



UNIVERSIDAD
DE LA REPUBLICA
URUGUAY



TRAYECTORIA NUTRICIONAL Y DESEMPEÑO ESCOLAR

Una aplicación de modelos conjuntos para datos
longitudinales y de sobrevivencia

María Alejandra Marroig Baldini

Programa de Posgrado en Ingeniería Matemática
Facultad de Ingeniería
Universidad de la República

Montevideo – Uruguay

Setiembre de 2017



UNIVERSIDAD
DE LA REPUBLICA
URUGUAY



TRAYECTORIA NUTRICIONAL Y DESEMPEÑO ESCOLAR

Una aplicación de modelos conjuntos para datos
longitudinales y de sobrevivencia

María Alejandra Marroig Baldini

Tesis de Maestría presentada al Programa de Posgrado en Ingeniería Matemática, Facultad de Ingeniería de la Universidad de la República, como parte de los requisitos necesarios para la obtención del título de Magister en Ingeniería Matemática.

Director:

Ph.D. Prof. Graciela Muniz Terrera

Director académico:

Ph.D. Prof. Paola Bermolen

Montevideo – Uruguay

Setiembre de 2017

Marroig Baldini, María Alejandra

TRAYECTORIA NUTRICIONAL Y DESEMPEÑO ESCOLAR
/ María Alejandra Marroig Baldini. - Montevideo:
Universidad de la República, Facultad de Ingeniería, 2017.
XIX, 116 p.: il.; 29, 7cm.

Director:

Graciela Muniz Terrera

Director académico:

Paola Bermolen

Tesis de Maestría – Universidad de la República,
Programa en Ingeniería Matemática, 2017.

Referencias bibliográficas: p. 89 – 95.

1. Clase Latente, 2. Modelo Conjunto, 3. Trayectoria
Nutricional, 4. Desempeño Escolar, 5. Uruguay.
I. Muniz Terrera, Graciela, . II. Universidad
de la República, Programa de Posgrado en Ingeniería
Matemática. III. Título.

INTEGRANTES DEL TRIBUNAL DE DEFENSA DE TESIS

Prof. Mathias Bourel

Prof. Gustavo Guerberoff

Prof. Andrea Mesa

Prof. Andrea Vigorito

Montevideo – Uruguay

Setiembre de 2017

Agradecimientos

Quisiera agradecer a mi tutora Graciela por sus aportes y la dedicación para la realización de este trabajo, por su hospitalidad en la estadía de Reino Unido y su enorme calidad humana. También a mi directora académica Paola por su invaluable acompañamiento durante el transcurso de mis estudios en el programa de maestría.

También quisiera agradecer a mis compañeros del Instituto de Economía y del Grupo de Estudios de Familia que apoyaron la realización de este trabajo, me dieron la oportunidad de hacer la estadía en Reino Unido y me han animado constantemente para que realice mis estudios de posgrado. A mis compañeros del Instituto de Estadística que han contribuido con la realización de esta tesis.

Agradezco también a mi familia y mis amigos por estar siempre.

La relación entre la libertad individual y el desarrollo social va más allá de la conexión constitutiva. Lo que pueden conseguir positivamente los individuos depende de las oportunidades económicas, las libertades políticas, las fuerzas sociales y las posibilidades que brindan la salud, la educación básica y el fomento y el cultivo de las iniciativas.

[\(Sen, 1999\)](#)

RESUMEN

El estudio del estado nutricional y su trayectoria a lo largo de la vida es relevante, pero cobra mayor importancia durante la infancia ya que los problemas en este período de la vida pueden tener efectos adversos de corto y largo plazo en varias dimensiones incluyendo, entre otras, la educación y la salud. Este trabajo analiza la trayectoria nutricional de una cohorte de niños y niñas de Uruguay en edad escolar y su vínculo con el desempeño educativo mediante la estimación de Modelos Mixtos de Clases Latentes y Modelos Mixtos Conjuntos de Clases Latentes. Esta metodología se enmarca en los modelos estructurales y permite analizar el cambio con la edad del estado nutricional, medido a través del Índice de Masa Corporal (IMC). Además, estos modelos suponen que no es posible capturar la heterogeneidad entre individuos por ninguna variable observada y, por tanto, ésta queda representada mediante la pertenencia a grupos latentes.

Los modelos ajustados identificaron tres grupos de niños y niñas según sus trayectorias nutricionales en el período escolar (entre 6 y 12 años aproximadamente). El 50 % de las niñas tiene un IMC dentro de los niveles normales para la edad, el 40 % presenta una trayectoria normal aunque con niveles más altos que la mediana y que en algunas edades supera el umbral de sobrepeso. Por último, el 10 % de niñas presenta problemas de sobrepeso en edad escolar. Por su parte, el 70 % de los niños tiene una trayectoria normal para la edad, el 20 % supera el umbral de sobrepeso para edades más altas y el 10 % el de obesidad en todo el tramo de edad considerado.

El grupo de niñas con sobrepeso se caracteriza por mayor peso y talla al nacer. Sin embargo, no se encuentran diferencias en términos del nivel educativo de las madres. Para los niños, el grupo por encima del umbral de obesidad tiene mayor peso al nacer que el resto de grupos pero la talla no presentó diferencias significativas. Además, hay una mayor proporción de madres con nivel educativo terciario en los grupos con sobrepeso y obesidad. En cuanto a la relación entre trayectoria nutricional y desempeño escolar, los evidencia muestra que los niños abandonan a edades más tempranas que las niñas. Sin embargo,

los grupos según trayectorias nutricionales no presentan diferencias en la edad de abandono escolar.

Los resultados de este trabajo indican que si bien ninguno de los grupos identificados muestra características de déficit nutricional o decrecimiento del IMC en el período escolar, sí se observa un conjunto de niñas y niños con problemas de sobrepeso u obesidad que podría ser objeto de estudio en futuras investigaciones.

Palabras claves:

Clase Latente, Modelo Conjunto, Trayectoria Nutricional, Desempeño Escolar, Uruguay.

ABSTRACT

The study of the nutritional status and its trajectory over time is relevant but becomes more important during childhood since problems in this period of life can have adverse short and long term effects in several dimensions including, among others, education and health. This work analyzes the nutritional trajectory of a cohort of uruguayan boys and girls in school age and its association with school performance. With this purpose, Latent Class Mixed Models and Joint Latent Class Mixed Models were fitted. This methodology is framed in the structural models and allows an analysis of nutritional status change with age of the cohort, measured through the Body Mass Index (BMI). In addition, the models assume that it is not possible to capture heterogeneity between individuals by any observed variable and, therefore, this is represented by latent groups.

The model identified three groups of boys and girls according to its nutritional trajectories during school (6 to 12 years old). These groups of girls show that 50 % is within normal levels of BMI for age, 40 % has a normal trajectory but with levels higher than the median and, in some ages, that exceeds the overweight threshold and, finally, 10 % of girls present problems of overweight during school age. 70 % of boys have a normal nutritional trajectory, 20 % exceeds the overweight threshold and the 10 % of obesity.

The group of overweight girls is characterized by greater weight and height for birth. However, there are no significant differences in mothers educational attainment. Amongst boys, the group of obese kids has heavier at birth than the other two groups, but size-at-birth did not show significant differences. In addition, there is a higher proportion of mothers with superior educational attainment in the obese and overweight groups. Regarding the relationship between nutritional trajectory and school performance, results indicates that boys drop out at younger ages than girls but nutritional groups do not differ in terms of school dropout age.

None of the groups identified exhibited a deficit or decrease in average BMI, although the group of children with overweight or obesity could be of concern

and the object of future research.

Keywords:

Latent class, Joint Model, Nutritional Trajectory, School Performance,
Uruguay.

Tabla de contenidos

Lista de figuras	XIV
Lista de tablas	XVI
Lista de siglas	XIX
1 Introducción	1
2 Estado nutricional y desempeño escolar	5
2.1 Marco conceptual	5
2.2 Antecedentes	7
2.3 Problema de investigación y relevancia	12
2.4 Objetivos y preguntas de investigación	14
3 Modelos Longitudinales y de Supervivencia	15
3.1 Modelos para Datos Longitudinales	16
3.1.1 Modelo Lineal Mixto	17
3.1.2 Modelo Mixto para Procesos Latentes	21
3.1.3 Modelo Mixto de Clases Latentes	24
3.2 Modelo de Supervivencia	28
3.3 Modelos Conjuntos para Datos Longitudinales y de Supervivencia	32
3.3.1 Modelo Mixto Conjunto de Clases Latentes	34
3.4 Cálculos Posteriores a la Estimación	37

3.4.1	Clasificación <i>a posteriori</i>	37
3.4.2	Elección de Grupos y Bondad de Ajuste	38
3.4.3	Predicciones	40
3.5	Aspectos Metodológicos: Trayectoria Nutricional y Desempeño Escolar	43
3.5.1	Base de Datos	44
3.5.2	Variables	46
3.5.3	Estrategia Empírica	48
4	Resultados	52
4.1	Estadísticas Descriptivas	52
4.2	Análisis de la Trayectoria Nutricional	56
4.2.1	Trayectoria Nutricional según Grupos de Niñas	65
4.2.2	Trayectoria Nutricional según Grupos de Niños	67
4.3	Asociación entre Trayectoria Nutricional y Desempeño Escolar .	69
4.3.1	Trayectoria Nutricional y Desempeño Escolar según Grupos de Niñas	74
4.3.2	Trayectoria Nutricional y Desempeño Escolar según Grupos de Niños	76
4.4	Análisis de Sensibilidad	79
4.4.1	Trayectoria Nutricional y Desempeño Escolar sin Repe- tición	79
5	Conclusiones	84
	Referencias bibliográficas	89
	Apéndices	96
Apéndice A	Resultados	97
A.1	Clasificación de Niñas según Trayectoria Nutricional con Fun- ción de Enlace Lineal y <i>Splines</i>	98

A.2	Clasificación de Niños según Trayectoria Nutricional con función de Enlace Lineal y <i>Splines</i>	100
A.3	Análisis de la Trayectoria Nutricional para Niñas con Función de Enlace <i>Splines</i>	102
A.4	Análisis de la Trayectoria Nutricional para Niños con Función de Enlace <i>Splines</i>	104
Anexos		106
Anexo A Anexo metodológico		107
A.1	Proceso Generador de Datos Faltantes para Datos Longitudinales	107
A.2	Método de Estimación por Máxima Verosimilitud	113
A.3	Algoritmo Marquardt	115

Lista de figuras

3.1	Grafo dirigido que define el LCMM	24
3.2	Grafo dirigido que define el JLCMM	34
4.1	Distribución según edades simples en los años de medición . . .	53
4.2	Trayectorias del IMC para la edad según género	54
4.3	Trayectoria nutricional según dos y tres grupos de niñas - Enlace Lineal	59
4.4	Trayectoria nutricional según cuatro y cinco grupos de niñas - Enlace Lineal	60
4.5	Trayectoria nutricional según dos y tres grupos de niños - Enlace Lineal	62
4.6	Trayectoria nutricional según cuatro y cinco grupos de niños - Enlace Lineal	63
4.7	Trayectoria nutricional media e individual según grupos de niñas	65
4.8	Trayectoria nutricional media e individual según grupos de niños	67
4.9	Trayectoria nutricional según tres y cuatro grupos de niñas - LCMM	71
4.10	Trayectoria nutricional según tres y cuatro grupos de niños - LCMM	73
4.11	Trayectoria nutricional media e individual según grupos de niñas	75
4.12	Riesgo de abandono según grupo de niñas	75
4.13	Trayectoria nutricional media e individual según grupos de niños	77
4.14	Riesgo de abandono según grupos de niños	77

4.15	Trayectoria nutricional media e individual según grupos de niñas	80
4.16	Riesgo de abandono según grupo de niñas	80
4.17	Trayectoria nutricional media e individual según grupos de niños	82
4.18	Riesgo de abandono según grupo de niños	82
A.1	Trayectoria nutricional según dos grupos de niñas - LCMM . . .	98
A.2	Trayectoria nutricional según tres grupos de niñas - LCMM . . .	98
A.3	Trayectoria nutricional según cuatro grupos de niñas - LCMM .	99
A.4	Trayectoria nutricional según cinco grupos de niñas - LCMM . .	99
A.5	Trayectoria nutricional según dos grupos de niños - LCMM . . .	100
A.6	Trayectoria nutricional según tres grupos de niños - LCMM . . .	100
A.7	Trayectoria nutricional según cuatro grupos de niños - LCMM .	101
A.8	Trayectoria nutricional según cinco grupos de niños - LCMM . .	101
A.9	Trayectoria nutricional según dos y tres grupos de niñas - <i>splines</i>	103
A.10	Trayectoria nutricional según cuatro y cinco grupos de niñas - <i>splines</i>	103
A.11	Trayectoria nutricional según dos y tres grupos de niños - <i>splines</i>	105
A.12	Trayectoria nutricional según cuatro y cinco grupos de niños - <i>splines</i>	105

Lista de tablas

2.1	Porcentaje de no asistencia al sistema educativo por edad y sexo	10
3.1	Tabla de clasificación	38
4.1	Distribución por género en los años de medición	53
4.2	Descriptivas del IMC en los años de medición	54
4.3	Tasa de abandono según repetición por género	55
4.4	Distribución de la tasa de abandono según edad por género . . .	55
4.5	Estadísticos y clasificación según grupos de niñas - LCMM - Lineal	57
4.6	Probabilidades <i>a posteriori</i> para tres grupos de niñas - Lineal .	58
4.7	Estadísticos y clasificación según grupos de niños - LCMM - Lineal	61
4.8	Probabilidades <i>a posteriori</i> para tres grupos de niños - Lineal .	61
4.9	Descriptivas para tres grupos de niñas - LCMM - Enlace Lineal	66
4.10	Descriptivas para tres grupos de niños - LCMM - Enlace Lineal	68
4.11	Estadísticos y clasificación según grupos de niñas - LCMM . . .	70
4.12	Estadísticos y clasificación según grupos de niños - LCMM . . .	72
4.13	Estadísticos y clasificación según grupos de niñas - JLCMM . .	74
4.14	Estadísticos y clasificación según grupos de niños - JLCMM . .	76
4.15	Estadísticos y clasificación según grupos de niñas - JLCMM - Sin Repetición	79

4.16 Estadísticos y clasificación según grupos de niños - JLCMM - Sin Repetición	81
A.1 Estadísticos y clasificación según grupos de niñas - LCMM - <i>splines</i>	102
A.2 Probabilidades <i>a posteriori</i> para tres grupos de niñas - <i>splines</i> .	102
A.3 Estadísticos y clasificación según grupos de niños - LCMM - <i>splines</i>	104
A.4 Probabilidades <i>a posteriori</i> para tres grupos de niños - <i>splines</i> .	104

Lista de siglas

- BIC** Criterio de Información Bayesiano
(*Bayesian Information Criterio*) 38, 39, 56, 57, 59, 60, 61, 63, 69, 70, 72, 74, 76, 79, 81, 102, 104
- CSIC** Comisión Sectorial de Investigación Científica 4, 45
- ECH** Encuesta Continua de Hogares 3, 10, 55
- ELBU** Estudio Longitudinal del Bienestar en Uruguay 4, 7, 8, 9, 10, 13, 43, 45, 46, 49, 52, 55, 85, 86, 87
- ENDIS** Encuesta de Nutrición, Desarrollo Infantil y Salud 7, 13
- ESF** Encuesta sobre Situaciones Familiares y Desempeños Sociales 13
- FCE** Fondo Clemente Estable 4
- GLMM** Modelos Lineales Mixtos Generalizados
(*Generalized Linear Mixed Models*) 16
- IMC** Índice de Masa Corporal 2, 7, 8, 12, 43, 45, 46, 48, 49, 50, 52, 53, 54, 58, 59, 62, 65, 67, 71, 72, 75, 76, 80, 81, 82, 84, 85, 87, 98, 99, 100, 101, 102, 103, 104, 105
- INE** Instituto Nacional de Estadística 7
- JLCMM** Modelo Mixto Conjunto de Clases Latentes
(*Joint Latent Class Mixed Models*) xiv, 33, 34, 35, 36, 37, 38, 39, 42, 46, 49, 50, 51, 52, 69, 72, 74, 76, 79, 80, 81, 84, 87
- LCMM** Modelos Mixtos de Clases Latentes
(*Latent Class Mixed Models*) xiv, xv, xvi, xvii, 16, 24, 26, 32, 35, 37, 38, 48, 49, 50, 51, 52, 56, 57, 69, 72, 74, 76, 84, 87, 98, 99, 100, 101, 102, 104
- LMM** Modelos Lineales Mixtos
(*Linear Mixed Models*) 16, 18, 20, 21, 24, 25, 26, 32, 33, 34, 113
- LPMM** Modelos Mixtos para Procesos Latentes
(*Latent Process Mixed Models*) 16, 22, 24, 25, 33, 34

MAR *Missing at Random* 12, 13, 109, 110, 111
MCAR *Missing Completely at Random* 109, 111
MLE Estimación por Máxima Verosimilitud
(*Maximum Likelihood Estimation*) 20, 22, 25, 35, 36, 87, 113
MNAR *Missing Not at Random* 12, 13, 109, 111
OMS Organización Mundial de la Salud 1, 2, 8, 43, 45, 46, 52, 53, 58, 65, 67,
87, 88
SEM Modelos de Ecuaciones Estructurales
(*Structural Equation Models*) 16
UCC Uruguay Crece Contigo 7
UdelaR Universidad de la República 7
Unicef Fondo Internacional de Emergencia de las Naciones Unidas para la
Infancia 4

Capítulo 1

Introducción

El estudio del estado nutricional de las personas es relevante en sí mismo pero, en particular, durante la infancia debido a que los problemas nutricionales en esta etapa de la vida pueden tener efectos adversos en otras dimensiones como, por ejemplo, los logros cognitivos de los niños o la asistencia escolar ([ACC/SCN, 2000](#)). Para el diseño de políticas públicas eficientes resulta de interés la asociación que muestra la literatura entre el déficit nutricional y los menores logros educativos alcanzados ([Glewwe, 2005](#); [Levinger, 1992](#); [Victoria *et al.*, 2008](#)). A su vez, las enfermedades crónicas asociadas al estado nutricional pueden conllevar problemas en lo laboral y consumen recursos de los servicios primarios de salud ([ACC/SCN, 2000](#)).

La transición epidemiológica¹ ha llevado a que en los países en desarrollo el foco de atención sea la desnutrición infantil, mientras que en los países desarrollados lo han sido la obesidad y el sobrepeso. Estos últimos problemas nutricionales se han reportado como factores de riesgo para la salud tanto de corto como largo plazo ([deOnis & Lobstein, 2010](#)) y recientemente han cobrado mayor atención. La prevalencia de obesidad en el mundo se ha más que duplicado entre 1980 y 2014 según la [Organización Mundial de la Salud \(OMS\)](#), por tanto, entender su trayectoria y las características de los individuos

¹ La transición epidemiológica consiste en un proceso de cambio dinámico de la morbilidad y mortalidad de la población. Esta transición, acompañada por la demográfica, se caracteriza por disminución de la prevalencia de enfermedades transmisibles y aumento de las no transmisibles, desplazamiento de la morbilidad y mortalidad de los más jóvenes hacia los grupos de mayor edad y aumento de los problemas asociados a la morbilidad en detrimento de la mortalidad. Los países en desarrollo en general se encuentran en una etapa previa de la transición epidemiológica respecto a los desarrollados ([Arredondo *et al.* \(2003\)](#), [Wikipedia](#)).

que transitan hacia problemas de obesidad y sobrepeso puede dar un mayor entendimiento del proceso y, por tanto, de las acciones a tomar (Stuart & Panico, 2016).

Si bien en Uruguay, existen trabajos que han analizado la situación nutricional en la niñez y sus determinantes (ver por ejemplo, Bérgholo *et al.*, 2006) no se encuentran estudios específicos sobre trayectorias nutricionales durante el período escolar. Por tanto, uno de los objetivos de este trabajo consiste en analizar la evolución temporal de las trayectorias para niños y niñas de Uruguay en edad escolar.

Por su parte, Failache *et al.* (2015) estudiaron los determinantes del desempeño escolar en Uruguay, específicamente analizando la probabilidad de repetición y abandono escolar. Los autores concluyen que el déficit nutricional afecta positivamente la probabilidad de repetición. Sin embargo, no reportan efectos directos de este déficit sobre la probabilidad de abandono del sistema educativo.

Estos resultados surgen de la estimación de un *probit* para la probabilidad de repetir y otro para la de abandonar, donde el estado nutricional se introduce para explicar dichas probabilidades¹. Además, los autores reportan que el nivel educativo y el ingreso del hogar así como características de los niños vinculadas a sus habilidades no cognitivas fueron significativos para explicar los desempeños considerados.

Los resultados reportados por este trabajo motivan el análisis de la asociación entre la trayectoria nutricional y el desempeño escolar que se aborda mediante una metodología diferente que permite analizar dicha asociación considerando a estos procesos dinámicamente en el tiempo .

El tema central de este trabajo consiste en analizar la trayectoria nutricional de una cohorte de niños y niñas de Uruguay y su vínculo con el desempeño escolar mediante la estimación de Modelos Mixtos de Clases Latentes y Modelos Mixtos Conjuntos de Clases Latentes.

Para este análisis se aplica un Modelo Mixto de Clases Latentes para el

¹ Failache *et al.* (2015) estiman un *probit* para la probabilidad de repetir y otro para la de abandonar, donde introducen la puntuación *z-score* (según OMS) del IMC del niño en 2009. Haber repetido corresponde a si el niño experimentó algún episodio de repetición durante el ciclo educativo previo, mientras que el abandono se relevó en 2011/12.

Índice de Masa Corporal (IMC)¹ de niños y niñas, que permite identificar grupos o clases a partir de la trayectoria de este índice según la edad (Muthén, 2004; Proust-Lima *et al.*, 2014). Este modelo no requiere contar con medidas longitudinales para las mismas edades o en los mismos momentos del tiempo para los niños y niñas, sino que considera la disparidad de medidas o momentos en los que se miden bajo el supuesto de “*missing at random*”. A su vez, se trata de un caso especial dentro de los Modelos de Ecuaciones Estructurales donde los parámetros que gobiernan la curva para cada niño o niña se estiman como variables latentes (Johnson, 2015).

El abandono del sistema educativo resulta un problema importante para Uruguay. Según información de la Encuesta Continua de Hogares (ECH) el porcentaje que no asiste al sistema educativo asciende a 3.36 %, 7.41 % y 12.96 % para 13, 14 y 15 años respectivamente en 2012. Esto muestra un problema importante vinculado a la no asistencia en este tramo de edad que empeora notablemente. Si bien para 2016 estas tasas disminuyen el problema persiste. La no asistencia según la información de la ECH 2016 se sitúa en 2.79 %, 5.8 % y 9.54 % para 13, 14 y 15 años respectivamente.

A su vez, los Modelos Conjuntos han cobrado mayor relevancia recientemente en el análisis biométrico. En la literatura internacional existen varias aplicaciones fundamentalmente en el área médica (Lin *et al.*, 2002; Proust-Lima *et al.*, 2009, 2015; Rizopoulos, 2012; Tsiatis & Davidian, 2004; Zajacova & Ailshire, 2012). Existe un único trabajo hecho en Uruguay que aplica esta metodología para el análisis del deterioro cognitivo en adultos mayores para un estudio longitudinal de Suecia (Massa, 2015).

El objetivo de los Modelos Conjuntos es responder a preguntas de investigación referentes a la asociación entre medidas a lo largo del tiempo de alguna variable de interés y el tiempo hasta la ocurrencia de un evento. Además, permiten incorporar distintos mecanismos que generan los datos faltantes como el supuesto de “*missing not at random*” (Rizopoulos, 2012).

En el presente trabajo, como desempeño escolar se considera el abandono o desvinculación del sistema educativo, que se estudia en el marco de un Modelo de Supervivencia donde la variable de interés es la edad en la que el niño deja

¹ El IMC es un cociente que relaciona el peso y la estatura de un individuo y se utiliza para evaluar el estado nutricional. En particular, la OMS cuenta con estándares de crecimiento por sexo según el IMC para el tramo de 5 a 19 años.

de estudiar en el sistema educativo formal. A la vez, se considera un Modelo Mixto de Clases Latentes para la trayectoria nutricional. Entonces, se estima un Modelo Mixto Conjunto de Clases Latentes que permite evaluar el vínculo entre la trayectoria nutricional y la edad de abandono escolar.

La información utilizada corresponde al [Estudio Longitudinal del Bienestar en Uruguay \(ELBU\)](#) realizado por el Instituto de Economía, con apoyo de la [Comisión Sectorial de Investigación Científica \(CSIC\)](#), [Unicef](#) y el [Fondo Clemente Estable](#). El estudio consiste en una muestra de niños y niñas que se encontraba cursando primer año en escuelas públicas de capitales departamentales del país en 2004. La encuesta se repite en años posteriores (2004, 2006 para Montevideo y área metropolitana y 2011/12). En 2015/16 con financiamiento de [CSIC](#) se realizó una nueva ola, en la que se entrevistó a los jóvenes con aproximadamente 17 y 18 años de edad. En este trabajo, se utiliza la información disponible hasta 2012¹.

El trabajo se organiza de la siguiente manera, en el [Capítulo 2](#) se presenta una revisión de la literatura acerca de la situación nutricional y el desempeño escolar. En el [Capítulo 3](#) se detallan los modelos, métodos de estimación así como la metodología utilizada para abordar el problema planteado. El [Capítulo 4](#) reseña los principales resultados y, por último, el [Capítulo 5](#) pone de relieve las principales consideraciones que se desprenden del análisis realizado, sus limitaciones y futuras líneas de investigación.

¹ Los datos definitivos de la ola realizada en 2015/16 no estaban disponibles cuando se realizó el análisis que se presenta en este trabajo.

Capítulo 2

Estado nutricional y desempeño escolar

En este capítulo se presentan aspectos teóricos de la trayectoria nutricional, particularmente durante la infancia, y su vínculo con el desempeño escolar. A su vez, se exponen los enfoques y resultados obtenidos por trabajos aplicados a esta temática en Uruguay así como en otros países. Por último, se presenta el problema de investigación del presente trabajo y su relevancia así como los principales objetivos e hipótesis que orientan el análisis.

2.1. Marco conceptual

La privación durante la infancia reporta particular interés en la discusión sobre el bienestar a lo largo del ciclo de vida, puesto que las capacidades accesibles durante esta etapa condicionan el espacio de funcionamientos a los que una persona podrá acceder en su vida adulta ([Sen, 1999](#)). La nutrición de las personas es uno de los funcionamientos elementales para sobrevivir y su privación puede generar incapacidades para alcanzar otros funcionamientos. Por ser una capacidad elemental del ser humano, resulta relevante el estudio en sí mismo del estado nutricional y su evolución a lo largo del tiempo.

A su vez, los logros nutricionales pueden ser un medio para alcanzar otras capacidades. Por tanto, la presencia de problemas en la nutrición puede tener efectos, incluso de muy largo plazo, sobre otras dimensiones ([ACC/SCN, 2000](#)). En particular, la literatura indica que el déficit nutricional se asocia con

menores logros educativos (Georgiadis *et al.*, 2016; Glewwe, 2005; Levinger, 1992; Victora *et al.*, 2008).

En los estudios de crecimiento y nutrición en la infancia, se pone de manifiesto la interacción de varios factores -hereditarios, referentes al nivel socio económico, al acceso a servicios y características de la comunidad en las que viven las familias- para determinar el estado y evolución nutricional (ACC/SCN, 2000; Alderman *et al.*, 1997; Behrman & Wolfe, 1987; Behrman, 1990). Asimismo, en la literatura biomédica se tiende a dar mayor relevancia a factores del entorno para condicionar las trayectorias de desarrollo y crecimiento de los niños (Barker *et al.*, 1989; deOnis *et al.*, 2004; Gluckman *et al.*, 2005). La interacción de genética con medio ambiente, da como resultado el crecimiento y nutrición alcanzados, afectando o favoreciendo, la salud a corto, mediano y largo plazo. Además, entender cómo se transita hacia los distintos problemas nutricionales puede dar un mejor entendimiento del proceso y, por tanto, de las acciones a tomar. En este sentido, la comprensión de esta evolución puede guiar la práctica clínica y poner de relieve los mecanismos que se producen antes de que las trayectorias queden “establecidas” (Engle *et al.*, 1997; Regnault & Gillman, 2014; Stuart & Panico, 2016; Victora *et al.*, 2010).

En cuanto a la vinculación entre trayectoria nutricional y desempeño escolar, Glewwe (2005) plantea tres períodos relevantes para el desarrollo del niño en donde su estado nutricional juega un rol importante en su desempeño educativo: desde el nacimiento hasta los dos años, a partir de entonces y hasta el momento de ingresar a la escuela (5 o 6 años) y el período de edad escolar hasta que se evalúan los desempeños. A su vez, indica que un conjunto de decisiones que toman los padres durante la primer infancia del niño, constituyen elementos centrales para comprender los desempeños posteriores de los hijos. En el trabajo de Zhao & Glewwe (2010) se resalta también la importancia de aspectos tales como la motivación del niño, sus gustos por la educación y las habilidades innatas. Además, Cueto & Chinen (2000) indican que un estado nutricional y de salud adecuados no son suficientes para que los niños obtengan logros en términos de aprendizaje.

2.2. Antecedentes

El estudio del estado nutricional puede realizarse a través de medidas antropométricas que representan adecuadamente el estado de salud y nutrición en una población. Estas medidas se utilizan en general para identificar problemas nutricionales o de salud de una población y realizar intervenciones correspondientes, así como para monitorear estos problemas y luego aplicar otras herramientas más específicas (deOnis & Habicht, 1996).

En Uruguay, dos estudios contienen medidas antropométricas a lo largo de la infancia para muestras representativas de una población: el ELBU y la Encuesta de Nutrición, Desarrollo Infantil y Salud (ENDIS). La ENDIS es una encuesta longitudinal llevada adelante por el Instituto Nacional de Estadística (INE), en conjunto con Uruguay Crece Contigo (UCC) y la Universidad de la República (UdelaR). Se realiza a hogares donde residen niños de 0 a 3 años y 11 meses de localidades urbanas (más de 5 mil habitantes), habiéndose llevado a cabo dos relevamientos: la primer entrevista en 2014 y otra en 2016.

El ELBU, fuente de información utilizada en el presente trabajo, es un estudio representativo de niños y niñas que cursaban primer año de educación primaria en escuelas públicas de capitales departamentales y el área metropolitana en 2004 y que fueron entrevistados posteriormente en otras tres instancias: 2006, 2011/12 y 2015/16.

A partir de información contenida en el ELBU, trabajos previos han contribuido a analizar el estado nutricional de los niños uruguayos y sus determinantes (Ver por ejemplo, Bérgho *et al.*, 2006). Sin embargo, los trabajos realizados para Uruguay utilizaron técnicas que se aplican a datos de tipo sección cruzada para el análisis del estado nutricional y sus determinantes.

Identificaron el peso al nacer, la historia reproductiva de la madre, la talla de la madre, el nivel educativo del hogar y en algunos casos la dotación de bienes públicos como principales factores que afectan la nutrición de los niños y niñas de Uruguay (Bérgho *et al.*, 2006). Además, la morbilidad del niño y la situación nutricional de la madre resultaron significativamente asociadas con sus logros nutricionales.

Por su parte, varios trabajos internacionales analizan la trayectoria nutricional en la infancia utilizando métodos para datos longitudinales. En particular, Stuart & Panico (2016) analizan la trayectoria nutricional de una cohorte

nacida entre 2000 y 2002 en Reino Unido. En el trabajo se estima la trayectoria latente de crecimiento para el **IMC** utilizando un modelo mixto semi paramétrico, de donde se desprende la existencia de grupos en función de la trayectoria nutricional de los niños. Además, se presentan modelos multinomiales para caracterizar estos grupos en función de características socio económicas y del comportamiento de las familias durante el embarazo y la infancia. Los autores identifican cuatro grupos de trayectorias: dos transitan en niveles del **IMC** dentro de un rango normal y representan en conjunto alrededor del 85 % de los niños de la cohorte; un segundo grupo representa el 14 % de los niños y se caracteriza por una trayectoria de sobrepeso; el último grupo con el 3 % de los niños mantiene una trayectoria en niveles de obesidad. A su vez, plantean que los perfiles socio económicos de los grupos sugieren que las diferencias pueden provenir de la infancia temprana.

Por otro lado, el trabajo de [Georgiadis et al. \(2016\)](#) vincula el estado nutricional en diferentes períodos de la infancia con los logros cognitivos alcanzados por los niños. Los autores indican que la evidencia respecto a la relación entre el estado nutricional, en diferentes períodos de la niñez, y los logros cognitivos es ambigua en la literatura y lo atribuyen en parte a las diferentes formas de modelar las trayectorias de crecimiento. Por tanto, utilizan un modelo general para aproximar la trayectoria nutricional con el fin de comparar distintas metodologías y explorar canales a través de los cuales se relacionan el estado nutricional y los logros cognitivos. Utilizando información proveniente del estudio *Young Lives*¹ de cuatro países (Etiopía, India, Perú y Vietnam), sus resultados sugieren que existe asociación entre los logros cognitivos de los niños a los 8 años y sus trayectorias de crecimiento en distintos períodos de la primera infancia.

Respecto al análisis del desempeño educativo en Uruguay [Failache et al. \(2015\)](#) analizan los factores asociados mediante la estimación de la probabilidad de repetir y abandonar, utilizando información contenida en el **ELBU**. En los modelos para estas probabilidades incluyen como variable explicativa la situación nutricional del individuo en el período previo. Esta variable se construye a partir del valor de la puntuación del *z-score* sobre el **IMC**, que se

¹ *Young Lives* es un estudio longitudinal de la pobreza infantil que realiza la Universidad de Oxford que sigue a niños de Etiopía, India (en los estados de Andhra Pradesh y Telangana), Perú y Vietnam, durante aproximadamente 15 años.

obtiene con los programas de cálculo que provee la [OMS](#)¹. Esta variable permite separar entonces a los niños en tres grupos según su estado nutricional: *z-score* por debajo de -1 con problemas de desnutrición, entre -1 y 1 rango de valores normales, es decir, sin problemas nutricionales serios, y por encima de 1 con sobrepeso u obesidad.

