \frac{1}{q_j} & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}, \\ \frac{\partial^2 d_{\phi_1}(\mathbf{p}, \mathbf{q})}{\partial q_i \partial q_j} &= \begin{cases} \phi_1''\left(\frac{p_j}{q_j}\right) \frac{p_j^2}{q_j^3} & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}, \\ \frac{\partial^2 d_{\phi_1}(\mathbf{p}, \mathbf{q})}{\partial p_i \partial q_j} &= \begin{cases} -\phi_1''\left(\frac{p_j}{q_j}\right) \frac{p_j}{q_j^2} & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}, \end{aligned}$$

se tiene

$$\begin{aligned} \left. \frac{\partial^2 d_{\phi_1}(\mathbf{p}, \mathbf{p}(\boldsymbol{\theta}_0))}{\partial \mathbf{p} \partial \mathbf{p}^T} \right|_{\mathbf{p}=\mathbf{p}(\boldsymbol{\theta}_0)} &= \left. \frac{\partial^2 d_{\phi_1}(\mathbf{p}(\boldsymbol{\theta}_0), \mathbf{q})}{\partial \mathbf{q} \partial \mathbf{q}^T} \right|_{\mathbf{q}=\mathbf{p}(\boldsymbol{\theta}_0)} = \left. \frac{\partial^2 d_{\phi_1}(\mathbf{p}, \mathbf{q})}{\partial \mathbf{q} \partial \mathbf{p}^T} \right|_{\mathbf{p}=\mathbf{p}(\boldsymbol{\theta}_0), \mathbf{q}=\mathbf{p}(\boldsymbol{\theta}_0)} \\ &= \phi_1''(1) \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1}. \end{aligned} \quad (4.10)$$

Por tanto,

$$\begin{aligned}
d_{\phi_1}(\widehat{\mathbf{p}}, \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2})) &= d_{\phi_1}(\mathbf{p}(\boldsymbol{\theta}_0), \mathbf{p}(\boldsymbol{\theta}_0)) + \left. \frac{\partial d_{\phi_1}(\mathbf{p}, \mathbf{p}(\boldsymbol{\theta}_0))}{\partial \mathbf{p}^T} \right|_{\mathbf{p}=\mathbf{p}(\boldsymbol{\theta}_0)} (\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) \\
&+ \left. \frac{\partial d_{\phi_1}(\mathbf{p}(\boldsymbol{\theta}_0), \mathbf{q})}{\partial \mathbf{q}^T} \right|_{\mathbf{q}=\mathbf{p}(\boldsymbol{\theta}_0)} (\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}) - \mathbf{p}(\boldsymbol{\theta}_0)) \\
&+ \frac{1}{2}(\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0))^T \left. \frac{\partial^2 d_{\phi_1}(\mathbf{p}, \mathbf{p}(\boldsymbol{\theta}_0))}{\partial \mathbf{p} \partial \mathbf{p}^T} \right|_{\mathbf{p}=\mathbf{p}(\boldsymbol{\theta}_0)} (\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) \\
&+ \frac{1}{2}(\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}) - \mathbf{p}(\boldsymbol{\theta}_0))^T \left. \frac{\partial^2 d_{\phi_1}(\mathbf{p}(\boldsymbol{\theta}_0), \mathbf{q})}{\partial \mathbf{q} \partial \mathbf{q}^T} \right|_{\mathbf{q}=\mathbf{p}(\boldsymbol{\theta}_0)} (\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}) - \mathbf{p}(\boldsymbol{\theta}_0)) \\
&+ (\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0))^T \left. \frac{\partial^2 d_{\phi_1}(\mathbf{p}, \mathbf{q})}{\partial \mathbf{q} \partial \mathbf{p}^T} \right|_{(\mathbf{p}=\mathbf{p}(\boldsymbol{\theta}_0), \mathbf{q}=\mathbf{p}(\boldsymbol{\theta}_0))} (\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}) - \mathbf{p}(\boldsymbol{\theta}_0)) \\
&+ o(\|\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)\|^2) + o(\|\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}) - \mathbf{p}(\boldsymbol{\theta}_0)\|^2) \\
&= \frac{1}{2} \phi_1''(1) (\widehat{\mathbf{p}} - \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}))^T \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} (\widehat{\mathbf{p}} - \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2})) + o(\|\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)\|^2).
\end{aligned}$$

Por el Teorema 4.1 se tiene que $o(N \|\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)\|^2) = o_p(1)$, y así

$$\frac{T^{\phi_1}(\mathbf{Y}, \widehat{\boldsymbol{\theta}}_{\phi_2})}{\widetilde{\vartheta}_{n,N}} = \frac{1}{\widetilde{\vartheta}_{n,N}} \frac{2Nn}{\phi_1''(1)} d_{\phi_1}(\widehat{\mathbf{p}} - \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2})) = \frac{1}{\widetilde{\vartheta}_{n,N}} \mathbf{Q}\mathbf{Q}^T + o_p(1), \quad (4.11)$$

lo que significa que $T^{\phi_1}(\mathbf{Y}, \widehat{\boldsymbol{\theta}}_{\phi_2})/\widetilde{\vartheta}_{n,N}$ y $X^2(\mathbf{Y}, \widehat{\boldsymbol{\theta}}_{\phi_2})/\widetilde{\vartheta}_{n,N}$ tienen la misma distribución asintótica, $\chi_{M-M_0-1}^2$, según el Corolario 4.1. ■

El siguiente resultado es un caso particular del Teorema anterior en el que $\phi_1 = x \log x - x + 1$.

Corolario 4.2 *El estadístico de contraste semiparamétrico del cociente de verosimilitudes para la bondad de ajuste para datos agrupados en clusters de igual tamaño, tiene la siguiente distribución asintótica*

$$\frac{G^2(\mathbf{Y}, \widehat{\boldsymbol{\theta}}_{\phi_2})}{\widetilde{\vartheta}_{n,N}} \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \chi_{M-M_0-1}^2, \quad (4.12)$$

donde

$$G^2(\mathbf{Y}, \widehat{\boldsymbol{\theta}}_{\phi_2}) = nN \sum_{r=1}^M \widehat{p}_r \log \frac{\widehat{p}_r}{p_r(\widehat{\boldsymbol{\theta}}_{\phi_2})}. \quad (4.13)$$

Observación 4.2 *El estimador del efecto del diseño denotado mediante $\tilde{\vartheta}_{n,N}$ puede ser bien, el estimador no-paramétrico $\tilde{\vartheta}_{n,N,\bullet}$, o bien, el semiparamétrico $\tilde{\vartheta}_{n,N,\phi_2}$, ya que, bajo el supuesto de que el modelo loglineal se verifique, se espera que el estadístico semiparamétrico tenga mejores propiedades, tanto respecto al nivel de significación exacto, como respecto a la potencia. Esta cuestión se estudiará posteriormente mediante simulación.*

A partir del Teorema 4.2, se rechazará la hipótesis nula, de que los datos que se están analizando a través de un muestreo por clusters o conglomerados se pueden ajustar mediante un modelo loglineal dado, si

$$T^{\phi_1}(\widehat{\boldsymbol{\theta}}_{\phi_2}, \tilde{\vartheta}_{n,N}) > \chi_{M-M_0-1, \alpha}^2. \quad (4.14)$$

Para determinar la potencia de dicho contraste, sea $\mathbf{q} = (q_1, \dots, q_M)^T$ un punto en la hipótesis alternativa, es decir se considera en la alternativa el modelo loglineal saturado asociado a \mathbf{q} . Si se supone que la hipótesis alternativa es cierta, entonces

$$\widehat{\mathbf{p}} \xrightarrow[N \rightarrow \infty]{p} \mathbf{q}.$$

Sea $\boldsymbol{\theta}_a$ el punto de Θ verificando que

$$\boldsymbol{\theta}_a = \arg \min_{\boldsymbol{\theta} \in \Theta} d_{\phi_2}(\mathbf{q}, \mathbf{p}(\boldsymbol{\theta})),$$

luego, debido a la consistencia de los estimadores de mínima divergencia se cumple que

$$\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}) \xrightarrow[N \rightarrow \infty]{p} \mathbf{p}(\boldsymbol{\theta}_a).$$

En lo que sigue, se tendrá en cuenta que

$$\sqrt{N}((\widehat{\mathbf{p}}, \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2})) - (\mathbf{q}, \mathbf{p}(\boldsymbol{\theta}_a))) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} N(\mathbf{0}, \boldsymbol{\Sigma}) \quad (4.15)$$

bajo la hipótesis alternativa, donde

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \quad \text{con} \quad \boldsymbol{\Sigma}_{11} = \text{diag}(\mathbf{q}) - \mathbf{q}\mathbf{q}^T \quad \text{y} \quad \boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}.$$

Teorema 4.3 *La potencia para el test dado por (4.14) en \mathbf{q} y suponiendo que se verifica la condición (4.15) se puede calcular de forma asintótica mediante la siguiente expresión*

$$\beta_{N,\phi_1}(\mathbf{q}) = 1 - \Phi_N \left(\frac{1}{\sigma_{\phi_1}(\mathbf{q})} \left(\frac{\phi_1''(1) \tilde{\vartheta}_{n,N}}{2n\sqrt{N}} \chi_{M-M_0-1,\alpha}^2 - \sqrt{N} d_{\phi_1}(\mathbf{q}, \mathbf{p}(\boldsymbol{\theta}_a)) \right) \right) \quad (4.16)$$

donde

$$\sigma_{\phi_1}^2 = \mathbf{Z}^T \boldsymbol{\Sigma}_{11} \mathbf{Z} + 2\mathbf{Z}^T \boldsymbol{\Sigma}_{12} \mathbf{S} + \mathbf{S}^T \boldsymbol{\Sigma}_{22} \mathbf{S},$$

siendo

$$\mathbf{Z} = \left. \frac{\partial d_{\phi_1}(\mathbf{v}, \mathbf{p}(\boldsymbol{\theta}_a))}{\partial \mathbf{v}^T} \right|_{\mathbf{v}=\mathbf{q}},$$

$$\mathbf{S} = \left. \frac{\partial d_{\phi_1}(\mathbf{q}, \mathbf{w})}{\partial \mathbf{w}^T} \right|_{\mathbf{w}=\mathbf{p}(\boldsymbol{\theta}_a)},$$

y $\Phi_N(x)$ una sucesión de funciones de distribución que convergen uniformemente a la función de distribución $\Phi(x)$ de una distribución normal de media cero y varianza 1.

Demostración El desarrollo en serie de Taylor de primer orden de la medida de divergencia en el punto $(\mathbf{q}, \mathbf{p}(\boldsymbol{\theta}_a))$, viene dado por

$$d_{\phi_1}(\widehat{\mathbf{p}}, \widehat{\mathbf{p}}(\widehat{\boldsymbol{\theta}}_{\phi_2})) = d_{\phi_1}(\mathbf{q}, \mathbf{p}(\boldsymbol{\theta}_a)) + \mathbf{Z}^T (\widehat{\mathbf{p}} - \mathbf{q}) + \mathbf{S}^T (\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}) - \mathbf{p}(\boldsymbol{\theta}_a)) + O_p(N^{-1/2}).$$

Entonces, en base a (4.15), se tiene

$$\sqrt{N} \left(d_{\phi_1}(\widehat{\mathbf{p}}, \widehat{\mathbf{p}}(\widehat{\boldsymbol{\theta}}_{\phi_2})) - d_{\phi_1}(\mathbf{q}, \mathbf{p}(\boldsymbol{\theta}_a)) \right) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} N \left(0, \sigma_{\phi_1}^2(\mathbf{q}) \right). \quad (4.17)$$

Este resultado se utiliza para encontrar la potencia asintótica del contraste, puesto que

$$\begin{aligned}\beta_{N,\phi_1}(\mathbf{q}) &= P\left(T^{\phi_1}(\widehat{\boldsymbol{\theta}}_{\phi_2}, \widetilde{\vartheta}_{n,N}) > \chi_{M-M_0-1,\alpha}^2 / H_1 : \mathbf{p} = \mathbf{q}\right) \\ &= P\left(\frac{2Nn}{\phi_1''(1)\widetilde{\vartheta}_{n,N}} d_{\phi_1}(\widehat{\mathbf{p}}, \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2})) > \chi_{M-M_0-1,\alpha}^2 / H_1 : \mathbf{p} = \mathbf{q}\right),\end{aligned}$$

o equivalentemente

$$\begin{aligned}\beta_{N,\phi_1}(\mathbf{q}) &= P\left(\frac{2Nn}{\phi_1''(1)\widetilde{\vartheta}_{n,N}} d_{\phi_1}(\widehat{\mathbf{p}}, \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2})) - \frac{2Nn}{\phi_1''(1)\widetilde{\vartheta}_{n,N}} d_{\phi_1}(\mathbf{q}, \mathbf{p}(\boldsymbol{\theta}_a))\right. \\ &\quad \left.> \chi_{M-M_0-1,\alpha}^2 - \frac{2Nn}{\phi_1''(1)\widetilde{\vartheta}_{n,N}} d_{\phi_1}(\mathbf{q}, \mathbf{p}(\boldsymbol{\theta}_a)) / H_1 : \mathbf{p} = \mathbf{q}\right).\end{aligned}$$

Para simplificar, a partir de ahora se supone que se verifica $H_1 : \mathbf{p} = \mathbf{q}$, pero no se explicita el condicionante en la fórmula. Luego, en esta situación

$$\begin{aligned}\beta_{N,\phi_1}(\mathbf{q}) &= P\left(\sqrt{N}\left(d_{\phi_1}(\widehat{\mathbf{p}}, \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2})) - d_{\phi_1}(\mathbf{q}, \mathbf{p}(\boldsymbol{\theta}_a))\right)\right. \\ &\quad \left.> \frac{\phi_1''(1)\widetilde{\vartheta}_{n,N}}{2n\sqrt{N}}\left(\chi_{M-M_0-1,\alpha}^2 - \frac{2Nn}{\phi_1''(1)\widetilde{\vartheta}_{n,N}} d_{\phi_1}(\mathbf{q}, \mathbf{p}(\boldsymbol{\theta}_a))\right)\right) \\ &= P\left(\frac{\sqrt{N}}{\sigma_{\phi_1}(\mathbf{q})}\left(d_{\phi_1}(\widehat{\mathbf{p}}, \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2})) - d_{\phi_1}(\mathbf{q}, \mathbf{p}(\boldsymbol{\theta}_a))\right)\right. \\ &\quad \left.> \frac{\phi_1''(1)\widetilde{\vartheta}_{n,N}}{2n\sqrt{N}}\left(\chi_{M-M_0-1,\alpha}^2 - \frac{2Nn}{\phi_1''(1)\widetilde{\vartheta}_{n,N}} d_{\phi_1}(\mathbf{q}, \mathbf{p}(\boldsymbol{\theta}_a))\right) \frac{1}{\sigma_{\phi_1}(\mathbf{q})}\right),\end{aligned}$$

y teniendo en cuenta (4.17) se tiene (4.16). ■

El resultado obtenido en el Teorema anterior no es fácil de aplicar por la dificultad que conlleva la obtención de $\boldsymbol{\Sigma}_{12}$ y $\boldsymbol{\Sigma}_{22}$. Por esta razón, se va a considerar una secuencia de alternativas contiguas que aproxima la hipótesis nula con una tasa $O_p(N^{-1/2})$. Sean las hipótesis alternativas contiguas

$$H_{1,N} : \mathbf{p}_N = \mathbf{p}(\boldsymbol{\theta}_0) + \frac{1}{\sqrt{N}} \mathbf{d}, \quad (4.18)$$

donde $\mathbf{d} = (d_1, \dots, d_M)^T$ es un vector fijo tal que $\sum_{r=1}^M d_r = 0$. Seguidamente se presentará una aproximación para la función de potencia del test de hipótesis dado en (4.14).

Teorema 4.4 *La distribución asintótica del estadístico de contraste $T^{\phi_1}(\widehat{\boldsymbol{\theta}}_{\phi_2}, \widetilde{\vartheta}_{n,N})$ bajo las hipótesis alternativas contiguas dadas en (4.18) es una ji-cuadrado con $M - M_0 - 1$ grados de libertad y parámetro de no centralidad*

$$\delta = \mathbf{d}^T \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} (\mathbf{I} - \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0)) \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \mathbf{d}.$$

Demostración Se tiene que

$$\begin{aligned} \sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) &= \sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p}_N + \mathbf{p}_N - \mathbf{p}(\boldsymbol{\theta}_0)) \\ &= \sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p}_N) + \sqrt{N}(\mathbf{p}_N - \mathbf{p}(\boldsymbol{\theta}_0)) \\ &= \sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p}_N) + \mathbf{d}, \end{aligned} \tag{4.19}$$

entonces bajo las alternativas (4.18) se tiene,

$$\sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}\left(\mathbf{d}, \frac{\vartheta_n}{n} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}\right),$$

porque bajo dichas alternativas

$$\sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p}_N) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \frac{\vartheta_n}{n} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}\right).$$

Teniendo en cuenta el resultado (2.17) del Capítulo 2

$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_{\phi_2} - \boldsymbol{\theta}_0) = (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) + o(\mathbf{1}_{M_0}),$$

y (4.19) se tiene

$$\begin{aligned}
\sqrt{N}(\widehat{\boldsymbol{\theta}}_{\phi_2} - \boldsymbol{\theta}_0) &= (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \sqrt{N} \left((\widehat{\mathbf{p}} - \mathbf{p}_N) + \mathbf{d}/\sqrt{N} \right) + o(\mathbf{1}_{M_0}) \\
&= (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \sqrt{N} (\widehat{\mathbf{p}} - \mathbf{p}_N) \\
&\quad + (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \mathbf{d} + o(\mathbf{1}_{M_0}).
\end{aligned}$$

Por otra parte,

$$\begin{aligned}
\sqrt{N}(\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}) - \mathbf{p}(\boldsymbol{\theta}_0)) &= \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{1/2} \mathbf{A}(\boldsymbol{\theta}_0) \sqrt{N}(\widehat{\boldsymbol{\theta}}_{\phi_2} - \boldsymbol{\theta}_0) + o(\mathbf{1}_M) \\
&= \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \sqrt{N} (\widehat{\mathbf{p}} - \mathbf{p}_N) \\
&\quad + \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{d} + o(\mathbf{1}_M).
\end{aligned}$$

Si se denota

$$\mathbf{L} = \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0),$$

se tiene

$$\sqrt{N} \left(\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}) - \mathbf{p}(\boldsymbol{\theta}_0) \right) = \sqrt{N} \mathbf{L} (\widehat{\mathbf{p}} - \mathbf{p}_N) + \mathbf{L} \mathbf{d} + o(\mathbf{1}_M)$$

y

$$\sqrt{N} \begin{pmatrix} \widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0) \\ \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}) - \mathbf{p}(\boldsymbol{\theta}_0) \end{pmatrix} = \begin{pmatrix} \mathbf{I}_M \\ \mathbf{L} \end{pmatrix} (\widehat{\mathbf{p}} - \mathbf{p}_N) + \begin{pmatrix} \mathbf{d} \\ \mathbf{L} \mathbf{d} \end{pmatrix} + \begin{pmatrix} \mathbf{0}_M \\ o(\mathbf{1}_M) \end{pmatrix}.$$

Así

$$\sqrt{N} \begin{pmatrix} \widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0) \\ \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}) - \mathbf{p}(\boldsymbol{\theta}_0) \end{pmatrix} \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N} \left(\begin{pmatrix} \mathbf{d} \\ \mathbf{L} \mathbf{d} \end{pmatrix}, \begin{pmatrix} \frac{\vartheta_n}{n} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} & \frac{\vartheta_n}{n} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{L}^T \\ \frac{\vartheta_n}{n} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{L}^T & \mathbf{L} \frac{\vartheta_n}{n} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{L}^T \end{pmatrix} \right),$$

y

$$\sqrt{N} \left(\widehat{\mathbf{p}} - \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}) \right) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N} \left((\mathbf{I}_M - \mathbf{L}) \mathbf{d}, \frac{\vartheta_n}{n} (\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} - \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{L}^T - \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{L}^T + \mathbf{L} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{L}^T) \right).$$

Por otro lado, tal y como se vió en (4.11), se verifica

$$T^{\phi_1}(\widehat{\boldsymbol{\theta}}_{\phi_2}) = \frac{2Nn}{\phi_1''(1) \vartheta_n} d_{\phi_1}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}_{\phi_2})) = \boldsymbol{Q}\boldsymbol{Q}^T + o_p(1),$$

donde

$$\boldsymbol{Q} = \sqrt{\frac{n}{\vartheta_n}} \boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}^{-1/2} \sqrt{N} \left(\widehat{\boldsymbol{p}} - \boldsymbol{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}) \right) \quad (4.20)$$

y en este caso, bajo las hipótesis alternativas contiguas se tiene que

$$\boldsymbol{Q} = \sqrt{\frac{n}{\vartheta_n}} \boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}^{-1/2} \sqrt{N} \left(\widehat{\boldsymbol{p}} - \boldsymbol{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}) \right) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}^{-1/2} (\boldsymbol{I}_M - \boldsymbol{L}) \boldsymbol{d}, \boldsymbol{V}(\boldsymbol{\theta}_0)). \quad (4.21)$$