Estos autores, en el modelo para la repetición reportan que el déficit nutricional observado en el período previo (*z-score* debajo de -1) se asocia positivamente con una mayor probabilidad de repetir. Esto es, aquéllos niños que presentan déficit nutricional tienen, en promedio, mayor probabilidad de repetir algún año del ciclo educativo que aquéllos que tienen un estado nutricional en el rango normal (*z-score* entre -1 y 1).

Luego, reportan que la repetición afecta significativamente el abandono y, además, plantean una posible endogeneidad de la repetición como regresor para explicar el abandono por lo que estiman mediante variables instrumentales. Sin embargo, plantean que no habría un efecto significativo directo del déficit nutricional sobre la probabilidad de abandono del sistema educativo. Asimismo, haber asistido a enseñanza preescolar, el ingreso del hogar, la educación de la madre y haber obtenido un puntaje en el rango de normalidad de pruebas psicométricas aparece negativamente relacionado con el abandono ([Failache et al., 2015](#)).

Por tanto, esto constituye una motivación del presente trabajo en donde uno de los objetivos consiste entonces en aportar al conocimiento de la relación entre la trayectoria nutricional y el desempeño educativo utilizando una metodología diferente a los trabajos previos. Si bien estos trabajos dan un buen indicio de la asociación entre estado nutricional y desempeño escolar de los niños en Uruguay, la metodología allí utilizada no explota el carácter longitudinal de los datos del [ELBU](#). En este sentido, el análisis realizado hasta el momento ha sido con técnicas para datos de sección cruzada donde no se considera la evolución temporal sino el estado nutricional de los niños en un momento dado.

Otros trabajos para Uruguay, han argumentado que las condiciones de los hogares donde crecen los niños son determinantes de la desvinculación del sistema educativo de los adolescentes. Estas condiciones, en general, se asocian

¹ Estos programas de cálculo se pueden encontrar en la web de la [OMS](#).

con carencias de ingreso o niveles educativos bajos de los padres (Boado & Fernández, 2010; Bucheli & Casacuberta, 2000).

El problema de desvinculación del sistema educativo es relevante y su estudio constituye un desafío para el caso de educación media ya que allí se encuentran tasas de desvinculación muy elevadas, en tanto, para primaria el problema más importante es la repetición. Según se observa en la Tabla 2.1 el porcentaje de varones y mujeres que no asisten al sistema educativo aumenta de forma muy importante con la edad, según datos del Censo 2011. Además, la no asistencia crece de forma más marcada con la edad para los varones que para las mujeres.

Tabla 2.1: Porcentaje de no asistencia al sistema educativo por edad y sexo

Edad	Mujer	Varón	Total
12	1.6	2.1	1.8
13	4.1	5.6	4.9
14	8.5	13.1	10.8
15	14.6	22.0	18.4
16	21.8	30.5	26.2
17	29.4	40.9	35.3
18	43.6	56.2	50.0
19	53.2	65.1	59.2
Total	21.9	29.0	25.5

Fuente: deMelo *et al.* (2015) en base al Censo 2011.

Para los niños que integran el ELBU, la tasa de abandono escolar en la tercer ola, realizada en 2011/12 cuando tenían entre 13 y 14 años aproximadamente, es del entorno de 9%. A su vez, según la ECH 2011 la tasa desciende hasta 6.3% para el tramo etario de 13 y 14 años que cursaron primaria pública (Failache *et al.*, 2015). Por tanto, si bien la cobertura en educación primaria es prácticamente universal el abandono se vuelve un problema serio en el tramo etario que cursa educación media en Uruguay.

Por otro lado, deMelo *et al.* (2015) muestran tres momentos destacados de desvinculación en educación media: primer año (13 y 14 años), culminación de ciclo medio básico (14 y 15 años) y culminación de ciclo medio superior (17 y 18 años), donde este último tramo sería el de mayor probabilidad de abandono. Por su parte, Aristimuño & deArmas (2012) hacen referencia a dos períodos

críticos de abandono, las franjas etarias donde se presentan las tasas más altas de abandono son 14-15 y 18-19 años.

2.3. Problema de investigación y relevancia

El trabajo consiste en el estudio de la trayectoria nutricional de niños de Uruguay y su vínculo con el desempeño escolar, medido por el abandono del sistema educativo. En primer lugar, el análisis de la trayectoria nutricional se realiza mediante la estimación de Modelos Mixtos de Clases Latentes, que permiten agrupar a los individuos según la evolución de su IMC para la edad y también capturar la heterogeneidad en las trayectorias dentro de los grupos. Esto supone que la estimación de una trayectoria media única no es adecuada, y por tanto, se utilizan grupos latentes con trayectorias específicas para representar la heterogeneidad de la población. Además, suponer que estos grupos latentes capturan la heterogeneidad, asume que ninguna variable observada puede capturarla de forma adecuada. Por último, en este contexto, la trayectoria nutricional se considera un proceso no observable directamente que se aproxima adecuadamente por la medida del IMC para la edad (Proust-Lima *et al.*, 2009, 2015).

Por su parte, el vínculo entre esta trayectoria y el abandono se aborda mediante la estimación de Modelos Mixtos Conjuntos de Clases Latentes¹ que agrupan a los individuos considerando la trayectoria nutricional pero incorporan conjuntamente un Modelo de Supervivencia para la edad de abandono.

Cabe señalar que, mientras el Modelo Mixto supone un proceso generador de datos faltantes *Missing at Random* (MAR), el Modelo Conjunto asume uno *Missing Not at Random* (MNAR) lo que generalmente resulta más ajustado a la realidad en este tipo de aplicaciones (Rizopoulos, 2012).

En este trabajo, donde se estudia la trayectoria nutricional y el desempeño escolar, la idea es que bajo el supuesto de MAR la distribución *predictiva*² de los valores no observados de esta trayectoria depende de la trayectoria efectivamente observada. Sin embargo, en el escenario de MNAR la distribución *predictiva* depende además de los valores que tendría la trayectoria nutricional en las ocasiones no observadas, si éstas pudiesen observarse. En definitiva,

¹ Estos modelos constituyen una alternativa a los Modelos Conjuntos para Datos Longitudinales y de Supervivencia conocidos como Modelos de Efectos Aleatorios Compartidos (*Shared Random-Effect Models*) (Proust-Lima *et al.*, 2009, 2014, 2015; Rizopoulos, 2012).

² La distribución *predictiva* refiere a la que tendría la variable medida longitudinalmente para aquellas ocasiones donde no se observa si éstas pudiesen observarse (Rizopoulos, 2012). Una explicación más detallada de estos supuestos y sus implicancias se presenta en el Anexo A.1.

bajo el supuesto [MNAR](#) la trayectoria nutricional tiene una distribución que depende de factores no observables.

La consecuencia fundamental es entonces que bajo el escenario [MAR](#) se puede estimar utilizando únicamente los valores observados de la trayectoria nutricional y no se cometen sesgos en la estimación, ya que los datos faltantes únicamente dependen de factores observables. Sin embargo, en un esquema [MNAR](#), los valores faltantes se asocian con factores observables así como no observables, por tanto, estimar utilizando únicamente la trayectoria observada genera sesgos en la estimación.

La relevancia del problema de investigación se fundamenta en entender los factores asociados a la problemática nutricional de los niños, su desempeño escolar y la vinculación entre ellos. Entender estos factores puede contribuir al diseño de políticas y programas públicos que atiendan dichos aspectos. El desarrollo y aplicación de Modelos Conjuntos no se ha hecho aún con este fin, lo cual puede aportar nueva evidencia. En la literatura internacional se encuentran varias aplicaciones de estos modelos, fundamentalmente en el área médica ([Proust-Lima *et al.*, 2009, 2015](#); [Rizopoulos, 2012](#); [Tsiatis & Davidian, 2004](#); [Wu *et al.*, 2012](#)). Por su parte, existe un único trabajo realizado en Uruguay que aplica esta metodología para el análisis del deterioro cognitivo en adultos mayores para un estudio longitudinal de Suecia ([Massa, 2015](#)).

Como ya fue mencionado, en Uruguay existen bases de datos longitudinales que pueden ser fuente de información para distintos estudios sociales, como la [ENDIS](#), el [ELBU](#) y la [Encuesta sobre Situaciones Familiares y Desempeños Sociales \(ESF\)](#). Esta última, es un panel de mujeres de dos olas: la primera realizada en 2001 encuestó una muestra de 1806 mujeres de entre 25 y 54 años residentes en el Área Metropolitana, la segunda en 2008 logró re entrevistar a 828 mujeres de la primera ola.

Por lo tanto, se tiene información que permite el estudio de distintos temas de importancia social e individual a los que se podría aplicar los modelos estudiados en este trabajo, entonces conocer su funcionamiento puede propiciar su aplicación a otros problemas de investigación. En consecuencia, tener conocimiento sustantivo de los Modelos Conjuntos, sus ventajas y desventajas, sus propiedades, diferentes extensiones y especificaciones puede extenderse a otros problemas de investigación.

2.4. Objetivos y preguntas de investigación

Objetivos generales

El estudio se propone analizar la evolución temporal del estado nutricional de los niños y niñas mediante la estimación de Modelos Mixtos de Clases Latentes. Además, indaga en la relación de esta trayectoria nutricional con el problema de la desvinculación escolar a través de la estimación de Modelos Mixtos Conjuntos de Clases Latentes.

Objetivos específicos

- a) Estimar Modelos Mixtos de Clases Latentes para las trayectorias nutricionales de niñas y niños y clasificarlos en grupos según estas trayectorias
- b) Caracterizar a las niñas y niños pertenecientes a los grupos según factores que se encuentran asociados al estado nutricional
- c) Estimar Modelos Mixtos Conjuntos de Clases Latentes para analizar la relación entre trayectoria nutricional y edad de abandono escolar

Principales preguntas

- a) ¿Es posible agrupar a las niñas y niños según sus trayectorias nutricionales en edad escolar?
- b) ¿Existen diferencias en los factores asociados al estado nutricional de las niñas y los niños según los grupos encontrados en base a la trayectoria nutricional?
- c) ¿Existe asociación entre la trayectoria nutricional y la edad de abandono escolar?

Capítulo 3

Modelos Longitudinales y de Supervivencia

En este capítulo se presentan los modelos aplicados en el trabajo, se detallan los principales supuestos en los que éstos se basan, los métodos y procedimientos de estimación utilizados generalmente y, en particular, los aplicados en este trabajo. En la sección 3.1 se presentan los Modelos para Datos Longitudinales, en la 3.2 los Modelos de Supervivencia y en la 3.3 los Modelos Conjuntos para Datos Longitudinales y de Supervivencia. En la sección 3.4 se detallan los cálculos realizados posteriormente a la estimación de estos modelos.

Por último, la sección 3.5 expone la metodología aplicada para lograr los objetivos de este trabajo en términos de la aplicación a la trayectoria nutricional y el desempeño escolar de los niños. En particular, se describe de forma detallada la información utilizada con las variables relevantes para el análisis y la estrategia empírica llevada adelante para lograr los objetivos planteados en esta investigación.

3.1. Modelos para Datos Longitudinales

Los Modelos Lineales Mixtos (**LMM**) se han convertido en una herramienta estándar de análisis de la trayectoria de una medida longitudinal gaussiana a lo largo del tiempo y son utilizados para analizar el efecto de covariables sobre dichas trayectorias ([Rizopoulos, 2012](#); [Verbeke & Lesaffre, 1996](#); [Verbeke & Molenberghs, 2000](#)). Sin embargo, debido a que la estructura de datos longitudinales puede ser compleja, que puede llevar al incumplimiento de los supuestos acerca de las distribuciones en que se basan los **LMM**, se ha planteado la utilización de modelos más generales ([Proust-Lima & Jacqmin-Gadda, 2005](#); [Proust-Lima *et al.*, 2014, 2015, 2016b](#)).

Los **LMM**, además, asumen que el efecto de las covariables es constante a lo largo del rango de valores de la medida longitudinal, por lo que surgen los Modelos Lineales Mixtos Generalizados (**GLMM**). Estos últimos son una extensión de los **LMM** para medidas binarias, ordinales o Poisson.

Los Modelos Mixtos de Clases Latentes (**LCMM**) resultan de utilidad para solucionar algunos problemas que presentan los **LMM** aplicados a medidas longitudinales. En particular, los **LCMM** resuelven los problemas generados por el supuesto de continuidad de la medida longitudinal y distribución gaussiana de las desviaciones aleatorias¹ de los **LMM** ([Proust-Lima *et al.*, 2016b](#)).

Los **LCMM** se enmarcan en la literatura de Modelos de Ecuaciones Estructurales (**SEM**). En ellos se separa el modelo estructural, que describe la trayectoria del proceso de interés (latente) a lo largo del tiempo en función de covariables, del modelo de medición que vincula este proceso con las observaciones. En este contexto, los Modelos Mixtos para Procesos Latentes (**LPMM**) suponen un **LMM** para el proceso latente y separan el modelo de medición (observable) del estructural (latente). Además, los **LCMM** incorporan la heterogeneidad de la población a través del supuesto de pertenencia a grupos que capturan la evolución diferencial del proceso latente, donde estos grupos se suponen internamente homogéneos en cuanto a la evolución de dicho proceso.

La aplicación de los **LCMM**, como los **GLMM**, constituye entonces una alternativa que flexibiliza los supuestos restrictivos de los **LMM**.

¹Las desviaciones aleatorias refieren de forma general a los efectos aleatorios, los errores correlacionados y los errores de medida.

En este contexto, se detallan los tres modelos antes mencionados que se aplican en el contexto de datos longitudinales: [LMM](#), [LPMM](#) y [LCMM](#). Se reseñan los principales supuestos, los métodos de estimación usualmente utilizados y los cálculos post-estimación relevantes para este trabajo.

3.1.1. Modelo Lineal Mixto

Se considera la variable de interés Y_{ij} , correspondiente al individuo i en la ocasión j , medido en el momento t_{ij} , con $i = 1, \dots, N$ y $j = 1, \dots, n_i$. La ocasión j no coincide con el tiempo t_{ij} para permitir diferenciar el momento así como la cantidad de ocasiones de medición entre individuos.

Por tanto, se tiene el vector \mathbf{Y}_i de dimensión n_i que contiene las medidas repetidas de la variable de respuesta para el individuo i en las n_i ocasiones, esto es, $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ij}, \dots, Y_{in_i})^\top$.

[Laird & Ware \(1982\)](#) plantean el **Modelo Lineal Mixto** de la siguiente forma:

$$\left\{ \begin{array}{l} \mathbf{Y}_i = \mathbf{X}_{Li}\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i \\ \mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D}) \\ \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Xi}_i), \quad i = 1, \dots, N \\ \mathbf{b}_1, \dots, \mathbf{b}_N, \boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_N \text{ independientes} \end{array} \right. \quad (3.1)$$

donde N es la cantidad de individuos, \mathbf{X}_{Li} y \mathbf{Z}_i son matrices de covariables conocidas de dimensión $(n_i \times p)$ y $(n_i \times q)$ respectivamente. $\boldsymbol{\beta}$ es un vector en \mathbb{R}^p que contiene los efectos fijos¹, mientras que \mathbf{b}_i es un vector en \mathbb{R}^q que contiene los efectos aleatorios. El vector $\boldsymbol{\epsilon}_i$, con n_i componentes, contiene las perturbaciones del modelo en cada ocasión j . Finalmente, \mathbf{D} es una matriz de covarianzas simétrica general de $(q \times q)$ y $\boldsymbol{\Xi}_i$ es una matriz de covarianzas de $(n_i \times n_i)$ que únicamente depende del individuo i en el n_i y, por tanto, los parámetros desconocidos en $\boldsymbol{\Xi}_i$ no dependen del individuo i .

¹ En este caso los efectos fijos corresponden a los parámetros de las covariables del modelo. La literatura económica de datos longitudinales o de panel considera a los efectos fijos con otro significado. Ver por ejemplo [Cameron & Trivedi \(2010\)](#); [Frees \(2004\)](#); [Wooldridge \(2002, 2010\)](#).

De la ecuación (3.1) se desprende que, $\mathbf{Y}_i|\mathbf{b}_i$ se distribuye normal con media $\mathbf{X}_{Li}\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i$ y matriz de covarianzas $\boldsymbol{\Xi}_i$. Además, considerando la distribución del vector de efectos aleatorios se llega a que $\mathbf{Y}_i \sim N(\mathbf{X}_{Li}\boldsymbol{\beta}, \mathbf{V}_i)$, donde $\mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top + \boldsymbol{\Xi}_i$ (Verbeke & Molenberghs, 2000).

La estructura de los residuos puede suponerse independiente, es decir, que $\boldsymbol{\Xi}_i = \sigma^2\mathbf{I}_{n_i}$, donde \mathbf{I}_{n_i} denota la matriz identidad de $(n_i \times n_i)$. En este caso el modelo (3.1) se llama **modelo independiente condicional**. Esto implica que las n_i respuestas del individuo i son independientes entre sí, condicional en \mathbf{b}_i y $\boldsymbol{\beta}$.

Bajo este supuesto, la estructura de covarianzas puede ser muy simple y no reflejar la realidad del vector de respuestas \mathbf{Y}_i , en particular cuando se introducen pocos efectos aleatorios, por ejemplo, únicamente en el intercepto. Sin embargo, la relajación de este supuesto en la práctica puede comprometer la estimación del modelo y, por tanto, generalmente se recomienda mantener el supuesto de independencia condicional en presencia de efectos aleatorios en otros coeficientes más allá del intercepto (Verbeke & Molenberghs, 2000).

Por su parte, se puede introducir correlación serial en el LMM, que supone al menos parte del perfil observado en la trayectoria individual de la variable responde a un proceso estocástico que varía con el tiempo y opera al interior de cada individuo. Verbeke & Molenberghs (2000) indican que existen varias propuestas para introducir este supuesto en el modelo que pueden resumirse en el planteo general hecho por Diggle *et al.* (1994) que se presenta a continuación.

En este modelo se descompone el término $\boldsymbol{\epsilon}_i$ de varianza constante como $\boldsymbol{\epsilon}_i = \boldsymbol{\omega}_i + \boldsymbol{\varepsilon}_i$ donde $\boldsymbol{\omega}_i$ es un componente de correlación serial que usualmente se supone función decreciente del tiempo de separación entre las medidas. Por su parte, $\boldsymbol{\varepsilon}_i$ es un componente de error que refleja la variabilidad que agrega el proceso de medición en sí mismo y que se supone independiente de $\boldsymbol{\omega}_i$.

Por tanto, el modelo resultante se escribe:

$$\left\{ \begin{array}{l} \mathbf{Y}_i = \mathbf{X}_{Li}\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\omega}_i + \boldsymbol{\varepsilon}_i \\ \mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D}) \\ \boldsymbol{\omega}_i \sim N(\mathbf{0}, \mathbf{R}_i = \sigma_\omega^2 \boldsymbol{\Omega}_i) \\ \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i = \sigma_\varepsilon^2 \mathbf{I}_{n_i}) \\ \mathbf{b}_1, \dots, \mathbf{b}_N, \boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_N, \boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_N \text{ independientes} \end{array} \right. \quad (3.2)$$

donde el modelo se completa asumiendo una estructura específica para la matriz de correlaciones $\boldsymbol{\Omega}_i$ de $(n_i \times n_i)$. Estas estructuras usualmente se toman de los análisis de series temporales o bien se considera que el efecto $\boldsymbol{\omega}_i$ es un fenómeno poblacional independiente del individuo. Esta matriz de correlaciones $\boldsymbol{\Omega}_i$ solamente depende del individuo i por la dimensión n_i y por los momentos en los que se tomaron las medidas t_{ij} . Además, se asume que el elemento (j, k) de la matriz $\boldsymbol{\Omega}_i$ es modelado como $\omega_{ijk} = g(|t_{ij} - t_{ik}|)$ para alguna función decreciente $g(\cdot)$ dada por parámetros en el vector $\boldsymbol{\rho}$, que cumpla $g(0) = 1$. Esto es, la correlación entre dos momentos depende únicamente del intervalo de tiempo entre las medidas y es decreciente a medida que la amplitud del intervalo se hace cada vez mayor (Verbeke & Molenberghs, 2000).

Análogamente a lo hecho para la ecuación (3.1), del modelo en (3.2) se desprende que la distribución del vector \mathbf{Y}_i condicional en los efectos aleatorios es una normal con media $\mathbf{X}_{Li}\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i$ y ahora la matriz de covarianzas $\mathbf{R}_i + \boldsymbol{\Sigma}_i$. Además, considerando la distribución del vector de efectos aleatorios se llega a que $\mathbf{Y}_i \sim N(\mathbf{X}_{Li}\boldsymbol{\beta}, \mathbf{V}_i)$, donde $\mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top + \mathbf{R}_i + \boldsymbol{\Sigma}_i$.

El modelo presentado en forma matricial en (3.2) se puede escribir:

$$Y_{ij} = \mathbf{X}_{Li}(t_{ij})^\top \boldsymbol{\beta} + \mathbf{Z}_i(t_{ij})^\top \mathbf{b}_i + \omega_i(t_{ij}) + \varepsilon_{ij} \quad (3.3)$$

donde $\mathbf{X}_{Li}(t_{ij})$ y $\mathbf{Z}_i(t_{ij})$ son dos vectores de covariables en el momento t_{ij} de tamaño p y q respectivamente. Los vectores $\boldsymbol{\beta}$ y \mathbf{b}_i son los mismos que fueron definidos previamente en (3.1). El componente ε_{ij} corresponde a la perturbación del modelo en cada ocasión con distribución gaussiana, independiente y

varianza constante σ_ε^2 según fue definido en (3.2). Finalmente, el proceso estocástico $\omega_i(t_{ij})$ tiene media cero y sigue distribución gaussiana como se supuso en (3.2), donde los parámetros que definen la matriz $\mathbf{\Omega}_i$ son los $\boldsymbol{\rho}$ de la función $g(\cdot)$.

Estimación del Modelo Lineal Mixto

El modelo especificado en (3.1) corresponde a la **formulación jerárquica** de los LMM, ya que se define a través de la distribución de la variable de respuesta, condicional en los efectos aleatorios y su distribución. De este modelo, se deduce la distribución marginal para \mathbf{Y}_i de donde surge la **formulación marginal** de los LMM. Sin embargo, ambos modelos no son equivalentes en el sentido que la estimación con una u otra formulación puede arrojar resultados diferentes (Verbeke & Molenberghs, 2000).

La estimación clásica del modelo LMM implica basarse en la distribución marginal y, por tanto, no toma en cuenta la estructura jerárquica del modelo (3.1)¹. El método de estimación más utilizado para estos modelos es el de Máxima Verosimilitud (MLE). Una descripción general del mismo se presenta en el Anexo A.2.

El vector de parámetros a estimar es $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \text{vec}(\mathbf{D})^\top, \sigma_\omega^2, \boldsymbol{\rho}^\top, \sigma_\varepsilon^2)^\top$, donde $\text{vec}(\mathbf{D})$ corresponde al vector de parámetros de la matriz simétrica y definida positiva \mathbf{D} . El número de parámetros dependerá de la especificación de dicha matriz, ya que puede especificarse definida positiva diagonal o definida positiva simétrica. Si se especifica \mathbf{D} diagonal, se tendrán q parámetros a estimar. Por el contrario, si \mathbf{D} es simétrica y definida positiva, contiene $\frac{q(q+1)}{2}$ parámetros a estimar mediante la transformación de Cholesky.

La función a maximizar por el método MLE es el logaritmo de la verosimilitud que se define como:

$$l(\boldsymbol{\theta}) = \sum_{i=1}^N \log(L_i(\boldsymbol{\theta})) \quad (3.4)$$

¹ Verbeke & Molenberghs (2000) plantean que la formulación jerárquica no necesariamente implica una formulación marginal apropiada, por tanto la optimización numérica con la distribución marginal puede dar resultados fuera del espacio paramétrico de la formulación jerárquica. Los autores ejemplifican algún problema que puede arribar con la estimación mediante la formulación marginal y recomiendan realizar análisis exploratorio de los datos antes de la estimación.

donde L_i es la contribución individual a la verosimilitud que, para el caso del LMM, es:

$$L_i(\boldsymbol{\theta}) = \phi_i(\mathbf{Y}_i; \boldsymbol{\theta})$$

donde ϕ_i es la función de densidad de \mathbf{Y}_i , una normal multivariada de media $\mathbf{X}_{L_i}\boldsymbol{\beta}$ y matriz de covarianzas $\mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top + \mathbf{R}_i + \boldsymbol{\Sigma}_i$ según se desprende de (3.2).

3.1.2. Modelo Mixto para Procesos Latentes

En ciertos análisis de datos longitudinales, puede resultar que el proceso de interés no sea directamente observable y sea necesario aproximarse mediante uno o varios indicadores o medidas para su representación. En estos casos, no es posible obtener una medida del proceso de interés directamente de los individuos. Lo que se puede obtener son medidas repetidas de un indicador del proceso y, por tanto, es pertinente la utilización de un Modelo Mixto para Procesos Latentes (Proust-Lima *et al.*, 2015, 2016b). La idea básica consiste en plantear, por un lado, un **modelo estructural** para el proceso latente y, por otro, un **modelo de medición** para los indicadores que aproximan al proceso de interés. En esta sección se presenta el caso donde se tiene un único indicador del proceso latente, puede consultarse en Proust-Lima *et al.* (2016b) el caso con múltiples indicadores.

El **modelo estructural** describe la evolución en el tiempo del proceso latente en función de covariables. Éste se dice latente ya que no es observable directamente de los individuos. Por su parte, el **modelo de medición** representa la evolución longitudinal de la medida o indicador observable que intenta capturar dicho proceso latente (Proust-Lima *et al.*, 2016b). El interés se encuentra entonces, en modelar la trayectoria temporal del proceso latente y entender los efectos de covariables, a partir de un marcador o indicador medido en los individuos repetidamente en el tiempo.

En el modelo estructural, el proceso latente puede aproximarse correctamente por un LMM estándar sin error de medida, que en tiempo continuo se denota $\Lambda_i(t)$ y se representa como:

$$\Lambda_i(t) = \mathbf{X}_{L_i}(t)^\top \boldsymbol{\beta} + \mathbf{Z}_i(t)^\top \mathbf{b}_i + \omega_i(t) \quad \forall t \in \mathbb{R}^+ \quad (3.5)$$

donde $\mathbf{X}_{Li}(t)$ es un vector de p covariables en t asociado a los efectos fijos $\boldsymbol{\beta}$, $\mathbf{Z}_i(t)$ es un vector de q covariables en t asociado a los efectos aleatorios \mathbf{b}_i (típicamente $\mathbf{Z}_i(t)$ incluye funciones del tiempo). El vector \mathbf{b}_i de q efectos aleatorios satisface los mismos supuestos que en la sección 3.1.1, al igual que el proceso estocástico ω_i pero ahora en tiempo continuo.

Para tomar en cuenta la diferencia entre la medida longitudinal y el proceso latente, se vincula $\Lambda_i(t_{ij})$ y Y_{ij} , en los momentos t_{ij} , mediante una función denominada **función de enlace**¹. Esto se representa como sigue:

$$Y_{ij} = H(\tilde{Y}_{ij}; \boldsymbol{\eta}) = H(\Lambda_i(t_{ij}) + \varepsilon_{ij}; \boldsymbol{\eta}) \quad (3.6)$$

donde ε_{ij} son los errores gaussianos independientes con varianza σ_ε^2 , la función de enlace (H) tiene parámetros $\boldsymbol{\eta}$ y \tilde{Y}_{ij} denota el proceso latente con error en el momento t_{ij} .

Si el indicador es una medida cuantitativa, H^{-1} es una función continua monótona creciente. Pero, también es posible especificar una función discreta para la estimación. En Proust-Lima *et al.* (2016b) se presentan alternativas para la función H .

Estimación del Modelo Mixto para Procesos Latentes

El método en este caso también corresponde al de [MLE](#) y dependerá de si la función H especificada es continua o discreta. A continuación, se expone únicamente el caso continuo y se refiere al trabajo de Proust-Lima *et al.* (2016b) para el caso discreto.

Para que el modelo esté identificado se requiere el cumplimiento de dos restricciones:

- (I) $\beta_0 = 0$, intercepto del proceso latente
- (II) $\sigma_\varepsilon^2 = 1$, escala del proceso latente

Considerando estas restricciones se tiene que el vector de parámetros a estimar es $\boldsymbol{\theta} = \left(\tilde{\boldsymbol{\beta}}^\top, \text{vec}(\mathbf{D})^\top, \sigma_\omega^2, \boldsymbol{\rho}^\top, \boldsymbol{\eta} \right)^\top$, donde $\text{vec}(\mathbf{D})$ es el ya definido en

¹ Función de enlace es una traducción del término en inglés *link function*.

la sección 3.1.1, $\tilde{\beta}$ es el vector β sin incluir la constante y η los parámetros que caracterizan la función H .

La función que maximiza el método MLE está representada en la ecuación (3.4) y la contribución individual a la verosimilitud para este caso, es decir para el LPMM, es:

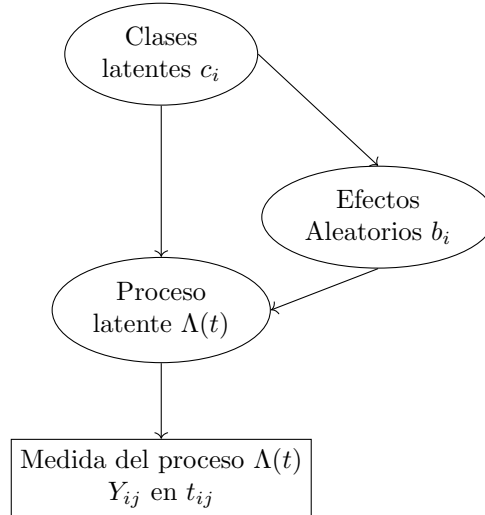
$$L_i(\theta) = \phi_i(\tilde{\mathbf{Y}}_i; \theta) \prod_{j=1}^{n_i} J(H^{-1}(Y_{ij}; \theta)) \quad (3.7)$$

donde ϕ_i es la misma función de densidad definida en la sección 3.1.1, J es el determinante de la matriz *jacobiana* de la inversa de la función de enlace (H^{-1}).

3.1.3. Modelo Mixto de Clases Latentes

Tanto el **LMM** como el **LPMM** asumen que la población de N individuos es homogénea y queda representada a nivel poblacional por un único perfil medio $\mathbf{X}_{Li}(t)^\top \boldsymbol{\beta}$ del proceso longitudinal. Sin embargo, esto puede no aproximarse a la realidad. El **LCMM** asume que la población es heterogénea y está compuesta por G grupos latentes de individuos caracterizados por un perfil particular de la trayectoria del proceso longitudinal. La existencia de grupos latentes que capturan la heterogeneidad entre individuos asume que esto no es posible a través de ninguna covariable observada (Proust-Lima *et al.*, 2016b).

Figura 3.1: Grafo dirigido que define el **LCMM**



Fuente: Proust-Lima *et al.* (2015)

La figura 3.1 muestra el esquema que supone un **LCMM** que parte de una estructura con G clases latentes. Cada individuo pertenece a un único grupo latente, por tanto la pertenencia a dichos grupos se define mediante una variable c_i que es igual a g si el individuo i pertenece al grupo g (con $g = 1, \dots, G$). El modelo asume que cada clase latente genera un perfil medio homogéneo del proceso latente ($\Lambda(t)$). A su vez, este proceso genera una variable de resultado observable (Y). Esta variable medida para el individuo i , en la ocasión j y en el momento t_{ij} se denota como Y_{ij} .

La variable latente (no observable) c_i sigue una distribución de probabilidad que se describe mediante un modelo logit multinomial en función de ciertas covariables \mathbf{X}_{Ci} :

$$\pi_{ig} = P(c_i = g | \mathbf{X}_{C_i}) = \frac{e^{\mathbf{X}_{C_i}^\top \boldsymbol{\lambda}_g}}{\sum_{l=1}^G e^{\mathbf{X}_{C_i}^\top \boldsymbol{\lambda}_l}} \quad (3.8)$$

donde $\boldsymbol{\lambda}_g$ es un vector de parámetros específicos del grupo g asociados a las covariables \mathbf{X}_{C_i} invariantes temporalmente (se incluye intercepto) de dimensión q_1 . Cuando no se introducen covariables este modelo se reduce a una probabilidad específica de cada grupo (solamente incluye un intercepto).

Ahora se permite especificar G perfiles longitudinales medios acordes al tiempo y las covariables en un **LMM** específico, por lo que, los efectos fijos y la distribución de los aleatorios podría diferir entre grupos. Por último, la incorporación de un **LPMM** disocia el modelo estructural del de medición según lo presentado en la sección 3.1.2.

Entonces, el **modelo estructural específico** queda representado en la siguiente ecuación en tiempo continuo:

$$\Lambda_i(t)|_{c_i=g} = \mathbf{X}_{L_i}(t)^\top \boldsymbol{\beta}_g + \mathbf{Z}_i(t)^\top \mathbf{b}_{ig} + \omega_i(t) \quad \forall t \in \mathbb{R}^+ \quad (3.9)$$

Luego, el **modelo de medición específico**, suponiendo una función de enlace lineal, se presenta en la siguiente ecuación:

$$Y_{ij}|_{c_i=g} = \mathbf{X}_{L_i}(t_{ij})^\top \boldsymbol{\beta}_g + \mathbf{Z}_i(t_{ij})^\top \mathbf{b}_{ig} + \omega_i(t_{ij}) + \varepsilon_{ij} \quad (3.10)$$

que difiere de la ecuación (3.3) en el vector de efectos fijos $\boldsymbol{\beta}_g$ y aleatorios \mathbf{b}_{ig} que ahora podrían diferir entre grupos. Entonces, ahora la matriz de covarianzas de los efectos aleatorios puede diferir entre grupos incorporando un coeficiente de proporcionalidad σ_g^2 . Por consiguiente, su distribución en este contexto es $\mathbf{b}_{ig} \sim N(\mathbf{0}, \mathbf{D}_g = \sigma_g^2 \mathbf{D})$ donde \mathbf{D} es la misma matriz definida en la sección 3.1.1.

Para incorporar la separación entre el proceso latente y la evolución de la medida longitudinal en la ecuación (3.10) se puede introducir una función H con parámetros $\boldsymbol{\eta}$ como en la sección 3.1.2.

Estimación del Modelo Mixto de Clases Latentes

En los modelos de grupos latentes el método aplicado de **MLE** depende de la función H especificada. En una primera instancia, se supone que esta función

es lineal como lo muestra el modelo de medición específico en la ecuación (3.10). Además, es necesario incluir restricciones sobre los parámetros para que el modelo esté identificado, en este caso:

- (I) $\boldsymbol{\lambda}_G = \mathbf{0}$, parámetros del modelo de pertenencia a grupos del último grupo
- (II) $\beta_{01} = 0$, intercepto del proceso latente del primer grupo
- (III) $\sigma_\varepsilon^2 = 1$, escala del proceso latente
- (IV) $\sigma_G^2 = 1$, coeficiente de proporcionalidad de matriz de covarianzas de efectos aleatorios del último grupo

El vector de parámetros a estimar incluye algunos dependientes del número de grupos y, dado éste igual a G se puede expresar como $\boldsymbol{\theta}_G = \left([\boldsymbol{\lambda}_g^\top]_{g=1}^{G-1}, \tilde{\boldsymbol{\beta}}_1^\top, [\boldsymbol{\beta}_g^\top]_{g=2}^G, \text{vec}(\mathbf{D})^\top, [\sigma_g^2]_{g=1}^{G-1}, \sigma_\omega^2, \boldsymbol{\rho}^\top \right)$, donde $\text{vec}(\mathbf{D})$, σ_ω^2 y $\boldsymbol{\rho}$ son los mismos que los definidos en la sección 3.1.1, $\tilde{\boldsymbol{\beta}}_1$ es el vector $\boldsymbol{\beta}_1$ sin incluir la constante y σ_g^2 es el coeficiente de proporcionalidad de la matriz de covarianzas de efectos aleatorios para el grupo g . Si en la ecuación (3.10) se incluye una función de enlace (H) al vector $\boldsymbol{\theta}_g$ se agregan los parámetros $\boldsymbol{\eta}_g$ que la definen y que podrían ser específicos de cada grupo.

En este caso, se maximiza la función de verosimilitud dado el número de grupos G como muestra la siguiente ecuación:

$$l(\boldsymbol{\theta}_G) = \sum_{i=1}^N \log(L_i(\boldsymbol{\theta}_G)) \quad (3.11)$$

La contribución individual a la verosimilitud en el LCMM es:

$$L_i(\boldsymbol{\theta}_G) = \sum_{g=1}^G \{ \pi_{ig} \phi_{ig}(\mathbf{Y}_i | c_i = g; \boldsymbol{\theta}_G) \} \quad (3.12)$$

donde π_{ig} está dada en la ecuación (3.8), ϕ_{ig} es la densidad de una normal multivariada con media $\mathbf{X}_{Li} \boldsymbol{\beta}_g$ y matriz de covarianzas $\mathbf{V}_{ig} = \mathbf{Z}_i \mathbf{D}_g \mathbf{Z}_i^\top + \mathbf{R}_i + \boldsymbol{\Sigma}_i$. Cabe notar que si el número de grupos se fija en $G = 1$, el planteo coincide exactamente con el presentado para LMM en la sección 3.1.1.

Por su parte, si se quiere disociar el proceso latente de la medida longitudinal se puede reemplazar ϕ_{ig} en (3.12) por la función dada para proceso latente en (3.7) con los adecuados parámetros específicos por grupo g :

$$L_i(\boldsymbol{\theta}_G) = \sum_{g=1}^G \left\{ \pi_{ig} \phi_{ig} \left(\tilde{\mathbf{Y}}_i | c_i = g; \boldsymbol{\theta}_G \right) \prod_{j=1}^{n_i} J \left(H^{-1}(Y_{ij} | c_i = g; \boldsymbol{\theta}_G) \right) \right\} \quad (3.13)$$

3.2. Modelo de Supervivencia

En este modelo la variable T_i^* corresponde al tiempo en que ocurre cierto evento de interés. Para describir su distribución se utiliza, en general, la función de supervivencia S que expresa la probabilidad de que el evento ocurra después de t , es decir, la probabilidad de sobrevivir hasta t :

$$S(t) = P(T^* > t) = \int_t^{\infty} p(s) ds$$

donde p corresponde a la función de densidad de probabilidad de T^* que, en este caso, se asume es una variable continua. La función de supervivencia debe ser no creciente a medida que t aumenta y debe verificar $S(0) = 1$.

Otra función que figura en el análisis de supervivencia es la h que describe el riesgo instantáneo de ocurrencia del evento en un intervalo $[t, t + dt)$ dado que se ha sobrevivido hasta t . Esta función se expresa como:

$$h(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T^* < t + dt \mid T^* \geq t)}{dt}, \quad t > 0$$

De esta forma, es posible establecer una relación entre las funciones anteriores:

$$S(t) = \exp \{-\mathcal{H}(t)\} = \exp \left\{ - \int_0^t h(s) ds \right\} \quad (3.14)$$

donde \mathcal{H} es la función de riesgo o *hazard* acumulado que describe el riesgo acumulado hasta el momento t de que ocurra del evento. Esta función también se ha interpretado como el número esperado de eventos que se espera observar hasta t .

Al estimar los modelos de supervivencia a partir de una muestra es necesario tomar en cuenta la censura. Por tanto, si C_i es el momento donde se presenta la censura, se define el tiempo observado en que ocurre el evento de interés T_i como el mínimo entre el verdadero tiempo y el momento de censura, de forma que $T_i = \min(T_i^*, C_i)$.

Además, se define $D_i = \mathbb{1}_{T_i^* \leq C_i}$ que corresponde a una variable que indica la ocurrencia del evento: vale uno si al individuo i le ocurrió el evento y cero

en caso contrario. El análisis de supervivencia se interesa entonces en la distribución de T_i^* utilizando únicamente la información disponible, es decir, T_i y D_i .

Estimación del Modelo de Supervivencia

El objetivo de esta sección consiste en exponer los principales métodos de estimación del modelo de supervivencia que tienen como finalidad principal encontrar un estimador para la función de supervivencia y la función de riesgo descritas.

Estimación no paramétrica

Uno de los estimadores más conocidos es el **Kaplan-Meier** ([Rizopoulos, 2012](#)), estimador no paramétrico que no requiere imponer supuestos acerca de la distribución del tiempo de ocurrencia del evento.

Para introducir este estimador se definen t_1, \dots, t_k los tiempos disponibles en la muestra de ocurrencia del evento de interés. Utilizando la ley de probabilidad total, se puede escribir la probabilidad de supervivencia en cualquier tiempo t como el producto de las probabilidades condicionales:

$$P(T^* > t) = P(T^* > t \mid T^* > t - 1) \times P(T^* > t - 1 \mid T^* > t - 2) \times \dots$$

Para estimar las probabilidades de supervivencia del evento en cada momento, se utiliza la expresión anterior y, en el cálculo de las probabilidades condicionales, se toma en cuenta la censura ajustando por el número de individuos que están en riesgo del evento (es decir, aquéllos a los que no les ha ocurrido aún el evento de interés), que deriva en el estimador:

$$\hat{S}_{KM}(t) = \prod_{i:t_i < t} \frac{r_i - d_i}{r_i}$$

donde r_i es el número de sujetos a los que no les ha ocurrido el evento en el momento t_i y d_i el número de eventos que han ocurrido en la muestra al momento t . Es posible obtener una estimación para la varianza y, por tanto,

se pueden construir intervalos de confianza. Además, a partir de la relación en (3.14) se obtiene un estimador para la función de riesgo acumulado $\hat{\mathcal{H}}_{KM}(t) = -\log \hat{S}_{KM}(t)$.

Otro estimador no paramétrico similar al anterior es el propuesto para la función de riesgo acumulada que se conoce como el estimador **Nelson-Aalen**:

$$\hat{\mathcal{H}}_{NA}(t) = \sum_{i:t_i \leq t} \frac{d_i}{r_i}$$

donde d_i y r_i tienen la misma interpretación que para el estimador **Kaplan-Meier**. Por tanto, basados en el Nelson-Aalen y utilizando otra vez la relación en (3.14), se deriva el estimador para la sobrevivencia propuesto por Breslow (Rizopoulos, 2012):

$$\hat{S}_B(t) = \exp \left\{ -\hat{\mathcal{H}}_{NA}(t) \right\} = \prod_{i:t_i \leq t} \exp(-d_i/r_i)$$

En este caso, también es posible construir intervalos de confianza a partir de la estimación de la varianza del estimador de la función de riesgo acumulada.

Los dos estimadores de la función de sobrevivencia antes descritos, el **Kaplan-Meier** y **Nelson-Aalen**, son asintóticamente equivalentes si bien pueden diferir en muestras finitas (Rizopoulos, 2012).

Estimación paramétrica

Cuando se supone una función de sobrevivencia con cierta forma paramétrica, es posible estimar por el método de máxima verosimilitud el modelo de sobrevivencia. En particular, se tienen $\{T_i, D_i\}$ para $i = 1, \dots, N$, la información de sobrevivencia de una muestra de tamaño N cuya función de distribución viene dada por P con parámetros $\boldsymbol{\theta}_S$. La función de densidad de probabilidad es $p(t; \boldsymbol{\theta}_S)$.

Para construir la función de verosimilitud se debe tomar en cuenta la censura. En este sentido, un individuo i contribuye con $p(T_i; \boldsymbol{\theta}_S)$ a la verosimilitud si le ha ocurrido el evento, mientras que contribuye $S_i(T_i; \boldsymbol{\theta}_S)$ en caso contrario, ya que lo que se sabe en este último caso es que el individuo ha sobrevivido hasta este momento.

Combinando la información de los individuos que les ha ocurrido el evento con los que no (censurados) se obtiene la función de verosimilitud censurada:

$$\ell(\boldsymbol{\theta}_S) = \sum_{i=1}^{i=N} D_i \log p(T_i; \boldsymbol{\theta}_S) + (1 - D_i) \log S_i(T_i; \boldsymbol{\theta}_S)$$

Si se utiliza que $h(t) = p(t)/S(t)$ y $S(t) = \exp\{-\mathcal{H}\}$ la verosimilitud se puede expresar:

$$\ell(\boldsymbol{\theta}_S) = \sum_{i=1}^{i=N} D_i \log h_i(T_i; \boldsymbol{\theta}_S) - \int_0^{T_i} h_i(s; \boldsymbol{\theta}_S) ds$$

donde se observa que la contribución de todos los individuos es igual a menos la función de riesgo acumulada evaluada en T_i y los individuos que han experimentado el evento adicionalmente contribuyen en una cantidad igual al logaritmo de la función de riesgo evaluada en T_i , donde $T_i = T_i^*$ en estos casos. Por tanto, las observaciones censuradas contribuyen con menor información para la inferencia en este modelo.

3.3. Modelos Conjuntos para Datos Longitudinales y de Supervivencia

En los estudios de seguimiento se suele tener medidas de diferentes resultados longitudinales de los individuos así como del tiempo que transcurre hasta la ocurrencia de un evento de interés particular. Las preguntas de investigación realizadas en base a estos estudios muchas veces requieren el análisis por separado de estos resultados. Sin embargo, en varias ocasiones el interés puede estar en la estructura de asociación entre la medida longitudinal y tiempo hasta que ocurre el evento. En la literatura biomédica, este ejemplo se encuentra de forma frecuente, ya que se realizan estudios clínicos que analizan la existencia de *biomarcadores* que sean buenos predictores del tiempo de ocurrencia del evento ([Rizopoulos, 2012](#)).

En general, en estos estudios se encuentra una característica inherente común: la naturaleza dinámica de los resultados. Es decir, que la tasa de cambio en el tiempo es diferente entre individuos pero además ésta puede cambiar a lo largo del tiempo para el mismo individuo. Por tanto, el verdadero potencial de una medida longitudinal para predecir el tiempo de ocurrencia del evento y su asociación con la supervivencia del individuo solamente se alcanza al evaluar medidas repetidas en el tiempo ([Rizopoulos, 2012](#)).

Para responder a las preguntas acerca de la asociación entre la medida longitudinal y el tiempo hasta el evento, se ha desarrollado una clase de modelos estadísticos conocidos como Modelos Conjuntos para Datos Longitudinales y de Supervivencia. Éstos han cobrado mayor relevancia recientemente particularmente en el análisis biométrico, en particular los trabajos de [Faucett & Thomas \(1996\)](#) y [Wulfsohn & Tsiatis \(1997\)](#) han introducido lo que hoy en día se conoce como el Modelo Conjunto Estándar y en los últimos años ha habido un gran desarrollo en este campo. Además, permiten incorporar distintos mecanismos que generan los datos faltantes como el supuesto de *missing not at random* ([Rizopoulos, 2012](#)). Esto resuelve problemas de sesgo que se generan con la estimación bajo el supuesto *missing at random* que se realiza en la estimación de la trayectoria longitudinal del proceso en los modelos para datos longitudinales como [LMM](#) o [LCMM](#).

La mayoría de los desarrollos estadísticos en el área de los Modelos Conjuntos se han focalizado en los modelos denominados de Efectos Aleatorios

Compartidos (*Shared Random-Effect Models*). En estos modelos se suele incluir alguna característica longitudinal como predictor en un modelo de Supervivencia, en donde el interés radica en analizar el tiempo hasta la ocurrencia de un evento. En general, el procedimiento de vinculación consiste en incluir, mediante alguna forma funcional, los efectos aleatorios del Modelo Lineal Mixto en el de Supervivencia (Rizopoulos, 2012).

Sin embargo, existe otra aproximación a los Modelos Conjuntos a partir de Clases Latentes, que constituye una extensión del modelo presentado en la sección 3.1.3 y se conoce como Modelo Mixto Conjunto de Clases Latentes (JLCMM).

En los JLCMM, el supuesto fundamental es que la estructura de grupos latentes captura la asociación entre la trayectoria longitudinal de un proceso y el riesgo de ocurrencia del evento. Debido a la flexibilidad del supuesto acerca de esta asociación y la posibilidad de inclusión de covariables del JLCMM, estos modelos resultan particularmente adecuados en problemas centrados en la predicción (Proust-Lima *et al.*, 2014).

En particular, este supuesto implica que cada grupo latente se caracteriza por una trayectoria específica del proceso longitudinal, que se puede aproximar por un LMM o LPMM, y por un proceso de supervivencia específico¹. La hipótesis clave del modelo, conocida como **independencia condicional**, es que la evolución de la medida longitudinal y el riesgo del evento son independientes condicional a los grupos latentes (Proust-Lima *et al.*, 2009, 2014).

En resumen, el JLCMM se compone de:

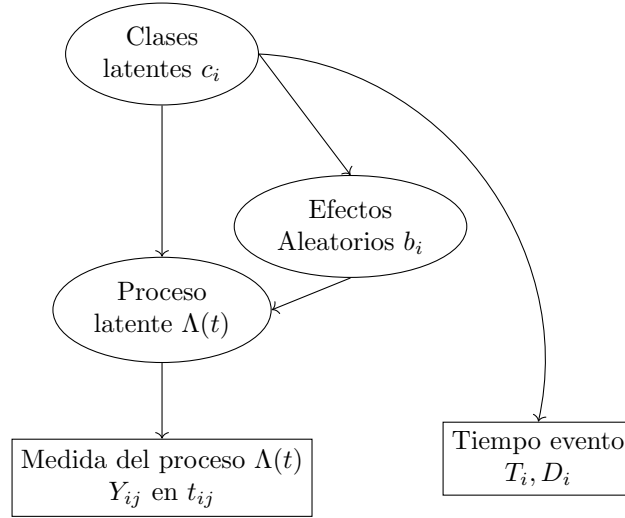
- (1) un modelo de pertenencia a los grupos latentes, en general un logit multinomial como el de la ecuación (3.8)
- (2) un modelo longitudinal específico de cada grupo que podría ser un LMM como el de la ecuación (3.10) o un LPMM que se compone de las ecuaciones (3.6), (3.9) y (3.10)
- (3) un modelo de supervivencia específico para cada grupo.

La figura 3.2 muestra un grafo dirigido que representa el esquema conceptual del JLCMM. La medida longitudinal Y_{ij} disponible para los momentos t_{ij}

¹Esto refiere a un riesgo de ocurrencia del evento específico para cada grupo.

con $j = 1, \dots, n_i$ está generada por el proceso latente $\Lambda(t)$, medido en tiempo continuo $t \in \mathbb{R}^+$. La flecha a cuadros desde los efectos aleatorios b_i hacia el tiempo del evento T_i corresponde si se incumple el supuesto de independencia entre la medida longitudinal y el riesgo del evento, condicional a la estructura de grupos latentes.

Figura 3.2: Grafo dirigido que define el **JLCMM**



Fuente: Proust-Lima *et al.* (2015)

3.3.1. Modelo Mixto Conjunto de Clases Latentes

En esta sección se presenta en detalle el modelo **JLCMM**, los modelos que lo componen, supuestos y métodos de estimación. Como ya fue mencionado, el **JLCMM** se compone de otros tres modelos, un logit multinomial y el **LMM** o **LPMM** presentados en forma detallada en la sección 3.1 y un modelo de sobrevivencia detallado en la sección 3.2. Ahora al especificar el Modelo de Sobrevivencia para el **JLCMM** se debe diferenciar entre los grupos. Por tanto, en el grupo latente g el riesgo del evento se especifica utilizando una función de riesgo proporcional como la siguiente:

$$h_i(t|c_i = g) = h_{0g}(t)e^{\mathbf{X}_{S_i}^T \delta_g} \quad (3.15)$$

donde \mathbf{X}_{S_i} es un vector de covariables asociadas al vector de parámetros δ_g que pueden tener efectos específicos en cada grupo g . La función $h_{0g}(t)$ corresponde a la función de riesgo de línea de base, que depende de ciertos parámetros ξ_g ,

la que puede ser igual o diferir entre grupos. Una forma de diferenciar esta función es especificarla de forma tal que $h_{0g}(t) = h_0(t; \boldsymbol{\xi}_g)$ o bien especificarla de forma proporcional entre grupos $h_{0g}(t) = h_0(t; \boldsymbol{\xi}^*)e^{\xi_g}$, donde ξ_g es el factor de proporcionalidad.

En resumen, el **JLCMM** compuesto por el modelo de pertenencia a los grupos, longitudinal y de supervivencia específicos se presenta en la ecuación (3.16):

$$\left\{ \begin{array}{l} \pi_{ig} = P(c_i = g | \mathbf{X}_{Ci}) = \frac{e^{\mathbf{X}_{Ci}^\top \boldsymbol{\lambda}_g}}{\sum_{l=1}^G e^{\mathbf{X}_{Ci}^\top \boldsymbol{\lambda}_l}} \\ Y_{ij} | c_i = g = \mathbf{X}_{Li}(t_{ij})^\top \boldsymbol{\beta}_g + \mathbf{Z}_i(t_{ij})^\top \mathbf{b}_{ig} + \omega_i(t_{ij}) + \varepsilon_{ij} \\ h_i(t | c_i = g) = h_{0g}(t) e^{\mathbf{X}_{Si}^\top \boldsymbol{\delta}_g} \end{array} \right. \quad (3.16)$$

donde $\boldsymbol{\lambda}_g$ es un vector de parámetros específicos del grupo g asociados a las covariables \mathbf{X}_{Ci} invariantes temporalmente (se incluye intercepto), $\boldsymbol{\beta}_g$ y \mathbf{b}_{ig} corresponden al vector de efectos fijos y aleatorios para cada grupo g y la matriz de covarianzas de los efectos aleatorios puede diferir entre grupos incorporando un coeficiente de proporcionalidad σ_g^2 . Por tanto, $\mathbf{b}_{ig} \sim N(\mathbf{0}, \mathbf{D}_g = \sigma_g^2 \mathbf{D})$ donde \mathbf{D} es la misma matriz definida en la sección 3.1.1. Se puede vincular el proceso latente con la evolución de la medida longitudinal agregando una función H con parámetros $\boldsymbol{\eta}$ como en la sección 3.1.3. Además, se incorpora una función de riesgo específico para cada grupo g explicada anteriormente.

Estimación del Modelo Mixto Conjunto de Clases Latentes

Los modelos conjuntos de grupos latentes se estiman por **MLE** y se supone una función de enlace (H) lineal en la ecuación (3.10). Constituyen una extensión de los **LCMM**, por tanto, el vector de parámetros y las restricciones a imponer para la identificación son similares a las expuestas en la sección 3.1.3. Se agregan los parámetros correspondientes al modelo de supervivencia y algunas restricciones cambian respecto al **LCMM**.

Las restricciones sobre los parámetros son:

- (I) $\boldsymbol{\lambda}_G = \mathbf{0}$, parámetros del modelo de pertenencia a grupos del último grupo

- (II) $\sigma_G^2 = 1$, coeficiente de proporcionalidad de matriz de covarianzas de efectos aleatorios del último grupo
- (III) $\boldsymbol{\xi}_G = \mathbf{0}$, coeficientes de la función de riesgo de línea de base del último grupo

Por consiguiente, el vector a estimar, dado el número de grupo G , es $\boldsymbol{\theta}_G = \left([\boldsymbol{\lambda}_g^\top]_{g=1}^{G-1}, [\boldsymbol{\xi}_g]_{g=1}^{G-1}, [\boldsymbol{\delta}_g^\top]_{g=1}^{G-1}, [\boldsymbol{\beta}_g^\top]_{g=1}^G, \text{vec}(\mathbf{D})^\top, [\sigma_g^2]_{g=1}^{G-1}, \sigma_\omega^2, \boldsymbol{\rho}^\top \right)$, donde $\text{vec}(\mathbf{D})$, σ_ω^2 y $\boldsymbol{\rho}$ son los ya definidos en la sección 3.1.1, $\boldsymbol{\beta}_g$ es el vector de efectos fijos específicos de cada grupo y σ_g^2 el coeficiente de proporcionalidad de la matriz de covarianzas de los efectos aleatorios para el grupo g .

El método MLE maximiza la función, dado el número de grupos G , como en la ecuación 3.11 y la contribución individual a la verosimilitud en el JLCMM es:

$$L_i(\boldsymbol{\theta}_G) = \sum_{g=1}^G \left\{ \pi_{ig} \phi_{ig}(\mathbf{Y}_i | c_i = g; \boldsymbol{\theta}_G) h_i(T_i | c_i = g; \boldsymbol{\theta}_G)^{D_i} S_i(T_i | c_i = g; \boldsymbol{\theta}_G) \right\} \quad (3.17)$$

donde π_{ig} está dada en la ecuación (3.8), $h_i(T_i | c_i = g; \boldsymbol{\theta}_G)$ es la función de riesgo específica de cada grupo definida en (3.15) y $S_i(T_i | c_i = g; \boldsymbol{\theta}_G)$ su correspondiente función de sobrevivencia. Por su parte, ϕ_{ig} es la densidad de una normal multivariada con media $\mathbf{X}_{Li} \boldsymbol{\beta}_g$ y matriz de covarianzas $\mathbf{V}_{ig} = \mathbf{Z}_i \mathbf{D}_g \mathbf{Z}_i^\top + \mathbf{R}_i + \boldsymbol{\Sigma}_i$.

3.4. Cálculos Posteriores a la Estimación

Una vez estimados los modelos se realiza una serie de cálculos posteriores que permiten validarlos y, en particular para los modelos de clases latentes, permiten elegir el número de grupos que mejor ajusta a los datos disponibles. En general, ésta debe basarse en un conjunto de consideraciones y no solamente en minimizar algún criterio de información (Proust-Lima *et al.*, 2014).

3.4.1. Clasificación *a posteriori*

Una vez estimados los modelos de clases latentes, se computan las probabilidades de pertenencia a las diferentes clases para cada individuo. Esto se basa en el cálculo *a posteriori* de probabilidades con la fórmula de Bayes y se utiliza para la clasificación de los sujetos así como para evaluar la bondad de ajuste del modelo (Proust-Lima *et al.*, 2016b).

La fórmula de cálculo difiere según se trate de un modelo de clases latentes simple o uno conjunto, longitudinal y sobrevivencia. Para los LCMM la probabilidad *a posteriori* se obtiene de la siguiente forma:

$$\hat{\pi}_{ig}^{(Y)} = P(c_i = g | \mathbf{X}_{Ci}, \mathbf{X}_{Li}, \mathbf{Y}_i; \hat{\boldsymbol{\theta}}_G) = \frac{\hat{\pi}_{ig} \phi_{ig}(\mathbf{Y}_i | c_i = g; \hat{\boldsymbol{\theta}}_G)}{\sum_{l=1}^G \hat{\pi}_{il} \phi_{il}(\mathbf{Y}_i | c_i = l; \hat{\boldsymbol{\theta}}_G)} \quad (3.18)$$

donde $\hat{\pi}_{ig}$ es la probabilidad de pertenencia a las clase g (con $g = 1, \dots, G$) definida en (3.8) y computada utilizando la estimación del vector de parámetros $\hat{\boldsymbol{\theta}}_G$ obtenido dado un número de clases G .

Por su parte, en el JLCMM las probabilidades *a posteriori* para cada individuo se calculan de la siguiente forma:

$$\begin{aligned} \hat{\pi}_{ig}^{(Y,T)} &= P(c_i = g | \mathbf{X}_{Ci}, \mathbf{X}_{Li}, (\mathbf{Y}_i, D_i); \hat{\boldsymbol{\theta}}_G) = \\ &= \frac{\hat{\pi}_{ig} \phi_{ig}(\mathbf{Y}_i | c_i = g; \hat{\boldsymbol{\theta}}_G) h_i(T_i | c_i = g; \hat{\boldsymbol{\theta}}_G)^{D_i} S_i(T_i | c_i = g; \hat{\boldsymbol{\theta}}_G)}{\sum_{l=1}^G \hat{\pi}_{il} \phi_{il}(\mathbf{Y}_i | c_i = l; \hat{\boldsymbol{\theta}}_G) h_i(T_i | c_i = l; \hat{\boldsymbol{\theta}}_G)^{D_i} S_i(T_i | c_i = l; \hat{\boldsymbol{\theta}}_G)} \end{aligned} \quad (3.19)$$

donde $\hat{\pi}_{ig}$ se define como en la ecuación (3.18) y $h_i(T_i|c_i = g; \boldsymbol{\theta}_G)$, $S_i(T_i|c_i = g; \boldsymbol{\theta}_G)$ y $\phi_{ig}(\mathbf{Y}_i|c_i = g; \boldsymbol{\theta}_G)$ se definen de igual forma en que en la sección 3.3.1.

A partir de estas probabilidades, cada sujeto se clasifica en la clase que maximiza su probabilidad de pertenencia *a posteriori*, esto es, que $\hat{c}_i = \arg \max_g \left(\hat{\pi}_{ig}^{(Y)} \right)$ para el LCMM o bien $\hat{c}_i = \arg \max_g \left(\hat{\pi}_{ig}^{(Y,T)} \right)$ para el JLCMM.

Una vez computadas las probabilidades *a posteriori* y realizada la clasificación para cada individuo según el criterio anterior, se pueden construir las **tablas de clasificación**. Estas tablas contienen la media de las probabilidades de pertenecer a cada clase latente entre los sujetos clasificados *a posteriori* en esa clase (Ver Tabla 3.1). Una clasificación perfecta proporcionaría unos en la diagonal y ceros fuera de ella, aunque en la práctica, se considera que altos valores en la diagonal y bajos fuera de ella indican buena discriminación de la población en las clases consideradas (Proust-Lima *et al.*, 2016b).

Tabla 3.1: Tabla de clasificación

Clase	Media de la probabilidad de pertenencia a cada clase					
Final \hat{c}_i	#	1	...	g	...	G
1	N_1	$\frac{1}{N_1} \sum_{i=1}^{N_1} \hat{\pi}_{i1}^{(\cdot)}$...	$\frac{1}{N_1} \sum_{i=1}^{N_1} \hat{\pi}_{ig}^{(\cdot)}$...	$\frac{1}{N_1} \sum_{i=1}^{N_1} \hat{\pi}_{iG}^{(\cdot)}$
...
g	N_g	$\frac{1}{N_g} \sum_{i=1}^{N_g} \hat{\pi}_{i1}^{(\cdot)}$...	$\frac{1}{N_g} \sum_{i=1}^{N_g} \hat{\pi}_{ig}^{(\cdot)}$...	$\frac{1}{N_g} \sum_{i=1}^{N_g} \hat{\pi}_{iG}^{(\cdot)}$
...
G	N_G	$\frac{1}{N_G} \sum_{i=1}^{N_G} \hat{\pi}_{i1}^{(\cdot)}$...	$\frac{1}{N_G} \sum_{i=1}^{N_G} \hat{\pi}_{ig}^{(\cdot)}$...	$\frac{1}{N_G} \sum_{i=1}^{N_G} \hat{\pi}_{iG}^{(\cdot)}$

$\hat{\pi}_{ig}^{(\cdot)}$ es $\hat{\pi}_{ig}^{(Y)}$ para LCMM y $\hat{\pi}_{ig}^{(Y,T)}$ para JLCMM
Fuente: Proust-Lima *et al.* (2016b).

3.4.2. Elección de Grupos y Bondad de Ajuste

Para la elección del número de clases se considera el **Criterio de Información Bayesiano (BIC)** que resulta usualmente utilizado en el contexto de modelos mixtos (Proust-Lima *et al.*, 2014, 2016b). El criterio consiste en minimizar el valor de:

$$BIC(\boldsymbol{\theta}_G) = -2L(\boldsymbol{\theta}_G) + k_{\boldsymbol{\theta}_G} \log(N)$$

donde $\boldsymbol{\theta}_G$ es el vector de parámetros del modelo dado el número de clases G , $L(\boldsymbol{\theta}_G)$ es el valor de la verosimilitud del modelo estimado, $k_{\boldsymbol{\theta}_G}$ es la cantidad de

parámetros que contiene ese vector y N el número de observaciones utilizadas en la estimación.

Para la correcta elección de las clases, además de valores bajos de **BIC**, debe lograrse una buena discriminación de los individuos. Para considerar este aspecto se calcula el **Índice de Entropía** (E) de la siguiente forma:

$$E = 1 + \frac{1}{N \log(G)} \left(\sum_{i=1}^N \sum_{g=1}^G P(c_i = g | \mathbf{X}_{C_i}) \log \left(P(c_i = g | \mathbf{X}_{C_i}) \right) \right)$$

donde c_i es la variable de clases latentes, \mathbf{X}_{C_i} el conjunto de covariables en el modelo y G el número de grupos en los que se clasifica a los individuos. $P(c_i = g | \mathbf{X}_{C_i})$ son las probabilidades *a posteriori* para cada individuo. Este índice se encuentra acotado en $[0, 1]$ y cuanto más cercano a uno, más clara es la identificación de las clases latentes. En la práctica, se considera que un valor en el entorno de 0.8 es suficiente para decir que la clasificación en esos grupos latentes es apropiada (Muthén, 2004).

Sin embargo, el valor de entropía depende del contexto, el objeto de estudio y las covariables utilizadas por tanto no se debe basar la elección del número de clases en su valor únicamente sino que resulta recomendable tener en cuenta las tablas de clasificación, ya que aportan mayor información sobre la calidad de clasificación (Muthén, 2004).

En el caso de los modelos conjuntos **JLCMM**, se evalúa el cumplimiento del **supuesto de independencia condicional** dado el número de clases. Una forma de hacerlo consiste en aplicar un contraste conocido como *Score Test* (Proust-Lima *et al.*, 2014). En este contraste la hipótesis alternativa plantea que el tiempo hasta la ocurrencia del evento y el modelo longitudinal tienen relación de dependencia a través de los efectos aleatorios además de las clases latentes. Por tanto, el modelo de sobrevivencia bajo la hipótesis alternativa comparte los efectos aleatorios con el longitudinal. En ese caso la ecuación (3.15) del modelo conjunto se reemplaza por:

$$h_i(t | c_i = g) = h_{0g}(t) e^{\mathbf{X}_{S_i}^T \delta_g + b_{ig} \gamma}$$

donde γ es un vector en \mathbb{R}^q que vincula los efectos aleatorios del modelo longitudinal en (3.10) con el de sobrevivencia. Por tanto, la evaluación del supuesto de independencia condicional consiste en contrastar las siguientes

hipótesis:

$$\begin{aligned} H_0 &: \boldsymbol{\gamma} = \mathbf{0} \\ H_1 &: \boldsymbol{\gamma} \neq \mathbf{0} \end{aligned}$$

En [Jacqmin-Gadda et al. \(2010\)](#) se presenta el estadístico utilizado para realizar el contraste:

$$U = \sum_{i=1}^N \sum_{g=1}^G \hat{\pi}_{ig} (D_i - \mathcal{H}_{ig}(T_i)) \hat{b}_{ig}$$

que se obtiene estimando la covarianza entre los residuos martingala¹ del modelo de sobrevivencia $(D_i - \mathcal{H}_{ig}(T_i))$ y las estimaciones *bayesianas* de los efectos aleatorios específicos de clase (\hat{b}_{ig}) , ponderados por la probabilidad *a posteriori* de pertenencia a las mismas $(\hat{\pi}_{ig})$.