Puesto que la matriz $\boldsymbol{V}(\boldsymbol{\theta}_0)$ es idempotente de rango $M - 1 - M_0$, si se comprueba que $\boldsymbol{V}(\boldsymbol{\theta}_0)\boldsymbol{\mu} = \boldsymbol{\mu}$ quedará demostrado que $T^{\phi_1}(\widehat{\boldsymbol{\theta}}_{\phi_2}, \widetilde{\vartheta}_{n,N})$, bajo las hipótesis alternativas contiguas dadas en (4.18), es una ji-cuadrado con $M - M_0 - 1$ grados de libertad y parámetro de no centralidad $\delta = \boldsymbol{\mu}^T \boldsymbol{\mu}$. Teniendo en cuenta que

$$\begin{aligned} \boldsymbol{V}(\boldsymbol{\theta}_0)\boldsymbol{\mu} &= \left(\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{p}(\boldsymbol{\theta}_0)} - \boldsymbol{A}(\boldsymbol{\theta}_0) (\boldsymbol{A}^T(\boldsymbol{\theta}_0) \boldsymbol{A}(\boldsymbol{\theta}_0))^{-1} \boldsymbol{A}^T(\boldsymbol{\theta}_0) \right) \boldsymbol{\mu} \\ &= \boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{p}(\boldsymbol{\theta}_0)} \boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}^{-1/2} (\boldsymbol{I}_M - \boldsymbol{L}) \boldsymbol{d} \\ &\quad - \boldsymbol{A}(\boldsymbol{\theta}_0) (\boldsymbol{A}^T(\boldsymbol{\theta}_0) \boldsymbol{A}(\boldsymbol{\theta}_0))^{-1} \boldsymbol{A}^T(\boldsymbol{\theta}_0) \boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}^{-1/2} (\boldsymbol{I}_M - \boldsymbol{L}) \boldsymbol{d} \\ &= \boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{p}(\boldsymbol{\theta}_0)} \boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}^{-1/2} \boldsymbol{d} \\ &\quad - \boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{p}(\boldsymbol{\theta}_0)} \boldsymbol{A}(\boldsymbol{\theta}_0) (\boldsymbol{A}^T(\boldsymbol{\theta}_0) \boldsymbol{A}(\boldsymbol{\theta}_0))^{-1} \boldsymbol{A}^T(\boldsymbol{\theta}_0) \boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}^{-1/2} \boldsymbol{d} \\ &\quad - \boldsymbol{A}(\boldsymbol{\theta}_0) (\boldsymbol{A}^T(\boldsymbol{\theta}_0) \boldsymbol{A}(\boldsymbol{\theta}_0))^{-1} \boldsymbol{A}^T(\boldsymbol{\theta}_0) \boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}^{-1/2} \boldsymbol{d} \\ &\quad + \boldsymbol{A}(\boldsymbol{\theta}_0) (\boldsymbol{A}^T(\boldsymbol{\theta}_0) \boldsymbol{A}(\boldsymbol{\theta}_0))^{-1} \boldsymbol{A}^T(\boldsymbol{\theta}_0) \boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}^{-1/2} \boldsymbol{d}, \end{aligned}$$

si se utilizan las relaciones (4.4) y (4.5) se obtiene

$$\begin{aligned} \boldsymbol{V}(\boldsymbol{\theta}_0)\boldsymbol{\mu} &= \boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{p}(\boldsymbol{\theta}_0)} \boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}^{-1/2} \boldsymbol{d} - \boldsymbol{A}(\boldsymbol{\theta}_0) (\boldsymbol{A}^T(\boldsymbol{\theta}_0) \boldsymbol{A}(\boldsymbol{\theta}_0))^{-1} \boldsymbol{A}^T(\boldsymbol{\theta}_0) \boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}^{-1/2} \boldsymbol{d} \\ &= \boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}^{-1/2} \left(\boldsymbol{I}_M - \boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}^{1/2} \boldsymbol{A}(\boldsymbol{\theta}_0) (\boldsymbol{A}^T(\boldsymbol{\theta}_0) \boldsymbol{A}(\boldsymbol{\theta}_0))^{-1} \boldsymbol{A}^T(\boldsymbol{\theta}_0) \boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}^{-1/2} \right) \boldsymbol{d} = \boldsymbol{\mu}. \end{aligned}$$

Luego, el parámetro de no centralidad es

$$\delta = \mathbf{d}^T \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} (\mathbf{I}_M - \mathbf{L}^T) (\mathbf{I}_M - \mathbf{L}) \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \mathbf{d}.$$

Obsérvese que

$$(\mathbf{I}_M - \mathbf{L}) = \mathbf{I}_M - \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0),$$

donde

$$\mathbf{U} = \mathbf{I}_M - \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0)$$

de modo que

$$\begin{aligned} \mathbf{U}^T \mathbf{U} &= \mathbf{I}_M - \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \\ &\quad - \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \\ &\quad + \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0) \\ &\quad \times (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \\ &= \mathbf{I}_M - \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \end{aligned}$$

y

$$\begin{aligned} \delta &= \mathbf{d}^T \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \mathbf{U} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \mathbf{d} \\ &= \mathbf{d}^T \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \left(\mathbf{I} - \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{A}^T(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \mathbf{A}^T(\boldsymbol{\theta}_0) \right) \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \mathbf{d}. \end{aligned}$$

■

Basado en el Teorema anterior, se tiene que la potencia para el contraste dado en (4.14) en $\mathbf{p}_N = \mathbf{p}(\boldsymbol{\theta}_0) + \frac{1}{\sqrt{N}}\mathbf{d}$, viene dada por

$$\begin{aligned}\beta_{N,\phi_1,\phi_2}(\mathbf{p}_N) &= P\left(T^{\phi_1}(\widehat{\boldsymbol{\theta}}_{\phi_2}, \widetilde{\vartheta}_{n,N}) > \chi_{M-M_0-1,\alpha}^2 / H_{1,N} : \mathbf{p} = \mathbf{p}_N\right) \\ &= 1 - \chi_{M-M_0-1;\delta}^2(\chi_{M-M_0-1,\alpha}^2),\end{aligned}$$

donde por $\chi_{M-M_0-1;\delta}^2(a)$, se denota la función de distribución de una ji-cuadrado no centrada con $M - M_0 - 1$ grados de libertad y parámetro de centralización δ evaluada en el punto a .

Observación 4.3 *El Teorema anterior se puede utilizar para obtener una aproximación de la función de potencia en un punto \mathbf{q} de la alternativa. En efecto, \mathbf{q} se puede escribir como*

$$\mathbf{q} = \mathbf{p}(\boldsymbol{\theta}_0) + \frac{1}{\sqrt{N}}\sqrt{N}(\mathbf{q} - \mathbf{p}(\boldsymbol{\theta}_0)) = \mathbf{p}(\boldsymbol{\theta}_0) + \frac{1}{\sqrt{N}}\mathbf{d},$$

siendo

$$\mathbf{d} = \sqrt{N}(\mathbf{q} - \mathbf{p}(\boldsymbol{\theta}_0)).$$

4.2. Aplicaciones numéricas.

En el Capítulo 2 se estudió una muestra formada por $N = 71$ parejas de hermanos, si se agrupan todos los individuos en una tabla de contingencia sin distinguir los clusters a los que pertenecen, se tiene la Tabla 2.6. Se desea determinar si el diagnóstico de Esquizofrenia (X_2) es independiente del sexo (X_1). Para ello se plantea el modelo loglineal no saturado con hipótesis de independencia

$$p_{ij}(\boldsymbol{\theta}) = \exp\{u + \theta_{1(i)} + \theta_{2(j)}\} \quad \text{con } i, j = 1, 2,$$

donde el efecto de la variable Sexo viene representado por $\theta_{1(i)}$, y el diagnóstico de la enfermedad por $\theta_{2(k)}$ con la condición

$$\sum_{i=1}^2 \theta_{1(i)} = \sum_{j=1}^2 \theta_{2(j)} = 0. \quad (4.22)$$

Con el modelo así planteado, los dos parámetros de interés, $\beta = (\rho^2, \mathbf{p}(\theta))^T$, se estiman por

$$\hat{\beta}_{n,N,\lambda} = (\hat{\rho}_{n,N,\lambda}^2, \mathbf{p}(\hat{\theta}_{\phi_\lambda}))^T = \left(\frac{\tilde{\vartheta}_{n,N,\lambda} - 1}{n - 1}, \frac{\exp\{\mathbf{W}\hat{\theta}_{\phi_\lambda}\}}{\mathbf{1}_M^T \exp\{\mathbf{W}\hat{\theta}_{\phi_\lambda}\}} \right)^T,$$

mediante los datos de las Tablas 2.5 y 2.6 respectivamente. Las estimaciones obtenidas se resumen en la Tabla 4.1.

Tabla 4.1. Estimaciones de las probabilidades y de ρ^2 para diferentes valores de λ y no paramétricas.

λ	$-\frac{1}{2}$	0	$\frac{2}{3}$	1	2	No paramétrico
$\rho_{n,N}^2(\tilde{\mathbf{Y}}, \hat{\theta}_\phi)$	0.2738	0.2670	0.2657	0.2638	0.2603	0.2926
$\hat{p}_{11}(\theta)$	0.1888	0.1927	0.1982	0.2014	0.2110	0.1056
$\hat{p}_{12}(\theta)$	0.2810	0.2791	0.2779	0.2773	0.2767	0.3662
$\hat{p}_{21}(\theta)$	0.2131	0.2157	0.2181	0.2193	0.2216	0.3028
$\hat{p}_{22}(\theta)$	0.3171	0.3124	0.3058	0.3020	0.2907	0.2253

El estadístico para contrastar la bondad del ajuste

$$T^{\phi_1}(\hat{\theta}_{\phi_2}, \tilde{\vartheta}_{n,N}) = \frac{2Nn}{\phi_1''(1)\tilde{\vartheta}_{n,N}} d_{\phi_1}(\hat{\mathbf{p}}, \mathbf{p}(\hat{\theta}_{\phi_2}))$$

se puede calcular teniendo en cuenta que el estimador del efecto del diseño $\tilde{\vartheta}_{n,N}$ puede ser bien, $\tilde{\vartheta}_{n,N,\bullet}$ (Tabla 4.2) o bien, $\tilde{\vartheta}_{n,N,\lambda_2}$ (Tabla 4.3).

Obsérvese que el estadístico de Pearson corregido por el efecto del diseño calculado por Cohen (1976), sobre estos mismos datos $W' = 13.751$, es uno de los estadísticos $T^{\phi_1}(\hat{\theta}_{\phi_2}, \tilde{\vartheta}_{n,N}) = 13.8362$ calculado cuando las funciones ϕ_1, ϕ_2 son $\phi_{\lambda=1}$ y $\phi_{\lambda=0}$ respectivamente. En este caso, debido a la fuerte dependencia que existe entre la enfermedad y

Tabla 4.2. Valores del estadístico $T^{\phi_1}(\hat{\theta}_{\phi_2}, \tilde{\vartheta}_{n,N})$ y su p-valor.

ϕ_1, ϕ_2	$\lambda_2 = -\frac{1}{2}$	$\lambda_2 = 0$	$\lambda_2 = \frac{2}{3}$	$\lambda_2 = 1$	$\lambda_2 = 2$
$\tilde{\vartheta}_{n,N,\bullet}$	1.2926	1.2926	1.2926	1.2926	1.2926
$\lambda_1 = -0.5$	14.6358 (0.0001)	14.6585 (0.0001)	14.7581 (0.0001)	14.8526 (0.0001)	15.2797 (0.0001)
$\lambda_1 = 0$	14.2990 (0.0002)	14.2775 (0.0002)	14.3130 (0.0002)	14.3657 (0.0002)	14.6509 (0.0001)
$\lambda_1 = \frac{2}{3}$	14.0116 (0.0002)	13.9392 (0.0002)	13.9046 (0.0002)	13.9128 (0.0002)	14.0502 (0.0002)
$\lambda_1 = 1$	13.9318 (0.0002)	13.8362 (0.0002)	13.7713 (0.0002)	13.7608 (0.0002)	13.8376 (0.0002)
$\lambda_1 = 2$	13.9289 (0.0002)	13.7672 (0.0002)	13.6231 (0.0002)	13.5655 (0.0002)	13.4965 (0.0002)

el sexo, aunque no se hubiera tenido en cuenta la corrección por el efecto del diseño, en cuyo caso el estadístico de Pearson $X^2 = 17.885$, el resultado del contraste es el mismo, es decir rechazamos la independencia, como ocurre para todos los estadísticos de bondad de ajuste utilizados.

Tabla 4.3. Valores del estadístico $T^{\phi_1}(\hat{\theta}_{\phi_2}, \tilde{\vartheta}_{n,N})$ y su p-valor..

ϕ_1, ϕ_2	$\lambda_2 = -\frac{1}{2}$	$\lambda_2 = 0$	$\lambda_2 = \frac{2}{3}$	$\lambda_2 = 1$	$\lambda_2 = 2$
$\tilde{\vartheta}_{n,N,\lambda_2}$	1.2738	1.2697	1.2657	1.2638	1.2602
$\lambda_1 = -0.5$	14.8518 (0.0001)	14.9231 (0.0001)	15.0720 (0.0001)	15.1911 (<0.0001)	15.6716 (<0.0001)
$\lambda_1 = 0$	14.5101 (0.0001)	14.5352 (0.0001)	14.6174 (0.0001)	14.6932 (0.0001)	15.0267 (0.0001)
$\lambda_1 = \frac{2}{3}$	14.2184 (0.0001)	14.1908 (0.0001)	14.2003 (0.0002)	14.2300 (0.0002)	14.4106 (0.0002)
$\lambda_1 = 1$	14.1375 (0.0002)	14.0859 (0.0002)	14.0642 (0.0002)	14.0745 (0.0002)	14.1925 (0.0002)
$\lambda_1 = 2$	14.1345 (0.0002)	14.0157 (0.0002)	13.9128 (0.0002)	13.8747 (0.0002)	13.8427 (0.0002)

En todos los casos se rechaza la hipótesis nula de independencia puesto que el valor $\chi_{M-M_0-1,\alpha}^2 = \chi_{1,0,05}^2 = 3.841$ y todos los p-valores son menores de 0.0002. Técnicas muy similares a esta son de gran interés en estudios oftalmológicos. Los procedimientos que

se proponen en esta memoria mejoran con diferencia los propuestos por J.R. Thompson (1993) "The χ^2 for data collected on eyes". British Journal of Ophtalmology, 77, 155 – 117. La motivación que presentan tales estudios es que, si se analizara sólo la observación correspondiente a un ojo podrían utilizarse las técnicas habituales, pero el uso del doble de observaciones convierte la muestra mucho más eficiente. Sin embargo, es necesario tener en cuenta la correlación existente entre las medidas realizadas en un mismo individuo, o cluster. En este sentido los modelos aquí presentados pueden ser de gran utilidad para estas situaciones.

4.3. Estadísticos de ϕ -divergencia para bondad de ajuste con clusters de diferente tamaño

En el Capítulo 3, se estudió el caso en el que los clusters tienen diferentes tamaños. Se modelizó esta situación considerando que se tienen G grupos de $\{N_g\}_{g=1}^G$ clusters de forma que todas las tablas de contingencia, $\mathbf{Y}^{(g,\ell)} = (Y_1^{(g,\ell)}, \dots, Y_M^{(g,\ell)})^T$, del mismo grupo de N_g clusters tienen el mismo tamaño muestral n_g , $\ell = 1, \dots, N_g$, $g = 1, \dots, G$, y el número total de clusters es $N = \sum_{g=1}^G N_g$. Se supone que al menos existe un índice g tal que $n_g > 1$. Si se sustituye la hipótesis $N \rightarrow \infty$ por $N_g \rightarrow \infty$ para cada uno de los grupos de clusters, entonces todos los resultados establecidos en la sección anterior se cumplen de forma separada para cada uno de los grupos de clusters ya que en éstos sí se cumple la condición de tamaños de cluster iguales.

En esta situación, el estimador no paramétrico de $\mathbf{p}(\boldsymbol{\theta})$, basado en el total de los G grupos de N_1, \dots, N_G clusters con tamaños muestrales n_1, \dots, n_G se puede expresar como $\hat{\mathbf{p}} = \sum_{g=1}^G w_g \hat{\mathbf{p}}^{(g)}$, donde $\hat{\mathbf{p}}^{(g)}$ son los estimadores dentro de cada grupo de cluster de igual tamaño y w_g , $g = 1, \dots, G$, con $\sum_{g=1}^G w_g = 1$, son los pesos asociados a cada grupo de clusters. Además, suponiendo que existe una sucesión $\{N_h^*\}_{h=1}^G$, tal que

$$\frac{N_h}{N} \xrightarrow[N \rightarrow \infty]{P} N_h^* \in (0, 1],$$

se llama el tamaño muestral medio de cluster ponderado por la proporción de clusters de cada tamaño a

$$\bar{n} = \sum_{g=1}^G N_g^* n_g,$$

y análogamente se define el tamaño muestral medio ponderado por la proporción de individuos en cada grupo de clusters

$$n^* = \sum_{g=1}^G w_g^* n_g,$$

donde

$$w_g^* = \frac{N_g^* n_g}{\sum_{h=1}^G N_h^* n_h} > 0, \quad g = 1, \dots, G.$$

Definición 4.1 *En el caso de tamaños de cluster diferentes, con el nuevo parámetro, $\vartheta_{n^*} = 1 + \rho^2 (n^* - 1) \in (1, n^*]$, donde $n^* = \sum_{g=1}^G w_g^* n_g$, establecido para medir la sobredispersión se define el estadístico de contraste de bondad de ajuste como*

$$T^{\phi_1}(\widehat{\boldsymbol{\theta}}_{\phi_2}) = \frac{2N\widehat{\bar{n}}}{\phi_1''(1)\widetilde{\vartheta}_{n^*}} d_{\phi_1}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}_{\phi_2})). \quad (4.23)$$

4.3.1. Distribución asintótica

En este apartado se obtendrá la distribución asintótica de la familia de estadísticos de contraste dada en (4.23) para el caso en que los clusters o conglomerados tienen diferente tamaño. El resultado esencial se presenta en el siguiente teorema.

Teorema 4.5 *Sea $\mathbf{Y}^{(g,\ell)} = (Y_1^{(g,\ell)}, \dots, Y_M^{(g,\ell)})^T$, con $\ell = 1, \dots, N_g$ y $g = 1, \dots, G$, de forma que el tamaño muestral n_g de los clusters de cada grupo es el mismo, para $N = \sum_{g=1}^G N_g$ suficientemente grande*

$$E \left[\mathbf{Y}^{(g,\ell)} \right] = n_g \boldsymbol{p}(\boldsymbol{\theta}) \quad y \quad Var \left[\mathbf{Y}^{(\ell)} \right] = \vartheta_{n_g} n_g \boldsymbol{\Sigma}_{\boldsymbol{p}(\boldsymbol{\theta})},$$

con $\vartheta_{n^*} = 1 + \rho^2 (n^* - 1) \in (1, n^*]$, donde $n^* = \sum_{g=1}^G w_g^* n_g$. Entonces, bajo la hipótesis nula de que $\mathbf{p}(\boldsymbol{\theta})$ sigue el modelo loglineal asociado a la matriz de diseño \mathbf{W} , cumpliéndose además las condiciones de regularidad de Birch establecidas en la Sección 2.1.1, el estadístico (4.23) se distribuye como una $\chi_{M-M_0-1}^2$

Demostración Por el Teorema 3.3 del Capítulo 3 se sabe que, para un modelo loglineal con muestreo por conglomerados y tamaños desiguales, se tiene

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_{\phi} - \boldsymbol{\theta}_0) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_{M_0}, \frac{\vartheta_{n^*}}{\bar{n}} (\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W})^{-1})$$

y

$$\sqrt{N}(\mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi}) - \mathbf{p}(\boldsymbol{\theta}_0)) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_M, \frac{\vartheta_{n^*}}{\bar{n}} \mathbf{W} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} (\mathbf{W}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W})^{-1} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W}^T).$$

Luego, se demuestra de forma idéntica al caso de tamaños muestrales iguales que

$$T^{\phi_1}(\hat{\boldsymbol{\theta}}_{\phi_2}) = \frac{2N\bar{n}}{\phi_1''(1)\vartheta_{n^*}} d_{\phi_1}(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2})) = \mathbf{Q}\mathbf{Q}^T + o_p(1),$$

donde

$$\mathbf{Q} = \sqrt{\frac{\bar{n}}{\vartheta_{n^*}}} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \sqrt{N}(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2})) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_M, \mathbf{V}(\boldsymbol{\theta}_0)), \quad (4.24)$$

siendo $\mathbf{V}(\boldsymbol{\theta}_0)$ la matriz idempotente simétrica que se definió en el caso de tamaños de cluster iguales cuyo rango es $M - M_0 - 1$. ■

El siguiente resultado establece que, teniendo en cuenta el hecho de que en la práctica ϑ_{n^*} es desconocido, dicha constante puede estimarse por un ϕ -estimador consistente, en particular por cualquiera de los dos definidos en las Secciones 3.2 y 3.3 del Capítulo 3.