Bajo la hipótesis nula $U^\top \text{Var}(U)^{-1} U$ sigue una *Chi-cuadrado* con m grados de libertad, donde m es el tamaño del vector $\boldsymbol{\gamma}$ ([Jacqmin-Gadda et al., 2010](#)). En varios trabajos de simulación se ha encontrado a este procedimiento como uno de los más potentes para someter a consideración el supuesto de independencia condicional ([Jacqmin-Gadda et al., 2010](#); [Proust-Lima et al., 2014](#)).

3.4.3. Predicciones

Luego de las estimaciones se realizan predicciones de la trayectoria del indicador longitudinal y del proceso latente y, además, predicciones dinámicas de la ocurrencia del evento en el caso del modelo conjunto. Dado que los modelos utilizados en este trabajo se enmarcan en el contexto de los Modelos Lineales Mixtos se derivan las *estimaciones empíricas bayesianas* ([Proust-Lima et al., 2014, 2016b](#)).

¹ Los residuos martingala en el modelo de sobrevivencia se definen como la diferencia entre el número observado de eventos ocurridos a un individuo y el número esperado de eventos ocurridos condicional en el modelo estimado, el momento de seguimiento y la trayectoria observada de las covariables temporales. Por más detalles consultar [Fleming & Harrington \(2013\)](#); [Therneau & Grambsch \(2000\)](#); [Therneau et al. \(1990\)](#).

Efectos aleatorios

Para los modelos presentados en las secciones 3.1.3 y 3.3.1 los efectos aleatorios se predicen de la siguiente forma:

$$\widehat{\mathbf{b}}_{ig} = \widehat{\sigma}_g^2 \widehat{\mathbf{D}} \mathbf{Z}_i^\top \widehat{\mathbf{V}}_{ig}^{-1} \left(\widehat{\mathbf{Y}}_i - \mathbf{X}_{Li} \widehat{\boldsymbol{\beta}}_g \right) \quad (3.20)$$

donde $\widehat{\mathbf{Y}}_i$ es el vector de valores de la medición transformados según la función de enlace $\widehat{Y}_{ij} = H^{-1}(Y_{ij}; \widehat{\boldsymbol{\eta}})$ para $j = 1, \dots, n_i$. Las estimaciones marginales se obtienen ponderando por las probabilidades *a posteriori* $\widehat{\mathbf{b}}_i = \sum_{g=1}^G \widehat{\pi}_{ig}^{(\cdot)} \widehat{\mathbf{b}}_{ig}$.

Predicciones longitudinales y residuos

Las predicciones de la medida longitudinal, del proceso latente y de los residuos se pueden calcular marginales o específicas para los individuos, en donde en el último caso se incorporan las predicciones de los efectos aleatorios (Proust-Lima *et al.*, 2016b).

Cuando se separan los modelos correspondientes a la medida longitudinal y al proceso latente, las predicciones marginales para cada clase se obtienen en la escala del proceso latente como:

$$\widehat{Y}_{ijg}^{(M)} = \mathbf{X}_{Li}(t_{ij})^\top \widehat{\boldsymbol{\beta}}_g$$

mientras que las predicciones específicas para los individuos

$$\widehat{Y}_{ijg}^{(SS)} = \mathbf{X}_{Li}(t_{ij})^\top \widehat{\boldsymbol{\beta}}_g + \mathbf{Z}_i(t_{ij})^\top \widehat{\mathbf{b}}_{ig}$$

Por su parte los residuos marginales se obtienen como $R_{ijg}^{(M)} = Y_{ij} - \widehat{Y}_{ijg}^{(M)}$, mientras que los específicos para los individuos $R_{ijg}^{(SS)} = Y_{ij} - \widehat{Y}_{ijg}^{(SS)}$.

Por último, se pueden obtener las predicciones marginales y específicas de los individuos promediadas entre clases y sus correspondientes residuos:

$$\widehat{Y}_{ij}^{(M)} = \sum_{g=1}^G \widehat{\pi}_{ig}^{(\cdot)} \widehat{Y}_{ijg}^{(M)} \quad \widehat{Y}_{ij}^{(SS)} = \sum_{g=1}^G \widehat{\pi}_{ig}^{(\cdot)} \widehat{Y}_{ijg}^{(SS)}$$

Predicción de la trayectoria media

Por su parte, es posible realizar una predicción de la trayectoria media de la medida longitudinal acorde a cierto perfil en las covariables. El cálculo es el mismo que antes pero ahora se obtiene la predicción sobre un individuo hipotético en las covariables X . Se realiza la predicción sobre el vector de $E(Y|X = x; \hat{\theta}_G)$ calculando el vector marginal de predicciones específicos de cada clase como se explicó anteriormente. Además, es posible obtener la predicción sobre el proceso latente en los casos donde se separe el modelo de medición del estructural.

Predicciones de sobrevivencia

En el caso del [JLCMM](#) es posible realizar predicciones de la sobrevivencia al evento de los individuos y, además, estas predicciones son específicas por clases. Luego de estimar el modelo en (3.16), mediante la relación en (3.14) se obtienen funciones de sobrevivencia estimadas para cada clase y se pueden representar gráficamente.

3.5. Aspectos Metodológicos: Trayectoria Nutricional y Desempeño Escolar

Una vez expuestos los modelos teóricos que se consideran en este trabajo, la presente sección tiene por objetivo explicitar algunas consideraciones metodológicas realizadas para aplicar estos modelos al problema concreto de análisis, la trayectoria nutricional y su relación con el desempeño escolar.

En primer lugar, se elige el [IMC](#) como indicador del estado nutricional de los niños y niñas. De las medidas antropométricas disponibles, el [IMC](#) para la edad se utiliza generalmente y presenta estándares de crecimiento dados por la [OMS](#) para niños y niñas en el rango de edad considerado en este trabajo ([Butte *et al.*, 2007](#); [deOnis & Habicht, 1996](#); [deOnis *et al.*, 2004, 2007](#); [deOnis & Lobstein, 2010](#)).

En segundo lugar, se mide el desempeño escolar como la edad de desvinculación del sistema educativo. Esto se basa en que el problema del abandono es importante y presenta tasas altas en Uruguay que comienzan a ser elevadas particularmente a partir de los 13-14 años, edad que tienen los niños del [ELBU](#) en la última ocasión en que se los entrevistó. Se considera únicamente el abandono y no otros desempeños relevantes, como la repetición por ejemplo, ya que esta información no está disponible en el [ELBU](#). Los Modelos Conjuntos analizan el momento en que se produce el evento, lo cual está disponible únicamente para el abandono. Si bien se preguntó sobre la repetición, desempeño escolar relevante para estudiar en Uruguay, no se tiene el momento en que el episodio se produjo y, por tanto, no se analiza mediante esta metodología¹. Sin embargo, el análisis de la relación entre la trayectoria nutricional y la repetición constituye un tema de importancia que podría ser abordado mediante la aplicación de otros modelos.

Por último, el análisis se hace para niñas y niños por separado dado que las tasas de crecimiento y desarrollo en el tramo de edad considerado difieren según género ([deOnis & Habicht, 1996](#); [deOnis *et al.*, 2007](#); [Stuart & Panico, 2016](#)).

¹ En el [ELBU](#) tampoco se tiene información de otros desempeños, como test de aprendizaje por ejemplo, que puedan ser objeto de análisis mediante la metodología propuesta en este trabajo.

3.5.1. Base de Datos

La información pertenece a las primeras rondas del ELBU. Éste constituye una muestra de niños que cursaba primer año en escuelas públicas en 2004 representativa de las capitales departamentales que fue seguida en años ulteriores (2004, 2006, 2011/12 y 2015/2016, aproximadamente a los 6, 8, 14 y 18 años). Dado que las tasas de asistencia en primer año escolar son universales, la muestra no tiene los problemas de sesgo de selección que habitualmente presentan los relevamientos realizados a estudiantes en la adolescencia. La muestra fue seleccionada en base al Tercer Censo de Talla Escolar de 2002 ([Amarante et al., 2007](#)). Es representativa de los niños que concurrían a primer año de la educación primaria pública -aproximadamente un 85 % de los niños que estaban en primer año en 2004- en capitales departamentales y el área metropolitana, si bien presenta un truncamiento en los estratos socio económicos más altos, ya que solo el 43 % de los niños del quintil más alto concurrían a escuela pública ese año ([Failache et al., 2015](#)). Las escuelas fueron seleccionadas de los departamentos Artigas, Canelones, Colonia, Florida, Montevideo, Paysandú, Rivera y Soriano.

En 2004 se obtuvieron medidas antropométricas para aproximadamente 3000 niños que cursaban primer año en escuelas públicas. Estas medidas fueron tomadas por antropometristas en las escuelas. Ese mismo año se entrevistaron también a los miembros de los hogares obteniendo información socio económica del hogar, niño y sus integrantes. En 2005, se volvieron a tomar medidas antropométricas de los niños de Montevideo y el Área Metropolitana en las escuelas y en 2006 nuevamente se entrevistaron las familias. En 2009, se tomó otra ronda de medidas antropométricas para los niños al finalizar el ciclo educativo primaria, ya que en este momento la mayoría de los niños se encontraban en sexto de escuela. Por último, en 2011/2012 se recabó información de los hogares y sus integrantes para cuando los niños de la cohorte tenían aproximadamente entre 13 y 14 años ([Amarante et al., 2007](#); [Failache et al., 2015](#)).

En la primera ola, se realizaron más de 3000 encuestas a hogares y para 2811 se dispuso de información socio económica y antropométrica. En 2011/12 se entrevistaron 2152 hogares, lo cual implica que se recuperó el 65 % del panel inicial (desgranamiento del 35 %)¹. Combinando ambas olas se cuenta con 1563

¹ De los casos entrevistados en la tercera ola, algunos no cuentan con información an-

casos con información antropométrica y socio económica para ambas. Por su parte, en 2006 fueron entrevistados 1160 hogares de Montevideo y Canelones, de los cuales 900 se recuperaron en 2011/12 (Arim *et al.*, 2014).

En el marco de un proyecto CSIC, se obtuvo financiamiento para la realización de una nueva ola de este estudio. Esto permitió obtener información tanto socio económica como antropométrica para cuando tienen entre 17 y 18 años. Aún no se cuenta con la base de datos final, pero con ella se podrán analizar aspectos poco explorados en este trabajo lo que constituye una línea de investigación a futuro.

El ELBU contiene la información requerida para este trabajo. En primer lugar, se tomaron medidas antropométricas en el primer año escolar, al año siguiente y, por último, en 2009 las medidas se tomaron al finalizar el ciclo primaria. En definitiva, se cuenta con una base de datos longitudinal con tres momentos de medición de peso y talla de los niños que darán cuenta de su trayectoria nutricional. Cabe mencionar que no todos los individuos cuentan las tres medidas debido a que la segunda fue realizada únicamente para los niños de Montevideo y Área Metropolitana y, además, el estudio presenta los problemas usuales de datos faltantes.

Con las medidas de peso y talla de niños se construye el IMC, indicador que aproxima la situación nutricional en este rango de edad, según las recomendaciones de la OMS. Además, se calculan las medidas de los *z-scores* para el IMC según la metodología de la OMS.

Por otro lado, se cuenta con información socio económica de los hogares a los que pertenecen los niños, así como información referente al nacimiento del niño, recogida de forma retrospectiva, características de los integrantes del hogar, particularmente la madre del niños. En las últimas olas se cuenta, además, con resultados de test psicológicos que intentan captar características asociadas a las habilidades cognitivas y no cognitivas de los niños¹. De aquí, se podrán tomar en cuenta otras dimensiones para explicar el desempeño educativo (Failache *et al.*, 2015).

tropométrica y otros niños adicionales fueron medidos en 2009 pero no se logró concretar la entrevista en el hogar en 2011/12.

¹ Las habilidades cognitivas se recogieron en la última ola del estudio en 2015/16 en tanto que las habilidades no cognitivas se exploraron en las dos últimas entrevistas, 2011/12 y 2015/16

3.5.2. Variables

La variable medida para el análisis de la trayectoria nutricional es el **IMC** que se calcula:

$$\text{IMC} = \frac{\text{Peso (kg)}}{[\text{Estatura (m)}]^2} \quad (3.21)$$

Se cuenta con la fecha de nacimiento y de toma de las medidas que permiten calcular el *z-score* de los individuos a lo largo de las distintas mediciones en el estudio. Estos *z-score* se utilizan para tomar valores que se encuentran dentro de los rangos considerados posibles según la **OMS**¹. En los modelos donde se analiza la evolución de la trayectoria nutricional las variables utilizadas son la edad y edad al cuadrado (en meses) y el **IMC** construido como en (3.21).

Para el análisis de la asociación entre trayectoria nutricional y desempeño escolar en el **JLCMM**, se tienen dos variables de resultados: la trayectoria nutricional y el desempeño escolar. Como ya fue mencionado, la trayectoria nutricional se aproxima mediante el **IMC** de igual forma que cuando se analiza la trayectoria nutricional únicamente. El desempeño escolar analizado es la edad de abandono. En la última ronda del estudio considerada (2011/12) se preguntó si el niño continúa estudiando en ese momento y, en caso que haya dejado los estudios, la edad de abandono. Esta variable es reportada por la madre del niño o niña ex-post a las mediciones de la trayectoria nutricional. Para el **JLCMM**, el desempeño escolar se analiza en el marco de un modelo de sobrevivencia, por tanto, es necesario tener la edad de abandono del niño, que es una variable disponible en esta ola del **ELBU**.

La revisión de antecedentes indica que en el análisis del desempeño escolar es importante considerar el fracaso, que en este estudio se introducen en el modelo mediante una variable dicotómica que indica si el niño repitió algún año durante todo el ciclo educativo previo al momento de la última entrevista. Esta información se tiene únicamente para los niños entrevistados en el 2011/12 y dado que la repetición es una covariable en el modelo conjunto, este análisis se realiza solamente con los niños y niñas que fueron entrevistados en 2011/12 en sus hogares y cuentan con la información de repetición previa.

¹ Los valores de *z-score* menores a -5 y mayores que 5 no se consideran humanamente posibles según la **OMS**. Por tanto, en las estimaciones del presente trabajo no fueron consideradas las ocasiones donde el *z-score* del **IMC** pasara los umbrales de valores posibles dados por la **OMS**

Otras variables relevantes en el [ELBU](#):

- Fracaso escolar:

Repitió: vale 1 si repitió algún año y 0 en otro caso

- Nivel educativo materno:

Primaria: vale 1 si tiene hasta nivel primaria completa y 0 en otro caso (referencia)

Secundaria: vale 1 si tiene hasta nivel secundaria completa y 0 en otro caso

Terciaria: vale 1 si tiene algún año de nivel terciario

- Características del niño:

Edad: edad en meses

Peso: peso al nacer (g)

Talla: talla al nacer (cm)

- Otras variables de control:

Género

3.5.3. Estrategia Empírica

En esta sección se presentan los modelos estimados y algunas consideraciones metodológicas en la estimación. Todas las estimaciones se realizaron utilizando **R**, particularmente las funciones de la librería `lcmm` que se describe en [Proust-Lima *et al.* \(2016a\)](#). Según lo expuesto anteriormente el desempeño escolar se aproxima mediante la edad de abandono y el estado nutricional por el **IMC**.

Análisis de la Trayectoria Nutricional

Para realizar este análisis se utilizan los modelos **LCMM** por separado para niñas y niños. Éstos indagan en la estructura que mejor logra agrupar a los niños y niñas según sus trayectorias nutricionales. Para ello se estiman cinco modelos fijando el número de grupos de uno hasta cinco y mediante los criterios de bondad de ajuste y clasificación *a posteriori* expuestos en la sección 3.4 se elige el número más adecuado. La estimación se realizó fijando como máximo cinco grupos dado que aumentarlo implica aumentar el número de parámetros y además se compromete la convergencia con gran cantidad de grupos.

Fijado el número de grupos, cada modelo se estima suponiendo una **función de enlace lineal** y una **I-spline cuadrática** con tres nodos en los cuantiles ([Proust-Lima *et al.*, 2016b](#); [Ramsay, 1988](#)). Por tanto, se tienen diez parametrizaciones por sexo.

Para los **LCMM** se especifica:

- trayectoria cuadrática según edad diferenciando coeficientes para cada grupo
- efectos aleatorios en el intercepto y coeficiente lineal de esta trayectoria
- matriz de covarianzas de los efectos aleatorios no diagonal (*unstructured*)
- matriz de covarianzas de los efectos aleatorios proporcional entre grupos

Se especifica una trayectoria cuadrática ya que resulta adecuado para representar la evolución del **IMC** para este rango de edad y se cuenta con tres medidas del indicador como máximo por individuo. El resultado de la especificación surge luego de considerar otras más complejas que no aportaban a los objetivos

de este trabajo y que comprometían la convergencia en el proceso de estimación. Por un lado, se quitaron los efectos aleatorios en el término cuadrático de la trayectoria nutricional para la edad ya que no arrojaba diferencias en los resultados, pero aumentaba significativamente el tiempo computacional para la estimación. Por otro lado, se especificaron matrices de covarianzas de los efectos aleatorios diferentes para los grupos ya que la especificación más sencilla con matrices iguales para todos los grupos no lograba una discriminación razonable de los individuos.

Por tanto, el **LCMM** finalmente estimado se escribe como:

$$\begin{cases} \pi_{ig} = P(c_i = g | X_{Ci}) = \frac{e^{\lambda_g}}{\sum_{l=1}^G e^{\lambda_l}} \\ \text{IMC}_{ij|c_i=g} = \beta_{0g} + \beta_{1g}\text{Edad}_{ij} + \beta_{2g}\text{Edad}_{ij}^2 + b_{0ig} + b_{1ig}\text{Edad}_{ij} + \varepsilon_{ij} \end{cases} \quad (3.22)$$

donde la ecuación de pertenencia a las clases no incluye covariables por lo que es una probabilidad específica de pertenencia pero no condicionada por ninguna característica de los individuos. El modelo de medición en la segunda ecuación representa la trayectoria del **IMC** para el individuo i en las ocasiones $j = 1, \dots, n_i$, donde se incluye la función de enlace (H) para la especificación con *splines*. El intercepto en esta ecuación se encuentra para cero meses de edad, es decir al nacimiento del niño o niña. Por último, en cuanto a los efectos aleatorios y la perturbación se supone que $\mathbf{b}_{ig} \sim N(\mathbf{0}, \sigma_g^2 \mathbf{D})$, $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ y son independientes.

Relación entre Trayectoria Nutricional y Desempeño Escolar

Este análisis se realiza mediante los **JLCMM** por separado para niñas y niños. Para la muestra de niños y niñas que tienen información sobre abandono y repetición escolar se vuelven a estimar los **LCMM**¹.

Los modelos que consideran únicamente las trayectorias nutricionales, permiten analizar la estructura de grupos y comparar los resultados con los del **JLCMM**. Dado que la muestra en los modelos conjuntos se restringe a los niños

¹ La pregunta sobre repetición y edad de abandono escolar en el estudio **ELBU** fue realizada en la entrevista del 2011/12, por tanto, la muestra para estimar son los niños y niñas que tienen información en la covariable *Repetició*

y niñas que tienen información en la covariable *Repitió*, el análisis de la trayectoria nutricional con el LCMM se vuelve a realizar para esta muestra con la especificación presentada anteriormente para el total de la muestra de niños y niñas.

Los JLCMM tienen las siguientes características:

- trayectoria cuadrática según edad diferenciando coeficientes para cada grupo
- efectos aleatorios en el intercepto y coeficiente lineal de dicha trayectoria
- matriz de covarianzas de efectos aleatorios no diagonal (*unstructured*)
- función de riesgo de línea de base *Weibull* y sin diferenciar entre grupos

La especificación correspondiente a la trayectoria nutricional es la misma que la utilizada para los LCMM. En cuanto al modelo de sobrevivencia se especifica una función de riesgo de línea de base sencilla sin diferencias entre grupos debido a que la convergencia y el tiempo computacional de estimación de estos modelos se compromete con especificaciones más complejas.

El modelo se escribe entonces:

$$\left\{ \begin{array}{l} \pi_{ig} = P(c_i = g | X_{Ci}) = \frac{e^{\lambda g}}{\sum_{l=1}^G e^{\lambda l}} \\ \text{IMC}_{ij|c_i=g} = \beta_{0g} + \beta_{1g}\text{Edad}_{ij} + \beta_{2g}\text{Edad}_{ij}^2 + b_{0ig} + b_{1ig}\text{Edad}_{ij} + \epsilon_{ij} \\ h_i(t|c_i = g) = h_0(t)e^{\{\gamma_{0g} + \gamma_{1g}\text{Repitio}_i\}} \end{array} \right. \quad (3.23)$$

donde la ecuación de pertenencia a las clases y la trayectoria para el IMC son análogas a las del modelo en (3.22) aunque aquí solamente se considera una función de enlace lineal. Por su parte, el modelo de sobrevivencia se estima incluyendo la covariable Repitio_i , $h_0(t)$ es la función de riesgo de línea de base.

Estimación de los Modelos y Valores Iniciales

Como fue mencionado anteriormente, la estimación de los modelos se realiza mediante el método de máxima verosimilitud. Para estimar los modelos se

utilizan las funciones `lcmm` y `jlcmm` de la librería `lcmm` de R. En la estimación de modelos se debe tener en cuenta que la función de verosimilitud posiblemente tenga óptimos locales (Muthén, 2004; Proust-Lima *et al.*, 2016b; Rizopoulos, 2012). Por tanto, se recomienda estimar el modelo varias veces imponiendo diferentes puntos iniciales y estudiar la convergencia de la estimación. Con el objetivo de disminuir la probabilidad de que se identifiquen óptimos locales al estimar los modelos, se realiza un procedimiento propuesto en Biernacki *et al.* (2003).

Por tanto, los resultados finales de la estimación de este trabajo `LCMM` así como `JLCMM` se obtienen luego de estimar 30 modelos que parten de diferentes puntos iniciales. De entre todos ellos, se elige aquél que maximiza la función de verosimilitud y se utilizan sus parámetros estimados como valor inicial de una nueva estimación que es la finalmente considerada para los resultados de este trabajo.

En los `LCMM` los 30 valores iniciales se obtienen a partir de la estimación del modelo análogo con una única clase latente. Estos 30 valores se obtienen a partir de una $N(\hat{\boldsymbol{\theta}}_{G=1}, \hat{V}(\hat{\boldsymbol{\theta}}_{G=1}))$ que corresponde a la distribución asintótica del vector de estimadores para una clase latente (Proust-Lima *et al.*, 2016b).

En el caso de los `JLCMM` los valores iniciales se obtienen a partir de la estimación del modelo `LCMM` estimado para el mismo número de grupos y la misma muestra de niños y niñas. A partir del vector de parámetros estimados para el `LCMM` se construyen otros 30 vectores utilizando la distribución uniforme. Una vez más, se estiman los 30 modelos, se elige el de mayor verosimilitud y, finalmente, se estima una última vez el modelo con valor inicial igual al vector de parámetros del modelo elegido según la verosimilitud. Esta última estimación constituye el resultado final presentado en este trabajo.

La estimación se basa en el método de Máxima verosimilitud y se optimiza mediante un algoritmo iterativo de Marquardt. En el anexo A.3 se presenta con mayor detalle el procedimiento de optimización seguido por este algoritmo que es el implementado en las funciones del paquete `lcmm` utilizado para estimar todos los modelos.

Capítulo 4

Resultados

Este capítulo presenta los principales resultados a los que ha llegado esta investigación. En la sección 4.1 muestra las principales descriptivas de la base de datos utilizada para las estimaciones en cuanto a las variables principales que se utilizarán en los modelos. La sección 4.2 presenta los resultados de los LCMM mientras que la 4.3 los referentes a la asociación entre trayectoria nutricional y desempeño escolar. Por último, la sección 4.4 muestra los resultados del análisis de sensibilidad.

4.1. Estadísticas Descriptivas

A continuación, se presenta una descripción de la información relevante para el análisis contenida en el ELBU. En primer lugar, dado que la cohorte de niños y niñas se conforma con los que estaban en primer año de escuela pública en 2004, las edades entre los participantes del estudio pueden diferir en cada momento de la medición. Además, las medidas antropométricas utilizadas para analizar la situación nutricional de los niños y niñas tienen estándares según la edad al momento de realizar las medidas. Por ello, la Figura 4.1 muestra la distribución según género de las edades simples de los participantes en los años donde se realizaron las medidas antropométricas, es decir en 2004, 2005 y 2009. La pérdida de casos a lo largo del estudio es casi un 35 %, resultando en una pérdida similar según género.

La muestra utilizada para estimar los LCMM y JLCMM es menor al conjunto de niños entrevistados en el ELBU ya que, como fue mencionado ante-

riormente, no se cuenta con información antropométrica para todos los niños y niñas entrevistados (Tabla 4.1). Además, fueron excluidas las ocasiones donde el *z-score* del IMC se encontraba fuera de los límites humanamente posibles, según la OMS. Por último, la muestra para el JLCMM excluye casos que no tengan dato en la covariable *Repetió* en la entrevista realizada en 2011/12.

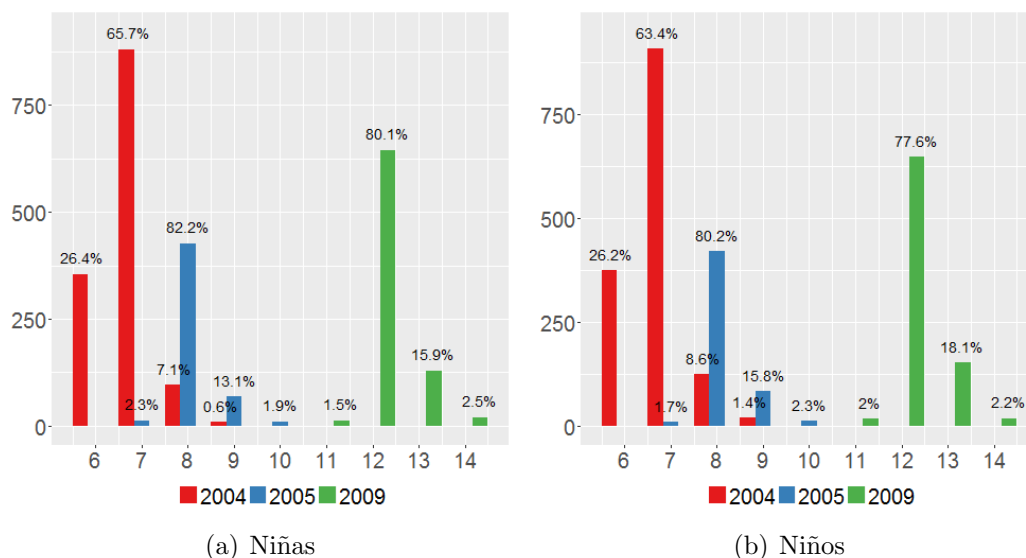
Tabla 4.1: Distribución por género en los años de medición

	Niñas	Niños
Medidas en 2004 ^a	1341	1432
Medidas en 2005 ^a	518	525
Medidas en 2009 ^a	805	835
Total participantes ELBU	1570	1691
Total muestra LCMM	1392	1492
Total muestra JLCMM	971	1026

Fuente: Elaboración propia

^a Cantidad de medidas válidas según OMS

Figura 4.1: Distribución según edades simples en los años de medición



Nota: Porcentaje de niños o niñas en esa edad del total de medidos ese año.

La Figura 4.1 muestra la distribución según edades simples de niños y niñas en los años donde fueron tomadas las medidas antropométricas que se utilizan para calcular el IMC. Se observa que en 2004, donde se tienen medidas de 1341 niñas y 1432 niños, más del 90 % tenía entre 6 y 7 años. En la segunda ronda de medición, realizada en 2005 únicamente para las escuelas de Montevideo y

el área metropolitana, más del 90 % tenía entre 8 y 9 años. Por último, en 2009 más del 90 % tenía entre 12 y 13 años.

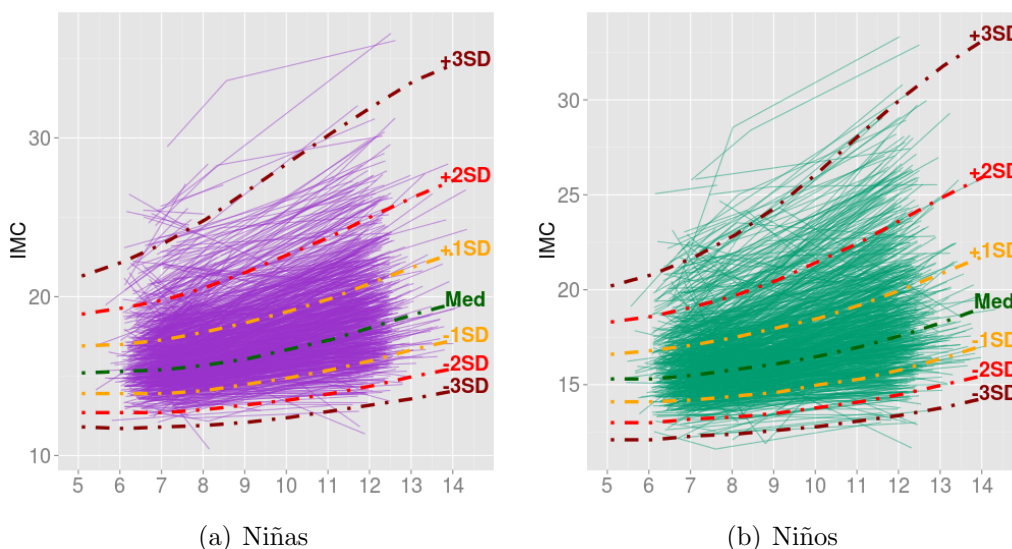
Tabla 4.2: Descriptivas del IMC en los años de medición

	Niñas			Niños		
	2004	2005	2009	2004	2005	2009
$z < -2^a$	0.8 %	1.0 %	2.4 %	1.5 %	1.5 %	3.4 %
$-2 < z < -1^b$	9.8 %	6.8 %	11.0 %	9.3 %	7.3 %	12.1 %
$-1 < z < 1^b$	65.9 %	61.5 %	58.5 %	66.7 %	63.0 %	52.3 %
$1 < z < 2^c$	16.5 %	21.2 %	21.1 %	14.0 %	17.2 %	21.3 %
$2 < z^d$	7.0 %	9.5 %	7.0 %	8.5 %	11 %	10.9 %
Media (SD)	16.1 (2.2)	17.0 (2.8)	19.5 (3.5)	16.1 (1.9)	16.8 (2.4)	19.1 (3.4)
N	1341	518	805	1432	525	835

Fuente: Elaboración propia

^a Bajo peso ^b Normal ^c Sobrepeso ^d Obesidad

Figura 4.2: Trayectorias del IMC para la edad según género



Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$, $\pm 2SD$ y $\pm 3SD$)

La Tabla 4.2 permite visualizar la proporción de niños y niñas en los distintos umbrales definidos por el *z-score* para la edad del IMC. Se observa que en 2004, más del 20 % presenta problemas de sobrepeso u obesidad, mientras que aproximadamente un 1 % tiene bajo peso tanto para niñas como niños. A su vez, en las siguientes olas del estudio el porcentaje de niñas y niños con bajo peso se incrementa y, al mismo tiempo lo hace el de sobrepeso. En efecto, en 2009 hay una mayor proporción de niños y niñas con problemas nutricionales

Tabla 4.3: Tasa de abandono según repetición por género

	No repitió	Repitió	Total
Niñas	2.4 %	16.6 %	7.3 %
Niños	3.8 %	19.6 %	10.9 %
Total	3.1 %	18.4 %	9.1 %

Fuente: Elaboración propia

de bajo peso y de sobrepeso, según el *z-score* para la edad. Por último, la Figura 4.2 muestra las trayectorias del IMC para la edad según género.

Por otro lado, se presenta la tasa de abandono de niños y niñas del ELBU según si repitieron o no algún año del ciclo educativo previo a la entrevista realizada en 2011/12. La Tabla 4.3 muestra la tasa de abandono reportada en 2011/12 donde para los participantes del ELBU asciende a 9 %, cuando más del 90 % tenía entre 13 y 14 años. Este valor resulta superior al que surge de la ECH para 2011 -6 % aproximadamente- en el mismo tramo etario. Además, resulta notable la diferencia en esta tasa según si experimentaron repeticiones previas ya que abandona un 3 % de los que no repitieron mientras que el abandono representa el 18.4 % de los que repitieron. Por su parte, las tasas de abandono también presentan diferencias según género y éstas se acentúan al considerar los episodios de repetición previa de los individuos.

Tabla 4.4: Distribución de la tasa de abandono según edad por género

Edad	11 años	12 años	13 años	14 años	15 años
Niñas	5.5 %	17.3 %	37.3 %	34.6 %	5.3 %
Niños	1.7 %	19.5 %	35.6 %	40.7 %	2.5 %
Total	3.1 %	18.7 %	36.3 %	38.3 %	3.6 %

Fuente: Elaboración propia

Por último, la Tabla 4.4 muestra distribución de las niñas y niños que declaran haber abandonado los estudios en 2011/12 según edades simples. Del total de abandonos, más del 70 % de ellos se dan a los 13 y 14 años de edad tanto para niñas como niños. Se aprecia que el porcentaje de abandonos para 15 años es muy bajo lo que se debe fundamentalmente a que el ELBU es una cohorte no representativa de este tramo etario¹.

¹ Si se calcula la tasa de no asistencia con la ECH 2011 para 15 años es del entorno del

4.2. Análisis de la Trayectoria Nutricional

En esta sección se presentan los resultados obtenidos en la estimación de los modelos [LCMM](#) que permiten analizar la trayectoria nutricional de niños y niñas en el rango de edad considerado.