Corolario 4.3 Sea $\mathbf{Y}^{(g,\ell)} = (Y_1^{(g,\ell)}, \dots, Y_M^{(g,\ell)})^T$, con $\ell = 1, \dots, N_g$ y $g = 1, \dots, G$, de forma que el tamaño muestral n_g de los clusters de cada grupo es el mismo, para $N = \sum_{g=1}^G N_g$ suficientemente grande

$$E \left[\mathbf{Y}^{(g,\ell)} \right] = n_g \mathbf{p}(\boldsymbol{\theta}) \quad y \quad Var \left[\mathbf{Y}^{(\ell)} \right] = \vartheta_{n_g} n_g \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})},$$

con $\vartheta_{n^*} = 1 + \rho^2 (n^* - 1) \in (1, n^*]$, donde $n^* = \sum_{g=1}^G w_g n_g$. Entonces, bajo la hipótesis nula de que $\mathbf{p}(\boldsymbol{\theta})$ sigue el modelo loglineal asociado a la matriz de diseño \mathbf{W} , cumpliéndose además las condiciones de regularidad de Birch establecidas en la Sección 2.1.1, el estadístico

$$T^{\phi_1}(\widehat{\boldsymbol{\theta}}_{\phi_2}, \widetilde{\vartheta}_{\widehat{n}^*, N}) = \frac{2N\widehat{n}^*}{\phi_1''(1)\widetilde{\vartheta}_{\widehat{n}^*, N}} d_{\phi_1}(\widehat{\mathbf{p}}, \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2})) \quad (4.25)$$

se distribuye como una $\chi_{M-M_0-1}^2$, donde $\widetilde{\vartheta}_{\widehat{n}^*, N}$ puede ser bien el estimador de Brier modificado $\widetilde{\vartheta}_{\widehat{n}^*, N, \bullet}$, o bien el semiparamétrico $\widetilde{\vartheta}_{\widehat{n}^*, N}(\widetilde{\mathbf{Y}}, \widehat{\boldsymbol{\theta}}_{\phi_\lambda})$.

Demostración Basta considerar el Teorema de Slutsky y recordar que por los Teoremas 3.4 y 3.5 del Capítulo 3

$$\widetilde{\vartheta}_{\widehat{n}^*, N} \xrightarrow[N \rightarrow \infty]{P} \vartheta_{n^*}, \quad \widetilde{\vartheta}_{\widehat{n}^*, N, \bullet} \xrightarrow[N \rightarrow \infty]{P} \vartheta_{n^*}, \quad \widetilde{\vartheta}_{\widehat{n}^*, N, \lambda} \xrightarrow[N \rightarrow \infty]{P} \vartheta_{n^*},$$

y

$$\widehat{n}^* = \sum_{g=1}^G w_g n_g \xrightarrow[N_1, \dots, N_G \rightarrow \infty]{P} n^*.$$

■

4.4. Aplicaciones numéricas.

En el Capítulo 3 se analizó el conjunto de datos formado por viviendas en Montevideo (Minnesota, EEUU), para estudiar la relación entre la satisfacción con el vecindario (X_1) y

el grado de satisfacción con la vivienda particular (X_2). Para ello, se obtuvo una muestra formada por $N = 20$ clusters (vecindarios), de forma que en $N_1 = 18$ de ellos se observan $n_1 = 5$ viviendas y en $N_2 = 2$, $n_2 = 3$ viviendas.

Bajo la hipótesis de que la clasificación de las familias de acuerdo a la satisfacción con la vivienda propia es independiente de su clasificación de acuerdo al nivel de satisfacción con el vecindario, se planteó el modelo loglineal para la tabla de contingencia $I \times J$ con $I = J = 3$ (Tabla 3.2). Para estos datos, se consideró el siguiente modelo loglineal, en el caso de independencia entre las dos variables estudiadas

$$p_{ij}(\boldsymbol{\theta}) = \exp\{u + \theta_{1(i)} + \theta_{2(j)}\}, \quad \text{con } i, j = 1, 3. \quad (4.26)$$

Bajo este modelo se estimaron los parámetros de interés $\boldsymbol{\beta} = (\rho^2, \mathbf{p}(\boldsymbol{\theta}))^T$ por

$$\widehat{\boldsymbol{\beta}}_{\widehat{n}^*, N, \lambda} = (\widehat{\rho}_{\widehat{n}^*, N, \lambda}^2, \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_\lambda}))^T = \left(\frac{\widetilde{\vartheta}_{\widehat{n}^*, N, \lambda} - 1}{\widehat{n}^* - 1}, \frac{\exp\{\mathbf{W}\widehat{\boldsymbol{\theta}}_{\phi_\lambda}\}}{\mathbf{1}_M^T \exp\{\mathbf{W}\widehat{\boldsymbol{\theta}}_{\phi_\lambda}\}} \right)^T,$$

donde $\widetilde{\vartheta}_{\widehat{n}^*, N}$ se calcula según (3.15), y $\widehat{\boldsymbol{\theta}}_{\phi_\lambda}$ utilizando la familia de ϕ -divergencias de Cressie y Read definidas en (2.4). Las estimaciones calculadas se presentaron en la Tabla 3.2. El estadístico para contrastar la bondad del ajuste definido en (4.23) se ha calculado teniendo en cuenta que el estimador del efecto del diseño $\widetilde{\vartheta}_{\widehat{n}^*, N}$ puede ser, bien el estimador de Brier modificado $\widetilde{\vartheta}_{\widehat{n}^*, N, \bullet}$, (Tabla 4.4), o bien el semiparamétrico $\widetilde{\vartheta}_{\widehat{n}^*, N}(\widetilde{\mathbf{Y}}, \widehat{\boldsymbol{\theta}}_{\phi_\lambda})$ (Tabla 4.5).

Tal y como hizo notar Brier (1980) en su artículo, el valor del coeficiente de sobre-dispersión, $\widetilde{\vartheta}_{\widehat{n}^*, N} = 1 + \widehat{\rho}^2(\widehat{n}^* - 1)$, puede llegar a ser 4.875, en el caso extremo de total dependencia, es decir cuando $\widehat{\rho}^2 = 1$. Esto significa que los estadísticos de bondad de ajuste, al ser corregidos por este coeficiente, disminuyen notablemente su valor y pueden llevar a no rechazar la hipótesis de independencia cuando esta se rechazaba en el caso de

Tabla 4.4. Valores del estadístico $T^{\phi_1}(\hat{\theta}_{\phi_2}, \tilde{\vartheta}_{\hat{n}^*, N})$ y su p-valor.

ϕ_1, ϕ_2	$\lambda_2 = -\frac{1}{2}$	$\lambda_2 = 0$	$\lambda_2 = \frac{2}{3}$	$\lambda_2 = 1$	$\lambda_2 = 2$
$\tilde{\vartheta}_{\hat{n}^*, N, \bullet}$	1.0653	1.0653	1.0653	1.0653	1.0653
$\lambda_1 = -\frac{1}{2}$	15.4857 (0.0038)	16.7482 (0.0022)	19.6280 (0.0006)	21.0340 (0.0003)	24.5659 (0.0001)
$\lambda_1 = 0$	15.8641 (0.0032)	14.4521 (0.0060)	15.3118 (0.0041)	16.0005 (0.0030)	17.9803 (0.0012)
$\lambda_1 = \frac{2}{3}$	21.2993 (0.0003)	15.3929 (0.0040)	14.1763 (0.0068)	14.3063 (0.0064)	15.2807 (0.0042)
$\lambda_1 = 1$	26.6618 (0.0001)	16.8059 (0.0022)	14.2648 (0.0064)	14.1116 (0.0069)	14.6608 (0.0055)
$\lambda_1 = 2$	68.5704 (0.0001)	26.1695 (0.0001)	16.3167 (0.0026)	14.9665 (0.0048)	14.0586 (0.0071)

Tabla 4.5. Valores del estadístico $T^{\phi_1}(\hat{\theta}_{\phi_2}, \tilde{\vartheta}_{\hat{n}^*, N})$ y su p-valor.

ϕ_1, ϕ_2	$\lambda_2 = -\frac{1}{2}$	$\lambda_2 = 0$	$\lambda_2 = \frac{2}{3}$	$\lambda_2 = 1$	$\lambda_2 = 2$
$\tilde{\vartheta}_{\hat{n}^*, N}(\hat{\mathbf{Y}}, \hat{\boldsymbol{\theta}}_{\phi_\lambda})$	2.1788	1.5871	1.3315	1.2706	1.2706
$\lambda_1 = -\frac{1}{2}$	7.5713 (0.1086)	11.2413 (0.0239)	15.7037 (0.0034)	17.6354 (0.0015)	22.1535 (0.0002)
$\lambda_1 = 0$	7.7563 (0.1009)	9.7004 (0.0458)	12.2504 (0.0156)	13.4152 (0.0094)	16.2145 (0.0027)
$\lambda_1 = \frac{2}{3}$	10.4138 (0.0340)	10.3318 (0.0351)	11.3419 (0.0230)	11.9948 (0.0174)	13.7802 (0.0080)
$\lambda_1 = 1$	13.0356 (0.0111)	11.2803 (0.0236)	11.4128 (0.0223)	11.8316 (0.0186)	13.2211 (0.0102)
$\lambda_1 = 2$	33.5257 (0.0001)	17.5652 (0.0015)	13.0544 (0.0110)	12.5483 (0.0137)	12.6780 (0.0130)

no aplicarse la corrección. En este sentido, se ha querido resaltar la importancia de la corrección por dicho coeficiente, $\tilde{\vartheta}_{\hat{n}^*, N}$, añadiendo una fila con su valor en las tablas en donde aparece el valor del estadístico de bondad de ajuste, así como su p-valor. Obsérvese que ϕ_2 , en el estadístico de bondad de ajuste $T^{\phi_1}(\hat{\boldsymbol{\theta}}_{\phi_2}, \tilde{\vartheta}_{\hat{n}^*, N})$, es la función de ϕ -divergencia utilizada para estimar tanto los parámetros del modelo loglineal como el parámetro de sobredispersión en la Tabla 4.5.

En el caso en el que el coeficiente de sobredispersión se ha estimado mediante el estimador de Brier modificado $\tilde{\vartheta}_{\hat{n}^*, N, \bullet}$, (Tabla 4.4), los p-valores obtenidos son todos muy similares y en todos los casos se rechaza la hipótesis nula para un nivel $\alpha = 0.01$. Sin embargo en la Tabla 4.5, como consecuencia de las diferencias que aparecen en los coeficientes de sobredispersión estimados utilizando el estimador semiparamétrico $\tilde{\vartheta}_{\hat{n}^*, N}(\tilde{\mathbf{Y}}, \hat{\boldsymbol{\theta}}_{\phi_\lambda})$, los p-valores obtenidos son muy diferentes llevando, en algunos casos a no rechazar la hipótesis nula de independencia (cuando $\lambda_1 = -1/2$ y $\lambda_2 = -1/2$ ó $\lambda_1 = 0$ y $\lambda_2 = -1/2$).

Los estadísticos, de Pearson obtenido cuando $\lambda_1 = 1$, y $\lambda_2 = 0$, y del cociente de verosimilitudes, que se obtiene cuando $\lambda_1, \lambda_2 = 0$, corregidos por el efecto del diseño calculado por Brier, se corresponden a los valores $T^{\phi_{\lambda_1}}(\mathbf{Y}, \hat{\boldsymbol{\theta}}_{\phi_{\lambda_2}}) = 11.2803$ y $T^{\phi_{\lambda_1}}(\mathbf{Y}, \hat{\boldsymbol{\theta}}_{\phi_{\lambda_2}}) = 9.7004$, respectivamente. Es de resaltar que, la constante K de Brier ha sido reemplazada por un estimador consistente de $\frac{n}{\hat{v}_n}$.

4.5. Estudio de simulación.

Para analizar el comportamiento de los nuevos estadísticos de bondad de ajuste en el caso de tamaños desiguales, a continuación se realiza un estudio de simulación Monte Carlo. Se generan $R = 10,000$ repeticiones de muestras con $G = 3$ grupos con $N_1 = 18$, $N_2 = 2$, $N_3 = 5$, clusters en cada grupo, en los que los tamaños de cluster son $n_1 = 5$, $n_2 = 3$, $n_3 = 7$, respectivamente, utilizando valores del coeficiente de correlación intracluster ρ^2 desde 0.15 a 0.95. Para generar las muestras se utilizarán las tres distribuciones: Dirichlet-

multinomial (DM), Multinomial n-inflada (NI) y la distribución Agrupada Aleatoriamente (RC). Se plantea el modelo loglineal y se contrasta la hipótesis de independencia, en esta situación, se estima la probabilidad de error de tipo I (α), determinando la frecuencia de rechazo de la hipótesis nula cuando esta es cierta para un nivel de significación $\alpha = 0.05$.

En las Figuras 4.1, 4.2 y 4.3 se compara el comportamiento de los estadísticos de contraste $T^{\phi_1}(\hat{\boldsymbol{\theta}}_{\phi_2}, \tilde{\vartheta}_{\hat{n}^*, N})$ teniendo en cuenta que el estimador del efecto del diseño $\tilde{\vartheta}_{\hat{n}^*, N}$ puede ser, bien el estimador de Brier modificado $\tilde{\vartheta}_{\hat{n}^*, N, \bullet}$, o bien el semiparamétrico $\tilde{\vartheta}_{\hat{n}^*, N}(\tilde{\mathbf{Y}}, \hat{\boldsymbol{\theta}}_{\phi_\lambda})$. De entre todos los estadísticos que se han calculado en las aplicaciones numéricas se han escogido aquellos de los que se espera un mejor comportamiento teniendo en cuenta los resultados obtenidos en los Capítulos 2 y 3 para el error cuadrático medio y el sesgo de los estimadores de ρ^2 , y en consecuencia del efecto del diseño. Además, Cressie, N, et al., (2003) realizaron un estudio de simulación encontrando que, para los modelos loglineales, los estadísticos de contraste de bondad de ajuste de la familia, T_{ϕ_λ, ϕ_0} cuando $\lambda = 2/3$, tienen un mejor comportamiento que los estadísticos de Pearson ($\lambda = 1$) y el de cociente de verosimilitudes ($\lambda = 0$). Por estas razones, y para confirmar el comportamiento de estos estadísticos en el caso de modelos con *sobredispersión*, hemos escogido para el estudio comparativo los $T^{\phi_{\lambda_1}}(\hat{\boldsymbol{\theta}}_{\phi_{\lambda_2}}, \tilde{\vartheta}_{\hat{n}^*, N})$ para los valores $\lambda_1 = 0, \lambda_2 = 0$, (Cociente de verosimilitudes), $\lambda_1 = 0, \lambda_2 = 2/3$, $\lambda_1 = 2/3, \lambda_2 = 0$ (Cressie-Read), $\lambda_1 = 2/3, \lambda_2 = 2$, y $\lambda_1 = 1, \lambda_2 = 0$, (Pearson).

Es importante destacar que, sea cual sea la distribución utilizada para generar los datos, el estadístico $T^{\phi_1}(\hat{\boldsymbol{\theta}}_{\phi_2}, \tilde{\vartheta}_{\hat{n}^*, N})$, en el que el estimador del efecto del diseño utilizado es el estimador de Brier $\tilde{\vartheta}_{\hat{n}^*, N, \bullet}$, no consigue corregir la influencia de la alta correlación intraccluster, puesto que la probabilidad de error de tipo I aumenta a medida que crece ρ^2 , llegando a valer hasta casi 0.25. Sin embargo, dicha probabilidad permanece estable aunque aumente ρ^2 , alrededor del nivel de significación marcado 0.05 en el caso de que el estadístico $T^{\phi_1}(\hat{\boldsymbol{\theta}}_{\phi_2}, \tilde{\vartheta}_{\hat{n}^*, N})$ utilice como estimador del efecto del diseño el estimador semiparamétrico

$\tilde{\vartheta}_{\hat{n}^*, N}(\tilde{\mathbf{Y}}, \hat{\boldsymbol{\theta}}_{\phi_\lambda})$. Para estos últimos, el que tiene un peor comportamiento es el del cociente de verosimilitudes.

Figura 4.1. Probabilidad de error de Tipo I estimada para $T^{\phi_1}(\hat{\boldsymbol{\theta}}_{\phi_2}, \tilde{\vartheta}_{\hat{n}^*, N})$ con datos generados con la distribución Dirichlet Multinomial.

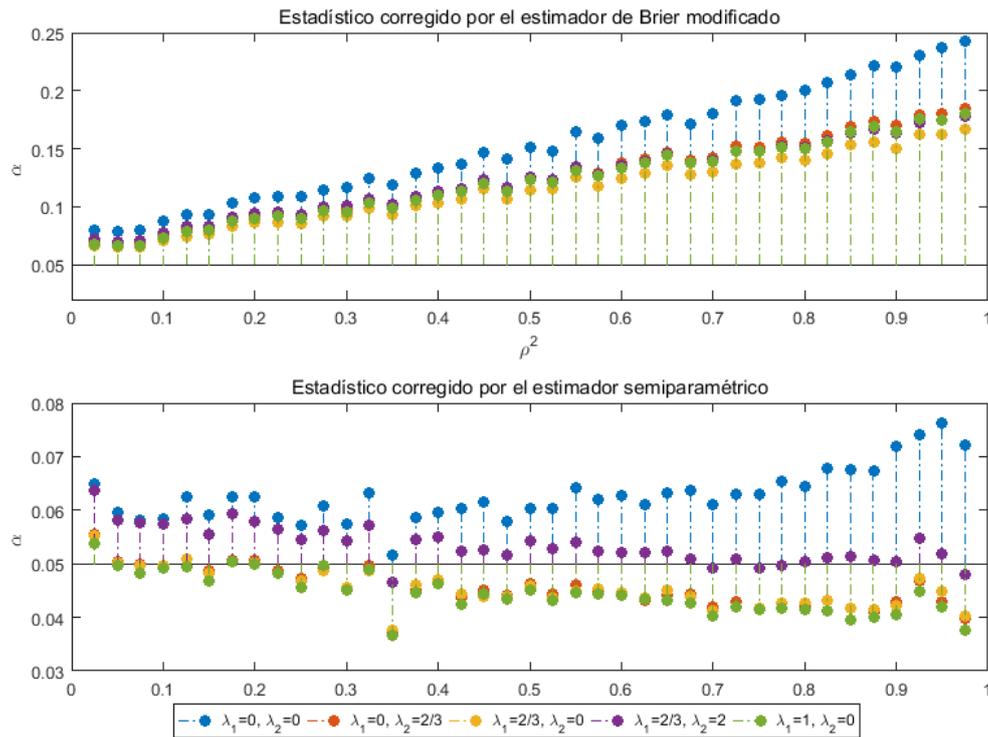


Figura 4.2. Probabilidad de error de Tipo I estimada para $T^{\phi_1}(\hat{\theta}_{\phi_2}, \tilde{\vartheta}_{\hat{n}^*, N})$ con datos generados según una distribución n-inflada.

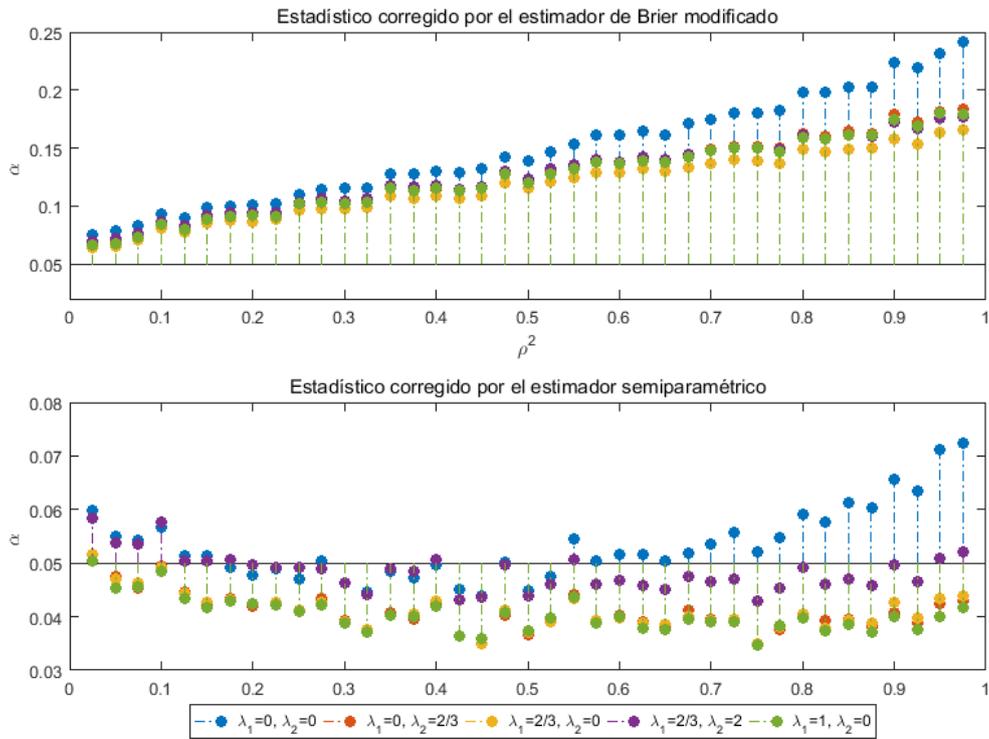
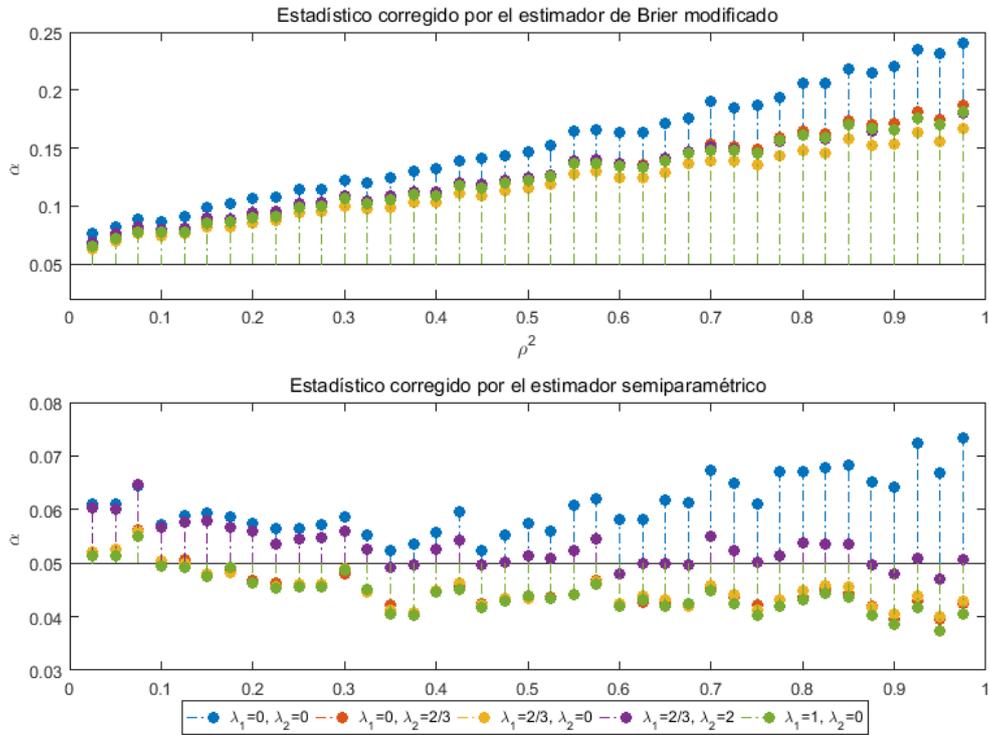


Figura 4.3. Probabilidad de error de Tipo I estimada para $T^{\phi_1}(\hat{\theta}_{\phi_2}, \tilde{\vartheta}_{\hat{n}^*, N})$ con datos generados según distribución Agrupada Aleatoriamente .