En primer lugar, se detallan los resultados relativos a la elección del número de grupos según la trayectoria nutricional. Se estiman los modelos, por separado para niñas y niños, con función de enlace lineal y con una *splines* cuadrática en el modelo de medición. En todos los casos, se fija el número de grupos de uno a cinco para estimar los modelos y se comparan indicadores de ajuste y las probabilidades *a posteriori* detallados en la sección [3.4](#). Se considera el porcentaje de casos en cada grupo y se sigue un criterio de parsimonia para la elección del número de grupos de niños y niñas. Este criterio refiere a que ante modelos estimados que arrojan resultados similares en términos de ajuste y distribución de los individuos entre grupos, se elige el más simple con menor número de parámetros a estimar.

Por último, una vez elegido el número de grupos se analizan diferencias entre los niños y niñas pertenecientes a diferentes grupos, en algunas características asociadas al estado nutricional según se ha reportado en otros trabajos.

Como criterio general para la presentación de los resultados, los grupos se etiquetan de forma ascendente en función del tamaño que tienen, una vez realizada la clasificación de los individuos en cada uno de ellos. De esta forma, “G 1” es el grupo con mayor porcentaje de niñas o niños del total de la muestra, “G 2” es el segundo grupo en términos de este porcentaje y así sucesivamente con los restantes.

Cabe destacar que los resultados de las estimaciones se obtienen luego de realizar 30 réplicas con distintos puntos iniciales, para disminuir la probabilidad de identificar óptimos locales, según se expuso en la sección [3.5.3](#).

Elección del Número de Grupos para Niñas

Los resultados con función de enlace lineal para niñas (Tabla [4.5](#)) muestran que al aumentar el número de grupos hasta tres disminuye el [BIC](#), luego

10%.

aumenta para cuatro y baja nuevamente para cinco. Esto indica que, en términos de ajuste, puede resultar adecuado clasificar a las niñas en tres grupos. Por su parte, la entropía para esta clasificación es 0.513, indicando que esta separación de niñas es medianamente nítida (Muthén, 2004). Las trayectorias nutricionales de las niñas según estos tres grupos puede observarse en la Figura 4.3 panel *b* donde se aprecia cierta superposición de las trayectorias de niñas en distintos grupos.

La Tabla 4.5 también contiene el porcentaje de niñas clasificadas posteriormente en cada grupo. Para dos grupos uno contiene más del 80 % de las niñas, para tres grupos los dos más grandes contienen un porcentaje de niñas similar y el más pequeño no alcanza el 10 %. Luego, para cuatro y cinco grupos la distribución de niñas entre ellos es muy similar, ya que el quinto grupo contiene menos del 2 % de las niñas.

Tabla 4.5: Estadísticos y clasificación según grupos de niñas - LCMM - Lineal

G	loglik ^a	BIC	Entropy	NP ^b	% Niñas en cada grupo				
					G 1	G 2	G 3	G 4	G 5
1	-5775.9	11602.5		7	100 %				
2	-5524.2	11135.4	0.644	12	86 %	14 %			
3	-5471.4	11065.9	0.513	17	49 %	42 %	9 %		
4	-5457.3	11073.9	0.564	22	57 %	32 %	8 %	3 %	
5	-5409.2	11013.8	0.538	27	51 %	34 %	8 %	5 %	2 %
N=1392									

Fuente: Elaboración propia

^a Logaritmo de la verosimilitud ^b Número de parámetros estimados

También debe observarse la Tabla de clasificación ya que aporta información acerca de la capacidad de discriminación del número de grupos. La Tabla 4.6 contiene información respecto a la clasificación para tres grupos: probabilidad media *a posteriori* (panel *a*) y porcentaje de niñas que superan cierto umbral en su probabilidad máxima (panel *b*). En el panel *a*, se observa que, dado el grupo al que pertenece la niña, la probabilidad de ser clasificada en él supera para todos los grupos 0.74 en promedio. Esto es, la diagonal de la Tabla de clasificación muestra sustento en la discriminación que se logra con tres grupos de niñas.

Por otro lado, en el panel *b* se observa que el “G 3” tiene más de la mitad de las niñas con probabilidad máxima superior a 0.9, lo que indica clara per-

Tabla 4.6: Probabilidades *a posteriori* para tres grupos de niñas - Lineal

	(a) Prob. media			(b) % >umbral ^a		
	G 1	G 2	G 3	> 0.7	> 0.8	> 0.9
G 1	0.745	0.239	0.016	69 %	37 %	2 %
G 2	0.134	0.785	0.081	70 %	55 %	27 %
G 3	0.005	0.138	0.857	78 %	69 %	56 %
N=1392						

Fuente: Elaboración propia

^a % de niñas con probabilidad *a posteriori* que supera el umbral

tenencia a este grupo. Para “G 2” la mitad de las niñas superan 0.8 mientras que en el “G 1” aproximadamente el 70 % supera 0.7 de probabilidad, si bien hay solo un 2 % que supera el 0.9 en este grupo. Por tanto, si bien la entropía es intermedia para la clasificación en tres grupos de niñas el análisis de las probabilidades *a posteriori* muestra que tres grupos de niñas logra una buena discriminación.

Por otro lado, en las Figuras 4.3 y 4.4 pueden verse las trayectorias nutricionales de cada niña según el grupo. Estas trayectorias se representan junto con los umbrales del IMC para la edad que proporciona la OMS, estos son, la mediana del IMC para la edad (*Med*) y los desvíos estándar ($\pm 1SD$, $\pm 2SD$ y $\pm 3SD$)¹.

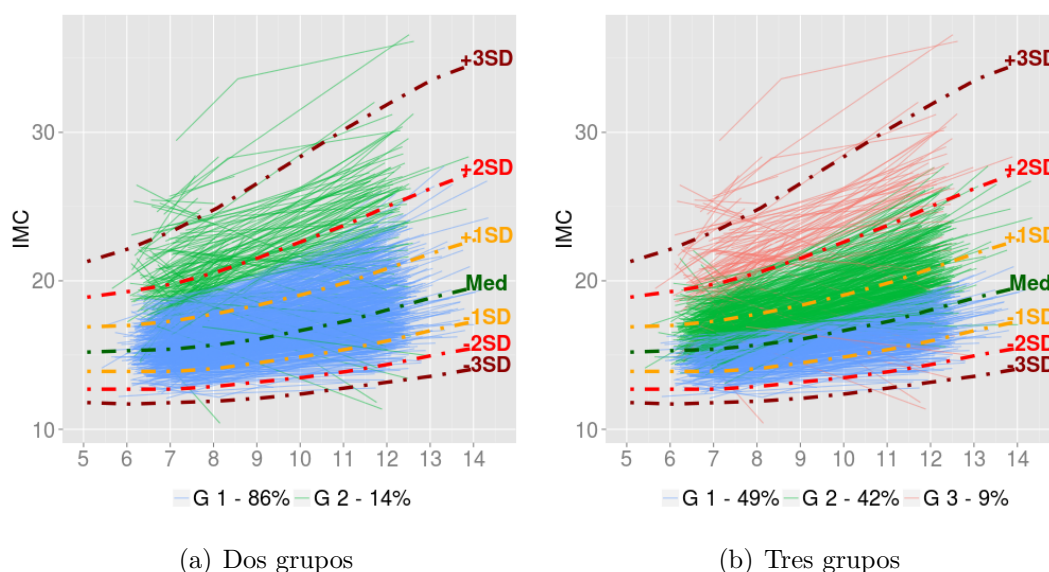
Como ya fue mencionado, existe cierta superposición de las trayectorias de las niñas pertenecientes a los diferentes grupos. Sin embargo, en todas las clasificaciones, el grupo con mayor cantidad de niñas contiene las trayectorias en los niveles más bajos del IMC para la edad, respecto a los restantes grupos. A su vez, para la clasificación en tres, cuatro y cinco grupos se observa que los tres primeros grupos (“G 1”, “G 2” y “G 3”) son similares en términos del porcentaje de niñas que lo integran y los niveles de IMC que representan. En este sentido, el “G 1” tiene los niveles más bajos del IMC respecto al resto de grupos. El “G 2” tiene niveles superiores de IMC que el “G 1”. Por último, el “G 3” con niveles altos del IMC o bien trayectorias que decrecen en el rango

¹ Los umbrales se obtienen del Programa *Growth reference 5-19 years* disponibles en la Web de la OMS. La interpretación de estos valores es que un IMC $>1SD$ (equivalente a un IMC de 25 kg/m² a los 19 años) corresponde a problemas de sobrepeso y si es $>2SD$ (equivalente a un IMC de 30 kg/m² a los 19 años) el problema es obesidad. Por su parte, un IMC $<-2SD$ conlleva problemas de desnutrición (*thinnes*) en tanto si el índice es $<-3SD$ el problema de desnutrición es severo.

de edad considerado y con poca proporción de niñas en él.

La clasificación en cuatro o cinco grupos no parece aportar en la separación de las trayectorias nutricionales de las niñas ya que los grupos “G 4” y “G 5” representan una proporción muy baja de niñas (apéndice A.1)¹.

Figura 4.3: Trayectoria nutricional según dos y tres grupos de niñas - Enlace Lineal



Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$, $\pm 2SD$ y $\pm 3SD$)

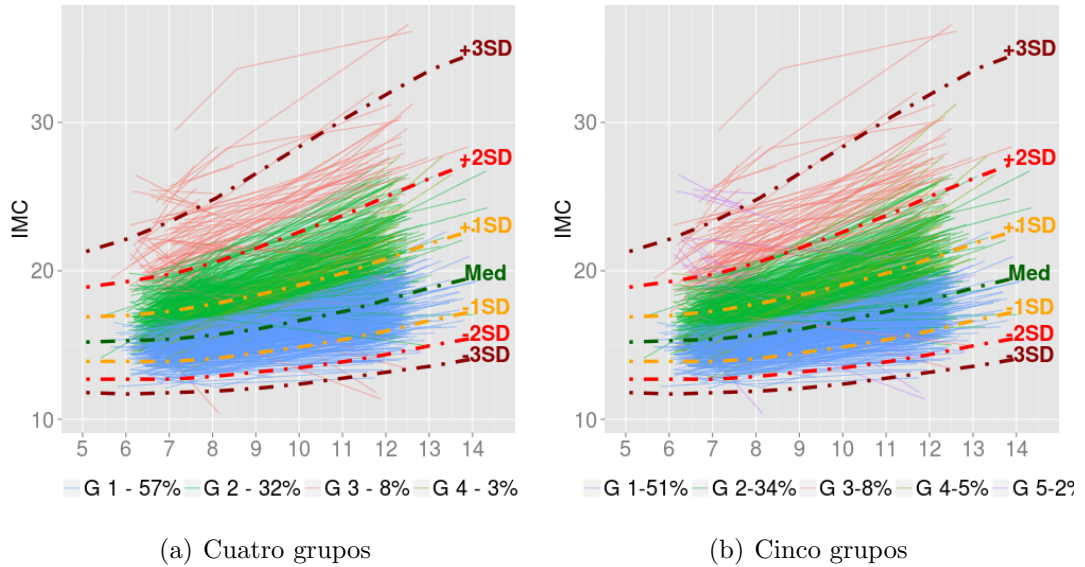
Por su parte, los resultados utilizando como función de enlace *spline* cuadrática no arrojan diferencias sustantivas con los obtenidos utilizando la función lineal (apéndice A.3). Sin embargo, con la *spline* el ajuste mejora ya que el BIC es menor al obtenido con la función lineal dado el mismo número de grupos.

En la especificación con *splines* se debe tener en cuenta que el modelo con cuatro grupos no converge², por tanto según el criterio de BIC tres grupos sería adecuado. Si bien luego con cinco vuelve a disminuir la diferencia entre los respectivos valores del BIC es muy pequeña. La entropía para la clasificación de niñas en tres grupos con *splines* es 0.582, levemente mayor a la obtenida

¹ En el apéndice A.1 se presentan los gráficos que separan las trayectorias según grupos para observar más claramente las diferencias.

² La no convergencia puede deberse a que la función *splines* si bien logra mayor ajuste del modelo aumenta su complejidad ya que se incrementan los parámetros a estimar o a la no identificación del modelo, en el sentido de que no habrían cuatro grupos claramente definidos para niñas

Figura 4.4: Trayectoria nutricional según cuatro y cinco grupos de niñas - Enlace Lineal



Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$, $\pm 2SD$ y $\pm 3SD$)

con función lineal (0.513) pero continúa siendo un valor medio respecto a las recomendaciones prácticas.

Incorporar esta función *splines* hace más flexible el modelo, ya que permite separar el estructural del correspondiente a la medida longitudinal observada, pero el aumento de los parámetros compromete la estimación, su interpretación y los resultados no aportan diferencias sustanciales respecto a la especificación con una función lineal.

En conclusión, del análisis previo se desprende que la clasificación de las niñas según sus trayectorias nutricionales en tres grupos resulta adecuada y el modelo especificado con una función de enlace lineal logra un ajuste adecuado.

Elección del Número de Grupos para Niños

Los resultados con función de enlace lineal para niños (Tabla 4.7) muestran que, en términos de ajuste resulta adecuado clasificarlos en tres grupos, ya que aumentar su número disminuye el BIC hasta tres, para cuatro permanece constante y baja nuevamente para cinco, aunque esta disminución es muy pequeña. Por su parte, la entropía para tres grupos es 0.556 mostrando que la

separación es medianamente nítida. La forma en que quedan clasificadas las trayectorias de los niños en tres grupos se observa en la Figura 4.5 panel b.

Por otro lado, la Tabla 4.7 muestra el porcentaje que queda clasificado *a posteriori* en cada grupo: para dos uno contiene casi el 80 % de los niños, para tres grupos uno de ellos tiene el 70 % mientras que los otros 21 y 8 %. Luego, con cuatro se tienen dos grandes grupos (40 y 34 %) y otros dos más pequeños (17 y 8 %). Por último, la estructura con cinco grupos es similar a la de cuatro ya que el quinto contiene menos del 1 % de los niños.

Tabla 4.7: Estadísticos y clasificación según grupos de niños - LCMM - Lineal

G	loglik ^a	BIC	Entropy	NP ^b	% Niños en cada grupo				
					G 1	G 2	G 3	G 4	G 5
1	-5915.3	11881.7		7	100 %				
2	-5622.6	11332.8	0.535	12	79 %	21 %			
3	-5530.2	11184.7	0.556	17	71 %	21 %	8 %		
4	-5511.9	11184.7	0.456	22	41 %	34 %	17 %	8 %	
5	-5446.4	11090.1	0.516	27	43 %	28 %	20 %	8 %	1 %

N=1492

Fuente: Elaboración propia

^a Logaritmo de la verosimilitud ^b Número de parámetros estimados

Tabla 4.8: Probabilidades *a posteriori* para tres grupos de niños - Lineal

	(a) Prob. media			(b) % >umbral ^a		
	G 1	G 2	G 3	> 0.7	> 0.8	> 0.9
G 1	0.823	0.022	0.154	85 %	66 %	33 %
G 2	0.018	0.826	0.155	69 %	60 %	47 %
G 3	0.146	0.112	0.742	56 %	40 %	24 %

N=1492

Fuente: Elaboración propia

^a % de niños con probabilidad *a posteriori* que supera el umbral

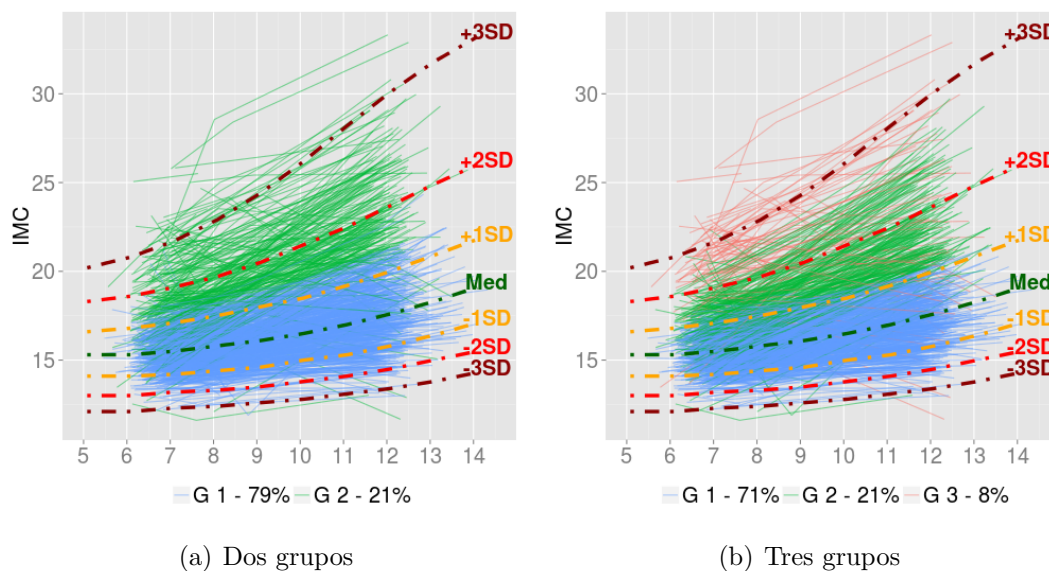
Para indagar en la calidad de clasificación que se tiene con tres grupos, se estudian las probabilidades *a posteriori*. La Tabla 4.8 muestra la probabilidad media de pertenencia a cada grupo (panel a) y el porcentaje de niños cuya probabilidad máxima supera cierto umbral (panel b). La probabilidad de ser clasificados en su propio grupo supera en promedio 0.74 para todos los grupos, por tanto, la diagonal de la Tabla de clasificación contiene valores cercanos a

uno y sustenta una buena discriminación de los niños en esos grupos latentes. Por su parte, se observa que el “G 1” y “G 2” tienen aproximadamente un 60% de los niños con probabilidad máxima mayor a 0.8 en tanto que el “G 3” alcanza un 56% de los niños con probabilidad máxima superior a 0.7. Por lo tanto, al considerar las probabilidades *a posteriori*, también es adecuada la clasificación en tres grupos de niños, con un índice de entropía en el entorno de 0.6.

A su vez, en las Figuras 4.5 y 4.6 se puede observar la trayectoria nutricional para cada niño según el grupo en que quedó clasificado, como ya fue mencionado, se aprecia cierta superposición de las trayectorias entre distintos grupos. A diferencia de lo que ocurría para las niñas, ahora los grupos mayoritarios de niños no siempre tienen menores niveles de IMC respecto a los restantes grupos. Esto último se cumple para la clasificación de niños en dos y tres grupos pero no para la clasificación en cuatro y cinco.

Cuando se clasifica en cuatro grupos las trayectorias con nivel de IMC más bajo son las del “G 2” (34%), le supera en nivel el “G 1” (41%) y luego el “G 3” (17%) y “G 4” (8%).

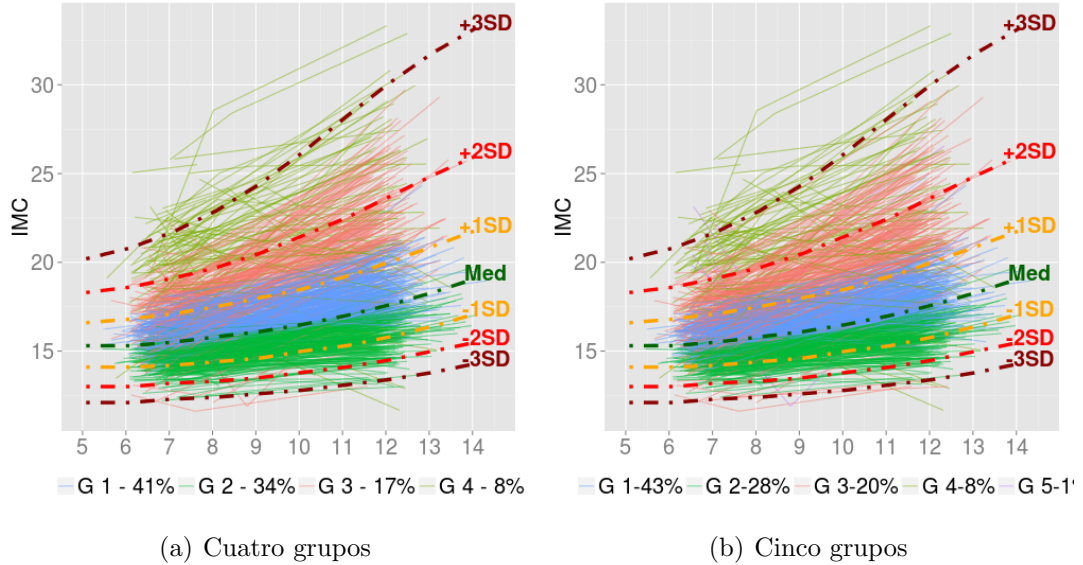
Figura 4.5: Trayectoria nutricional según dos y tres grupos de niños - Enlace Lineal



Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$, $\pm 2SD$ y $\pm 3SD$)

Por su parte, en la clasificación en cinco grupos los cuatro primeros tienen trayectorias similares a los de la clasificación en cuatro grupos, ya que el quinto

Figura 4.6: Trayectoria nutricional según cuatro y cinco grupos de niños - Enlace Lineal



Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$, $\pm 2SD$ y $\pm 3SD$)

contiene menos del 1% de los niños con trayectorias “raras” (apéndice A.2)¹.

Por otro lado, los resultados utilizando una función de enlace *splines* cuadrática (apéndice A.4) son similares respecto a los obtenidos utilizando la función lineal y, para el mismo número de grupos, el BIC es menor con *splines*.

Cuando se introduce una función *splines*, el modelo con cinco grupos no converge² y el BIC se minimiza en la especificación con cuatro grupos. La distribución de niños *a posteriori* con esta clasificación es similar a la obtenida con tres grupos, ya que el cuarto grupo contiene solamente un 1% de los niños. El valor de la entropía para cuatro grupos es 0.504, situándose en valores medios respecto a las recomendaciones prácticas, mayor al valor para cuatro grupos con función lineal (0.456), pero menor que la obtenida con tres grupos (0.541 con *splines* y 0.556 con lineal).

Al comparar las trayectorias de los niños en función de los grupos obtenidos

¹ En el apéndice A.2 se presentan estos mismos gráficos separando las trayectorias según grupos por paneles para observar este tipo de comportamientos.

² Las razones de la no convergencia son similares a las expuestas para las niñas. Una razón es el aumento del número de parámetros y además un problema de identificación del modelo, ya que puede exigirse identificar un número de grupos que podría ser no adecuado para clasificar a los niños.

con enlace lineal o *splines* (apéndice A.2), la clasificación en cinco grupos separa en el grupo adicional, un conjunto de trayectorias “raras” (1%). La clasificación en cuatro grupos con *splines*, muestra uno con trayectorias en forma de “V”, al igual que las clasificaciones en cinco grupos con función lineal o *splines*. Al comparar las trayectorias de la clasificación en cuatro grupos con función lineal y con *splines* se observa que en esta última se separa un cuarto grupo con trayectorias en forma de “V” y los grupos “G 1” y “G 2” del caso lineal parecen fusionarse en el “G 1” del caso con *splines* (apéndice A.2).

Como ya fue mencionado, incorporar la función *splines* hace más flexible el modelo pero compromete la estimación y su interpretación. Los resultados, presentan leves diferencias en el agrupamiento que resulta de aplicar la función lineal o la *splines*, pero en las clasificaciones con cuatro y cinco grupos. En particular, el grupo de niños que presenta trayectorias nutricionales en forma de “V” no se identifica de igual forma con ambas funciones. Sin embargo, se considera un ajuste adecuado con tres grupos de niños según medidas de bondad de ajuste y, además, las probabilidades *a posteriori* arrojan una estructura adecuada para tres grupos, tanto para el caso lineal como *splines* (apéndice A.3).

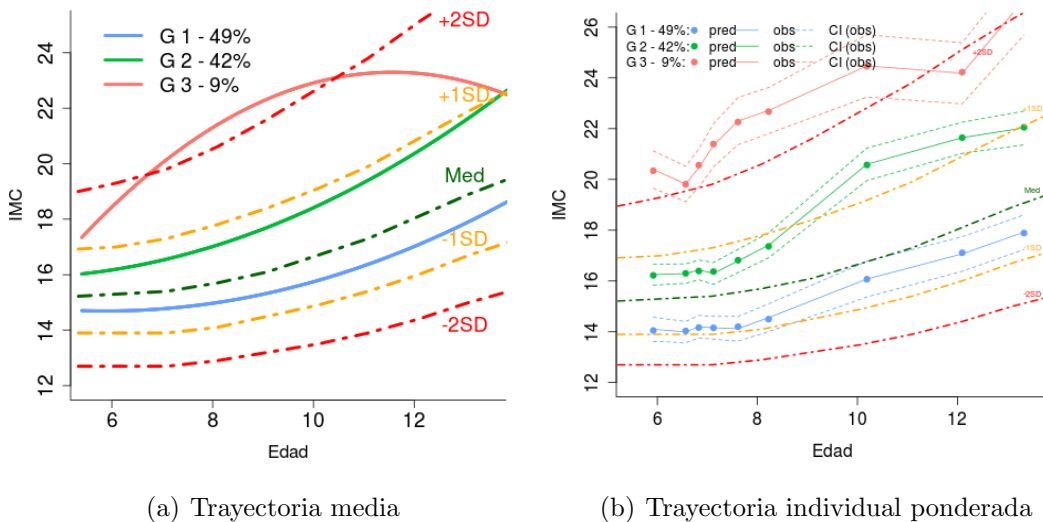
En definitiva, para el análisis de la trayectoria nutricional de los niños se opta por una clasificación en tres grupos que resulta adecuada, en términos de ajuste y probabilidades *a posteriori*, y se presentan los resultados del modelo especificado con función de enlace lineal.

4.2.1. Trayectoria Nutricional según Grupos de Niñas

En esta sección se muestran características de las niñas según el grupo en el que fueron clasificadas. El objetivo es caracterizar los grupos anteriormente identificados, según factores que se asocian con el estado nutricional. Para ello, se presentan las trayectorias predichas para cada grupo de niñas y niños. A su vez, se muestran descriptivas sobre el peso y la talla al nacer y educación materna de las niñas y niños según el grupo al que pertenecen.

En la Figura 4.7 se presentan las trayectorias medias (panel *a*) para cada grupo de niñas y específicas de los individuos ponderadas¹ con sus intervalos de confianza (panel *b*) y los umbrales de la OMS. Se observa que el “G 1” tiene una trayectoria creciente por debajo de la mediana. El “G 2” presenta una trayectoria en niveles superiores, que sobrepasa en algún tramo el umbral de normalidad (+1SD) y que crece con tendencia similar a la del “G 1”. El “G 3” tiene trayectoria en forma de “U” invertida y por encima del límite de sobrepeso (+1SD) e incluso obesidad (+2SD).

Figura 4.7: Trayectoria nutricional media e individual según grupos de niñas



Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$ y $\pm 2SD$)

Por tanto, se identifican dos grupos que representan el 90 % de las niñas que en general transitan dentro de los umbrales de normalidad: el grupo normal-bajo, con casi 50 % de las niñas, tiene una trayectoria por debajo de la mediana;

¹ La forma de obtener estas predicciones se presenta en la sección 3.4.3.

el normal-alto, con 40 %, transita por encima de la mediana y, en cierto rango de edad, supera el umbral de sobrepeso (+1SD). El último grupo con sobrepeso, representa menos del 10 % de niñas, sigue una trayectoria en forma de “U” invertida que se sitúa por encima del umbral de sobrepeso (+1SD) e incluso obesidad (+2SD).

Tabla 4.9: Descriptivas para tres grupos de niñas - LCMM - Enlace Lineal

	Normal-bajo	Normal-alto	Sobrepeso	Total
% niñas	49 %	42 %	9 %	
N	679	588	125	1392
Educación materna				Total
Primaria	48 %	46 %	44 %	47 %
Secundaria	45 %	47 %	49 %	46 %
Terciaria	6 %	7 %	8 %	7 %
N	569	558	84	1211
$\chi^2=1.4$ df=9 p=0.99				
Peso al nacer (g)				Total
Media (SD)	3111 (570)	3246 (557)	3329 (542)	3188 (567)
N	635	556	115	1306
ANOVA ^a F=12.6 df=(2,1303) p=0.00***				
Talla al nacer (cm)				Total
Media (SD)	48.4 (4.3)	48.9 (3.4)	48.9 (2.7)	48.7 (3.8)
N	452	420	94	966
ANOVA ^b F=2.4 df=(2,963) p=0.08*				

* p<0.10 ** p<0.05 *** p<0.01

^a Contraste igualdad de varianzas (Bartlett's), no rechaza que las varianzas son iguales ($K^2=0.61$ df=2 p=0.74)

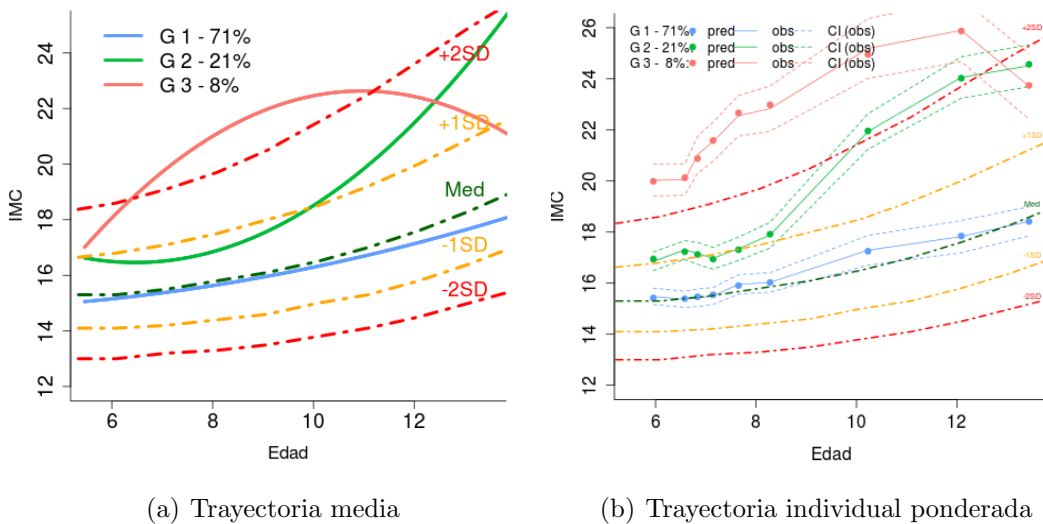
^b Contraste igualdad de varianzas (Bartlett's), rechaza que las varianzas son iguales ($K^2=42.51$ df=2 p=0.00***)

Además, se indaga en las posibles diferencias entre los grupos según nivel de educación materna, el peso y la talla al nacer de las niñas. Respecto a la educación materna, no se observan diferencias significativas entre los tres grupos de niñas (Tabla 4.9). En cuanto al peso y la talla al nacer, sí se aprecian diferencias significativas entre los grupos. Por tanto, el peso al nacer de las niñas del grupo normal-bajo en promedio menor a las del grupo normal-alto y éste, a su vez, menor en promedio a las del grupo de sobrepeso. En tanto que, la talla al nacer es en promedio más alta para el grupo de sobrepeso, le sigue el grupo de normal-alto y el normal-bajo (Tabla 4.9).

4.2.2. Trayectoria Nutricional según Grupos de Niños

En cuanto al análisis de las características para los niños la Figura 4.8 muestra las trayectorias medias (panel *a*) para cada grupo de niños y específicas de los individuos ponderadas¹ con sus intervalos de confianza (panel *b*) y los umbrales de la OMS. El “G 1” tiene una trayectoria creciente que transita entorno a la mediana, el “G 2” con trayectoria creciente pero pendiente más inclinada que el “G 1” transita en los niveles normales y por encima del umbral de sobrepeso (+1SD) y, el “G 3” tiene trayectoria en forma de “U” invertida por encima del nivel de obesidad (+2SD).

Figura 4.8: Trayectoria nutricional media e individual según grupos de niños



Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$ y $\pm 2SD$)

Por tanto, se identifica un grupo normal, que representa más del 70 % de los niños que transita dentro de los umbrales de normalidad con trayectoria creciente entorno a la mediana. Un grupo normal-sobrepeso con el 20 % de los niños, con crecimiento más pronunciado y por encima del umbral de sobrepeso para algún tramo de edad. Un pequeño grupo con obesidad que contiene el 8 % de los niños.

En cuanto a las características de estos grupos de niños, existen diferencias en los niveles de educación materna (significativas al 10 %). En este sentido,

¹ La forma de obtener estas predicciones se presenta en la sección 3.4.3.

los niños del grupo normal tienen más proporción de madres con nivel educativo hasta primaria que los niños del grupo normal-sobrepeso y con obesidad. Mientras tanto, el grupo con obesidad tiene mayor proporción de niños con madres de nivel terciario que los otros dos grupos (Tabla 4.10).

Por otro lado, en cuanto al peso y la talla al nacer de los niños existe una diferencia significativa en el peso medio al nacer entre los grupos aunque la talla no presenta diferencias significativas. Por tanto, el peso al nacer de los niños del grupo normal es en promedio menor al de los grupos normal-sobrepeso y obesidad, mientras que la talla al nacer promedio entre los grupos no difiere de forma significativa (Tabla 4.10).

Tabla 4.10: Descriptivas para tres grupos de niños - LCMM - Enlace Lineal

	Normal	Normal-Sobrepeso	Obesidad	Total
% niños	71 %	21 %	8 %	
N	1064	310	118	1492
Educación materna				Total
Primaria	52 %	44 %	39 %	49 %
Secundaria	44 %	47 %	51 %	45 %
Terciaria	4 %	9 %	10 %	6 %
N	938	270	104	1312
$\chi^2=16.1$ df=9 p=0.06*				
Peso al nacer (g)				Total
Media (SD)	3306 (574)	3372 (645)	3427 (578)	3330 (590)
N	999	295	113	1407
ANOVA ^a F=3.1 df=(2,1404) p=0.04**				
Talla al nacer (cm)				Total
Media (SD)	49.4 (4.0)	49.9 (3.3)	49.9 (3.5)	49.6 (3.9)
N	713	241	90	1044
ANOVA ^b F=2.1 df=(2,1041) p=0.12				

* p<0.10 ** p<0.05 *** p<0.01

^a Contraste igualdad de varianzas (Bartlett's), rechaza que las varianzas son iguales ($K^2=6.5$ df=2 p=0.04**)

^b Contraste igualdad de varianzas (Bartlett's), rechaza que las varianzas son iguales ($K^2=14.8$ df=2 p=0.00***)

4.3. Asociación entre Trayectoria Nutricional y Desempeño Escolar

En esta sección se presentan los resultados del estudio de la vinculación entre los grupos según trayectoria nutricional y la edad de abandono escolar. Para ello se estiman modelos conjuntos de clases latentes especificando de uno a cinco el número de grupos, igual que en la sección 4.2.

Como fue mencionado anteriormente, en el modelo de sobrevivencia de los JLCMM se incluye como covariable si el niño o niña repitió¹ algún año. Esto hace que la muestra de niños y niñas que se emplea para el JLCMM no sea la misma que la utilizada para los LCMM en la sección 4.2. Por tanto, para la elección del número de grupos y la comparación entre el JLCMM y el LCMM se estiman los LCMM considerando la misma muestra de niños y niñas que para los JLCMM.

Se realiza un análisis análogo al de la sección 4.2 para elegir el número adecuado de grupos. Es decir, se estiman los modelos fijando el número de grupos de uno a cinco y se comparan indicadores de bondad de ajuste, las probabilidades *a posteriori* y el porcentaje de casos en cada grupo. También se sigue un criterio de parsimonia para la elección del número de grupos una vez estimados los modelos conjuntos para niños y niñas. Además, se presentan resultados de los JLCMM que permiten analizar la relación entre desempeño escolar y trayectoria nutricional. A su vez, se comparan los resultados de los LCMM con JLCMM, particularmente respecto al número de grupos que mejor identifica trayectorias nutricionales de niños y niñas.

Por último, se destaca que los resultados de las estimaciones tanto para los modelos LCMM como los JLCMM de esta sección, se obtienen luego de realizar 30 réplicas con distintos puntos iniciales, para disminuir la probabilidad de convergencia a óptimos locales, según se expuso en la sección 3.5.3.

Elección del Número de Grupos para Niñas

En primer lugar, se estiman modelos LCMM para las niñas con la misma muestra utilizada para los JLCMM. De esta forma, se pueden comparar los

¹ La variable incluida *Repitió* toma el valor uno si el niño tuvo algún episodio de repetición durante el ciclo escolar y cero en caso contrario.

resultados entre ambos modelos. En la Tabla 4.11, se observa que el BIC se minimiza con cinco grupos de niñas, sin embargo, el indicador no es tanto más bajo que el obtenido con cuatro. Además, la entropía es mayor con cuatro grupos, lo que indica una mejor separación de la clasificación en cuatro que cinco.

Por su parte, con tres grupos de trayectorias para niñas se logra una clasificación muy parecida a la de cuatro y cinco grupos, en términos de la proporción de niñas clasificadas *a posteriori* en cada grupo. En este sentido, con cuatro y cinco grupos existe uno que contiene muy pocas niñas (solamente un 1 % de la muestra) mientras que los tres grupos con mayor proporción de niñas tienen una distribución similar en las tres clasificaciones (tres, cuatro y cinco grupos). Este pequeño grupo con 1 % de las niñas para las estructuras de cuatro y cinco grupos, parece aislar algunas trayectorias nutricionales “raras”, por lo que no aportaría en términos del objetivo de este trabajo.

Tabla 4.11: Estadísticos y clasificación según grupos de niñas - LCMM

G	loglik ^a	BIC	Entropy	NP ^b	% Niñas en cada grupo				
					G 1	G 2	G 3	G 4	G 5
1	-4519.4	9086.9		7	100 %				
2	-4339.9	8762.4	0.691	12	88 %	12 %			
3	-4295.6	8708.1	0.541	17	46 %	45 %	9 %		
4	-4263.5	8678.3	0.610	22	49 %	40 %	10 %	1 %	
5	-4245.7	8677.2	0.537	27	46 %	31 %	12 %	10 %	1 %

N=971

Fuente: Elaboración propia

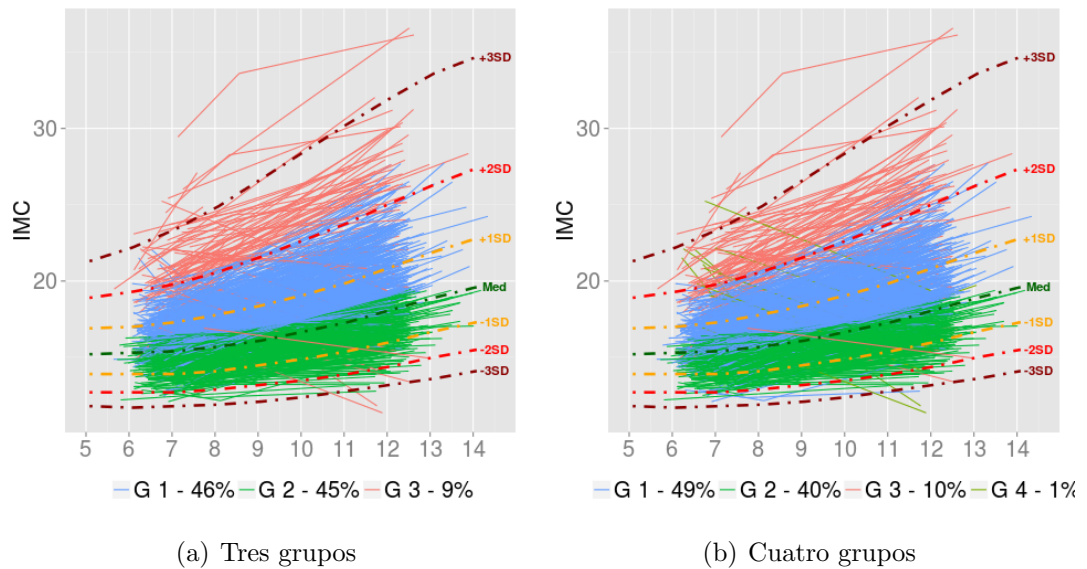
^a Logaritmo de la verosimilitud ^b Número de parámetros estimados

Para decidir entonces la estructura de grupos se observan, además, las trayectorias según la clasificación en los grupos *a posteriori* (Tabla 4.9). Si bien el BIC se minimiza con cinco esta estructura se descarta ya que no aporta diferencias sustanciales con la estructura de cuatro grupos.

En la Figura 4.9, se observa que las trayectorias de los tres primeros grupos (99 % de niñas) son similares en las estructuras de tres y cuatro grupos, donde el cuarto grupo (1 % de niñas) separa trayectorias “raras”. En consecuencia, tomando en consideración las medidas de ajuste (BIC y entropía), la proporción de niñas clasificadas en los grupos según las distintas estructuras

propuestas y las trayectorias nutricionales según la clasificación realizada, resulta adecuada una estructura de tres grupos para analizar la relación entre trayectoria nutricional y abandono escolar para las niñas.

Figura 4.9: Trayectoria nutricional según tres y cuatro grupos de niñas - LCMM



Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$, $\pm 2SD$ y $\pm 3SD$)

Elección del Número de Grupos para Niños

Al igual que para las niñas, se estiman modelos **LCMM** para la elección del número de grupos para los niños sobre la muestra para los que se estiman los modelos conjuntos. También se comparan los resultados de los **LCMM** y **JLCMM** para esta muestra de niños.

La Tabla 4.12 muestra los resultados de la estimación para distinta cantidad de grupos. El modelo con cinco grupos no converge por tanto no se considera. Entonces, el **BIC** se minimiza con cuatro grupos, aunque es levemente inferior a la clasificación en tres y, además, la entropía es mayor con tres grupos.

Se observan las trayectorias según la clasificación *a posteriori* de los niños en tres y cuatro grupos. La Figura 4.9 muestra que el “G 1” de la clasificación en tres grupos se separa en dos cuando se imponen cuatro grupos, el “G 1” y “G 2”. A la vez, los otros dos grupos mantienen una proporción similar de niños cuando se pasa de clasificar en tres a cuatro grupos (26-22 % y 7%-7 %, respectivamente).

Lo anterior podría evidenciar una mejor clasificación de los niños en cuatro grupos. Sin embargo, dado que el **BIC** no es tanto más bajo y la entropía es mayor con tres, una clasificación de los niños en tres grupos se considera adecuada¹.

Tabla 4.12: Estadísticos y clasificación según grupos de niños - LCMM

G	loglik ^a	BIC	Entropy	NP ^b	% Niñas en cada grupo				
					G 1	G 2	G 3	G 4	G 5
1	-4712.0	9472.5		7	100 %				
2	-4509.0	9101.1	0.514	12	72 %	28 %			
3	-4426.3	8970.4	0.593	17	67 %	26 %	7 %		
4	-4407.5	8967.6	0.511	22	40 %	31 %	22 %	7 %	
5	-4345.0	8877.1	0.565	27	40 %	26 %	25 %	8 %	1 %

N=1026

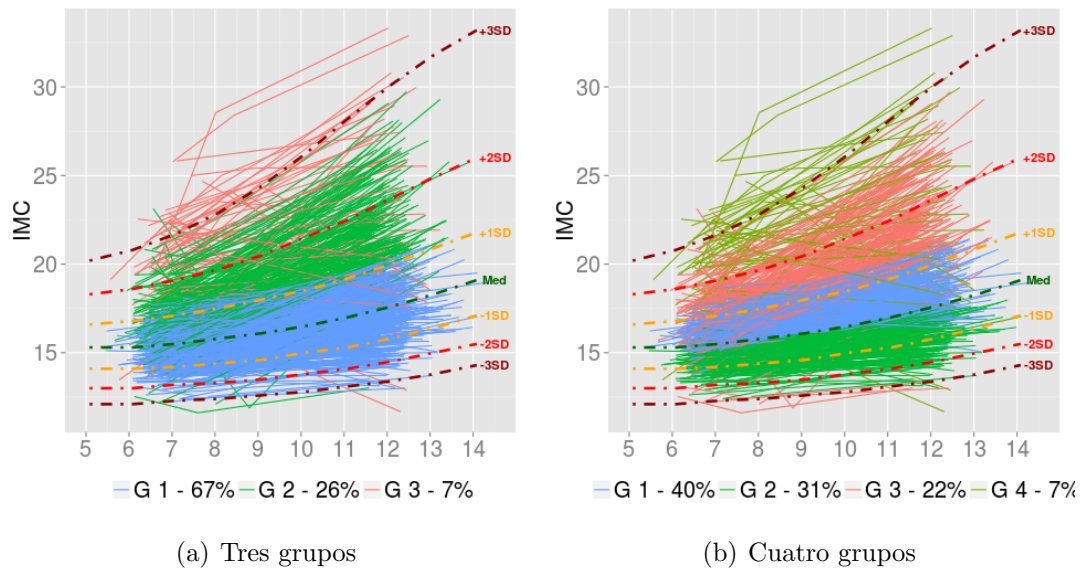
Fuente: Elaboración propia

^a Logaritmo de la verosimilitud

^b Número de parámetros estimados

¹ De todas formas, el análisis del **JLCMM** se realiza también para las otras estructuras de grupos, en particular para el caso de niños se comparan los resultados en términos del desempeño escolar con una estructura de tres y cuatro grupos según trayectoria nutricional y no hay diferencias sustantivas en los resultados.

Figura 4.10: Trayectoria nutricional según tres y cuatro grupos de niños - LCMM



Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$, $\pm 2SD$ y $\pm 3SD$)

4.3.1. Trayectoria Nutricional y Desempeño Escolar según Grupos de Niñas

A continuación, se presentan los resultados de los **JLCMM** estimados para niñas. En primer lugar, se presentan los estadísticos y clasificación *a posteriori* en grupos según trayectorias nutricionales que resulta de los modelos conjuntos. En la clasificación de grupos, los resultados no difieren sustancialmente con los obtenidos en los **LCMM**. Como se observa en la Tabla 4.13 el **BIC** se minimiza con cinco grupos, sin embargo para cuatro este estadístico es muy similar. Además, la entropía es mayor para cuatro y tres grupos que para cinco.

Tabla 4.13: Estadísticos y clasificación según grupos de niñas - JLCMM

G	loglik ^a	BIC	Entropy	NP ^b	% Niñas en cada grupo				
					G 1	G 2	G 3	G 4	G 5
1	-4971.1	10011.0		10	100 %				
2	-4791.6	9686.5	0.691	15	88 %	12 %			
3	-4747.3	9632.2	0.542	20	46 %	45 %	9 %		
4	-4715.2	9602.4	0.610	25	49 %	40 %	10 %	1 %	
5	-4697.5	9601.3	0.537	30	46 %	31 %	12 %	10 %	1 %

N=971

Fuente: Elaboración propia

^a Logaritmo de la verosimilitud ^b Número de parámetros estimados

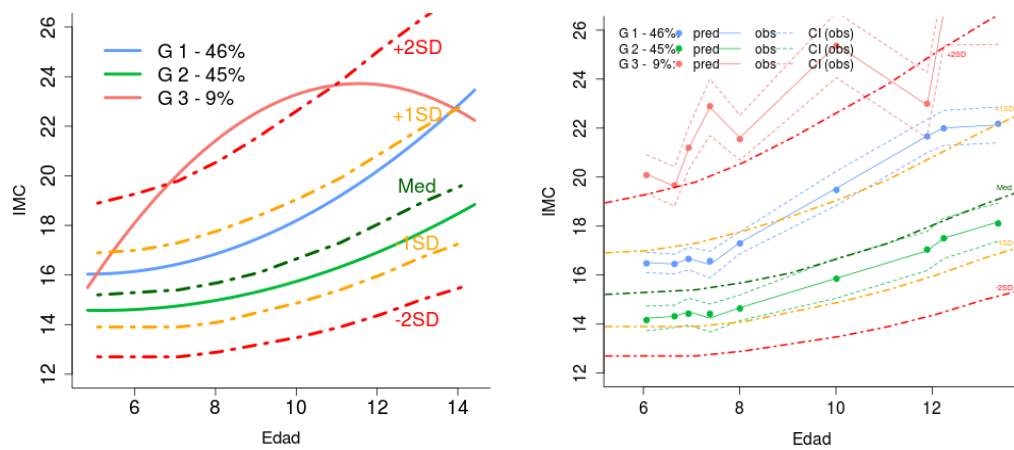
Por último, respecto a la proporción de niñas en cada grupo, las clasificaciones en mayor número de grupos parecen separar trayectorias “raras”, al igual que en el caso de los **LCMM**, pero no cambian sustancialmente las proporciones contenidas en los tres grupos más grandes. Por tanto, en el modelo conjunto, para el análisis del desempeño escolar y su relación con la trayectoria nutricional, es adecuado agrupar a las niñas en tres grupos según sus trayectorias nutricionales.

Las trayectorias predichas en cada grupo (Figura 4.11), según el **JLCMM**, muestran que el “G 1” (46 % de las niñas) está por encima de la trayectoria mediana y umbral de sobrepeso para cierto rango (Normal-alto) y el “G 2” sigue una trayectoria por debajo de la media con 45 % de las niñas (Normal-bajo). El grupo más pequeño, 9 % de las niñas presenta una trayectoria de sobrepeso e incluso obesidad.

Por último, la probabilidad de continuar en el sistema educativo, en distintas edades y para los distintos grupos se presenta en la Figura 4.12. Se

aprecia que la probabilidad de continuar los estudios se reduce con la edad y este comportamiento no difiere entre grupos. Por tanto, las diferentes trayectorias nutricionales de las niñas, que se capturan por estos grupos latentes, no resultan asociadas con la edad de abandono escolar de las niñas.

Figura 4.11: Trayectoria nutricional media e individual según grupos de niñas

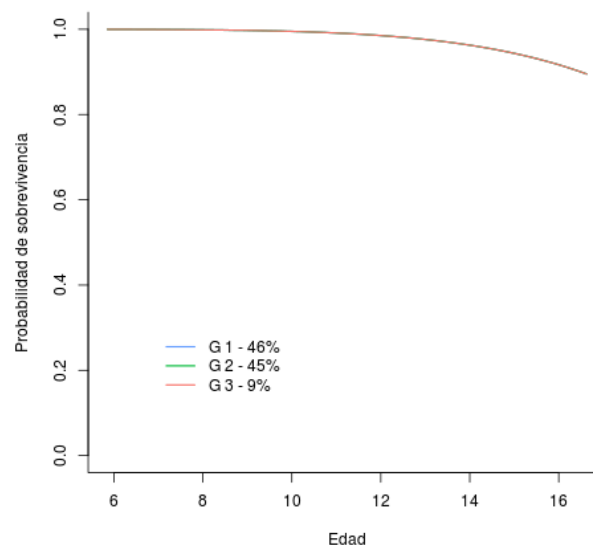


(a) Trayectoria media

(b) Trayectoria individual ponderada

Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$ y $\pm 2SD$)

Figura 4.12: Riesgo de abandono según grupo de niñas



4.3.2. Trayectoria Nutricional y Desempeño Escolar según Grupos de Niños

A continuación se presentan los resultados de los modelos conjuntos **JLCMM** estimados para niños. Se presentan los estadísticos y clasificación *a posteriori* en los grupos según trayectorias nutricionales que resulta de los modelos conjuntos y se compara con los resultados de los **LCMM**.

Tabla 4.14: Estadísticos y clasificación según grupos de niños - JLCMM

G	loglik ^a	BIC	Entropy	NP ^b	% Niños en cada grupo				
					G 1	G 2	G 3	G 4	G 5
1	-5371.0	10811.3		10	100 %				
2	-5168.0	10439.9	0.514	15	72 %	28 %			
3	-5085.3	10309.2	0.593	20	67 %	26 %	7 %		
4	-5019.8	10213.0	0.671	25	66 %	26 %	7 %	1 %	
5 ^c	-5003.9	10215.9	0.656	30	40 %	26 %	25 %	8 %	1 %

N=1026

Fuente: Elaboración propia

^a Logaritmo de la verosimilitud

^b Número de parámetros estimados

^c No converge

En la clasificación de grupos por **JLCMM**, los resultados no muestran diferencias significativas con los **LCMM**, al igual que ocurría para las niñas. La Tabla 4.14 muestra que el **BIC** se minimiza con cinco grupos, pero este modelo no converge, por tanto no es considerado. Entre las demás estructuras el mínimo **BIC** y el mayor índice de entropía se alcanza con cuatro grupos.

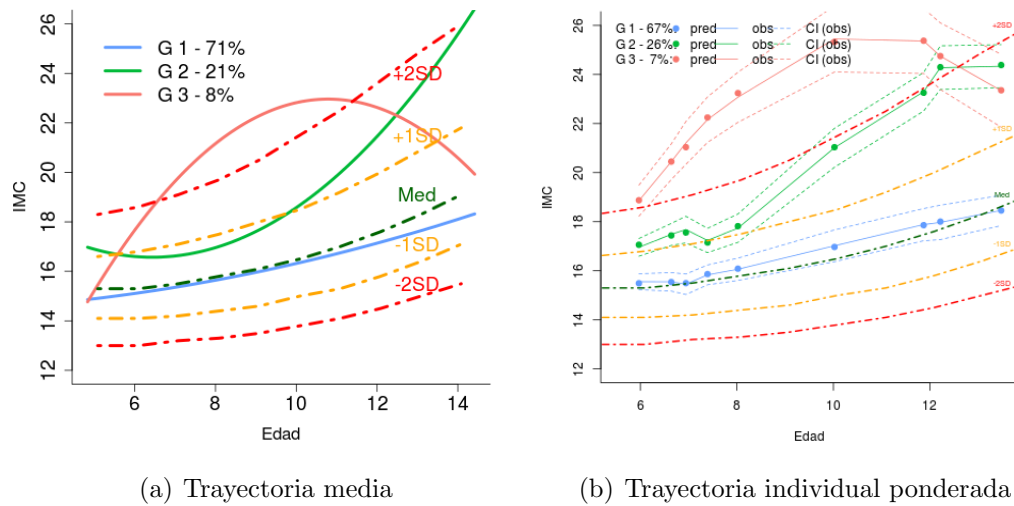
Por último, respecto a la proporción de niños en cada grupo, las clasificaciones en mayor número de grupos parecen separar trayectorias “raras”, al igual que en el caso de los **LCMM**, pero no cambian sustancialmente las proporciones contenidas en los grupos más grandes.

Por tanto, en el modelo conjunto y para el análisis del desempeño escolar y su relación con la trayectoria nutricional, es adecuado agrupar a los niños en tres grupos según sus trayectorias nutricionales.

Las trayectorias predichas en cada grupo (Figura 4.13), según el **JLCMM**, muestran que el 67 % de los niños sigue una trayectoria entorno a la mediana (grupo normal) y un 26 % está en el entorno de +1SD (grupo normal-sobrepeso). El grupo más pequeño con 7 % de los niños presenta una trayec-

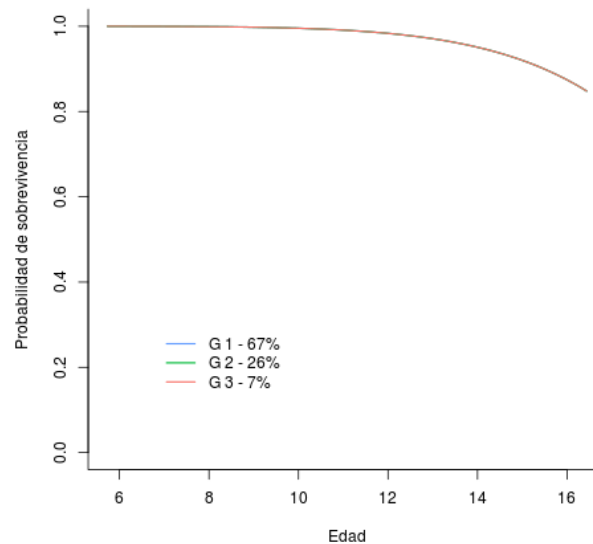
toría por encima de +2SD (grupo con obesidad), si bien para edades finales el **IMC** disminuye¹.

Figura 4.13: Trayectoria nutricional media e individual según grupos de niños



Nota: Valores de referencia del **IMC** para la edad: mediana (Med) y desvíos ($\pm 1SD$ y $\pm 2SD$)

Figura 4.14: Riesgo de abandono según grupos de niños



Por último, se muestra la probabilidad de continuar en el sistema educativo, en distintas edades, para los grupos (Figura 4.14). La probabilidad de continuar

¹Esto puede ser por el crecimiento en estatura muy alto que tienen los niños a esta edad y que puede mostrar disminución del **IMC**

se reduce con la edad y, al igual que para las niñas, este comportamiento no difiere significativamente entre grupos. Sin embargo, a diferencia de las niñas, la probabilidad de continuar en el sistema educativo decrece a edades más tempranas.

4.4. Análisis de Sensibilidad

Para el análisis de sensibilidad de los resultados se optó por estimar el modelo conjunto sin incluir la variable *Repitió*. Si bien los antecedentes y el análisis descriptivo de los datos muestran que esta variable es relevante para discriminar entre las tasas de abandono la forma en que fue relevada en la base de datos utilizada hace que la muestra se vea restringida para la estimación de estos modelos. Por tanto, se estiman modelos conjuntos sin incorporar esta covariable en el modelo de sobrevivencia.

4.4.1. Trayectoria Nutricional y Desempeño Escolar sin Repetición

Los resultados de la estimación de los JLCMM para niñas no muestran diferencias importantes con las obtenidas incluyendo la repetición previa como covariable en el modelo de sobrevivencia. Los estadísticos para los modelos y la clasificación *a posteriori* en grupos se presentan en la Tabla 4.15. Se observa que el BIC se minimiza con cinco grupos, sin embargo para cuatro este estadístico es muy similar. Además, la entropía es mayor para cuatro y tres grupos que para cinco.

Tabla 4.15: Estadísticos y clasificación según grupos de niñas - JLCMM - Sin Repetición

G	loglik ^a	BIC	Entropy	NP ^b	% Niñas en cada grupo				
					G 1	G 2	G 3	G 4	G 5
1	-6250.2	12565.5		9	100				
2	-5999.5	12100.3	0.644	14	86	14			
3	-5947.4	12032.4	0.506	19	48	43	9		
4	-5902.5	11978.7	0.552	24	50	38	10	2	
5	-5882.6	11975.1	0.514	29	40	37	18	3	2
N=1392									

Fuente: Elaboración propia

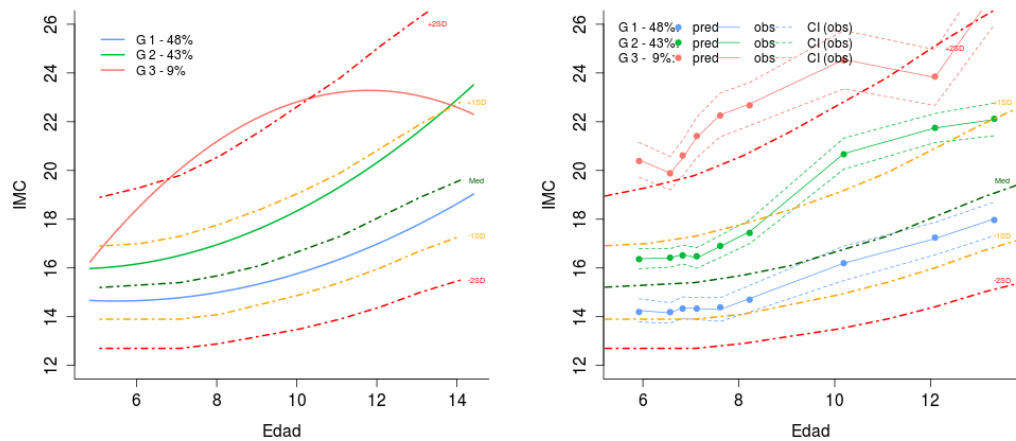
^a Logaritmo de la verosimilitud

^b Número de parámetros estimados

Por último, respecto a la proporción de niñas en cada grupo, las clasificaciones en mayor número de grupos parecen separar trayectorias “raras” pero no cambian sustancialmente las proporciones contenidas en los tres grupos más grandes. Por tanto, estos resultados son consistentes con una clasificación de

las niñas en tres grupos para analizar el desempeño escolar y su relación con la trayectoria nutricional.

Figura 4.15: Trayectoria nutricional media e individual según grupos de niñas

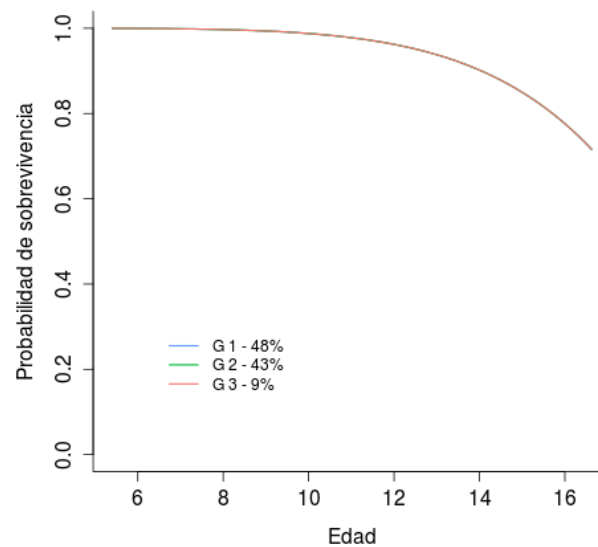


(a) Trayectoria media

(b) Trayectoria individual ponderada

Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$ y $\pm 2SD$)

Figura 4.16: Riesgo de abandono según grupo de niñas



Las trayectorias predichas en cada grupo (Figura 4.15), según el JLCMM, muestran que el “G 1” (48 % de las niñas) está por encima de la trayectoria mediana y umbral de sobrepeso para cierto rango (Normal-alto) y el “G 2”

sigue una trayectoria por debajo de la media con 43% de las niñas (Normal-bajo). El grupo más pequeño, 9% de las niñas presenta una trayectoria de sobrepeso e incluso obesidad.

Por último, la probabilidad de continuar en el sistema educativo para distintas edades y según grupos se presenta en la Figura 4.16. Como ocurría en el modelo que incluye la Repetición como covariable para la sobrevivencia, la probabilidad de continuar los estudios se reduce con la edad y el comportamiento entre grupos no difiere significativamente.

Los resultados de los modelos conjuntos JLCMM estimados para niños tampoco difieren respecto a la especificación que incluye la covariable *Repetición*. Los estadísticos y clasificación *a posteriori* en grupos se presentan en la Tabla 4.16. Se observa que el BIC se minimiza con cuatro aunque la diferencia con el BIC de tres grupos es pequeña. A su vez, la entropía para cuatro grupos es 0.634.

En este caso, respecto a la proporción de niños en cada grupo, también se observa que las clasificaciones con mayor número de grupos separan trayectorias “raras” pero no cambian sustancialmente las proporciones contenidas en los grupos más grandes. Además, la distribución de niños entre los grupos es similar a la que obtenida sin incluir *Repetición*.

Tabla 4.16: Estadísticos y clasificación según grupos de niños - JLCMM - Sin Repetición

G	loglik ^a	BIC	Entropy	NP ^b	% Niños en cada grupo				
					G 1	G 2	G 3	G 4	G 5
1	-6601.3	13268.3		9	100				
2	-6309.2	12720.7	0.535	14	79	21			
3	-6213.8	12566.6	0.553	19	71	21	8		
4	-6144.8	12465.0	0.634	24	70	22	8	1	
5	-6130.2	12472.2	0.673	29	43	27	21	8	1
N=1492									

Fuente: Elaboración propia

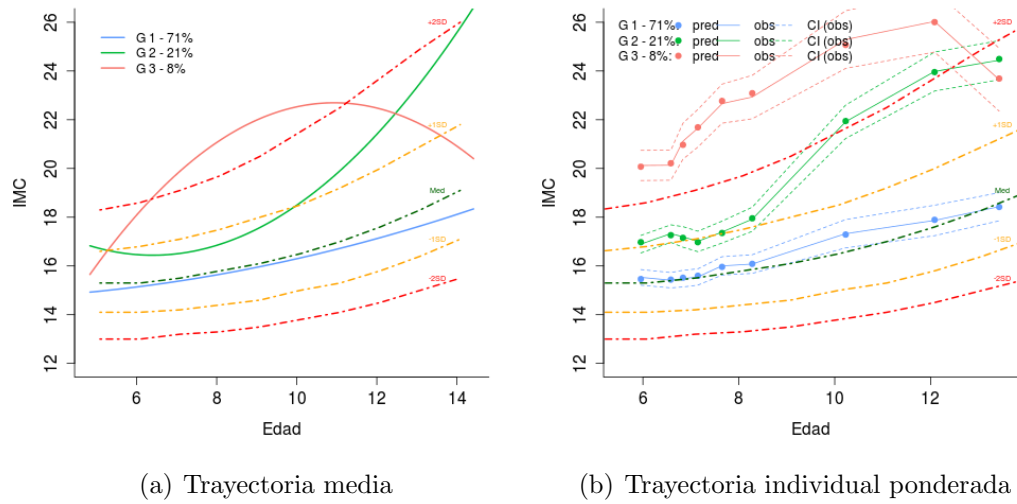
^a Logaritmo de la verosimilitud

^b Número de parámetros estimados

Las trayectorias predichas (Figura 4.17) según la clasificación en tres grupos, muestran que el 71% de los niños sigue una trayectoria entorno a la mediana (grupo normal) y un 21% está por encima de +1SD (grupo con sobrepeso). El grupo más pequeño con 8% de los niños presenta una trayectoria

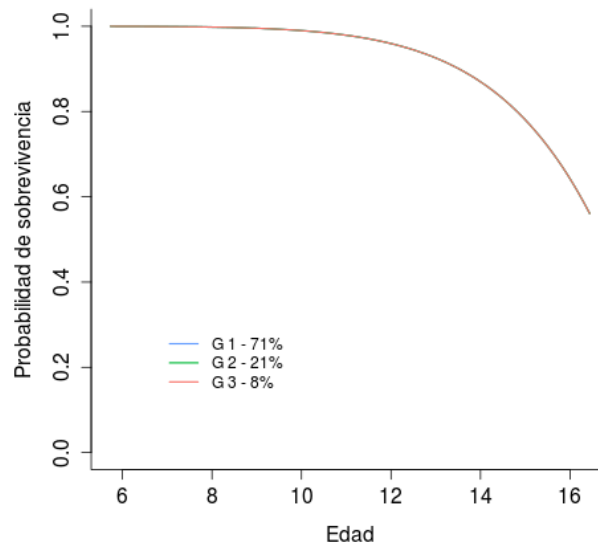
por encima de +2SD (grupo con obesidad), si bien para edades finales el IMC cae.

Figura 4.17: Trayectoria nutricional media e individual según grupos de niños



Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$ y $\pm 2SD$)

Figura 4.18: Riesgo de abandono según grupo de niños



Para finalizar, la Figura 4.18 muestra la probabilidad de no abandonar el sistema educativo para distintas edades y según los tres grupos de trayectorias nutricionales de niños. En este caso, también se observa que dicha probabilidad

disminuye con la edad pero no se ve modificada por la pertenencia a los grupos nutricionales.