Para hacer una comparación utilizando la misma escala se representan en la Figura (4.4) en un mismo gráfico los estadísticos $T^{\phi_1}(\widehat{\boldsymbol{\theta}}_{\phi_2}, \widetilde{\vartheta}_{\widehat{n}^*, N})$ corregidos por el estimador de Brier modificado $\widetilde{\vartheta}_{\widehat{n}^*, N, \bullet}$, y los corregidos por el estimador semiparamétrico $\widetilde{\vartheta}_{\widehat{n}^*, N}(\widetilde{\mathbf{Y}}, \widehat{\boldsymbol{\theta}}_{\phi_\lambda})$, en los casos de $\lambda_1 = 2/3$ y $\lambda_2 = 0$ y $\lambda_1 = 2/3$ y $\lambda_2 = 2$. Dicha figura evidencia el mejor comportamiento entre los nuevos estadísticos presentados, los cuales consiguen mantener probabilidad de error de tipo I del contraste aunque exista una alta correlación intracluster, mientras que con el de Brier aumenta hasta alcanzar un valor de 0.25.

En cuanto a la influencia de la distribución utilizada para generar los datos, se confirma que el comportamiento de estos estadísticos es independiente de la distribución subyacente, tal y como se puede observar en la Figura 4.5.

Figura 4.4. Probabilidad de error de Tipo I estimada para $T^{\phi_1}(\hat{\theta}_{\phi_2}, \tilde{\vartheta}_{\hat{n}^*, N})$ por distribuciones .

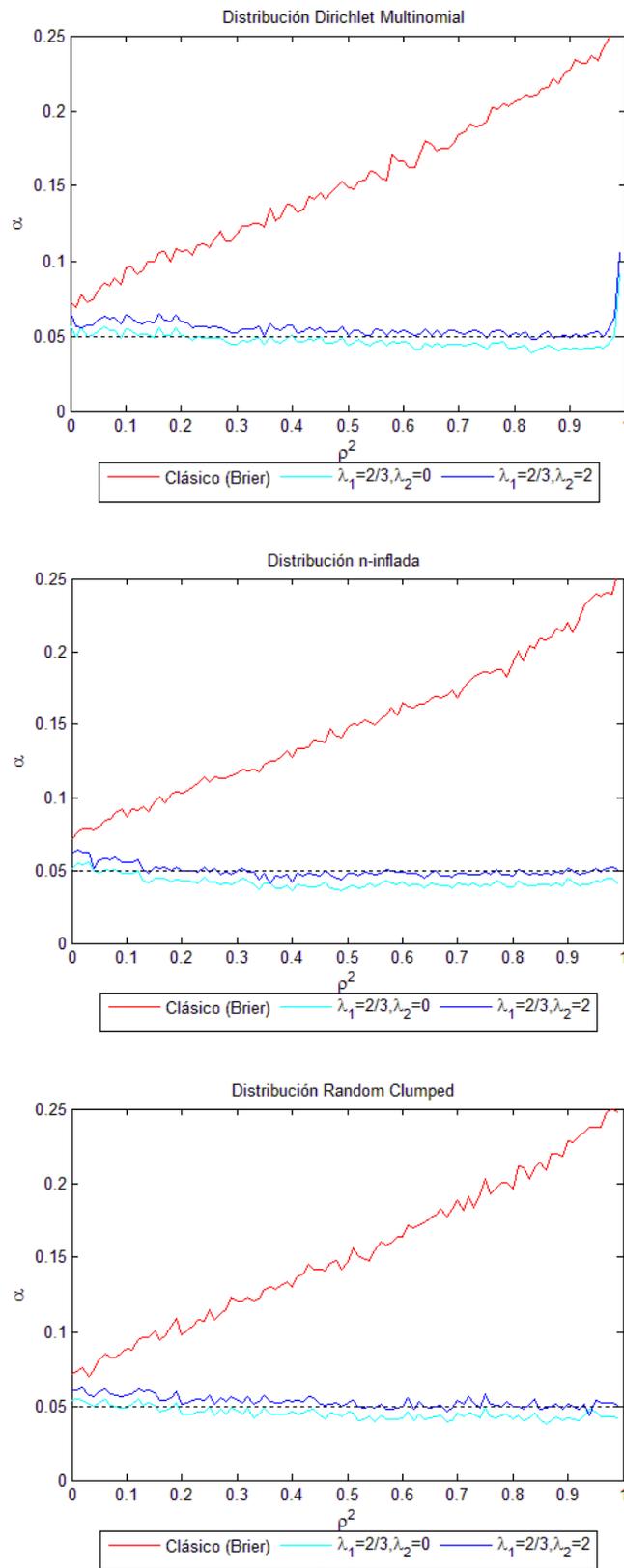
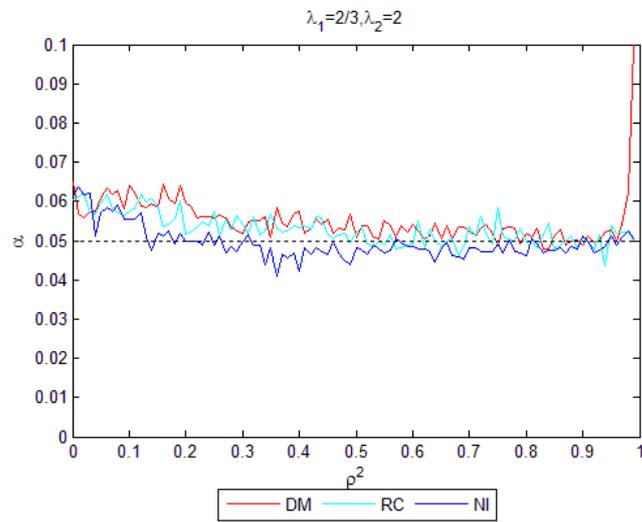
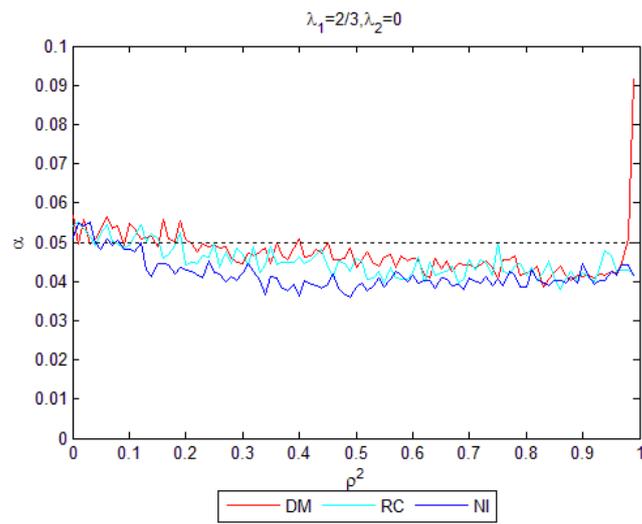
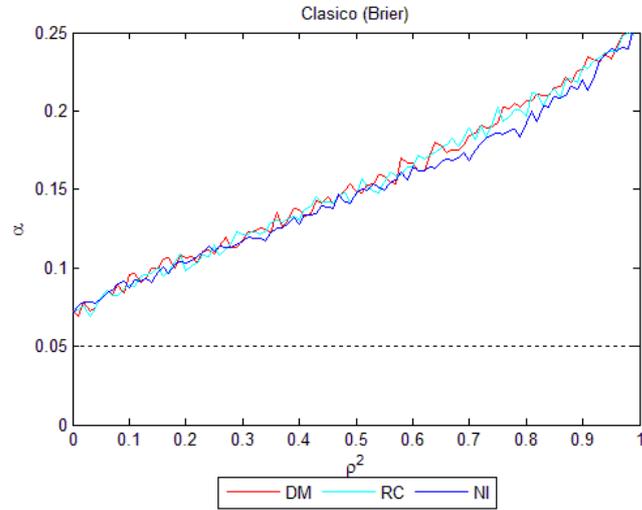


Figura 4.5. Probabilidad de error de Tipo I estimada para $T^{\phi_1}(\hat{\theta}_{\phi_2}, \tilde{\vartheta}_{\hat{n}^*, N})$, resumen .



Se ha realizado también un estudio para comparar las funciones de potencia de los nuevos estadísticos propuestos. Para lo cual se han generado $R = 10,000$ repeticiones de muestras con la misma distribución en clusters utilizada para calcular la probabilidad de error de tipo I, con valores del parámetro distanciándose de la hipótesis nula de independencia, de forma que el vector de probabilidad de las distribuciones Dirichlet Multinomial, Aleatoriamente Agrupada y n-inflada viene dado por $\mathbf{p}(\boldsymbol{\theta}^\delta) = (\mathbf{p}(\boldsymbol{\theta}_0) + \delta \mathbf{v}) / (1 + 3\delta)$, para valores de $\delta = \{0, 0.05, 0.1, \dots, 0.9\}$ y siendo el vector $\mathbf{v} = (1, 0, 0, 0, 1, 0, 0, 0, 1)^T$. Además para reflejar la influencia del grado de correlación intracluster se han generado estos escenarios para valores de $\rho^2 = \{0.15, 0.45, 0.75\}$. En estas condiciones se ha simulado el comportamiento de las funciones de potencia para los estadísticos con mejor comportamiento en su probabilidad de error de tipo I, $T^{\phi_1}(\widehat{\boldsymbol{\theta}}_{\phi_2}, \widetilde{\vartheta}_{\widehat{n}^*, N})$ corregidos por el estimador de Brier modificado $\widetilde{\vartheta}_{\widehat{n}^*, N, \bullet}$, y los corregidos por el estimador semiparamétrico $\widetilde{\vartheta}_{\widehat{n}^*, N}(\widetilde{\mathbf{Y}}, \widehat{\boldsymbol{\theta}}_{\phi_\lambda})$, en los casos de $\lambda_1 = 2/3$ y $\lambda_2 = 0$ y $\lambda_1 = 2/3$ y $\lambda_2 = 2$. Dado que las funciones de potencia para estos dos últimos estadísticos son prácticamente idénticas, se representa sólo el caso $\lambda_1 = 2/3$ y $\lambda_2 = 0$.

Sea cual sea la distribución utilizada para generar los datos e independientemente del estadístico utilizado, la potencia disminuye a medida que aumenta la correlación intracluster. Evidentemente, al tener mayor probabilidad de error de tipo I el estadístico clásico de Brier (obsérvese que en todas ellas $\alpha > 0.1$), tiene funciones de potencia con mayor probabilidad para detectar cambios en la hipótesis de partida pero la diferencia es muy pequeña.

Figura 4.6. Probabilidad de error de Tipo II estimada para $T^{\phi_1}(\hat{\theta}_{\phi_2}, \tilde{\vartheta}_{\hat{n}^*, N})$. Distribución Dirichlet Multinomial.

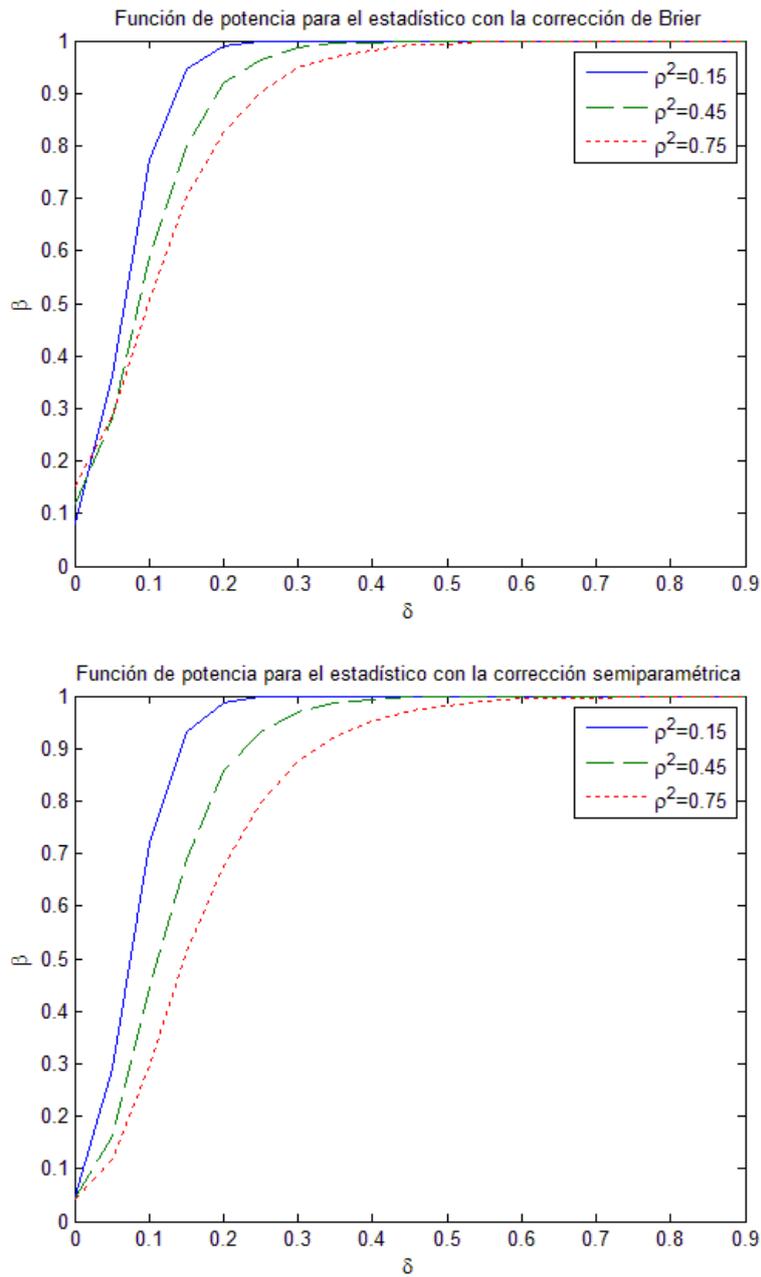


Figura 4.7. Probabilidad de error de Tipo II estimada para $T^{\phi_1}(\hat{\theta}_{\phi_2}, \tilde{\vartheta}_{\tilde{n}^*, N})$. Distribución Multinomial n-inflada.

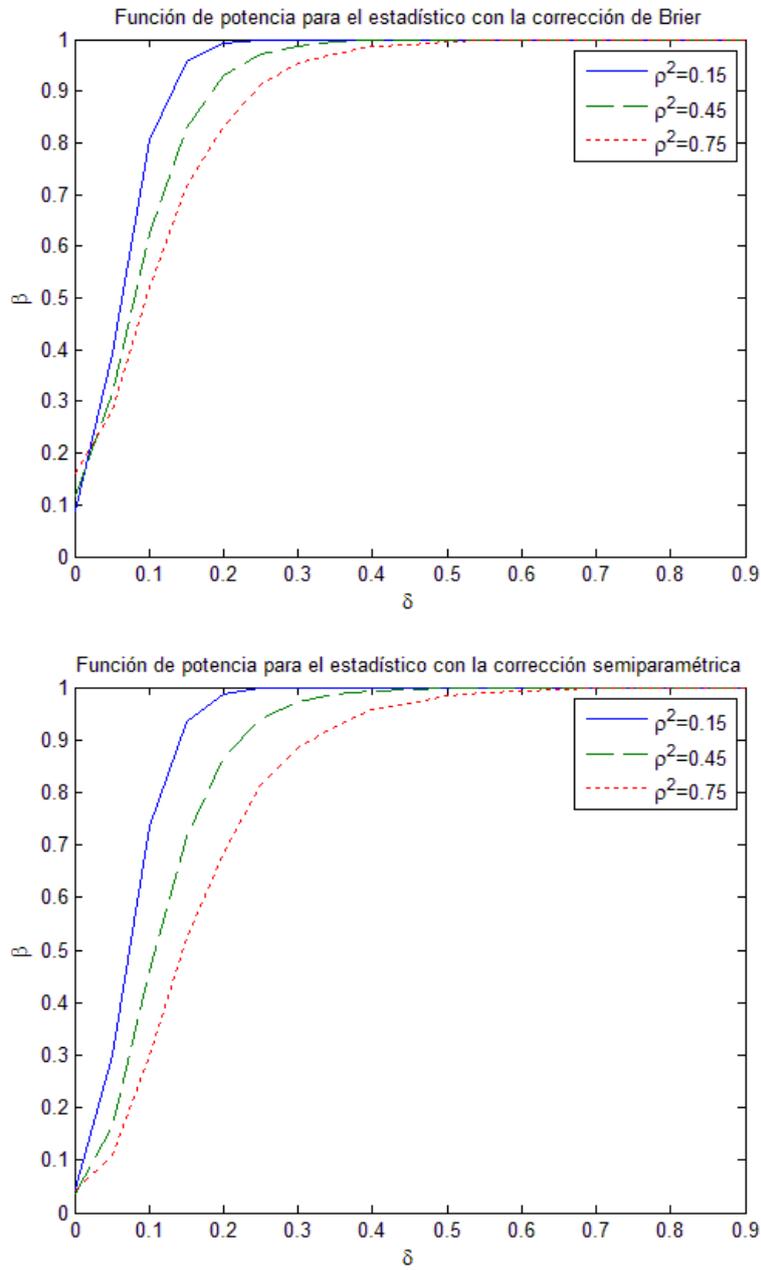
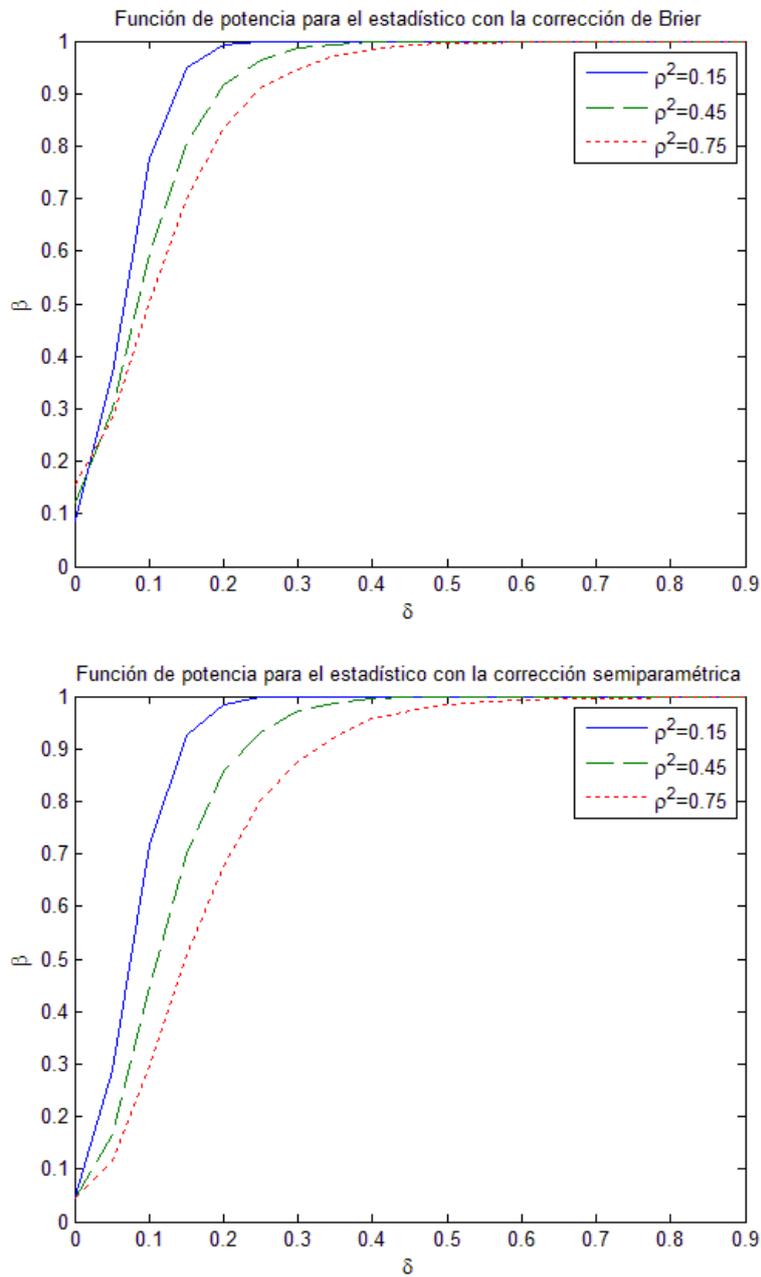


Figura 4.8. Probabilidad de error de Tipo II estimada para $T^{\phi_1}(\hat{\theta}_{\phi_2}, \tilde{\vartheta}_{\hat{n}^*, N})$. Distribución Agrupada Aleatoriamente.