Capítulo 5

Conclusiones

La realización de este trabajo permitió aplicar una técnica estadística moderna a un problema de relevancia en Uruguay y ha dado lugar a nuevas interrogantes y desafíos respecto a los temas vinculados con la trayectoria nutricional y el desempeño escolar de los niños y niñas. La aplicación de esta metodología constituye un avance ya que los trabajos existentes en Uruguay utilizan técnicas para datos de corte transversal para estudiar este tema. Los [LCMM](#) y [JLCMM](#) utilizados en el presente trabajo han permitido analizar la trayectoria nutricional con una perspectiva temporal, es decir, estudiar el cambio con la edad del estado nutricional y no únicamente el estado en una edad determinada.

Del análisis de la trayectoria nutricional, se desprende la existencia de grupos de niños y niñas en edad escolar, entre 6 y 12 años aproximadamente, con una evolución diferente del [IMC](#). La diferencia en la evolución del índice a lo largo del tiempo según grupos puede tener consecuencias sobre otros aspectos, como por ejemplo enfermedades que se han encontrado vinculadas a problemas nutricionales durante la infancia. Los resultados de los modelos estimados muestran que tanto niños como niñas pueden caracterizarse adecuadamente en términos de su trayectoria nutricional en tres grupos.

Estos tres grupos de niñas se caracterizan por:

- **Normal-bajo:** representa un 50 % de las niñas, con una trayectoria por debajo de la mediana
- **Normal-alto:** representa un 40 % de las niñas, transita por encima de la mediana y en algún rango de edad por encima del umbral de sobrepeso

- **Sobrepeso:** representa 10 %, con trayectoria en forma de “U” invertida en niveles de sobrepeso e incluso obesidad

Mientras que los tres grupos de niños se caracterizan por:

- **Normal:** representa un 70 % de los niños, con una trayectoria en el entorno de la mediana
- **Normal-Sobrepeso:** representa un 20 % de los niños, transita por encima de la mediana con tendencia creciente más inclinada y en algún rango de edad por encima del umbral de sobrepeso
- **Obesidad:** representa 10 %, con trayectoria en forma de “U” invertida en niveles de obesidad

Los resultados muestran diferencias de los grupos en otras características relevantes. Para las niñas las diferencias son significativas en el peso y la talla al nacer. En este sentido, las niñas con trayectoria nutricional normal-baja presentan peso y talla al nacer más bajos que los grupos normal-alto y con sobrepeso. Sin embargo, no son significativas las diferencias en los niveles de educación materno entre grupos. Por su parte, para los niños se encuentran diferencias significativas en el peso, pero no en la talla al nacer, y también en los niveles educativos maternos. El grupo normal se compone de niños con menor peso al nacer y cuyas madres tienen menor nivel educativo. Asimismo, para los niños con obesidad se encuentra que hay una mayor presencia de madres con nivel educativo más alto.

Los resultados de este trabajo para Uruguay no difieren sustancialmente de los reportados por [Stuart & Panico \(2016\)](#) para Reino Unido, si bien se trata de otro rango de edad y para un país desarrollado. [Stuart & Panico \(2016\)](#) identifican grupos de niños y niñas con sobrepeso y obesidad pero no con problemas de bajo peso o trayectorias del **IMC** decrecientes con la edad. Surge entonces como desafío el análisis de la trayectoria nutricional mediante la metodología propuesta en esta investigación para otros países en desarrollo. De esta forma, se pueden comparar los resultados obtenidos en el presente trabajo con estudios para países con similar nivel de desarrollo.

En cuanto al análisis de la relación entre trayectoria nutricional y desempeño escolar la evidencia muestra que los niños en Uruguay tienden a abondar

el sistema educativo a edades más tempranas que las niñas. Además, la pertenencia a los grupos según trayectoria nutricional no modifica la probabilidad de continuar en la escuela de los niños y las niñas en el rango de edad analizado. Por tanto, la evidencia no arroja una vinculación entre la trayectoria nutricional y la edad de abandono escolar de los niños y niñas del [ELBU](#).

Este resultado puede estar afectado por el rango de edad que queda representado en el [ELBU](#). Este estudio longitudinal de niños y niñas es una cohorte que abarca el tramo de edades entre seis y trece o catorce años en 2011/12. Como las tasas de abandono son crecientes con la edad y es a partir de los trece o catorce años que tienen un aumento pronunciado, la sub representación de las edades donde el problema de abandono se agrava puede afectar los resultados.

Esta debilidad, da lugar a futuros trabajos que incorporen la cuarta ola del [ELBU](#) que contendrá el rango de edad completo en el sistema educativo formal, de seis a diecisiete años aproximadamente. Además, se pueden estudiar otros desempeños educativos mediante otras técnicas estadísticas, que permitan entender la vinculación entre la trayectoria nutricional y el desempeño escolar medido por la repetición, por ejemplo.

A su vez, ninguno de los grupos identificados tiene una trayectoria nutricional caracterizada por el déficit nutricional o que tenga un comportamiento decreciente con la edad¹, si bien existe variabilidad en las trayectorias de los niños al interior de los grupos. En este sentido, otros trabajos reportaron que los problemas nutricionales de déficit se asocian con un peor desempeño escolar, particularmente con problemas de aprendizaje que pueden finalmente resultar en el abandono.

Una de las principales limitaciones de la investigación se encuentra en la elección del número de grupos mediante el análisis de medidas de bondad de ajuste y probabilidades *a posteriori* de los modelos estimados. Este procedimiento no cuenta con sustento formal que permita detectar de forma unívoca el número de grupos que mejor se ajusta a los datos. Sin embargo, en este trabajo se han estimado los modelos de varias formas² y no se han encontrado diferencias importantes en los resultados obtenidos. Además, la evidencia

¹ Salvo los grupos con trayectorias “raras”, ninguno de los grupos identificados con una cantidad relevante de niños o niñas tiene este tipo de trayectoria nutricional.

² Con distintas funciones de enlace para distintas muestras de niños y niñas.

internacional en otros trabajos muestra resultados similares a los encontrados en esta investigación.

Un desafío que resultó de la aplicación de esta metodología fue el proceso de estimación de los modelos, que resultó muy demandante computacionalmente. El método de estimación **MLE** que aplica la librería de **R** utilizada optimiza mediante un algoritmo de Newton-Raphson extendido, conocido como Algoritmo de Marquardt. Este último, resulta exigente en los criterios de convergencia utilizados¹. En este sentido, algunos de los modelos presentados no lograron converger aunque se realizaron varios ajustes en las estimaciones para hacer más sencilla la parametrización. De todas maneras, no se logró en algunos casos la convergencia ya que los modelos tienen muchos parámetros, particularmente cuando el número de grupos es alto y se especifican variaciones entre grupos².

Otra discusión que surge de los resultados del análisis de las trayectorias nutricionales proviene de observar las gráficas del **IMC** para la edad según las distintas clasificaciones entre grupos y los umbrales de la **OMS**. En los gráficos, parece que podrían diferenciarse las trayectorias de niñas y niños por los cuantiles de la distribución del **IMC** ya que no se evidencian claras diferencias en las tasas de cambio del **IMC** para la edad según los grupos.

Este resultado puede deberse en parte por la propia estructura de datos. El **ELBU** contiene solamente tres momentos de medición longitudinal, por lo que se cuenta con una baja cantidad de medidas para el **IMC** por individuo. Esto da poco margen para la identificación de trayectorias con distintos comportamientos en la pendiente lineal o cuadrática, lo que puede resultar en la separación de los grupos por el nivel del índice más que por la tasa de cambio según la edad de los niños y niñas. En general, se observa que las diferencias en los grupos provienen por estar encima o no de la mediana y los umbrales de obesidad y sobrepeso de la **OMS**. Por su parte, en futuros trabajos se puede realizar un análisis más detallado de las trayectorias identificadas como raras, para ver de qué forma pueden estar condicionando los resultados.

Se destaca que la implementación de estos modelos ha generado experiencia práctica que puede ser de utilidad en futuras aplicaciones. La estimación de

¹ Ver en el anexo **A.3** un detalle del algoritmo y las condiciones de convergencia que deben cumplirse en simultáneo.

² Además, los criterios de convergencia del algoritmo son rigurosos (**Proust-Lima et al., 2016b**).

los [LCMM](#) y [JLCMM](#) resulta exigente computacionalmente y se ha logrado implementar de forma tal que la experiencia puede ser útil en la aplicación de otros trabajos. En esta investigación se ha indagado en distintas formas de parametrización de los modelos. En particular, cuando se tenga disponible la nueva ronda del [ELBU](#), se contará con un tramo de edad donde el abandono escolar es más severo y, por tanto, resulta de interés analizar su vinculación con la trayectoria nutricional aplicando la misma metodología que en el presente trabajo.

La identificación del grupo con trayectoria nutricional de sobrepeso o incluso obesidad, según los estándares de la [OMS](#), podría dar lugar a futuras investigaciones en términos de su vínculo con otros aspectos de la salud de los niños, características socioeconómicas del entorno u otras, que den cuenta de las características particulares de los niños y niñas con estos problemas nutricionales durante el período escolar. Identificar las características de los grupos con sobrepeso u obesidad resulta relevante para entender la población de niños y niñas que enfrentan estos problemas nutricionales que pueden asociarse también con otros desempeños de los niños y niñas.

Referencias bibliográficas

- ACC/SCN (2000). *Fourth Report on the World Nutrition Situation*. Geneva: Administrative Committee on Coordination - Sub-Committee on Nutrition in collaboration with IFPRI, WHO.
- Alderman, H., Behrman, J., Lavy, V., & Menon, R. (1997). *Child Nutrition, Child Health, and School Enrollment*. Policy Research Working Paper, The World Bank.
- Amarante, V., Arim, R., Severi, C., Vigorito, A., & Aldabe, I. (2007). *El estado nutricional de los niños/as y las políticas alimentarias*. PNUD, Uruguay.
- Arim, R., Failache, E., Salas, G., Severi, C., & Vigorito, A. (2014). *Trayectorias del bienestar multidimensional en la infancia. Un estudio en base a datos de panel*. Proyectos de Investigación e Innovación Orientados a la Inclusión Social. Llamado 2010. Modalidad 1. CSIC. IECON, UdelaR, Uruguay.
- Aristimuño, A. & deArmas, G. (2012). *La transformación de la educación media en perspectiva comparada: Tendencias y experiencias innovadoras para el debate en Uruguay*. UNICEF, Uruguay.
- Arredondo, J., Carranza, N., Vázquez, M., & Rodríguez, M. (2003). *Transición epidemiológica*. Acta Pediátrica de México, 24(1):46-56.
- Barker, D., Winter, P., Osmond, C., Margetts, B., & Simmonds, S. (1989). *Weight in infancy and death from ischaemic heart disease*. Lancet, 9;2(8663):577-80.
- Behrman, J. (1990). *The Action of Human Resources and Poverty on One Another. What We Have Yet to Learn*. Working Paper no. 74. The World Bank.

- Behrman, J. & Wolfe, B. (1987). *Investments in Schooling in Two Generations in Pre-Revolutionary Nicaragua: The Roles of Family Background and School Supply*. Journal of Development Economics, 27(1-2): 395-420.
- Bérgolo, M., Leites, M., & Salas, G. (2006). *Privaciones nutricionales: su vínculo con la pobreza y el ingreso monetario*. Serie Documentos de Trabajo, DT 03/2006, Instituto de Economía, FCEA, UdelaR.
- Biernacki, C., Celeux, G., & Govaert, G. (2003). *Choosing Starting Values for the EM Algorithm for Getting the Highest Likelihood in Multivariate Gaussian Mixture models*. Computational Statistics and Data Analysis, 41(3-4), 561-575.
- Boado, M. & Fernández, T. (2010). *Trayectorias académicas y laborales de los jóvenes uruguayos. El panel PISA 2003-2007*. Facultad de Ciencias Sociales, UdelaR.
- Bucheli, M. & Casacuberta, C. (2000). *Asistencia escolar y participación en el mercado de trabajo de los adolescentes en Uruguay*. El Trimestre Económico, Vol. LXVII (3), México.
- Butte, N., Garza, C., & deOnis, M. (2007). *Evaluation of the Feasibility of International Growth Standards for School-Aged Children and Adolescents*. Journal of Nutrition, 137:153-7.
- Cameron, A. & Trivedi, P. (2010). *Microeconometrics using STATA*. STATA Press, Revised edition.
- Cueto, S. & Chinen, M. (2000). *Impacto Educativo de un Programa de Desayunos Escolares en Escuelas Rurales del Perú*. Documento de Trabajo 34., Grupo de Análisis para el Desarrollo, GRADE. Perú.
- deMelo, G., Failache, E., & Machado, A. (2015). *Adolescentes que no asisten a Ciclo Básico: caracterización de su trayectoria académica, condiciones de vida y decisiones de abandono*. Serie Documentos de Trabajo, DT 04/2015, Instituto de Economía, FCEA, UdelaR.
- Dempster, P. & Rubin, N. L. D. (1977). *Maximum Likelihood from Incomplete Data via the EM Algorithm*. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 39, Num. 1, pp.1-38.

- deOnis, M., Blössner, M., & Borghi, E. (2004). *Methodology for estimating regional and global trends of child malnutrition*. International Journal of Epidemiology, 33:1260-1270.
- deOnis, M. & Habicht, J.-P. (1996). *Anthropometric reference data for international use: recommendations from a World Health Organization Expert Committee*. The American Journal of Clinical Nutrition, Vol. 64:650-8.
- deOnis, M. & Lobstein, T. (2010). *Defining obesity risk status in general childhood population: Which cut-offs should we use?* International Journal of Pediatric Obesity, 5:458-460.
- deOnis, M., Onyango, A., Borghi, E., Siyam, A., Nishida, C., & Siekmann, J. (2007). *Development of a WHO growth reference for school-aged children and adolescents*. Bulletin of the World Health Organization, 85:660-7.
- Diggle, P., Liang, K.-Y., & Zeger, S. (1994). *Analysis of Longitudinal Data*. Clarendon Press.
- Engle, P., Menon, P., & Haddad, L. (1997). *Care and Nutrition: Concepts and Measurement*. International Food Policy Research Institute, Washington, D.C.
- Failache, E., Salas, G., & Vigorito, A. (2015). *Trayectorias educativas de los adolescentes en Uruguay. Un estudio en base a datos de panel*. Serie Documentos de Trabajo, DT 06/2015, Instituto de Economía, FCEA, UdelaR.
- Faucett, C. & Thomas, D. (1996). *Simultaneously modelling censored survival data and repeatedly measured covariates: A Gibbs sampling approach*. Statistics in Medicine 15, 1663-1685, doi:[10.1002/\(SICI\)1097-0258\(19960815\)15:15<1663::AID-SIM294>3.0.CO;2-1](https://doi.org/10.1002/(SICI)1097-0258(19960815)15:15<1663::AID-SIM294>3.0.CO;2-1).
- Fleming, T. & Harrington, D. (2013). *Counting Processes and Survival Analysis*. Wiley,.
- Frees, E. (2004). *Longitudinal and Panel Data: Analysis and Applications for the Social Sciences*. Cambridge University Press.
- Georgiadis, A., Benny, L., Crookston, B., Duc, L. T., Hermida, P., Mani, S., Woldehanna, T., Stein, A., & Behrman, J. (2016). *Growth trajectories from conception through middle childhood and cognitive achievement at*

- age 8 years: Evidence from four low- and middle-income countries*. SSM - Population Health, Elsevier, Vol. 2 pp. 43-54.
- Glewwe, P. (2005). *The impact of child health and nutrition on education in developing countries: Theory, econometric issues, and recent empirical evidence*. Food and Nutrition Bulletin, Vol. 26, no.2 (supplement 2). The United Nations University.
- Gluckman, P., Hanson, M., & Pinal, C. (2005). *The developmental origins of adult disease*. Matern Child Nutr. 1(3):130-41.
- Jacqmin-Gadda, H., Proust-Lima, C., Taylor, J., & Commenges, D. (2010). *Score Test for Conditional Independence Between Longitudinal Outcome and Time to Event Given the Classes in the Joint Latent Class Model*. Biometrics 66, 11–19, doi:[10.1111/j.1541-0420.2009.01234.x](https://doi.org/10.1111/j.1541-0420.2009.01234.x).
- Johnson, W. (2015). *Analytical Strategies in Human Growth Research*. American Journal of Human Biology. 27:69-83, Wiley.
- Knight, K. (2000). *Mathematical Statistics*. Chapman and Hall. CRC.
- Laird, N. & Ware, J. (1982). *Random-effects models for longitudinal data*. Biometrics 38, 963-974. The International Biometric Society.
- Levinger, B. (1992). *Nutrition, Health and Learning: Current Issues and Trends*. Education Development Center and UNDP.
- Lin, H., Turnbull, B., McCulloch, C., & Slate, E. (2002). *Latent Class Models for Joint Analysis of Longitudinal Biomarker and Event Process Data: Application to Longitudinal Prostate-Specific Antigen Readings and Prostate Cancer*. Journal of the American Statistical Association. Vol. 97, No. 457, pp.53-65.
- Massa, F. (2015). *Efecto de valores faltantes en estudios longitudinales en adultos mayores*. Trabajo de Tesis para la Maestría en Ingeniería Matemática, Facultad de Ingeniería, UdelaR.
- Molenberghs, G., Beunckens, C., Sotito, C., & Kenward, M. (2008). *EEvery missingness not at random model has a missingness at random counterpart with equal fit*. Journal of the Royal Statistical Society, Series B, 70, 371-388.

- Muthén, B. (2004). *Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data*. In D. Kaplan (ed.), *Handbook of quantitative methodology for the social sciences*, Newbury Park, CA: Sage Publications, in press.
- Proust-Lima, C., Dartigues, J.-F., & Jacqmin-Gadda, H. (2015). *Joint modelling of repeated multivariate cognitive measures and competing risks of dementia and death: a latent process and latent class approach*. *Statistics in Medicine*, 35(3). arXiv:1409.7598.
- Proust-Lima, C. & Jacqmin-Gadda, H. (2005). *Estimation of linear mixed models with a mixture of distribution for the random-effects*. *Computer Methods and Programs in Biomedicine*, Elsevier, Vol. 78 (2), pp.165-73.
- Proust-Lima, C., Joly, P., Dartigues, J.-F., & Jacqmin-Gadda, H. (2009). *Joint modelling of multivariate longitudinal outcomes and a time-to-event: A non-linear latent class approach*. *Computational Statistics and Data Analysis*, 53 1142-1154.
- Proust-Lima, C., Phillipps, V., Diakite, A., & Liqueet, B. (2016a). *Package 'lcmm'*. <https://cran.r-project.org/web/packages/lcmm/lcmm.pdf>.
- Proust-Lima, C., Phillipps, V., & Liqueet, B. (2016b). *Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: The R Package lcmm*. arXiv:1503.00890.
- Proust-Lima, C., Séne, M., Taylor, J., & Jacqmin-Gadda, H. (2014). *Joint latent class models for longitudinal and time-to-event data: A review*. *Statistical Methods in Medical Research*, Vol. 23(1) 74-90.
- Ramsay, J. (1988). *Monotone Regression Splines in Action*. *Statistical Science*, Vol. 3, No. 4, pp. 425-441.
- Regnault, N. & Gillman, M. (2014). *Importance of Characterizing Growth Trajectories*. *Annals of Nutrition and Metabolism*. 65:110-113.
- Rizopoulos, D. (2012). *Joint Models for Longitudinal and Time-to-event Data*. Chapman and Hall. CRC Biostatistics Series.
- Sen, A. (1999). *Desarrollo y libertad*. Planeta (1ra edición en español 2000).

- Stuart, B. & Panico, L. (2016). *Early-childhood BMI trajectories: evidence from a prospective, nationally representative British cohort study*. Nutrition and Diabetes, doi:[10.1038/nutd.2016.6](https://doi.org/10.1038/nutd.2016.6).
- Therneau, T. & Grambsch, P. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag.
- Therneau, T., Grambsch, P., & Fleming, T. (1990). *Martingale-Based Residuals for Survival Models*. Biometrika. Vol. 77 (1), pp. 147-160.
- Tsiatis, A. & Davidian, M. (2004). *Joint modeling of longitudinal and time-to-event data: An overview*. Statistica Sinica, 14:809–834.
- Verbeke, G. & Lesaffre, E. (1996). *A Linear Mixed-Effects Model with Heterogeneity in the Random-Effects Population*. Journal of the American Statistical Association, 91(433), 217–221.
- Verbeke, G. & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer-Verlag, New York.
- Victora, C., Adair, L., Fall, C., Hallal, P., Martorell, R., Richter, L., & Sachdev, H. (2008). *Maternal and child undernutrition: consequences for adult health and human capital*. Lancet. 371:340–57, Series 2. doi:[10.1016/S0140-6736\(07\)61692-4](https://doi.org/10.1016/S0140-6736(07)61692-4).
- Victora, C., DeOnis, M., Hallal, P., Blössner, M., & Shrimpton, R. (2010). *Worldwide timing of growth faltering: Revisiting implications for interventions*. Pediatrics. Vol. 125 (3), pp. e473-e480.
- Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press: Cambridge, MA.
- Wooldridge, J. (2010). *Correlated Random Effects Models with Unbalanced Panels*. Department of Economics. Michigan State University.
- Wu, L., Liu, W., Yi, G., & Huang, Y. (2012). *Analysis of Longitudinal and Survival Data: Joint Modeling, Inference Methods, and Issues*. Journal of Probability and Statistics, Vol. 2012, Article ID 640153, 17 pages. Doi:[10.1155/2012/640153](https://doi.org/10.1155/2012/640153).

- Wulfsohn, M. & Tsiatis, A. (1997). *A joint model for survival and longitudinal data measured with error*. Biometrics 53, 330-339.
- Zajacova, A. & Ailshire, J. (2012). *Body Mass Trajectories and Mortality Among Older Adults: A Joint Growth Mixture-Discrete-Time Survival Analysis*. The Gerontologist Vol. 54, No. 2, 221-231.
- Zhao, M. & Glewwe, P. (2010). *What determines basic school attainment in developing countries? Evidence from rural China*. Economics of Education Review, 29(3): 451-460.

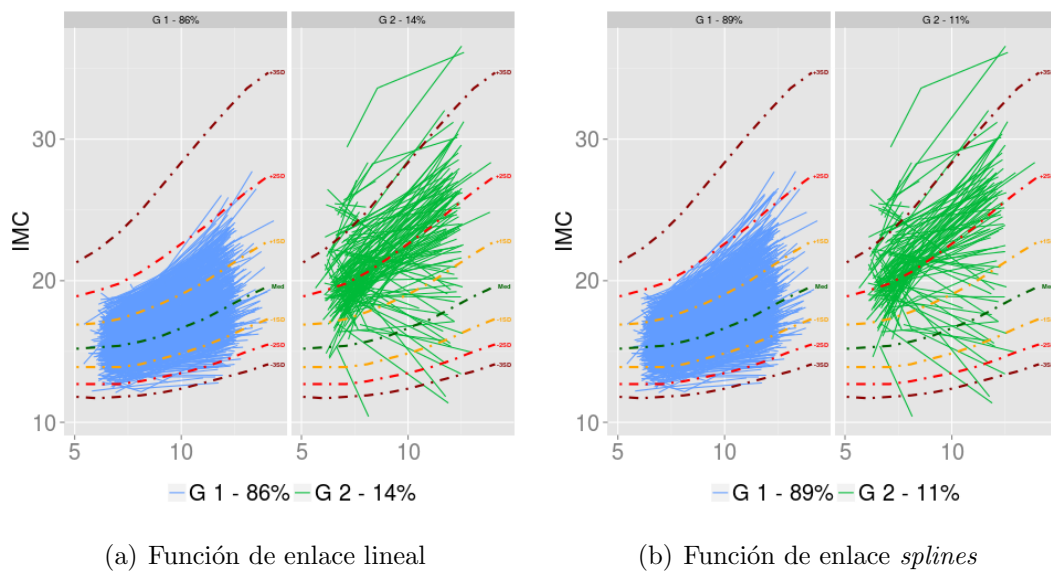
APÉNDICES

Apéndice A

Resultados

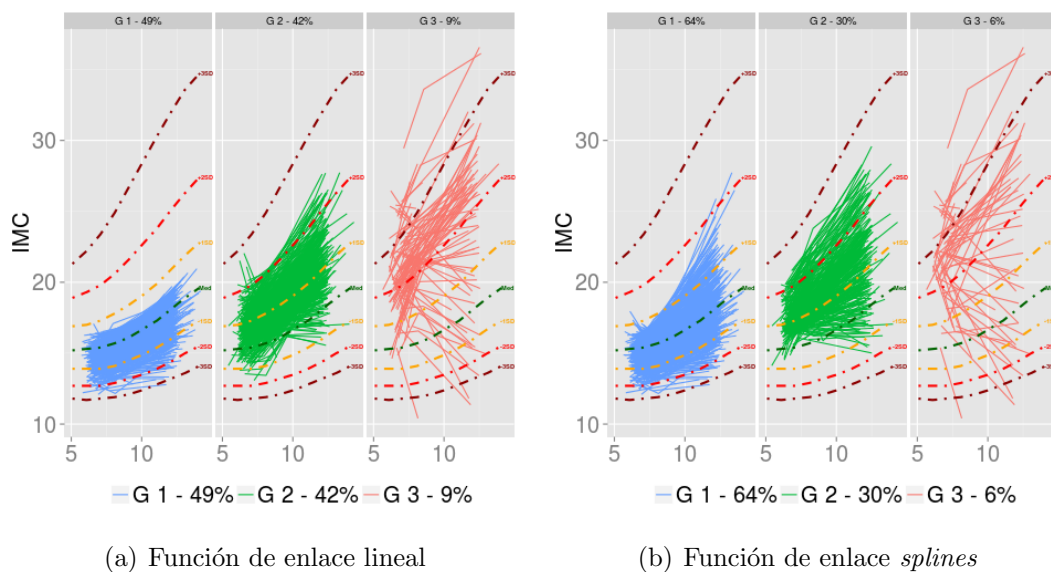
A.1. Clasificación de Niñas según Trayectoria Nutricional con Función de Enlace Lineal y *Splines*

Figura A.1: Trayectoria nutricional según dos grupos de niñas - LCMM



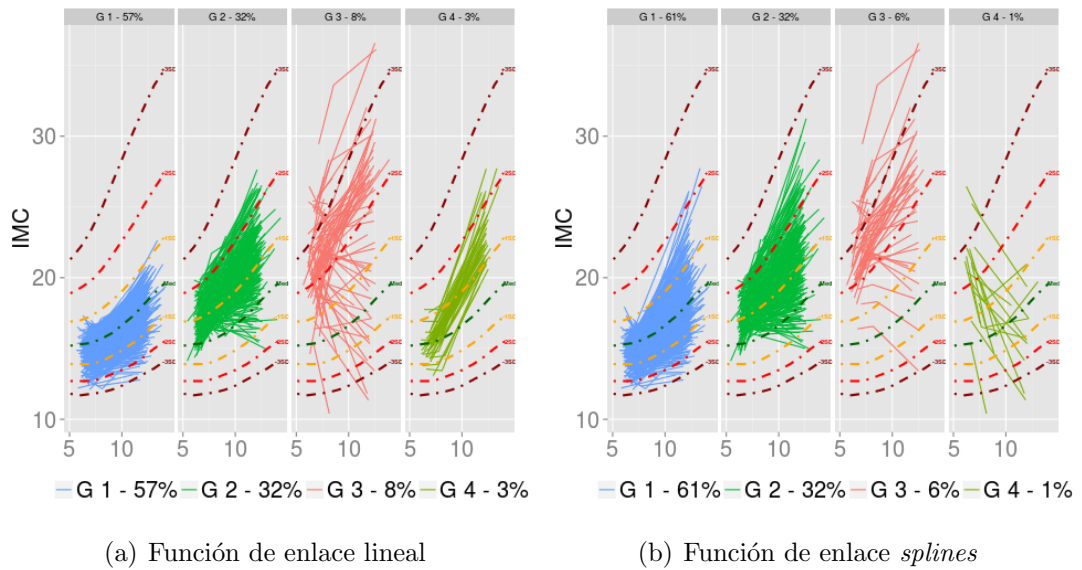
Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$, $\pm 2SD$ y $\pm 3SD$)

Figura A.2: Trayectoria nutricional según tres grupos de niñas - LCMM



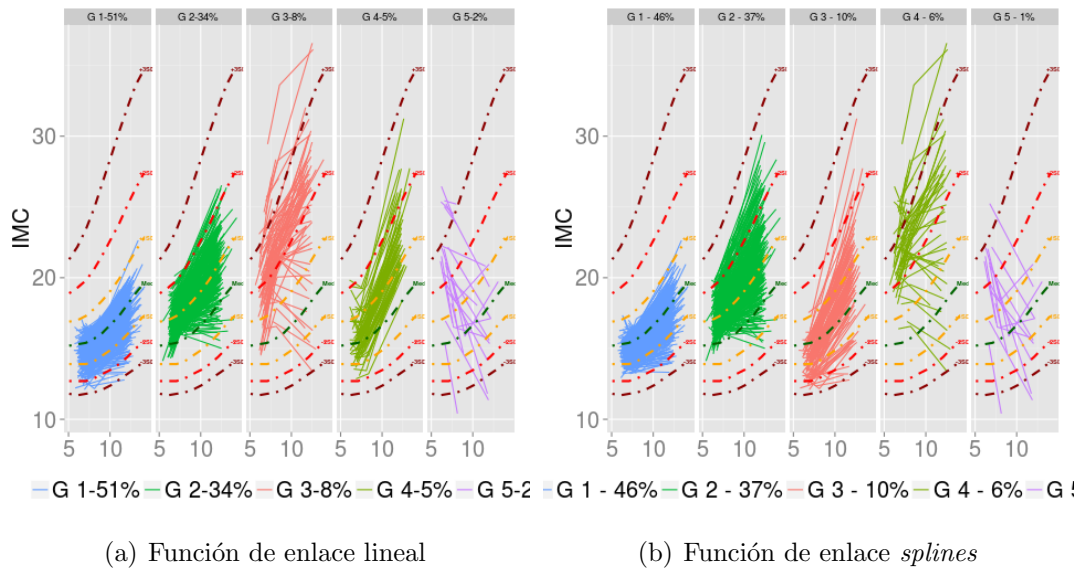
Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$, $\pm 2SD$ y $\pm 3SD$)

Figura A.3: Trayectoria nutricional según cuatro grupos de niñas - LCMM



Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$, $\pm 2SD$ y $\pm 3SD$)

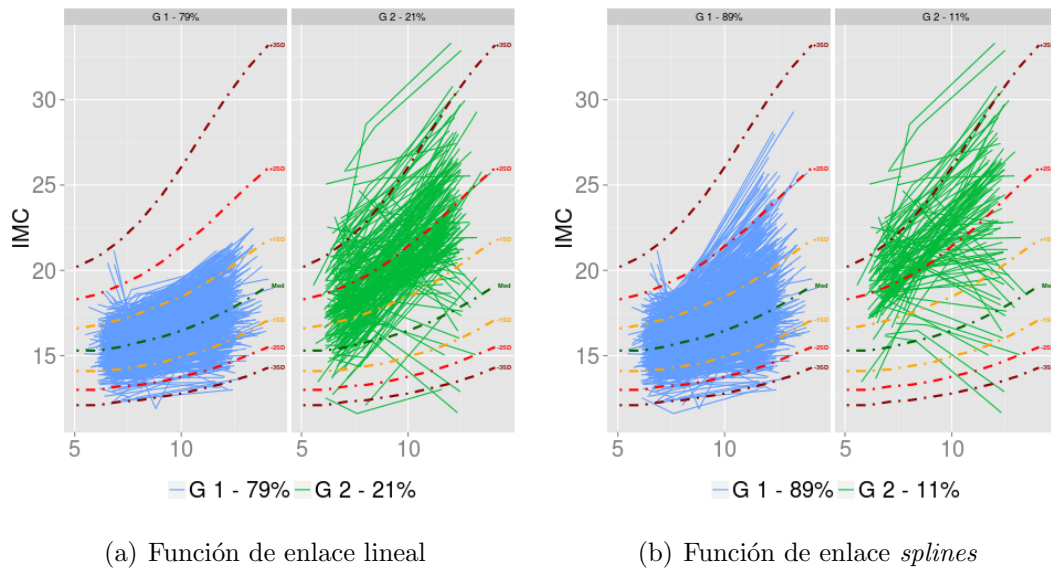
Figura A.4: Trayectoria nutricional según cinco grupos de niñas - LCMM



Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$, $\pm 2SD$ y $\pm 3SD$)

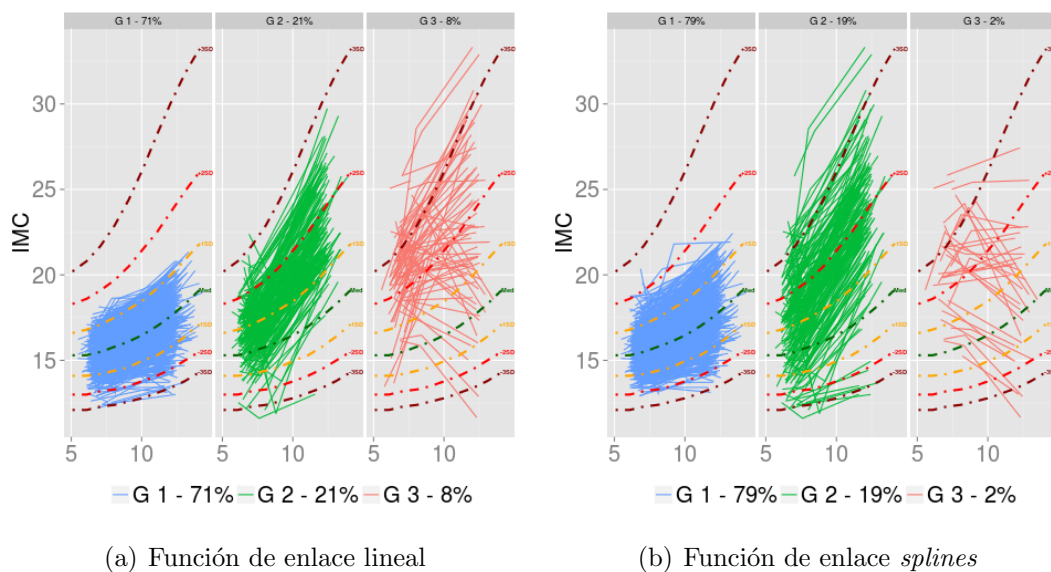
A.2. Clasificación de Niños según Trayectoria Nutricional con función de Enlace Lineal y *Splines*