CAPÍTULO 5

SECUENCIAS ANIDADAS DE MODELOS LOGLINEALES JERARQUIZADOS PARA DATOS MULTINOMIALES CON SOBREDISPERSIÓN

5.1. Planteamiento del problema

En el Capítulo 4 se ha visto la forma de encontrar la parametrización del modelo loglineal que mejor se ajusta a unos datos multinomiales con sobredispersión de forma que quede determinada la dependencia existente entre las variables que generan los datos multinomiales. En este Capítulo, se partirá de que se tiene un modelo loglineal que justifica convenientemente los datos multinomiales con sobredispersión y el objetivo será ver si se puede reducir la dimensión del espacio paramétrico o, dicho de otra forma, si se puede prescindir de alguna dependencia entre las variables que originalmente se tenía. En este estudio, hay dos aspectos importantes que es conveniente clarificar: por un lado el concepto de “secuencia anidada de modelos loglineales” y por otro “modelos loglineales jerarquizados”.

Sea $\{H_i\}$ una secuencia de hipótesis, verificando $\mathbf{p}(\boldsymbol{\theta})$ un modelo loglineal y $H_i : \mathbf{p} = \mathbf{p}(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta_i \subseteq \mathbb{R}^{M_0}, i = 1, \dots, m \leq M_0 < M$. Se dice que $\{H_i\}_{i=1, \dots, m}$ es una *secuencia anidada* de modelos loglineales si cada subsiguiente modelo de la sucesión es más restrictivo que el precedente, o dicho de otra forma, todas las interacciones que se anulan con H_i , son también cero con $H_{i+1}, i = 1, \dots, m - 1$. El objetivo de este Capítulo es introducir familias de estadísticos para abordar el problema de contrastar

$$H_{Nula} : H_{i+1} \quad \text{frente} \quad H_{Alternativa} : H_i, \quad (5.1)$$

teniendo en cuenta la *sobredispersión* que pueden presentar los datos. Una forma de designar una secuencia jerarquizada de modelos loglineales es escribir

$$H_m \subset H_{m-1} \subset \dots \subset H_2 \subset H_1.$$

En los *modelos loglineales jerarquizados* se hace referencia a la regla que establece que, si aparece en el parámetro $\boldsymbol{\theta}$ del modelo loglineal un término correspondiente a una interacción de un determinado orden, deben aparecer también en el modelo todas las interacciones correspondientes de orden inferior. Así, cuando $\theta_{12(ij)}$ forme parte del modelo, necesariamente $\theta_{1(i)}$ y $\theta_{2(j)}$ también deben formar parte del modelo.

5.2. Estadísticos de ϕ -divergencia para secuencias anidadas de modelos loglineales jerarquizados

Antes de introducir las familias de estadísticos de ϕ -divergencia para abordar el contraste planteado en (5.1), se introducirá alguna notación que será necesaria en lo sucesivo. De forma más concreta, el problema que se abordará en este Capítulo supone la existencia de una secuencia de hipótesis

$$H_i : \mathbf{p} = \mathbf{p}(\boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \Theta_i \subseteq \mathbb{R}^{M_0}, \quad i = 1, \dots, m \quad m \leq M_0 < M - 1 \quad (5.2)$$

donde

$$\Theta_m \subset \Theta_{m-1} \subset \dots \subset \Theta_2 \subset \Theta_1 \subseteq \Theta \subseteq \mathbb{R}^{M_0},$$

y

$$\dim(\Theta_i) = d_i, \quad i = 1, \dots, m \quad \text{con} \quad d_m < d_{m-1} < \dots < d_1 = M_0. \quad (5.3)$$

En este contexto, existe un entero m^* ($1 \leq m^* \leq m$) para el cual H_{m^*} es cierto, pero H_{m^*+1} no es cierta. El problema que se plantea es hacer inferencias sobre m^* . Para ello, la estrategia que se seguirá a lo largo de este Capítulo es la de contrastar sucesivamente

$$H_{Nula} : H_{i+1} \quad \text{frente} \quad H_{Alternativa} : H_i, \quad i = 1, \dots, m-1, \quad (5.4)$$

mientras que la hipótesis nula no sea rechazada y se estimará m^* como el primer i para el cual la hipótesis H_{i+1} sea rechazada como hipótesis nula.

Sean $\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)}$ y $\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)}$ los estimadores de mínima ϕ -divergencia bajo las hipótesis H_i y H_{i+1} , respectivamente. Con el fin de contrastar (5.4) se utilizará la familia de estadísticos de ϕ -divergencia

$$T_{\phi_1, \phi_2}^{(i)} = \frac{2Nn}{\phi_1''(1) \vartheta_n} d_{\phi_1} \left(\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)}), \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)}) \right). \quad (5.5)$$

Para abordar el problema de contraste dado en (5.4), es posible también considerar la familia de estadísticos de ϕ -divergencia dada por

$$S_{\phi}^{(i)} = \frac{2Nn}{\phi''(1) \vartheta_n} \left[d_{\phi}(\widehat{\mathbf{p}}, \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi}^{(i)})) - d_{\phi}(\widehat{\mathbf{p}}, \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi}^{(i+1)})) \right]. \quad (5.6)$$

Se observará que, en la familia $S_{\phi}^{(i)}$ solamente interviene una medida de ϕ -divergencia ya que se estima y se hace el contraste con la misma función ϕ , a diferencia de la familia $T_{\phi_1, \phi_2}^{(i)}$ en la que aparecen dos medidas de ϕ -divergencia, una para resolver el problema de estimación, ϕ_2 , y otra para construir los estadísticos de contraste, ϕ_1 . El hecho de que, en la familia $S_{\phi}^{(i)}$ únicamente intervenga una medida de ϕ -divergencia es para tener garantizada, como pone de manifiesto el siguiente Teorema, la no negatividad de la familia $S_{\phi}^{(i)}$.

Teorema 5.1 *Se considera un modelo loglineal verificando*

$$E \left[\mathbf{Y}^{(\ell)} \right] = n\mathbf{p}(\boldsymbol{\theta}) \quad y \quad Var \left[\mathbf{Y}^{(\ell)} \right] = \vartheta_n n \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})},$$

y la secuencia de hipótesis anidadas dadas por (5.2) y (5.3). Si se elige $\phi \in \Phi^*$, para contrastar

$$H_{Nula} : H_{i+1} \quad \text{frente} \quad H_{Alternativa} : H_i, \quad i = 1, \dots, m-1,$$

la familia de estadísticos de mínima ϕ -divergencia $S_\phi^{(i)}$, definida en (5.6) es no negativa.

Demostración Es claro que $S_\phi^{(i)}$, $i = 1, \dots, m-1$, es no negativa ya que

$$d_\phi(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}}_\phi^{(i+1)})) = \inf_{\boldsymbol{\theta} \in \Theta_{i+1}} d_\phi(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})) \geq \inf_{\boldsymbol{\theta} \in \Theta_i} d_\phi(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})) = d_\phi(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}}_\phi^{(i)}))$$

debido a que $\Theta_{i+1} \subset \Theta_i$. ■

Téngase en cuenta que si se consideraran, al definir el estadístico $S_\phi^{(i)}$, dos funciones ϕ_1 y ϕ_2 no se podría garantizar que $S_{\phi_1, \phi_2}^{(i)} \geq 0$.

Si se toma en $T_{\phi_1, \phi_2}^{(i)}$, $\phi_1(x) = \frac{1}{2}(1-x)^2$ y $\phi_2(x) = x \log x - x + 1$ se tiene

$$X^2 = \frac{nN}{\vartheta_n} \sum_{r=1}^M \frac{(p_r(\hat{\boldsymbol{\theta}}^{(i)}) - p_r(\hat{\boldsymbol{\theta}}^{(i+1)}))^2}{p_r(\hat{\boldsymbol{\theta}}^{(i+1)})}, \quad (5.7)$$

siendo $\hat{\boldsymbol{\theta}}^{(i)}$ y $\hat{\boldsymbol{\theta}}^{(i+1)}$ los estimadores de máxima verosimilitud de $\boldsymbol{\theta}$ bajo las hipótesis H_i y H_{i+1} , respectivamente. Si se considera $\phi_1(x) = \phi_2(x) = x \log x - x + 1$ se tiene el estadístico análogo al cociente de verosimilitudes dado por

$$G^2 = 2 \frac{nN}{\vartheta_n} \sum_{r=1}^M p_r(\hat{\boldsymbol{\theta}}^{(i)}) \log \frac{p_r(\hat{\boldsymbol{\theta}}^{(i)})}{p_r(\hat{\boldsymbol{\theta}}^{(i+1)})}.$$

5.3. Distribución asintótica del estadístico ϕ -divergencia con datos multinomiales con sobredispersión bajo la hipótesis nula

El objetivo esencial de esta Sección es establecer la distribución asintótica de la familia de estadísticos

$$T_{\phi_1, \phi_2}^{(i)} \quad \text{y} \quad S_\phi^{(i)},$$

definidos respectivamente en (5.5) y (5.6), bajo la hipótesis de que el modelo correcto es el dado por la hipótesis nula H_{i+1} . En el primer teorema se establecerá la distribución asintótica de $T_{\phi_1, \phi_2}^{(i)}$.

Teorema 5.2 *Se considera un modelo loglineal verificando*

$$E \left[\mathbf{Y}^{(\ell)} \right] = n\mathbf{p}(\boldsymbol{\theta}) \quad y \quad Var \left[\mathbf{Y}^{(\ell)} \right] = \vartheta_n n \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})},$$

y la secuencia de hipótesis anidadas dadas por (5.2) y (5.3). Si se eligen dos funciones ϕ_1 y $\phi_2 \in \Phi^*$, para contrastar

$$H_{Nula} : H_{i+1} \quad \text{frente} \quad H_{Alternativa} : H_i,$$

la familia de estadísticos de contraste $T_{\phi_1, \phi_2}^{(i)}$, definida en (5.5) bajo la hipótesis nula, sigue una distribución ji-cuadrado con $d_i - d_{i+1}$ grados de libertad, $i = 1, \dots, m - 1$.

Demostración El desarrollo en serie de Taylor de segundo orden de la función $d_{\phi_1 \phi_2} = d_{\phi_1}(\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)}), \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)}))$ alrededor del punto $(\mathbf{p}(\boldsymbol{\theta}_0), \mathbf{p}(\boldsymbol{\theta}_0))$ viene dado por

$$\begin{aligned} d_{\phi_1 \phi_2} &= \frac{1}{2} (\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)}) - \mathbf{p}(\boldsymbol{\theta}_0))^T \left. \frac{\partial^2 d_{\phi_1}(\mathbf{p}, \mathbf{p}(\boldsymbol{\theta}_0))}{\partial \mathbf{p} \partial \mathbf{p}^T} \right|_{\mathbf{p}=\mathbf{p}(\boldsymbol{\theta}_0)} (\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)}) - \mathbf{p}(\boldsymbol{\theta}_0)) \\ &+ \frac{1}{2} (\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)}) - \mathbf{p}(\boldsymbol{\theta}_0))^T \left. \frac{\partial^2 d_{\phi_1}(\mathbf{p}(\boldsymbol{\theta}_0), \mathbf{q})}{\partial \mathbf{q} \partial \mathbf{q}^T} \right|_{\mathbf{q}=\mathbf{p}(\boldsymbol{\theta}_0)} (\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)}) - \mathbf{p}(\boldsymbol{\theta}_0)) \\ &+ (\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)}) - \mathbf{p}(\boldsymbol{\theta}_0))^T \left. \frac{\partial^2 d_{\phi_1}(\mathbf{p}, \mathbf{q})}{\partial \mathbf{p} \partial \mathbf{q}^T} \right|_{(\mathbf{p}=\mathbf{p}(\boldsymbol{\theta}_0), \mathbf{q}=\mathbf{p}(\boldsymbol{\theta}_0))} (\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)}) - \mathbf{p}(\boldsymbol{\theta}_0)) \\ &+ o(\|\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)}) - \mathbf{p}(\boldsymbol{\theta}_0)\|^2) + o(\|\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)}) - \mathbf{p}(\boldsymbol{\theta}_0)\|^2) \\ &= \frac{1}{2} \phi_1''(1) (\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)}) - \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)}))^T \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} (\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)}) - \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)})) \\ &+ o(\|\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)}) - \mathbf{p}(\boldsymbol{\theta}_0)\|^2) + o(\|\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)}) - \mathbf{p}(\boldsymbol{\theta}_0)\|^2), \end{aligned} \tag{5.8}$$

ya que, al ser $\mathbf{p}(\boldsymbol{\theta}) = \mathbf{p}(\boldsymbol{\theta}_0)$ y $\phi_1(1) = \phi_1'(1) = 0$, las derivadas primeras del desarrollo se anulan y utilizando la expresión de las derivadas segundas calculada en (4.10) se llega a (5.8). Por otra parte, en el Teorema 2.3 se demostró (2.19), luego bajo la hipótesis nula

$$\sqrt{N}(\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)}) - \mathbf{p}(\boldsymbol{\theta}_0)) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_M, \frac{\vartheta_n}{n} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W}_{(i+1)} (\mathbf{W}_{(i+1)}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W}_{(i+1)})^{-1} \mathbf{W}_{(i+1)}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}), \quad (5.9)$$

donde $\mathbf{W}_{(i+1)}$, es la matriz de diseño del modelo loglineal válido bajo la hipótesis $H_{i+1} : \boldsymbol{\theta}_0 \in \Theta_{i+1}$. Así, $o(\|\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)}) - \mathbf{p}(\boldsymbol{\theta}_0)\|^2) = O_p(N^{-1})$ y puesto que se supone que $\boldsymbol{\theta}_0 \in \Theta_{i+1} \subset \Theta_i$, también se verifica que $o(\|\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)}) - \mathbf{p}(\boldsymbol{\theta}_0)\|^2) = O_p(N^{-1})$, en consecuencia

$$N \left(o(\|\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)}) - \mathbf{p}(\boldsymbol{\theta}_0)\|^2) + o(\|\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)}) - \mathbf{p}(\boldsymbol{\theta}_0)\|^2) \right) = o(1).$$

Así, la distribución asintótica del estadístico $T_{\phi_1, \phi_2}^{(i)}$ se puede obtener si se expresa como

$$T_{\phi_1, \phi_2}^{(i)} = \frac{2Nn}{\phi_1''(1) \vartheta_n} d_{\phi_1}(\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)}), \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)})) = \mathbf{Z}^T \mathbf{Z} + o_p(1),$$

donde

$$\mathbf{Z} = \sqrt{\frac{n}{\vartheta_n}} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \sqrt{N} \left(\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)}) - \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)}) \right). \quad (5.10)$$

Por otra parte,

$$\begin{aligned} \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)}) - \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)}) &= (\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W}_{(i+1)} (\mathbf{W}_{(i+1)}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W}_{(i+1)})^{-1} \mathbf{W}_{(i+1)}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \\ &\quad - \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W}_{(i)} (\mathbf{W}_{(i)}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W}_{(i)})^{-1} \mathbf{W}_{(i)}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}) \\ &\quad \times \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} (\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) \\ &\quad + o(\mathbf{1}_M). \end{aligned}$$

Si se denota

$$\mathbf{A}_{(i)} = \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W}_{(i)} (\mathbf{W}_{(i)}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W}_{(i)})^{-1} \mathbf{W}_{(i)}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2}, \quad (5.11)$$

que es una matriz simétrica, el vector aleatorio \mathbf{Z} se puede expresar como

$$\begin{aligned} \mathbf{Z} &= \sqrt{\frac{n}{\vartheta_n}} \sqrt{N} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} (\mathbf{A}_{(i+1)} - \mathbf{A}_{(i)}) (\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) \\ &\quad + o(\|\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)} - \boldsymbol{\theta}_0\|) + o(\|\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)} - \boldsymbol{\theta}_0\|). \end{aligned}$$

Luego, se verifica

$$\mathbf{Z} = \sqrt{\frac{n}{\vartheta_n}} \sqrt{N} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \left(\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)}) - \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)}) \right) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_M, \mathbf{V}^*(\boldsymbol{\theta}_0)),$$

donde

$$\mathbf{V}^*(\boldsymbol{\theta}_0) = (\mathbf{A}_{(i+1)} - \mathbf{A}_{(i)}) \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} (\mathbf{A}_{(i+1)} - \mathbf{A}_{(i)}).$$

Antes de comprobar si $\mathbf{V}^*(\boldsymbol{\theta}_0)$ es idempotente, se estudian algunas propiedades de las matrices $\mathbf{A}_{(i)}$. Primeramente se observa que $\mathbf{A}_{(i+1)}$ es una submatriz de $\mathbf{A}_{(i)}$ luego existe una matriz \mathbf{B} tal que $\mathbf{A}_{(i+1)} = \mathbf{A}_{(i)}\mathbf{B}$, y como consecuencia

$$\begin{aligned} \mathbf{A}_{(i)}\mathbf{A}_{(i+1)} &= \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W}_{(i)} (\mathbf{W}_{(i)}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W}_{(i)})^{-1} \mathbf{W}_{(i)}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \\ &\quad \times \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W}_{(i+1)} (\mathbf{W}_{(i+1)}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W}_{(i+1)})^{-1} \mathbf{W}_{(i+1)}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \\ &= \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W}_{(i)} (\mathbf{W}_{(i)}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W}_{(i)})^{-1} \mathbf{W}_{(i)}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W}_{(i)} \mathbf{B} \\ &\quad \times (\mathbf{W}_{(i+1)}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{W}_{(i+1)})^{-1} \mathbf{W}_{(i+1)}^T \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \\ &= \mathbf{A}_{(i+1)}. \end{aligned}$$

Además, como $\mathbf{A}_{(i)}$ es una matriz simétrica e idempotente se comprueba que

$$(\mathbf{A}_{(i+1)} - \mathbf{A}_{(i)}) (\mathbf{A}_{(i+1)} - \mathbf{A}_{(i)}) = \mathbf{A}_{(i)} - \mathbf{A}_{(i+1)},$$

y teniendo en cuenta que

$$\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{A}(\boldsymbol{\theta}_i) = \mathbf{A}(\boldsymbol{\theta}_i),$$

se puede simplificar la expresión de $\mathbf{V}^*(\boldsymbol{\theta}_0)$ como

$$\begin{aligned}\mathbf{V}^*(\boldsymbol{\theta}_0) &= (\mathbf{A}_{(i+1)} - \mathbf{A}_{(i)}) \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} (\mathbf{A}_{(i+1)} - \mathbf{A}_{(i)}) \\ &= \mathbf{A}_{(i+1)} (\mathbf{A}_{(i+1)} - \mathbf{A}_{(i)}) - \mathbf{A}_{(i)} (\mathbf{A}_{(i+1)} - \mathbf{A}_{(i)}) \\ &= (\mathbf{A}_{(i+1)} - \mathbf{A}_{(i)}) (\mathbf{A}_{(i+1)} - \mathbf{A}_{(i)}) = \mathbf{A}_{(i)} - \mathbf{A}_{(i+1)}.\end{aligned}$$

Luego, la matriz $\mathbf{V}^*(\boldsymbol{\theta}_0)$ es idempotente, y entonces el estadístico $T_{\phi_1, \phi_2}^{(i)}$ tendría una distribución asintótica ji-cuadrado cuyos grados de libertad vienen dados por la traza o el rango de la matriz $\mathbf{V}^*(\boldsymbol{\theta}_0)$, y para obtener su rango basta considerar el resultado de Harville (2008, pg 27),

$$\text{rg}(\mathbf{V}^*(\boldsymbol{\theta}_0)) = \text{rg}(\mathbf{A}_{(i)}) - \text{rg}(\mathbf{A}_{(i+1)}) = d_i - d_{i+1}.$$

■

Bajo las mismas hipótesis del Teorema 5.2 se puede demostrar (se omite la demostración por su analogía con la anterior) que la familia de estadísticos $S_{\phi}^{(i)}$ sigue, bajo la hipótesis nula de que el verdadero modelo es el de la hipótesis H_{i+1} , una distribución ji-cuadrado con $d_i - d_{i+1}$ grados de libertad.

Para contrastar las hipótesis anidadas $\{H_i\}_{i=1, \dots, m-1}$ dadas en (5.4), se lleva a cabo secuencialmente

$$H_{Nula} : H_{i+1} \quad \text{frente} \quad H_{Alternativa} : H_i, \quad i = 1, \dots, m,$$

haciendo uso de las familias de estadísticos $T_{\phi_1, \phi_2}^{(i)}$ ó $S_{\phi}^{(i)}$. Cuando los estadísticos de contraste sean grandes se deberá rechazar la hipótesis nula, es decir, se rechazará cuando $T_{\phi_1, \phi_2}^{(i)} > c_1$ ó $S_{\phi}^{(i)} > c_2$, donde tanto c_1 como c_2 se calcularán imponiendo que el tamaño del test sea α , es decir,

$$P\left(T_{\phi_1, \phi_2}^{(i)} > c_1 / H_{i+1}\right) = \alpha \quad \text{ó} \quad P\left(S_{\phi}^{(i)} > c_2 / H_{i+1}\right) = \alpha.$$

Según se ha visto $T_{\phi_1, \phi_2}^{(i)}$ y $S_{\phi}^{(i)}$ se distribuyen asintóticamente según una ji-cuadrado con $d_i - d_{i+1}$ grados de libertad, luego $c_1 = c_2 = \chi_{d_i - d_{i+1}, \alpha}^2$.