Figura A.5: Trayectoria nutricional según dos grupos de niños - LCMM



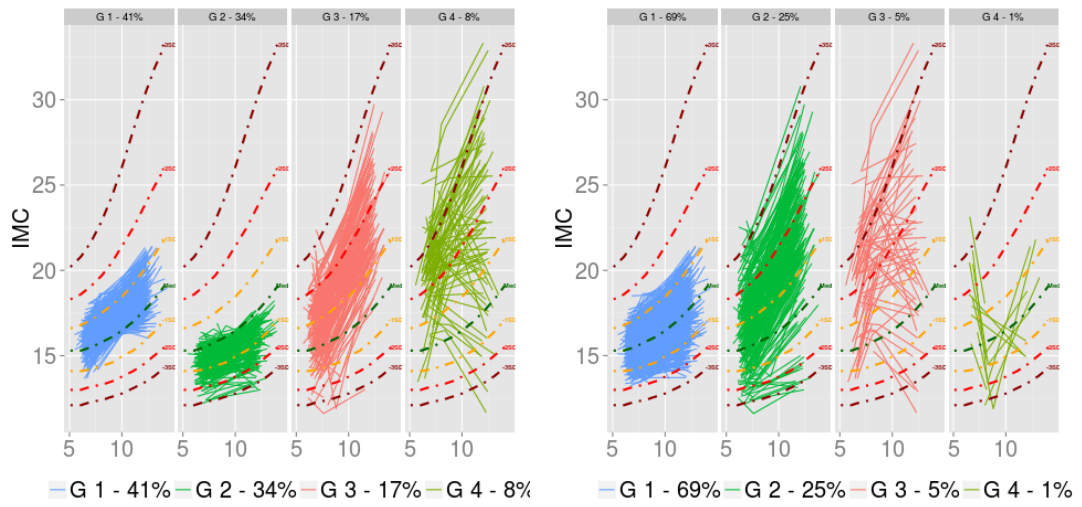
Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$, $\pm 2SD$ y $\pm 3SD$)

Figura A.6: Trayectoria nutricional según tres grupos de niños - LCMM



Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$, $\pm 2SD$ y $\pm 3SD$)

Figura A.7: Trayectoria nutricional según cuatro grupos de niños - LCMM

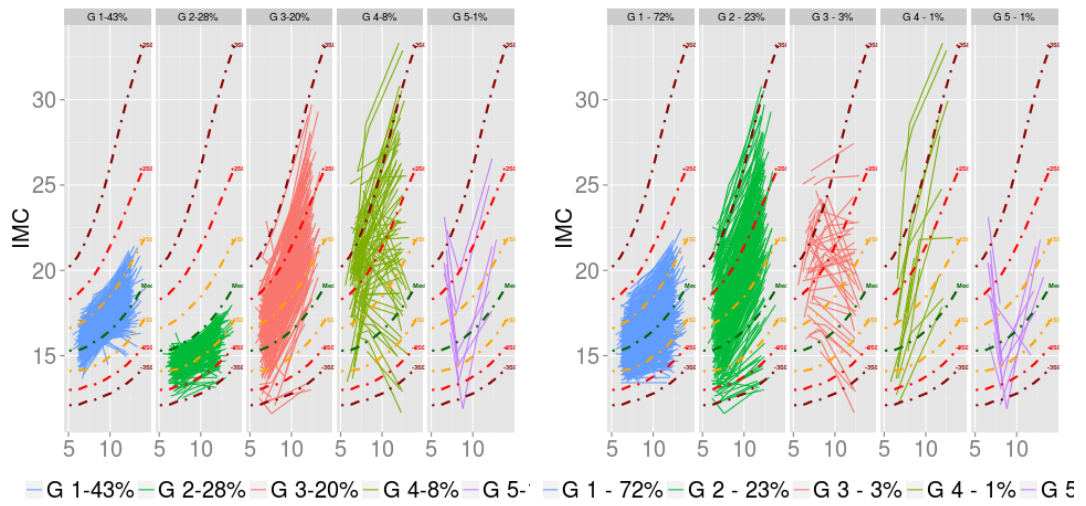


(a) Función de enlace lineal

(b) Función de enlace *splines*

Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$, $\pm 2SD$ y $\pm 3SD$)

Figura A.8: Trayectoria nutricional según cinco grupos de niños - LCMM



(a) Función de enlace lineal

(b) Función de enlace *splines*

Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$, $\pm 2SD$ y $\pm 3SD$)

A.3. Análisis de la Trayectoria Nutricional para Niñas con Función de Enlace *Splines*

Tabla A.1: Estadísticos y clasificación según grupos de niñas - *LCMM - splines*

G	loglik ^a	BIC	Entropy	NP ^b	% Niñas en cada grupo				
					G 1	G 2	G 3	G 4	G 5
1	-5532.1	11136.7		10	100				
2	-5454.0	11016.6	0.523	15	89	11			
3	-5420.8	10986.4	0.582	20	64	30	6		
4 ^c	-5397.3	10975.5	0.660	25	61	32	6	1	
5	-5384.0	10985.1	0.539	30	46	37	10	6	1

N=1392

Fuente: Elaboración propia

^a Logaritmo de la verosimilitud ^b Número de parámetros estimados ^c No Converge

Tabla A.2: Probabilidades *a posteriori* para tres grupos de niñas - *splines*

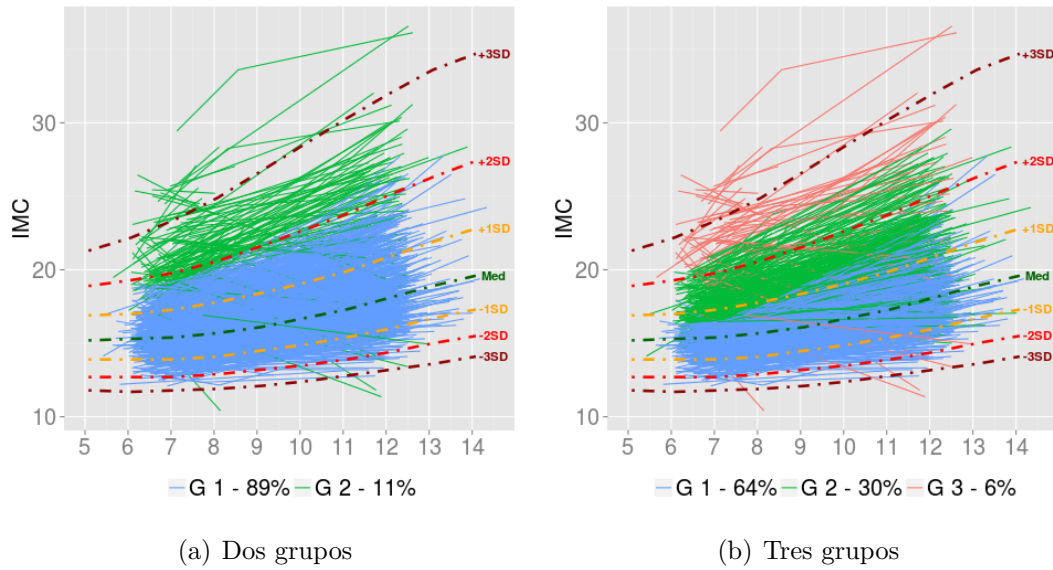
	(a) Prob. media			(b) % >umbral ^a		
	G 1	G 2	G 3	> 0.7	> 0.8	> 0.9
G 1	0.750	0.150	0.099	68 %	42 %	10 %
G 2	0.129	0.851	0.020	83 %	71 %	53 %
G 3	0.138	0.004	0.859	77 %	71 %	58 %

N=1392

Fuente: Elaboración propia

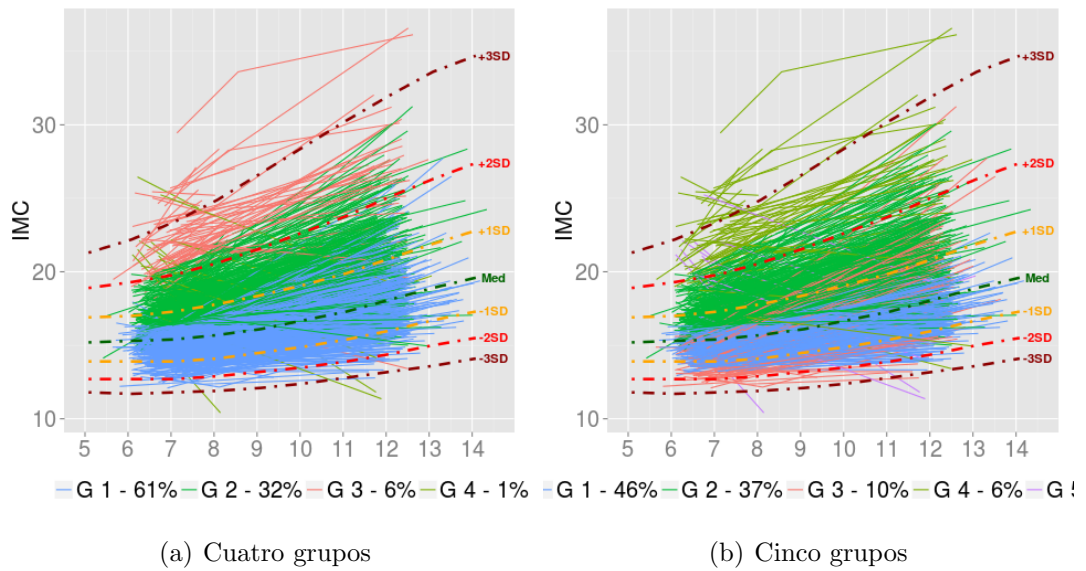
^a % de niñas con probabilidad *a posteriori* que supera el umbral

Figura A.9: Trayectoria nutricional según dos y tres grupos de niñas - *splines*



Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$, $\pm 2SD$ y $\pm 3SD$)

Figura A.10: Trayectoria nutricional según cuatro y cinco grupos de niñas - *splines*



Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$, $\pm 2SD$ y $\pm 3SD$)

A.4. Análisis de la Trayectoria Nutricional para Niños con Función de Enlace *Splines*

Tabla A.3: Estadísticos y clasificación según grupos de niños - *LCMM - splines*

G	loglik ^a	BIC	Entropy	NP ^b	% Niñas en cada grupo				
					G 1	G 2	G 3	G 4	G 5
1	-5592.0	11257.1		10	100				
2	-5509.7	11129.0	0.628	15	89	11			
3	-5483.0	11112.1	0.541	20	79	19	2		
4	-5418.8	11020.3	0.504	25	69	25	5	1	
5 ^c	-5408.9	11036.9	0.586	30	72	23	3	1	1

N=1492

Fuente: Elaboración propia

^a Logaritmo de la verosimilitud ^b Número de parámetros estimados ^c No Converge

Tabla A.4: Probabilidades *a posteriori* para tres grupos de niños - *splines*

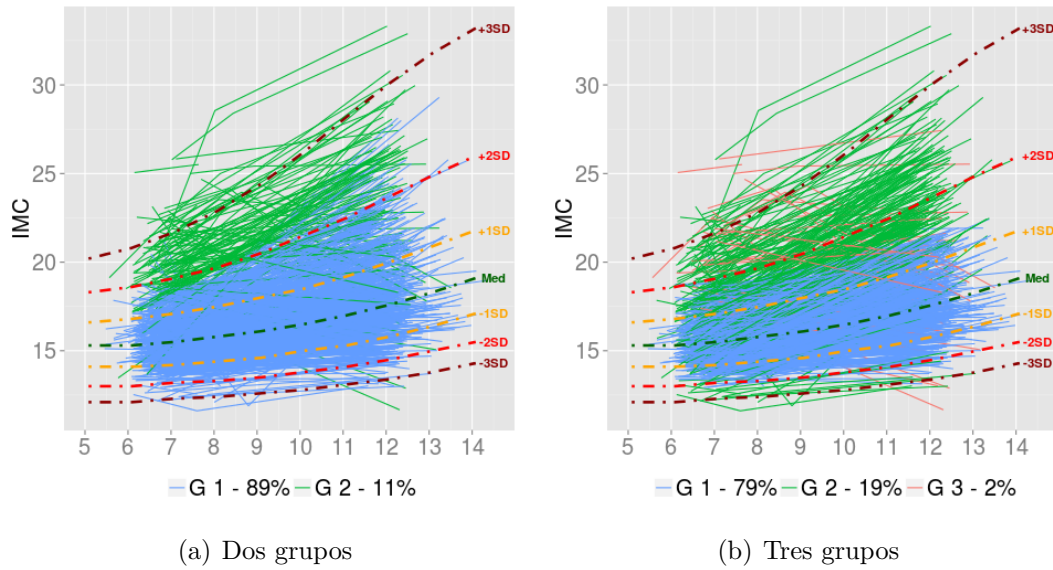
	(a) Prob. media			(b) % >umbral ^a		
	G 1	G 2	G 3	> 0.7	> 0.8	> 0.9
G 1	0.806	0.014	0.179	84 %	59 %	51 %
G 2	0.061	0.774	0.166	63 %	51 %	40 %
G 3	0.177	0.078	0.746	20 %	41 %	29 %

N=1492

Fuente: Elaboración propia

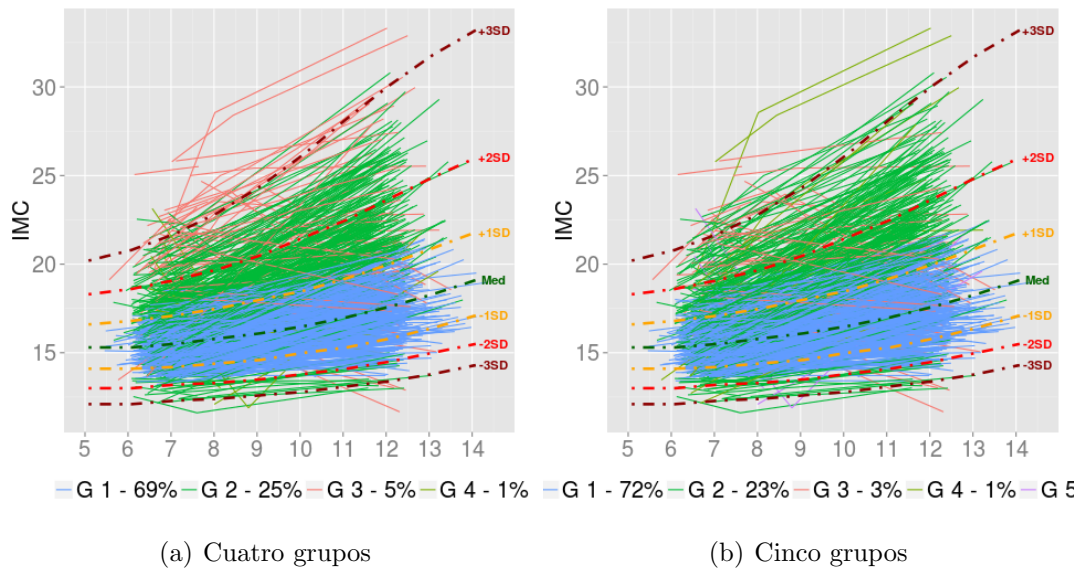
^a % de niños con probabilidad *a posteriori* que supera el umbral

Figura A.11: Trayectoria nutricional según dos y tres grupos de niños - *splines*



Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$, $\pm 2SD$ y $\pm 3SD$)

Figura A.12: Trayectoria nutricional según cuatro y cinco grupos de niños - *splines*



Nota: Valores de referencia del IMC para la edad: mediana (Med) y desvíos ($\pm 1SD$, $\pm 2SD$ y $\pm 3SD$)

ANEXOS

Anexo A

Anexo metodológico

A.1. Proceso Generador de Datos Faltantes para Datos Longitudinales

En los estudios longitudinales suelen presentarse problemas de datos faltantes y resulta un desafío el tratamiento de los mismos en las estimaciones. Los mecanismos que generan esta falta de datos pueden ser clave para determinar los métodos de estimación que proporcionan inferencias válidas. A continuación se presenta una descripción de los supuestos que se utilizan en general para afrontar este problema (ver por ejemplo, [Dempster & Rubin, 1977](#); [Molenberghs *et al.*, 2008](#); [Rizopoulos, 2012](#)).

[Rizopoulos \(2012\)](#) clasifica los datos faltantes de los estudios longitudinales en dos grandes grupos: monótonos y no monótonos. Los primeros, corresponden a problemas de *attrition* o *dropout*. Por su parte, los no monótonos corresponden cuando faltan observaciones para un individuo de forma intermitente a lo largo del estudio.

Las consecuencias de los datos faltantes en un estudio dependen del mecanismo que genera no observar dicho dato, sin embargo, se plantean tres consecuencias generales importantes:

- **Pérdida de eficiencia:** en los casos donde hay datos faltantes las estimaciones serán menos eficientes que con información completa, se necesitan muestras más grandes para lograr la misma eficiencia que con información completa.

- **Información no balanceada:** no todos los individuos en el estudio tendrán el mismo número de medidas para un conjunto dado de ocasiones (en los que se realizó el estudio). Esto puede generar problemas si se aplican técnicas que requieran información balanceada. Sin embargo, para la estimación de modelos lineales mixtos esto no es problema ya que no requieren datos balanceados.
- **Sesgos:** en algunas circunstancias y si no se maneja adecuadamente la presencia de datos faltantes, las estimaciones pueden ser sesgadas y por tanto dar lugar a inferencias erróneas.

Dados estos problemas, el tercero de ellos conlleva las peores consecuencias para la estimación, por lo tanto se deben identificar las circunstancias en las que se produce este problema.

Mecanismos generadores de datos faltantes

Estos mecanismos pueden pensarse como el modelo de probabilidad que describe la relación entre los datos faltantes y la variable de respuesta. Según el tipo de mecanismo en Rizopoulos (2012) se plantean las consecuencias sobre la estimación del modelo en presencia de datos faltantes.

Para introducir los conceptos formalmente se supone que un individuo i se mide en las ocasiones $j = 1, \dots, n_i$ y se espera obtener un vector de medidas de la variable de respuesta $y_i = (y_{i1}, \dots, y_{in_i})^\top$. Para distinguir este vector del efectivamente observado se define r_{ij} , un indicador de dato faltante, de la siguiente forma:

$$r_{ij} = \begin{cases} 1 & \text{si se observa } y_{ij} \\ 0 & \text{en caso contrario} \end{cases}$$

Por tanto, se particiona el vector de respuesta en el efectivamente observado y_{oi} que contiene los valores y_{ij} para cuando $r_{ij} = 1$ y el sub vector de datos faltantes y_{mi}

El vector $r_i = (r_{i1}, \dots, r_{in_i})^\top$ y su proceso generador se denominan *Proceso de datos faltantes* (*missing data process*). Para identificar este proceso se define

la función de densidad del proceso generador de datos faltantes condicional a la variable de respuesta:

$$p(r_i|y_{oi}, y_{mi}; \theta_r)$$

donde θ_r es el vector de parámetros del proceso generador de datos faltantes. Para determinar la forma de esta densidad condicional se proponen tres mecanismos: *Missing Completely at Random* (MCAR), *Missing at Random* (MAR) y *Missing Not at Random* (MNAR).

***Missing Completely at Random* (MCAR)**

La probabilidad de ser dato faltante no se relaciona con los valores que tendría la variable de respuesta si se hubiesen observado los casos faltantes y_{im} ni tampoco con los valores observados de la variable de respuesta y_{io} . Por tanto, la función de densidad condicional del proceso de datos faltantes se puede escribir de la siguiente forma:

$$p(r_i|y_{oi}, y_{mi}; \theta_r) = p(r_i; \theta_r)$$

De esta forma, los valores de y_{oi} pueden considerarse una muestra aleatoria de y_i , es decir, la distribución de y_{oi} coincide con la de y_i . En este contexto, se pueden obtener inferencias válidas utilizando cualquier procedimiento estadístico válido, ignorando el proceso generador de datos faltantes.

***Missing at Random* (MAR)**

Bajo este supuesto la probabilidad de ser dato faltante depende del conjunto de valores observados, pero no de los valores que tendría la variable de respuesta si se hubiesen observado los casos faltantes. Por tanto, la función de densidad condicional del proceso r_i es:

$$p(r_i|y_{oi}, y_{mi}; \theta_r) = p(r_i|y_{oi}; \theta_r)$$

En este contexto, dado que el mecanismo generador de datos faltantes depende de los datos observados de la variable y_{oi} la distribución de este vector

no coincide con la del vector completo y_i , los datos observados no pueden considerarse una muestra aleatoria de y_i . Por tanto, bajo el supuesto de **MAR** los momentos muestrales no estiman de forma adecuada los poblacionales. Sin embargo, la inferencia basada en métodos de verosimilitud pueden proporcionar estimaciones válidas aún cuando no se considere el proceso generador de datos faltantes. Esta propiedad de ignorar la verosimilitud del proceso generador de los datos faltantes se conoce como *ignorability*.

Otra forma de ver el supuesto de **MAR** es a través de la distribución predictiva de los valores que faltan y_{mi} condicional a los observados y_{oi} y el proceso generador de datos faltantes r_i . Básicamente, esta distribución es igual a la predictiva de y_{mi} condicional en y_{oi} :

$$p(y_{mi}|y_{oi}; \theta_r) = p(y_{mi}|y_{oi})$$

Un ejemplo es cuando el protocolo de un estudio exige que los pacientes salgan del mismo cuando el valor de la variable de respuesta supera cierto umbral. En estos casos, los datos faltantes están bajo el control del investigador y se relacionan con el componente observable de y_i . Dado que el proceso generador de datos faltantes depende de y_{oi} , entonces la distribución de los valores observados no coincide con la completa y_i . Por tanto, los valores observados no son una muestra aleatoria del vector poblacional de la variable y_i .

Sin embargo, se cumple que el valor faltante para un individuo i , es decir, y_{mi} condicional en sus valores observados y_{oi} , tiene la misma distribución que el vector completo y_i . Por tanto, los datos faltantes pueden predecirse de forma válida usando los datos observados de y_i , bajo un modelo para la distribución conjunta de y_{oi} y y_{mi} . Esto implica que bajo el supuesto de **MAR** los momentos muestrales no son estimaciones insesgadas de los poblacionales. Sin embargo, los análisis basados en la verosimilitud observada pueden proporcionar inferencias válidas aún si se ignora la contribución de r_i , dado que el modelo para y_i está correctamente especificado. Esto se desprende también observando la contribución a la verosimilitud de los datos completos $y_i = (y_{oi}, y_{mi}, r_i)$ para el individuo i puede factorizarse como sigue:

$$\begin{aligned}
L_i(\theta) &= \\
&= \int p(y_i, r_i; \theta) dy_{mi} \\
&= \int p(y_{oi}, y_{mi}; \theta_r) p(r_i | y_{oi}, y_{mi}; \theta_r) dy_{mi} \\
&= \int p(y_{oi}, y_{mi}; \theta_r) p(r_i | y_{oi}; \theta_r) dy_{mi} \\
&= p(y_{oi}; \theta_y) p(r_i | y_{oi}; \theta_r) \\
&= L_i(\theta_y) \times L_i(\theta_r)
\end{aligned}$$

Entonces, si θ_y y θ_r son disjuntos, en el sentido que $\theta = (\theta_y^\top, \theta_r^\top)^\top$ es el producto de los espacios paramétricos de θ_y y θ_r , respectivamente, se puede hacer inferencia válida sobre θ_y basada en la distribución marginal de los valores observados $p(y_{oi}; \theta_y)$ ignorando la distribución del proceso generador de datos faltantes (propiedad de *ignorability*).

Missing Not at Random (MNAR)

En este caso, la probabilidad de ser dato faltante depende de lo que se habría observado si el dato no hubiese sido faltante. En particular, la distribución de r_i depende por lo menos de algunos elementos de y_{mi} aún si se condiciona por y_{oi} :

$$p(r_i | y_{mi}; \theta_r) \text{ o bien } p(r_i | y_{mi}, y_{oi}; \theta_r)$$

Como en el caso de **MAR**, y_{oi} no constituye una muestra aleatoria de y_i . Sin embargo, la distribución de y_{mi} condicional a y_{oi} no es la misma que la poblacional, sino que depende de ambos y_{oi} y $p(r_i | y_i)$. Por tanto, el supuesto sobre el proceso que genera los datos faltantes es crucial y debe ser incluido en el análisis.

En definitiva, en presencia de datos faltantes del tipo **MNAR** y **MAR** si se estima únicamente con los datos observados ignorando el mecanismo se obtienen sesgos en la evolución estimada respecto a la real. Además, en el caso de **MNAR** estos sesgos pueden ser peores y es necesario incorporar en la estimación el modelo que se supone genera los datos faltantes.

Por último, puede ocurrir que la probabilidad de ser dato faltante dependa de alguna covariable y no de la variable de resultado, lo cual indicaría que

estamos frente a un mecanismo de **MCAR**. Pero si en el estudio no se condiciona por dicha covariable entonces el supuesto de **MCAR** ya no será válido

A.2. Método de Estimación por Máxima Verosimilitud

A continuación, se presenta una descripción del método de estimación por Máxima Verosimilitud (MLE) en el contexto de los modelos LMM extraída de Verbeke & Molenberghs (2000).

La inferencia se realiza en base a la distribución marginal de la variable de respuesta, es decir:

$$\mathbf{Y}_i \sim N(\mathbf{X}_{Li}\boldsymbol{\beta}, \mathbf{V}_i)$$

$$\mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top + \mathbf{R}_i + \boldsymbol{\Sigma}_i$$

Sea $\boldsymbol{\alpha}$ el vector que contiene los parámetros contenidos en la matriz de varianzas y covarianzas \mathbf{V}_i . El vector $\boldsymbol{\alpha}$ contiene $q(q+1)/2$ elementos diferentes de \mathbf{D} y todos los parámetros de \mathbf{R}_i y $\boldsymbol{\Sigma}_i$.

Sea $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top)^\top$ el vector en \mathbb{R}^s que contiene todos los parámetros a estimar y sea $\Theta = \Theta_\beta \times \Theta_\alpha$ el espacio paramétrico de $\boldsymbol{\theta}$, con Θ_β y Θ_α los espacios paramétricos de los efectos fijos y de los componentes de la varianza respectivamente. Notar que $\Theta_\beta = \mathbb{R}^p$ y Θ_α es el conjunto de valores para $\boldsymbol{\alpha}$ tal que \mathbf{D} , \mathbf{R}_i y $\boldsymbol{\Sigma}_i$ son (semi) definidas positivas.

La inferencia se realiza maximizando la función de verosimilitud $\mathbb{L}(\boldsymbol{\theta})$ respecto de $\boldsymbol{\theta}$:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^N \left\{ (2\pi)^{-n_i/2} |\mathbf{V}_i(\boldsymbol{\alpha})|^{-\frac{1}{2}} \times \exp\left(-\frac{1}{2}(\mathbf{Y}_i - \mathbf{X}_{Li}\boldsymbol{\beta})^\top \mathbf{V}_i^{-1}(\boldsymbol{\alpha})(\mathbf{Y}_i - \mathbf{X}_{Li}\boldsymbol{\beta})\right) \right\} \quad (\text{A.1})$$

Si $\boldsymbol{\alpha}$ es conocido, el estimador máximo verosímil de $\boldsymbol{\beta}$ se obtiene maximizando (A.1) condicional en $\boldsymbol{\alpha}$, y está dado por (Laird & Ware, 1982):

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \left(\sum_{i=1}^N \mathbf{X}_{Li}^\top \mathbf{V}_i^{-1} \mathbf{X}_{Li} \right)^{-1} \sum_{i=1}^N \mathbf{X}_{Li}^\top \mathbf{V}_i^{-1} \mathbf{y}_i \quad (\text{A.2})$$

Cuando $\boldsymbol{\alpha}$ no es conocido, pero se dispone de una estimación $\widehat{\boldsymbol{\alpha}}$, se puede hacer $\widehat{\mathbf{V}}_i = \mathbf{V}_i(\widehat{\boldsymbol{\alpha}})$ y estimar $\boldsymbol{\beta}$ usando la expresión en (A.2) en donde \mathbf{V}_i se reemplaza por $\widehat{\mathbf{V}}_i$.

La estimación máximo verosímil de $\boldsymbol{\alpha}$ se obtiene al maximizar la ecuación (A.1) con respecto a $\boldsymbol{\alpha}$, luego de reemplazar $\boldsymbol{\beta}$ por su expresión en (A.2). Esto último coincide con la estimación simultánea de $\boldsymbol{\beta}$ y $\boldsymbol{\alpha}$ a partir de maximizar la verosimilitud en (A.1). Por último, se destaca que la maximización de (A.1) se puede obtener aplicando logaritmo a la función de verosimilitud.

A.3. Algoritmo Marquardt

El algoritmo utilizado para maximizar la función de verosimilitud corresponde a una extensión de los algoritmos de la familia Newton-Raphson, que se denomina **Algoritmo de Marquardt**. El procedimiento de estimación de la librería `lcm` consiste en optimizar la función de verosimilitud para el modelo correspondiente utilizando este algoritmo. La elección de dicho procedimiento se fundamenta en su mejor velocidad de convergencia y rapidez reportada en análisis previos que el método EM (Proust-Lima & Jacqmin-Gadda, 2005; Proust-Lima *et al.*, 2016b).

Por tanto, en este procedimiento de optimización el vector de parámetros a estimar θ_G en el paso $l + 1$ se obtiene de la siguiente forma:

$$\theta_G^{(l+1)} = \theta_G^{(l)} - \delta \left(\tilde{\mathbf{H}}^{(l)} \right)^{-1} \nabla \left(L(\theta_G^{(l)}) \right)$$

El tamaño del paso δ por defecto vale uno, pero se modifica internamente para asegurar que se mejora la verosimilitud en cada paso, $\tilde{\mathbf{H}}$ es la *Hessiana* con la diagonal “inflada” para asegurar que sea definida positiva en cada paso. Por último, $\nabla \left(L(\theta_G^{(l)}) \right)$ es el gradiente de la log-verosimilitud en el paso l .

La matriz $\tilde{\mathbf{H}}$ se obtiene para cada paso l de la siguiente forma:

$$\begin{cases} \tilde{\mathbf{H}}_{ii}^{(l)} = \mathbf{H}_{ii}^{(l)} + \lambda \left[(1 - \eta) \left| \mathbf{H}_{ii}^{(l)} \right| + \eta \text{tr}(\mathbf{H}) \right] \\ \tilde{\mathbf{H}}_{ij}^{(l)} = \mathbf{H}_{ij}^{(l)} \text{ si } i \neq j \end{cases}$$

Los valores iniciales de λ y η se imponen en $\lambda = 0.01$ y $\eta = 0.01$ y se reducen cuando $\tilde{\mathbf{H}}_{ii}^{(l)}$ es definida positiva e incrementan en otro caso.

Además, para asegurar que la matriz de covarianzas de los efectos aleatorios sea definida positiva, se maximiza la verosimilitud para los valores distintos de cero de la matriz \mathbf{U} , correspondiente al factor de Cholesky de \mathbf{D} (esto es, $\mathbf{D} = \mathbf{U}^\top \mathbf{U}$).

Además, para dar contemplar la restricción de que la suma de las probabilidades de pertenencia a cada grupo debe ser uno, se maximiza sobre los parámetros transformados de la siguiente forma $\nu_g = \ln(\pi_g/\pi_G)$.

Los errores estándar de los elementos de \mathbf{D} y $(\pi_g)_{g=1, G-1}$ se estiman por el método *Delta* (Knight, 2000) mientras que los errores estándar de los demás parámetros se computan directamente usando la inversa de la matriz *Hessiana* observada.

Para alcanzar la convergencia en la estimación se deben cumplir simultáneamente tres condiciones, estas son:

$$\text{Estabilidad en los parámetros} \quad \sum_{j=1}^{n_\theta} (\theta_G(j)^{(l)} - \theta_G(j)^{(l-1)})^2 \leq \epsilon_\theta$$

$$\text{Estabilidad en la log-verosimilitud:} \quad |L^{(l)} - L^{(l-1)}| \leq \epsilon_L$$

$$\text{Tamaño de las derivadas:} \quad \frac{\nabla \left(L(\theta_G^{(l)}) \right)^\top (\mathbf{H}^{(l)})^{-1} \nabla \left(L(\theta_G^{(l)}) \right)}{n_\theta} \leq \epsilon_d$$

donde n_θ es el tamaño de θ_G

Los valores que se utilizan por defecto para los criterios de convergencia son: $\epsilon_\theta = \epsilon_L = \epsilon_d = 10^{-4}$. Los valores pueden parecer un poco elevados pero las condiciones deben satisfacerse simultáneamente y la condición sobre el tamaño de las derivadas es muy rigurosa, por tanto se asegura una buena convergencia aunque estos valores sean relativamente altos.

Una debilidad de otros algoritmos radica en que imponen como condición de convergencia únicamente la estabilidad de los parámetros o de la verosimilitud. En análisis complejos como los modelos de clases latentes o modelos conjuntos de clases latentes, la log-verosimilitud puede ser relativamente “aplanada” en algunas regiones del espacio paramétrico por lo que estas dos condiciones únicamente no aseguran la convergencia a un máximo global (Proust-Lima *et al.*, 2016b).