5.4. Distribución asintótica de los estadísticos ϕ -divergencia con datos multinomiales con sobredispersión bajo hipótesis alternativas contiguas

En general, no es fácil obtener resultados teóricos que puedan ser aplicados de forma sencilla para calcular la potencia de los contrastes realizados con las familias de estadísticos $T_{\phi_1, \phi_2}^{(i)}$. Una excepción es cuando se dispone de una sucesión de hipótesis alternativas contiguas que se aproximan a la hipótesis nula H_{i+1} , tal como se definió en (4.18). Sean las hipótesis alternativas contiguas

$$H_{i+1, N} : \mathbf{p}_N = \mathbf{p}(\boldsymbol{\theta}) + \frac{1}{\sqrt{N}} \mathbf{d}, \quad \text{con } \boldsymbol{\theta} \in \Theta_{i+1}, \quad (5.12)$$

donde $\mathbf{d} = (d_1, \dots, d_M)^T$ es un vector fijo tal que $\sum_{r=1}^M d_r = 0$. Cuando $N \rightarrow \infty$, el vector de probabilidad \mathbf{p}_N converge a un vector de probabilidad $\mathbf{p}(\boldsymbol{\theta})$ con $\boldsymbol{\theta}$ perteneciente a la hipótesis nula Θ_{i+1} .

Si se considera el contraste

$$H_{Nula} : H_{i+1} \quad \text{frente} \quad H_{Alternativa} : H_{i+1, N},$$

utilizando los estadísticos $T_{\phi_1, \phi_2}^{(i)}$ ó $S_{\phi}^{(i)}$, la potencia de estos contrastes viene dada por

$$\pi_N^{1(i)} = P \left(T_{\phi_1, \phi_2}^{(i)} > c_1 / H_{i+1, N} \right)$$

o

$$\pi_N^{2(i)} = P \left(S_{\phi}^{(i)} > c_2 / H_{i+1, N} \right).$$

El teorema siguiente establecerá que bajo las hipótesis alternativas $H_{i+1,N}$ y cuando $N \rightarrow \infty$, tanto $T_{\phi_1, \phi_2}^{(i)}$ como $S_{\phi}^{(i)}$ convergen a una ji-cuadrado no centrada con $d_i - d_{i+1}$ grados de libertad y parámetro de no centralidad μ , cuya expresión se dará en el Teorema 5.3. En consecuencia

$$\pi_N^{1(i)} \quad y \quad \pi_N^{2(i)} \rightarrow P \left(\chi_{d_i - d_{i+1}}^2(\mu) > x \right).$$

Teorema 5.3 *Se considera un modelo loglineal verificando*

$$E \left[\mathbf{Y}^{(\ell)} \right] = n\mathbf{p}(\boldsymbol{\theta}) \quad y \quad Var \left[\mathbf{Y}^{(\ell)} \right] = \vartheta_n n \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta})}.$$

La distribución asintótica del estadístico $T_{\phi_1, \phi_2}^{(i)}$, bajo las hipótesis alternativas contiguas dadas en (5.12), es una ji-cuadrado con $d_i - d_{i+1}$ grados de libertad y parámetro de no centralidad

$$\delta = \mathbf{d}^T \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \left(\mathbf{A}^T(\boldsymbol{\theta}_i) - \mathbf{A}(\boldsymbol{\theta}_{i+1}) \right) \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \mathbf{d}.$$

Demostración Por el Teorema 5.2 se sabe que

$$\begin{aligned} T_{\phi_1, \phi_2}^{(i)} &= \frac{2Nn}{\phi_1''(1) \vartheta_n} d_{\phi_1} \left(\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)}), \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)}) \right) \\ &= \mathbf{Z}^T \mathbf{Z} + o(\|\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)}) - \mathbf{p}(\boldsymbol{\theta}_0)\|^2) + o(\|\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)}) - \mathbf{p}(\boldsymbol{\theta}_0)\|^2), \end{aligned}$$

donde

$$\mathbf{Z} = \sqrt{\frac{n}{\vartheta_n}} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \sqrt{N} \left(\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)}) - \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)}) \right).$$

Pero, bajo la hipótesis H_{i+1} , $o(\|\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)}) - \mathbf{p}(\boldsymbol{\theta}_0)\|^2) = O_p(N^{-1})$ y puesto que se supone que $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}_{i+1} \subset \boldsymbol{\Theta}_i$, también se verifica que $o(\|\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)}) - \mathbf{p}(\boldsymbol{\theta}_0)\|^2) = O_p(N^{-1})$, en consecuencia

$$N \left(o(\|\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)}) - \mathbf{p}(\boldsymbol{\theta}_0)\|^2) + o(\|\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)}) - \mathbf{p}(\boldsymbol{\theta}_0)\|^2) \right) = o(1).$$

Luego la distribución del estadístico $T_{\phi_1, \phi_2}^{(i)}$ se puede obtener si se expresa como

$$T_{\phi_1, \phi_2}^{(i)} = \frac{2Nn}{\phi_1''(1) \vartheta_n} D_{\phi_1} \left(\mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)}), \mathbf{p}(\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)}) \right) = \mathbf{Z}^T \mathbf{Z} + o_p(1).$$

Por otra parte, cuando las hipótesis alternativas son contiguas se comprobó en la demostración del Teorema 5.2 que

$$\begin{aligned} \mathbf{Z} &= \sqrt{\frac{n}{\vartheta_n}} \sqrt{N} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} (\mathbf{A}_{(i+1)} - \mathbf{A}_{(i)}) (\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) \\ &\quad + o(\|\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)} - \boldsymbol{\theta}_0\|) + o(\|\widehat{\boldsymbol{\theta}}_{\phi_2}^{(i)} - \boldsymbol{\theta}_0\|). \end{aligned}$$

Luego bajo las alternativas (4.18) se tiene,

$$\sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{d}, \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}),$$

y así

$$\mathbf{Z} \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\boldsymbol{\mu}, \mathbf{V}^*(\boldsymbol{\theta}_0)),$$

donde $\boldsymbol{\mu} = (\mathbf{A}_{(i+1)} - \mathbf{A}_{(i)}) \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \mathbf{d}$ y

$$\begin{aligned} \mathbf{V}^*(\boldsymbol{\theta}_0) &= \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} (\mathbf{A}_{(i+1)} - \mathbf{A}_{(i)}) \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} (\mathbf{A}_{(i+1)} - \mathbf{A}_{(i)}) \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \\ &= (\mathbf{A}_{(i+1)} - \mathbf{A}_{(i)}) \end{aligned}$$

es una matriz idempotente de rango $d_i - d_{i+1}$. Si se comprueba que $\mathbf{V}^*(\boldsymbol{\theta}_0) \boldsymbol{\mu} = \boldsymbol{\mu}$, el parámetro de no centralidad sería $\boldsymbol{\delta} = \boldsymbol{\mu}^T \boldsymbol{\mu}$ y quedaría demostrado el teorema. Puesto que,

$$\begin{aligned} \mathbf{V}^*(\boldsymbol{\theta}_0) \boldsymbol{\mu} &= (\mathbf{A}_{(i+1)} - \mathbf{A}_{(i)}) \boldsymbol{\mu} \\ &= (\mathbf{A}_{(i+1)} - \mathbf{A}_{(i)}) (\mathbf{A}_{(i+1)} - \mathbf{A}_{(i)}) \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \mathbf{d} \\ &= (\mathbf{A}_{(i+1)} - \mathbf{A}_{(i)}) \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \mathbf{d} = \boldsymbol{\mu}, \end{aligned}$$

se tiene

$$\boldsymbol{\delta} = \mathbf{d}^T \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} (\mathbf{A}_{(i+1)} - \mathbf{A}_{(i)})^T \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \mathbf{d}.$$

■

Observación 5.1 *Para obtener una aproximación de la función de potencia en un punto \mathbf{q} de la alternativa se puede utilizar el Teorema 5.3. En efecto, \mathbf{q} se puede escribir como*

$$\mathbf{q} = \mathbf{p}(\boldsymbol{\theta}_0) + \frac{1}{\sqrt{N}} \sqrt{N} (\mathbf{q} - \mathbf{p}(\boldsymbol{\theta}_0)) = \mathbf{p}(\boldsymbol{\theta}_0) + \frac{1}{\sqrt{N}} \mathbf{d}$$

siendo

$$\mathbf{d} = \sqrt{N} (\mathbf{q} - \mathbf{p}(\boldsymbol{\theta}_0)).$$

5.5. Aplicación de modelos loglineales jerárquicos para medir el grado de acuerdo entre evaluadores (rater agreement)

Cuando se presenta una nueva técnica de diagnóstico es imprescindible medir su fiabilidad comparando con un estándar ya existente, en otras ocasiones es necesario medir el grado de acuerdo entre los diagnósticos proporcionados por varios especialistas sobre un conjunto de individuos. En estas situaciones se plantea la necesidad de definir coeficientes que midan de una forma objetiva el grado de acuerdo. La situación más sencilla se presenta cuando dos médicos evalúan y clasifican a n individuos en una de las M categorías (diagnósticos), para medir el grado de acuerdo entre estos evaluadores se define el coeficiente Kappa, κ , (Cohen, 1960; Krippendorff, 1980). En concreto, se define

$$\theta_1 = \sum_{r=1}^M \pi_{rr},$$

suma de las probabilidades donde ambos evaluadores coinciden y

$$\theta_2 = \sum_{r=1}^M \pi_{r\bullet} \pi_{\bullet r},$$

la suma de éstas probabilidades en el caso que los diagnósticos de los evaluadores fueran independientes y sus coincidencias se debieran al azar. El coeficiente Kappa definido como

$$\kappa = \frac{\theta_1 - \theta_2}{1 - \theta_1},$$

mide la proporción de individuos en donde los dos evaluadores están de acuerdo, sin tener en cuenta las coincidencias que se producen debidas al azar. Un estimador de este coeficiente en el caso de muestreo multinomial viene dado por

$$\hat{\kappa} = \frac{n \sum_{r=1}^M n_{rr} - \sum_{r=1}^M n_{r\bullet} n_{\bullet r}}{n^2 - \sum_{r=1}^M n_{r\bullet} n_{\bullet r}},$$

donde n es el número de individuos evaluados. Las características de esta medida del grado de acuerdo son las siguientes: valores negativos del coeficiente indican un grado de acuerdo inferior al del azar, el coeficiente es cero en caso de que el acuerdo sea el mismo que el debido al azar (aunque esto no significa que los evaluadores sean independientes) y $\kappa = 1$ indica que las celdas en donde no hay acuerdo entre los evaluadores están vacías, es decir, el acuerdo es total. Sin embargo, este coeficiente no recoge la información sobre la importancia que tiene coincidir en algunas categorías. En este sentido, Cohen (1968) introduce el coeficiente Kappa ponderado por diferentes pesos que midan esta diferencias. Así, se define

$$\theta_1^* = \sum_{r=1}^M \sum_{s=1}^M w_{rs} \pi_{rs}, \quad \text{con } 0 \leq w_{rs} \leq 1,$$

$$\theta_2^* = \sum_{r=1}^M \sum_{s=1}^M w_{rs} \pi_{r\bullet} \pi_{\bullet s}$$

y el coeficiente κ_w ponderado

$$\kappa_w = \frac{\theta_1^* - \theta_2^*}{1 - \theta_2^*}.$$

Algunos autores, como Agresti (1988) aseguran que un único estadístico puede ser insuficiente para medir el grado de acuerdo en algunas situaciones. Por ejemplo cuando las variables son ordinales y el acuerdo implica no solo medir las coincidencias exactas, sino también la asignación de puntuaciones similares. A continuación, se presentan modelos loglineales para medir el grado de acuerdo entre evaluadores que intentan explicar la naturaleza de ese acuerdo. Inicialmente se presenta un modelo base, que recoge el modelo de independencia entre evaluadores. Por ejemplo, en el caso de dos evaluadores A y B el modelo base será

$$\log m_{ij} = \theta + \theta_i^A + \theta_j^B, \quad i, j = 1, \dots, M - 1,$$

donde θ_i^A , ($i = 1, \dots, M - 1$), representa el efecto del evaluador A y θ_j^B ($j = 1, \dots, M - 1$), representa el efecto del evaluador B . Los siguientes modelos incorporan nuevos parámetros a este modelo base para medir el grado de acuerdo, es decir, se presentarán modelos loglineales jerárquicos. Además, existen varios modelos base posibles dependiendo del número de categorías, por ejemplo Brennan y Prediger (1981) proponen un modelo base en el que los dos parámetros θ^A , θ^B son cero. En el Capítulo 1 se definió el modelo loglineal,

$$\log \mathbf{m}(\boldsymbol{\theta}) = \mathbf{X}\boldsymbol{\theta}^*, \quad \boldsymbol{\theta}^* = (u, \boldsymbol{\theta}^T)^T, \quad \boldsymbol{\theta}^T = (\theta_1, \dots, \theta_{M_0})^T.$$

Sea \mathbf{X} una matriz de dimensión $M \times (M_0 + 1)$ con $rg(\mathbf{X}) = M_0 + 1$, de la forma,

$$\mathbf{X} = \begin{pmatrix} 1 & w_{11} & \cdots & w_{1M_0} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & w_{M1} & \cdots & w_{MM_0} \end{pmatrix} = (\mathbf{1}_M, \mathbf{W}).$$

Para modelos loglineales en los que se mide el grado de acuerdo, la matriz \mathbf{W} se puede descomponer en submatrices $[\mathbf{1}_M, \mathbf{W}_b, \mathbf{W}_\delta, \mathbf{W}_\beta, \mathbf{W}_C]$, donde $[\mathbf{1}_M, \mathbf{W}_b]$ corresponde al modelo base y el resto de las submatrices modelizan diferentes aspectos del grado de acuerdo.

Específicamente \mathbf{W}_δ es la submatriz de diseño que contiene indicadores para aquellas celdas en donde se produce un total acuerdo entre los evaluadores, éstas son las celdas de la diagonal de la tabla de contingencia. \mathbf{W}_β es la submatriz de diseño que refleja el grado de acuerdo cuando las categorías de evaluación son ordinales y el acuerdo, no solo se mide con las coincidencias sino también con la distancia entre evaluaciones. \mathbf{W}_C es la submatriz de diseño que recoge la influencia en el acuerdo de una covariable C , en el caso de que exista. Así mismo, los parámetros del modelo serían $\boldsymbol{\theta}^T = (\boldsymbol{\theta}_b^T, \boldsymbol{\theta}_\delta^T, \boldsymbol{\theta}_\beta^T, \boldsymbol{\theta}_C^T)$. Cualquier selección de submatrices da lugar a un modelo diferente para explicar el acuerdo. El modelo general sería

$$\log p_{ij}(\boldsymbol{\theta}) = \theta + \theta_i^A + \theta_j^B + \delta I(i = j) + \beta \mu_i \mu_j + \theta_{ij}^C \quad i, j = 1, \dots, M - 1,$$

donde $\beta \mu_i \mu_j$ asigna pesos proporcionales al producto de los valores ordinales de las categorías que representa cada celda, es decir, $\{\mu_i\}_{i,j} = \{1, \dots, M\}$. A continuación, se presentan con detalle cada uno de los submodelos que miden diferentes grados de acuerdo sobre un conjunto de datos, en donde, además aparecen dos grupos de individuos (clusters).

5.5.1. Aplicación sobre categorías de tipo ordinal

Se utilizarán los datos de diagnóstico de esclerosis múltiple analizados por Landis y Koch (1977). En estos datos se recogen dos grupos de pacientes, uno en Winnipeg, Manitoba y otro grupo en New Orleans, Louisiana. Cada uno de ellos es examinado por un neurólogo en su lugar de origen, seguidamente los neurólogos se intercambian de ciudad y analizan a los otros pacientes sin conocer el análisis de su predecesor. Los diagnósticos se agruparon en las siguientes categorías: 1. Seguro, 2. Probable, 3. Posible y 4. Dudoso o nulo. Los datos se recogen en la Tabla 5.1.

Cuando se utilizan los modelos loglineales de tipo jerárquico, se parte del modelo más completo para realizar los contrastes de hipótesis hasta conseguir el modelo más sencillo

Tabla 5.1. Datos de Esclerósis (Landis y Koch).

		Neurólogo W							
Pacientes		W				N.O.			
Neurólogo N.O		1	2	3	4	1	2	3	4
1		38	5	0	1	5	3	0	0
2		33	11	3	0	3	11	4	0
3		10	14	5	6	2	13	3	4
4		3	7	3	10	1	4	4	14

que explica adecuadamente los datos. El modelo más completo (Modelo 1) para medir el grado de acuerdo del que partimos en este caso, dado que no tenemos una covariable C es,

$$M1 : \log p_{ij}(\boldsymbol{\theta}) = \theta + \theta_i^A + \theta_j^B + \delta I(i = j) + \beta \mu_i \mu_j \quad i, j = 1, \dots, M - 1.$$

Cuya matriz de diseño asociada es

$$[\mathbf{1}_M, \mathbf{W}_b, \mathbf{W}_\delta, \mathbf{W}_\beta] = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 2 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 3 \\ 1 & 1 & 0 & 0 & -1 & -1 & -1 & 0 & 4 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 2 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 4 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 6 \\ 1 & 0 & 1 & 0 & -1 & -1 & -1 & 0 & 8 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 3 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 6 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 9 \\ 1 & 0 & 0 & 1 & -1 & -1 & -1 & 0 & 12 \\ 1 & -1 & -1 & -1 & 1 & 0 & 0 & 0 & 4 \\ 1 & -1 & -1 & -1 & 0 & 1 & 0 & 0 & 8 \\ 1 & -1 & -1 & -1 & 0 & 0 & 1 & 0 & 12 \\ 1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 16 \end{bmatrix}.$$

Los siguientes dos modelos anidados dependiendo del tipo de acuerdo serán: el Modelo 2 y el Modelo 2*. El Modelo 2, específico para categorías de tipo ordinal es

$$M2 : \log p_{ij}(\boldsymbol{\theta}) = \theta + \theta_i^A + \theta_j^B + \beta\mu_i\mu_j \quad i, j = 1, \dots, M - 1,$$

donde $\beta\mu_i\mu_j$ asigna pesos proporcionales al producto de los valores ordinales de las categorías que representa cada celda, es decir, $\{\mu_i\}_{i,j} = \{1, \dots, M\}$, y cuya matriz de diseño está formada por las submatrices $[\mathbf{1}_M, \mathbf{W}_b, \mathbf{W}_\beta]$. El Modelo 2*

$$M2^* : \log p_{ij}(\boldsymbol{\theta}) = \theta + \theta_i^A + \theta_j^B + \delta I(i = j) \quad i, j = 1, \dots, M - 1,$$

donde $\delta I(i = j)$ asigna pesos a las celdas de la diagonal principal, es decir cuando los dos evaluadores están de acuerdo,

$$\delta I(i = j) = \begin{cases} \delta, & \text{si } i = j \\ 0, & \text{si } i \neq j \end{cases}$$

y su matriz asociada es $[\mathbf{1}_M, \mathbf{W}_b, \mathbf{W}_\delta]$. Por último, el modelo base (Modelo 3) en el caso de independencia

$$M3 : \log p_{ij}(\boldsymbol{\theta}) = \theta + \theta_i^A + \theta_j^B, \quad i, j = 1, \dots, M - 1,$$

cuya matriz asociada es $[\mathbf{1}_M, \mathbf{W}_b]$. Para medir el grado de acuerdo entre los dos neurólogos se realizarán los contrastes sucesivos hasta encontrar el modelo más sencillo que explica adecuadamente los datos. En primer lugar se contrastará

$$\begin{aligned} H_N : H_2 & \quad \text{es decir} & \quad M2 : \log p_{ij}(\boldsymbol{\theta}) = \theta + \theta_i^A + \theta_j^B + \beta \mu_i \mu_j \\ H_{Alt} : H_1 & \quad \text{es decir} & \quad M1 : \log p_{ij}(\boldsymbol{\theta}) = \theta + \theta_i^A + \theta_j^B + \delta I(i = j) + \beta \mu_i \mu_j. \end{aligned}$$

Haciendo uso de la familia de estadísticos $T_{\phi_1, \phi_2}^{(i)}$ cuando los estadísticos de contraste sean grandes se deberá rechazar la hipótesis nula, es decir, se rechazará cuando $T_{\phi_1, \phi_2}^{(i)} > \chi_{d_i - d_{i+1}, \alpha}^2$ para que el tamaño del test sea α . El estadístico para contrastar la hipótesis planteada,

$$T_{\phi_1, \phi_2}^{(i)} = \frac{2Nn}{\phi_1''(1) \tilde{\vartheta}_n} d_{\phi_1} \left(\mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^{(i)}), \mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)}) \right),$$

se calcula utilizando como estimador del efecto del diseño $\tilde{\vartheta}_{\hat{n}^*, N}(\tilde{\mathbf{Y}}, \hat{\boldsymbol{\theta}}_{\phi_\lambda})$ (Tabla 5.2).

Con todos los estadísticos utilizados se llega a la misma conclusión, de que no podemos rechazar la hipótesis nula, luego el Modelo 1 no supone una mejora significativa en la explicación de los datos. Además, si se contrasta la bondad de ajuste del Modelo 2, utilizando

Tabla 5.2. Valores del estadístico $T_{\phi_1, \phi_2}^{(i)}$ y su p-valor.

ϕ_1, ϕ_2	$\lambda_2 = -\frac{1}{2}$	$\lambda_2 = 0$	$\lambda_2 = \frac{2}{3}$	$\lambda_2 = 1$	$\lambda_2 = 2$
$\tilde{\vartheta}_{\hat{n}^*, N}(\mathbf{Y}, \hat{\boldsymbol{\theta}}_{\phi_\lambda})$	4.7852	4.3100	4.2569	4.2492	4.2556
$\lambda_1 = -\frac{1}{2}$	0.0012 (0.9722)	0.3878 (0.5334)	0.0143 (0.9046)	0.0161 (0.8991)	0.0066 (0.9354)
$\lambda_1 = 0$	0.0012 (0.9722)	0.3822 (0.5364)	0.0143 (0.9046)	0.0161 (0.8991)	0.0066 (0.9354)
$\lambda_1 = \frac{2}{3}$	0.0012 (0.9722)	0.3752 (0.5402)	0.0143 (0.9046)	0.0161 (0.8991)	0.0066 (0.9354)
$\lambda_1 = 1$	0.0012 (0.9722)	0.3719 (0.5420)	0.0144 (0.9046)	0.0161 (0.8991)	0.0066 (0.9354)
$\lambda_1 = 2$	0.0012 (0.9722)	0.3629 (0.5469)	0.0144 (0.9046)	0.0161 (0.8991)	0.0066 (0.9354)

Tabla 5.3. Valores del estadístico $T^{\phi_1}(\hat{\theta}_{\phi_2}, \tilde{\vartheta}_{\hat{n}^*, N})$ y su p-valor.

ϕ_1, ϕ_2	$\lambda_2 = -\frac{1}{2}$	$\lambda_2 = 0$	$\lambda_2 = \frac{2}{3}$	$\lambda_2 = 1$	$\lambda_2 = 2$
$\tilde{\vartheta}_{\hat{n}^*, N}(\mathbf{Y}, \hat{\boldsymbol{\theta}}_{\phi_\lambda})$	4.7980	4.2448	4.2505	4.2458	4.2541
$\lambda_1 = -\frac{1}{2}$	3.1133 (0.9270)	5.1515 (0.7412)	4.7436 (0.7846)	5.1122 (0.7455)	5.8954 (0.6589)
$\lambda_1 = 0$	3.0376 (0.9320)	3.43353 (0.9043)	3.4252 (0.9049)	3.4327 (0.9043)	3.4259 (0.9048)
$\lambda_1 = \frac{2}{3}$	4.4091 (0.8184)	2.8235 (0.9449)	2.7787 (0.9474)	2.8136 (0.9455)	3.0040 (0.9341)
$\lambda_1 = 1$	6.1372 (0.6318)	2.7151 (0.0509)	2.7396 (0.9496)	2.7107 (0.9512)	2.8146 (0.9454)
$\lambda_1 = 2$	30.3544 (0.0002)	2.7526 (0.9489)	3.1002 (0.9279)	2.7727 (0.9478)	2.5666 (0.9585)

los estadísticos de ϕ -divergencia introducidos en el Capítulo 4, se comprueba que el modelo explica adecuadamente los datos, tal y como se muestra en la Tabla 5.3.

Obsérvese que el Modelo 2 tiene en cuenta los acuerdos entre evaluadores considerando el orden que presentan las categorías. La otra posibilidad para medir el grado de acuerdo es teniendo en cuenta sólo las coincidencias en la diagonal (coincidencias completas). Esta situación es la descrita por el Modelo 2*. Se plantea el contraste

$$\begin{aligned}
 H_N : H_{2^*} & \quad \text{es decir} & M2^* : \log p_{ij}(\boldsymbol{\theta}) &= \theta + \theta_i^A + \theta_j^B + \delta I(i = j) \\
 H_{Alt} : H_1 & \quad \text{es decir} & M1 : \log p_{ij}(\boldsymbol{\theta}) &= \theta + \theta_i^A + \theta_j^B + \delta I(i = j) + \beta \mu_i \mu_j,
 \end{aligned}$$

y los resultados obtenidos se muestran en la Tabla 5.4.

Tabla 5.4. Valores del estadístico $T_{\phi_1, \phi_2}^{(i)}$ y su p-valor.

ϕ_1, ϕ_2	$\lambda_2 = -\frac{1}{2}$	$\lambda_2 = 0$	$\lambda_2 = \frac{2}{3}$	$\lambda_2 = 1$	$\lambda_2 = 2$
$\vartheta_{\hat{n}^*, N}(\widehat{\mathbf{Y}}, \widehat{\boldsymbol{\theta}}_{\phi_\lambda})$	5.7471	5.2829	5.2784	5.2729	5.2866
$\lambda_1 = -\frac{1}{2}$	15.7862 (0.00007)	15.7202 (0.00007)	13.0010 (0.0003)	12.6100 (0.0004)	12.0422 (0.0005)
$\lambda_1 = 0$	19.9697 (0.00007)	17.8061 (0.00007)	16.3195 (0.00007)	15.7449 (0.00007)	14.8404 (0.0001)
$\lambda_1 = \frac{2}{3}$	39.9697 (< 0.0001)	27.5890 (< 0.0001)	26.6813 (< 0.0001)	25.0346 (< 0.0001)	22.3707 (< 0.0001)
$\lambda_1 = 1$	73.7300 (< 0.0001)	41.1032 (< 0.0001)	38.0782 (< 0.0001)	34.7229 (< 0.0001)	29.4711 (< 0.0001)
$\lambda_1 = 2$	1275.4717 (< 0.0001)	294.0189 (< 0.0001)	185.7738 (< 0.0001)	145.7274 (< 0.0001)	93.7054 (< 0.0001)

Evidentemente con cualquiera de los estadísticos calculados se rechaza la hipótesis nula, luego el Modelo 2* no explica lo suficiente los datos como para reducir el Modelo 1. Sin embargo, el Modelo 2, sí explicaba adecuadamente los datos, por lo que el siguiente contraste a plantear para ver si el modelo puede ser reducido será,

$$\begin{aligned}
 H_N : H_3 & \quad \text{es decir} & M3 : \log p_{ij}(\boldsymbol{\theta}) &= \theta + \theta_i^A + \theta_j^B \\
 H_{Alt} : H_2 & \quad \text{es decir} & M2 : \log p_{ij}(\boldsymbol{\theta}) &= \theta + \theta_i^A + \theta_j^B + \beta \mu_i \mu_j.
 \end{aligned}$$

Los resultados obtenidos para este contraste se muestran en la Tabla 5.5.

Tabla 5.5. Valores del estadístico $T_{\phi_1, \phi_2}^{(i)}$ y su p-valor.

ϕ_1, ϕ_2	$\lambda_2 = -\frac{1}{2}$	$\lambda_2 = 0$	$\lambda_2 = \frac{2}{3}$	$\lambda_2 = 1$	$\lambda_2 = 2$
$\widehat{\vartheta}_{\widehat{n}^*, N}(\widehat{\mathbf{Y}}, \widehat{\boldsymbol{\theta}}_{\phi_\lambda})$	5.7471	5.2829	5.2784	5.2729	5.2866
$\lambda_1 = -\frac{1}{2}$	16.6090 (< 0.0001)	21.4857 (< 0.0001)	20.4278 (< 0.0001)	20.1588 (< 0.0001)	19.6022 (< 0.0001)
$\lambda_1 = 0$	20.1994 (< 0.0001)	28.8061 (< 0.0001)	28.3625 (< 0.0001)	27.8154 (< 0.0001)	26.8006 (< 0.0001)
$\lambda_1 = \frac{2}{3}$	41.3602 (< 0.0001)	55.2513 (< 0.0001)	60.5860 (< 0.0001)	58.0873 (< 0.0001)	53.1615 (< 0.0001)
$\lambda_1 = 1$	79.4423 (< 0.0001)	90.8462 (< 0.0001)	107.2805 (< 0.0001)	99.9457 (< 0.0001)	86.8667 (< 0.0001)
$\lambda_1 = 2$	1681.9200 (< 0.0001)	767.5475 (< 0.0001)	1212.6568 (< 0.0001)	989.8017 (< 0.0001)	680.3003 (< 0.0001)

Puesto que, todos los p-valores son prácticamente cero se rechaza la hipótesis nula. Es decir, el modelo más sencillo que explica adecuadamente los datos es el que incluye la medida del grado de acuerdo que tiene en cuenta las coincidencias en el orden de las categorías, M2.

5.6. Estudio de simulación.

A continuación se realiza un estudio de simulación Monte Carlo para estudiar el comportamiento de los estadísticos de contraste asociados a modelos loglineales jerárquicos. Se estudia el caso particular de modelos loglineales jerárquicos construidos para medir el grado de acuerdo entre evaluadores cuando los individuos evaluados están agrupados en clusters y por lo tanto hay que corregir el valor de los estadísticos por el efecto del diseño. Para ello se generan $R = 1,000$ repeticiones de muestras con 20 clusters ($N = 20$) de tamaño cuarenta ($n = 40$), utilizando valores del coeficiente de correlación intracluster ρ^2 desde 0.15 a 0.95. Para generar las muestras se utilizarán las tres distribuciones que modelizan la dependencia dentro de los clusters: Dirichlet-multinomial (DM), Multinomial

n-inflada (NI) y la distribución Agrupada Aleatoriamente (RC). Se plantea el modelo loglineal jerárquico para determinar el modelo más sencillo que describe de forma correcta el grado de acuerdo que presentan los evaluadores, se realizan los contrastes sucesivos, con un nivel de significación $\alpha = 0.05$. Los siguientes modelos anidados dependiendo del tipo de acuerdo serán:

<i>Modelo</i>	<i>Matriz asociada</i>
$M1 : \quad \log m_{ij} = \theta + \theta_i^A + \theta_j^B + \delta I(i = j) + \beta \mu_i \mu_j$	$[\mathbf{1}_M, \mathbf{W}_b, \mathbf{W}_\delta, \mathbf{W}_\beta]$
$M2 : \quad \log m_{ij} = \theta + \theta_i^A + \theta_j^B + \beta \mu_i \mu_j$	$[\mathbf{1}_M, \mathbf{W}_b, \mathbf{W}_\beta]$
$M2^* : \quad \log m_{ij} = \theta + \theta_i^A + \theta_j^B + \delta I(i = j)$	$[\mathbf{1}_M, \mathbf{W}_b, \mathbf{W}_\delta]$
$M3 : \quad \log m_{ij} = \theta + \theta_i^A + \theta_j^B$	$[\mathbf{1}_M, \mathbf{W}_b]$.

Obsérvese que la dimensión del modelo $M2$ es igual a la del modelo $M2^*$. Se tiene entonces dos secuencias anidadas de contrastes tal y como se realizaron en la aplicación sobre los datos de Landis y Kock. Haciendo uso de la familia de estadísticos

$$T_{\phi_1, \phi_2}^{(i)} = \frac{2Nn}{\phi_1''(1) \tilde{\vartheta}_{n,N}} d_{\phi_1} \left(\mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^{(i)}), \mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^{(i+1)}) \right),$$

donde el estimador del efecto del diseño se calculará utilizando el estimador clásico de Brier, $\tilde{\vartheta}_{n,N}(\tilde{\mathbf{Y}})$, o utilizando el estimador semiparámetro basado en el modelo loglineal planteado en la hipótesis nula, $\tilde{\vartheta}_{n,N}(\tilde{\mathbf{Y}}, \hat{\boldsymbol{\theta}}_{\phi_\lambda})$. En primer lugar se contrastará la secuencia de modelos anidados que contiene al modelo $M2$

$$\begin{array}{ll} H_N : H_2 & \text{Si } T_{\phi_1, \phi_2}^{(1)} < \chi_{1, \alpha}^2 \text{ se contrastará } H_N : H_3 \\ H_{Alt} : H_1 & H_{Alt} : H_2 \end{array}$$

Dado el carácter condicionado de los contrastes sucesivos en una secuencia anidada, el nivel de significación total de cada secuencia de contrastes, depende del nivel de significación utilizado en cada contraste (α) y del número total de contrastes que se realizan

en la secuencia (t) . Martín, N (2007), en su tesis doctoral, “Medidas de divergencia en el análisis estadístico de modelos loglineales con restricciones lineales” demostró que el nivel de significación total de una secuencia anidada de contrastes en modelos loglineales de tipo jerárquico cuando se utilizan los estadísticos $T_{\phi_1, \phi_2}^{(i)}$, se puede calcular como

$$\alpha_t = 1 - (1 - \alpha)^t.$$

En nuestro estudio será $\alpha_t = 1 - (1 - \alpha)^2 = 0.0925$, si el nivel de significación utilizado en cada uno de los dos contrastes que la componen es $\alpha = 0.05$. Este nivel de significación total es el que se ha calculado en las simulaciones para los diferentes estadísticos. Para comparar mejor el comportamiento de los diferentes estadísticos se marca en los gráficos una línea con el nivel de significación teórico ($\alpha_t = 0.0925$). Además hay que tener en cuenta que los niveles de significación de estos contrastes están calculados a partir de la distribución asintótica del estadístico. A continuación se muestran los resultados obtenidos para los niveles de significación de las dos posibles secuencias de contrastes. Los estadísticos que se han escogido para estudiar su comportamiento son los mismos que analizamos en el Capítulo 4.

En el caso de la secuencia anidada que contrasta el modelo de grado de acuerdo con categorías de tipo ordinal (Modelo 2), los resultados obtenidos aparecen resumidos en las figuras 5.1, 5.2 y 5.3, para las tres distribuciones utilizadas en la generación de muestras que modelizan los diferentes niveles de correlación dentro de los clusters. Cabe destacar que los estadísticos corregidos por el estimador semiparámetro no se ven afectados por el aumento de la interdependencia dentro de los clusters, mientras que, la corrección por el coeficiente de Brier modificado no consigue eliminar el efecto de la correlación intracluster especialmente a partir de $\rho^2 = 0.5$. Obsérvese que el estadístico que presenta un peor comportamiento es el del cociente de verosimilitudes ($\lambda_1 = \lambda_2 = 0$).

Figura 5.1. Probabilidad de error de Tipo I estimada para $T_{\phi_1, \phi_2}^{(i)}$ con datos generados con la distribución Dirichlet Multinomial en función del coeficiente de correlación. M2.

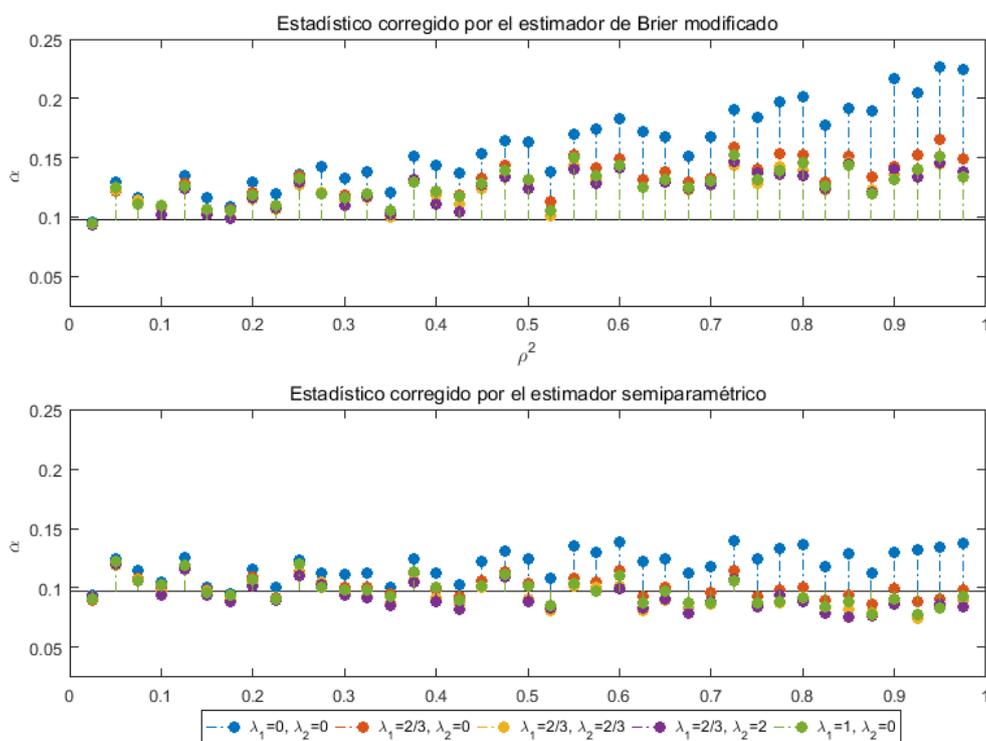


Figura 5.2. Probabilidad de error de Tipo I estimada para $T_{\phi_1, \phi_2}^{(i)}$ con datos generados con la distribución Agrupada Aleatoriamente en función del coeficiente de correlación. M2.

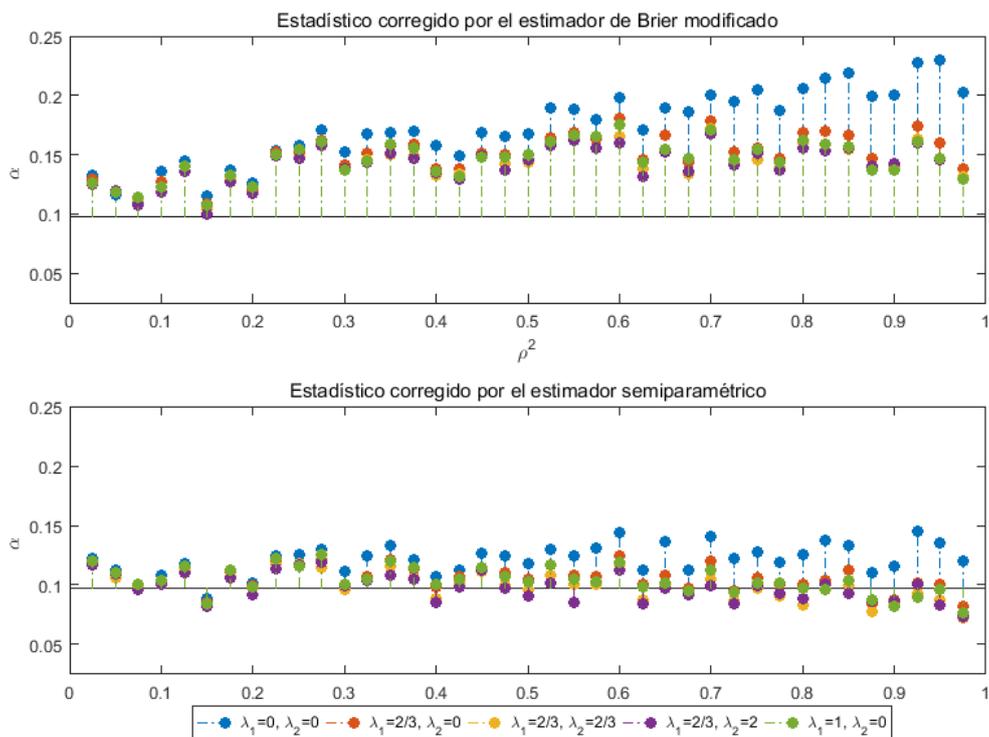
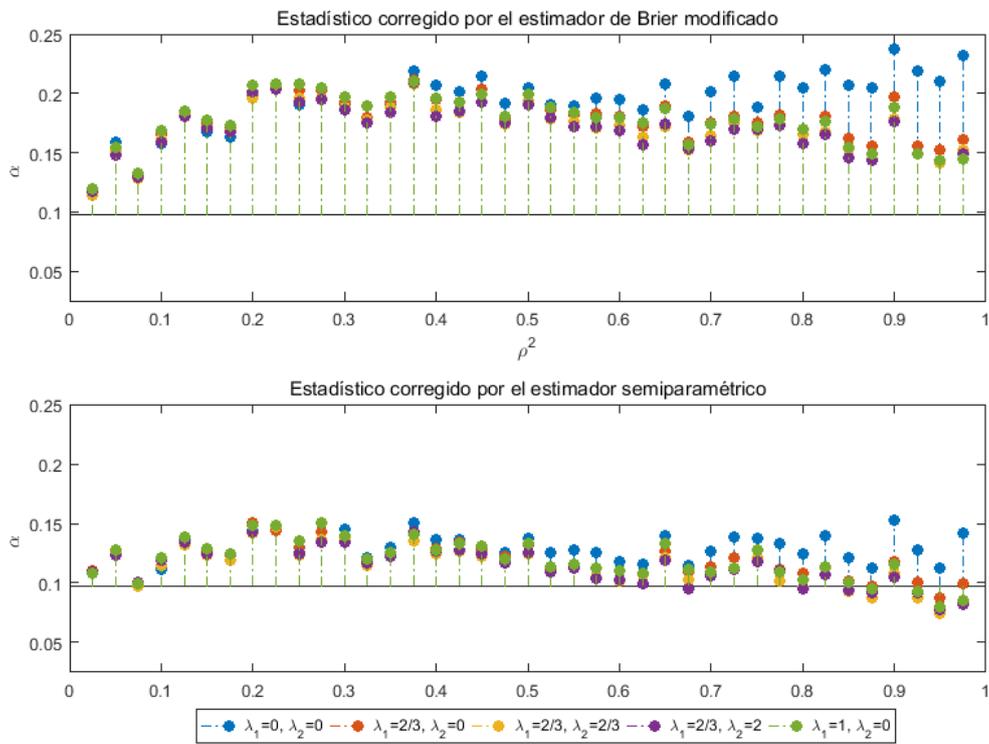


Figura 5.3. Probabilidad de error de Tipo I estimada para $T_{\phi_1, \phi_2}^{(i)}$ en función del coeficiente de correlación con datos generados con la distribución n-inflada. M2.



En segundo lugar, se contrastará la secuencia que contiene al modelo que mide el grado de acuerdo teniendo en cuenta sólo las coincidencias en la diagonal (Modelo 2*),

$$\begin{array}{ll}
 H_N : H_{2^*} & \text{Si } T_{\phi_1, \phi_2}^{(1)} < \chi_{1, \alpha}^2 \text{ se contrastará } H_N : H_3 \\
 H_{Alt} : H_1 & H_{Alt} : H_{2^*}.
 \end{array}$$

El comportamiento de los estadísticos en los contrastes de esta secuencia anidada es idéntico al obtenido en la anterior, como era de esperar. Los resultados se presentan en las figuras 5.4, 5.5 y 5.6.

Figura 5.4. Probabilidad de error de Tipo I estimada para $T_{\phi_1, \phi_2}^{(i)}$ con datos generados con la distribución Dirichlet Multinomial en función del coeficiente de correlación. $M2^*$.

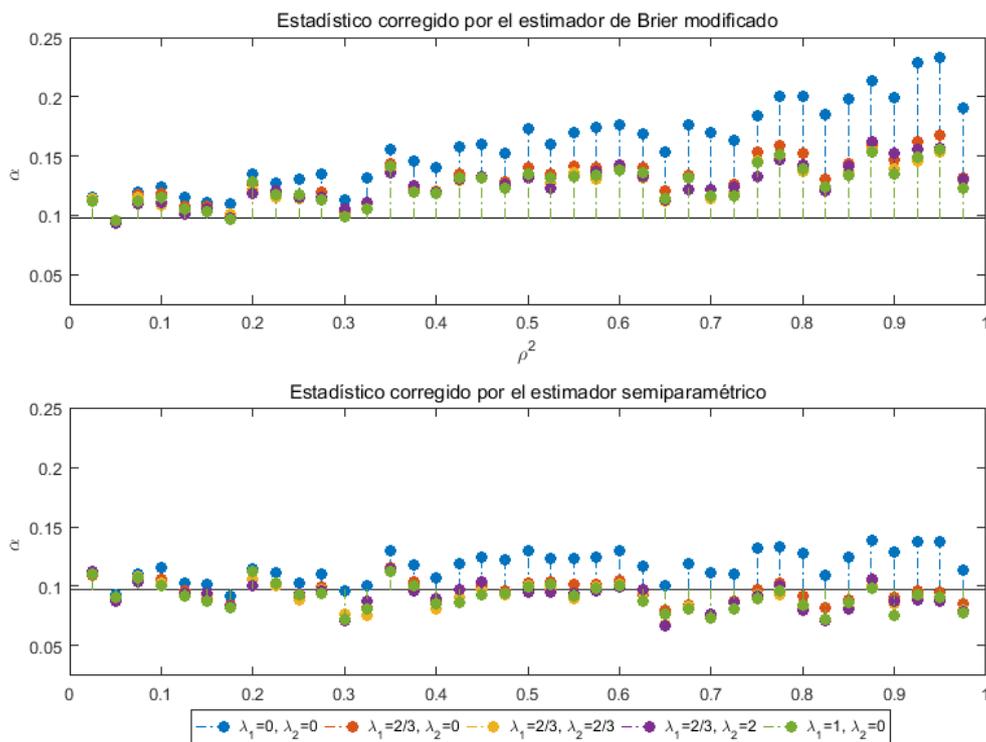


Figura 5.5. Probabilidad de error de Tipo I estimada para $T_{\phi_1, \phi_2}^{(i)}$ con datos generados con la distribución Agrupada Aleatoriamente en función del coeficiente de correlación. $M2^*$.

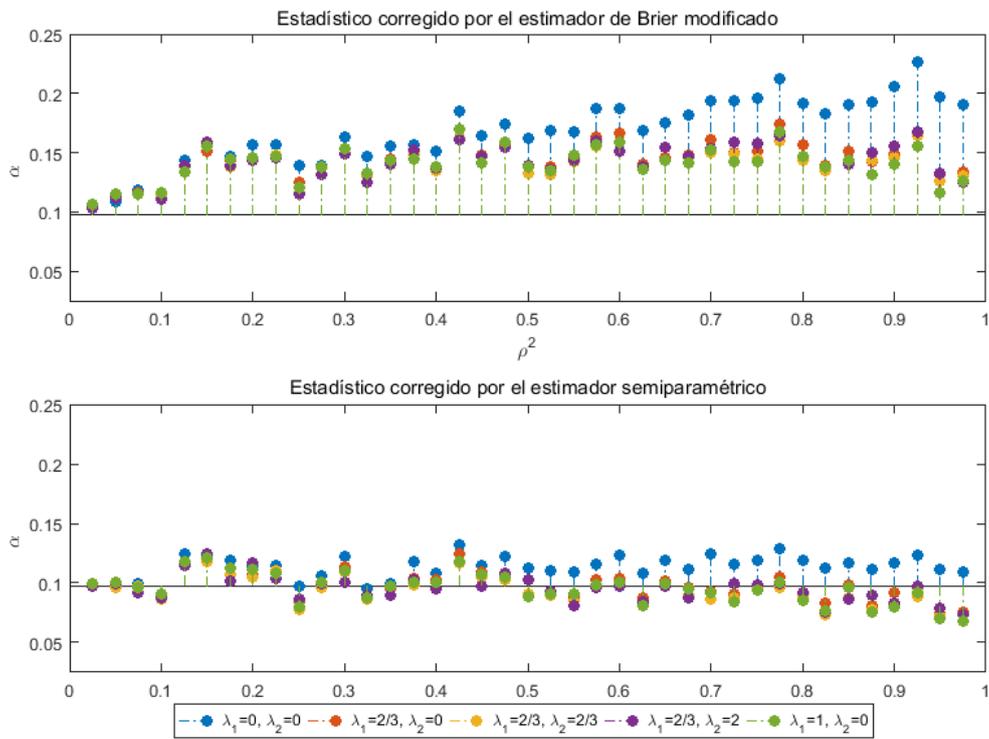
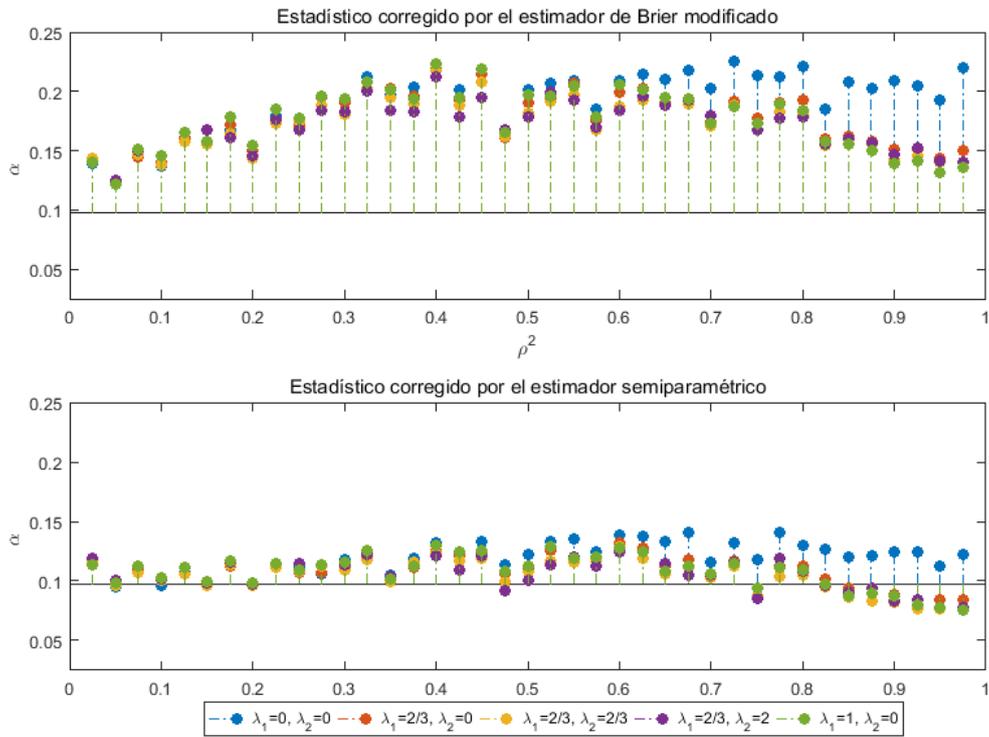


Figura 5.6. Probabilidad de error de Tipo I estimada para $T_{\phi_1, \phi_2}^{(i)}$ con datos generados con la distribución n-inflada en función del coeficiente de correlación. $M2^*$.



CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

En el desarrollo de esta tesis se ha estudiado el comportamiento de los estimadores de mínima ϕ -divergencia para modelos loglineales sobre datos multinomiales con sobredispersión. No se ha supuesto ninguna distribución específica sobre los datos, solamente sobre los dos primeros momentos, por esta razón los estimadores obtenidos para el modelo loglineal son *estimadores de mínima cuasi ϕ -divergencia* ($EMC\phi$), y dado que estos son una generalización de los estimadores de máxima cuasi-verosimilitud ($EMCV$), los resultados obtenidos sobre la distribución asintótica son generalizables también a estos. Además, se han presentado nuevos estimadores del efecto del diseño y del coeficiente de correlación intracluster para el caso de clusters del mismo tamaño cuyo comportamiento, en cuanto a su sesgo y error cuadrático medio es sensiblemente mejor que el de los estimadores clásicos presentados por Brier. En el caso de clusters de tamaños diferentes, se ha introducido un nuevo coeficiente de sobredispersión asociado al tamaño muestral medio y estimadores específicos para el caso de tamaños de cluster diferentes y grandes. Estos resultados han sido publicados en el artículo, “New improved estimators for overdispersion in models with clustered multinomial data and unequal cluster sizes”. *Statistics and Computing* (2017).

Un vez encontrados los estimadores, se plantea el problema de bondad de ajuste de los datos bajo consideración, a un modelo loglineal determinado. A tal fin, se ha introducido una familia de estadísticos de contraste basada en medidas de ϕ -divergencia en la que los parámetros desconocidos se estiman a través del estimador de cuasi-mínima ϕ -divergencia y se corrige el efecto del diseño con los nuevos estimadores presentados, tanto para tamaños de cluster iguales como distintos. En el estudio de simulación se ha comprobado que los

nuevos estadísticos de contraste de bondad de ajuste corregidos por el nuevo estimador del efecto del diseño mantienen la probabilidad de error de tipo I en el nivel del tamaño exacto del contraste fijado inicialmente, incluso cuando aumenta la correlación intracluster. Los resultados obtenidos se han recogido en el artículo: “Semiparametric clustered overdispersed multinomial goodness-of-fit for log-linear models”

Por último, se ha encontrado la distribución asintótica de los estadísticos de bondad de ajuste basados en medidas de ϕ -divergencia y corregidos por el efecto del diseño, para el caso de modelos loglineales de tipo jerárquico. Además se ha estudiado el comportamiento de los nuevos estadísticos para el caso particular, presentado por Agresti (1988), de su aplicación al estudio del grado de acuerdo entre evaluadores, comprobando que mantienen su buen comportamiento aunque la correlación intracluster aumente. En este momento se está redactando un trabajo que recogerá los resultados originales obtenidos en este capítulo.

Como futuras líneas de investigación se contempla la posibilidad de considerar la clase de modelos loglineales generalizados que considera Lang (1996). Esta consideración permitirá analizar no sólo nuevos modelos, sino también utilizar nuevas distribuciones

BIBLIOGRAFÍA

- [1] Agresti, A. (2002). *Categorical data Analysis*. 2^a Ed John Wiley and Sons.
- [2] Agresti, A. (1988). A Model for Agreement Between Ratings on an Ordinal Scale. *Biometrics*, **44**, No. 2,. 539-548.
- [3] Ahn, H. and James, J. C. (1995). Generation of Over-Dispersed and Under-Dispersed Binomial Variates. *Journal of Computational and Graphical Statistics*, **4**, 55–64.
- [4] Alonso-Revenga, J. M., Martín, N. and Pardo, L. (2017). New improved estimators for overdispersion in models with clustered multinomial data and unequal cluster sizes. *Statistics and Computing*, **27**,193-217.
- [5] Alonso-Revenga, J. M., Martín, N. and Pardo, L. (2017). Semiparametric clustered overdispersed multinomial goodness-of-fit for log-linear models. ArXiv:1609.07330.
- [6] Altham, P. M. E. (1976). Discrete variable analysis for individuals grouped into families. *Biometrika*, **63**, 263–269.
- [7] Becker, M.P., Agresti, A. (1992). Log-linear Modelling of pairwise interobserver agreement on a categorical scale. *Statistics in Medicine*. **11**, 104-114.
- [8] Bedrick, E. J. (1983). Adjusted chi-squared tests for cross-classified tables of survey data. *Biometrika*, **70**, 591–595.
- [9] Bishop, Y.M.M, Fienberg, S.E., & Holland, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice*, Cambridge , MA: MIT Press.
- [10] Brier, S. S. (1980). Analysis of contingency tables under cluster sampling. *Biometrika*, **67**, 591–596.

- [11] Bokossa, M. (1999). *Parameter estimation in overdispersion models*. Unpublished Ph.D. thesis.
- [12] Budowle, B. and Moretti, T. R. (1999). Genotype profiles for six population groups at the 13 CODIS short tandem repeat core loci and other PCR-based loci. *Forensic Science Communications* 1999.
- [13] Butler, (2006). Genetics and Genomics of Core STR Loci Used in Human Identity Testing, *Journal of Forensic Sciences*, 1-48.
- [14] Choi J.W. and McHugh R. B. (1989). A reduction factor in goodness-of-fit and independence tests for clustered and weighted observations. *Biometrics*, **45**, 979–996.
- [15] Cohen J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*.; **20**(1):37–46.
- [16] Cohen J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*. ; **70**(4):213–220.
- [17] Cohen, J. E. (1976). The distribution of the chi-squared statistic under clustered sampling from contingency tables. *Journal of American Statistical Association*, **71**, 665–670.
- [18] Cressie, N., Pardo, L. (2000). Minimum ϕ -divergence estimator and hierarchical testing in loglinear models. *Statistica Sinica*, **10**, 867–884.
- [19] Cressie, N., Pardo, L. (2002). Model checking in loglinear models using ϕ -divergences and MLEs. *Journal of Statistical Planning and Inference*, **103**, 437–453.
- [20] Cressie, N., Pardo, L., Pardo, M. C. (2003). Size and power considerations for testing loglinear models using ϕ -divergence test statistics. *Statistica Sinica*, **13**, 555–570.
- [21] Christensen, R. (1997). *Log-Linear Models and Logistic Regression*. 2^o Ed Springer-Verlag.

- [22] Dale J. R: (1986). Asymptotic Normality of Goodness-of-Fit Statistics for Sparse Product Multinomials. *Journal of de Royal Statistical Society. Serie B.* **48**, 48-59.
- [23] Donner, A. (1989). Statistical methods in ophthalmology : An adjusted Chi-square approach, *Biometrics*, **45**, 605-611.
- [24] Eldridge, S.M., Ukoumunne, O.C and .Carlin, J B.(2009). The Intra-Cluster Correlation Coefficient.in Cluster Randomized Trials: A Review of Definitions *International Statistical Review*, **77**, 3, 378–394.
- [25] Eldridge, S. Kelly, S.(2012). *A practical guide to cluster ramdomised trials in health Services Research*. Ed. John Wiley and Sons .
- [26] Fay, R. E. (1985). Complex samples. *Journal of the American Statistical Association*, **80**, 148–157.
- [27] Fellegi, I. P. (1980). Approximate tests of independence and goodness of fit based upon stratified multistage samples. *Journal of the American Statistical Association*, **75**, 261–268.
- [28] Fienberg, S. E. (1979). The use of chi-square statistics for categorical data problems. *Journal of the Royal Statistical Society, Series B*, **41**, 54–64.
- [29] Harville, D.A.(2008). *Matrix Algebra From a Statistician's Perspective*. Spinger.
- [30] Holt, D., Scott, A. J. and Ewings, P. O. (1980). Chi-squared tests with survey data. *Journal of the Royal Statistical Society, Series A*, **143**, 302–320.
- [31] Kroonenberg, P.M.(2008). *Applied Multiway Data Analisis*. John Wiley & Sons, Inc.
- [32] Krippendorff, K.(1980). *Content analysis: An introduction to its methodology*. Sage Publications; Beverly Hills, CA.
- [33] Landis JR, Koch GG.(1977). The measurement of observer agreement for categorical data. *Biometrics*, **33** (1):159–174.

- [34] Landis, J. R., Lepkowski, J. M., Eklund, S. A. and Stehouwer, S. A. (1984). *A Statistical Methodology for Analyzing Data from a Complex Survey: The First National Health and Nutrition Examination Survey*. Series 2 , No. 92. Hyattsville, Maryland: The National Center for Health Statistics.
- [35] Lang, J.B. (1996). Maximum likelihood methods for a generalized class of log-linear models. *The Annals of Statistics*, **24**, 726-752.
- [36] Martín, N. and Pardo, L. (2008a). New families of estimators and test statistics in log-linear models. *Journal of Multivariate Analysis*, **99**, 1590–1609.
- [37] Martín, N. and Pardo, L. (2008b). Minimum phi-divergence estimators for loglinear models with linear constraints and multinomial sampling, *Statistical Papers*, **49**, 15–36.
- [38] Martín, N. (2007). Medidas de divergencia en el análisis estadístico de modelos log-lineales con restricciones lineales Tesis UCM.
- [39] Martín, N. and Pardo, L. (2010). A new measure of leverage cells in multinomial loglinear models. *Communications in Statistics - Theory and Methods*, **39**, 517–530.
- [40] Martín, N. and Pardo, L. (2011). Fitting DNA sequences through log-linear modelling with linear constraints. *Statistics*, **45**, 605–621.
- [41] Martín, N. and Pardo, L. (2012). Poisson loglinear modeling with linear constraints on the expected cell frequencies. *Sankhya: The Indian Journal of Statistics*, **74B**, 238–267.
- [42] Menéndez, M. L., Morales, D., Pardo, L. and Vajda, I. (1995). Divergence-based estimation and testing of statistical models of classification. *Journal of Multivariate Analysis*, **54**, 329–354.

- [43] Menéndez, M. L., Morales, D., Pardo, L. and Vajda, I. (1996). About divergence-based goodness-of-fit tests in the Dirichlet-multinomial model. *Communications in Statistics - Theory and Methods*, **25**, 1119–1133.
- [44] Morel, J.G. (1992). A simple algorithm for generating multinomial random vectors with extra-variation. *Communication in Statistics-Simulation and Computation*, **21**, 1255-1268.
- [45] Morel, J.G. and Nagaraj, N.K. (1993). A finite mixture distribution for modelling multinomial extra variation. *Biometrika*, **80**, 363–371.
- [46] Morel, J.G, and Neerchal, N.K (2012). *Overdispersion Models in SAS*. SAS Institute Inc., Cary, NC, USA.
- [47] Mosimann, J. E. (1962). On the compound multinomial distributions, the multivariate β -distribution and correlation among proportions. *Biometrika*, **49**, 65–82.
- [48] Neerchal, N.K. and Morel, J.G. (1998). Large cluster results for two parametric multinomial extra variation models. *Journal of the American Statistical Association*, **93**, 1078–1087.
- [49] Pardo, L. (2006). *Statistical inference based on divergence measures*. Chapman & Hall/CRC.
- [50] Raim, A. M. , Neerchal, N. K. and Morel, J. G. (2015). Modeling overdispersion in *R*. Technical Report HPCI-2015-1 UMBC High Performance Computing Facility, University of Maryland.
- [51] Rao, J. N. K. and Scott, A. J. (1981). The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit and independence in two-way tables. *Journal of the American Statistical Association*, **76**, 221–230.

- [52] Rao, J. N. K. and Scott, A. J. (1984). On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. *The Annals of Statistics*, **12**, 46–60.
- [53] Stokes, M. E., Davis, Ch. S., Koch, G. (2000). *Categorical data Analysis using SAS System*. 2^a Ed. SAS Institute. Cary NC.
- [54] Sutradhar S. C., Neerchal, N.K. and Morel, J.G. (2008). A goodness-of-fit test for overdispersed binomial (or multinomial) models. *Journal of Statistical Planning and Inference*, **138**, 1450-1471.
- [55] Tanner, M. and Young, M. (1985). Modeling Agreement Among Raters. *Journal of the American Statistical Association*, **80**, 389, 175-180.
- [56] Tvedebrink, T. (2010). Statistical Aspects of Forensic Genetics. Models for Qualitative and Quantitative STR Data. Ph.D. thesis.
- [57] Tvedebrink, T. (2010). Overdispersion in allelic counts and θ -correction in forensic genetics. *Theoretical Population Biology*, **78**, 200-210.
- [58] Vos, P. W. (1992). Minimum f-divergence estimators and quasi-likelihood functions. *Annals of the Institute of Statistical Mathematics*, **44**, 261–279.
- [59] Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalised linear models, and the Gauss-Newton method. *Biometrika*, **61**, 439–447.
- [60] Weir, B. S. and Hill, W. G. (2002). Estimating F-statistics. *Annual Review of Genetics*, **36**, 721–750.
- [61] Wilson, J.R. (1989). Chi-square tests for overdispersion with multiparameter estimates. *Applied Statistics*, **38**. 441-453.