



FACULTAD DE
CIENCIAS ECONÓMICAS
Y DE ADMINISTRACIÓN



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

**UNIVERSIDAD DE LA REPÚBLICA
FACULTAD DE CIENCIAS ECONÓMICAS Y DE ADMINISTRACIÓN**

**TRABAJO FINAL DE GRADO PARA OBTENER EL TÍTULO DE
LICENCIADO EN ESTADÍSTICA**

**Estimaciones computacionales en modelos mixtos lineales
generalizados (GLMM)**

Federico Cescato

Tutor: Dr. Marco Scavino

Montevideo

URUGUAY

2021

UNIVERSIDAD DE LA REPÚBLICA

El tribunal docente integrado por los abajo firmantes aprueba el trabajo final de grado:

**Estimaciones computacionales en modelos mixtos lineales
generalizados (GLMM)**

Federico Cescato

Tutor:

Dr. Marco Scavino

Licenciatura en Estadística

Puntaje

Tribunal

Profesor Dr. Ramón Álvarez-Vaz(firma).

Profesor Dr. Juan J. Goyeneche(firma).

Profesor Dr. Marco Scavino(firma).

Fecha

Resumen

En este trabajo se busca explorar los modelos mixtos lineales generalizados (GLMM) y sus desafiantes estimaciones de parámetros por máxima verosimilitud.

La función de verosimilitud asociada a estos modelos tiene sus valores funcionales expresados por medio de integrales de Riemann evaluadas en subconjuntos de espacios euclídeos de altas dimensiones. Por lo tanto, hallar las estimaciones por máxima verosimilitud es encontrar los valores que maximizan las integrales de las cuales depende la verosimilitud.

En general, las integrales obtenidas de modelos GLMM aplicados a casos reales no son calculables analíticamente, debido a que una primitiva no es expresable por funciones elementales, o bien no hay métodos de cálculo que permitan obtenerla de forma efectiva. En las aplicaciones reales no es suficiente saber que existe la estimación máximo verosímil, sino que es necesario encontrar su valor al menos de forma aproximada.

Los métodos de estimación Monte Carlo se presentan como alternativas atractivas ya que suelen ser simples de implementar y convergen al verdadero valor del parámetro. Sin embargo, en muchos casos la convergencia es lenta con respecto al tamaño muestral. Por lo tanto, las técnicas Monte Carlo pueden demandar mucho esfuerzo computacional y tiempo de ejecución para obtener estimaciones cercanas a los valores verdaderos y desconocidos de los parámetros. Esto ha motivado la búsqueda de estrategias para acelerar su convergencia.

En el presente trabajo se explora la estrategia de Knudson (2016), la cual consiste en aproximar la entera función de verosimilitud de un modelo GLMM a través de un muestreo de importancia con una densidad elegida adecuadamente y dependiente de los datos. El trabajo de Knudson se basa en estrategias similares y muy bien estudiadas (Sung y Geyer, 2007; Geyer, 1994), en las que radica el método Monte Carlo de aproximación de la verosimilitud (MCLA).

El conjunto de datos utilizado en Knudson (2016) corresponde a un experimento que se

realizó en 1986 por científicos del Departamento de Ecología y Evolución en la Universidad de Chicago para estimar las probabilidades de apareamiento efectivo entre salamandras de diferentes especies. Dichas probabilidades son relevantes para un biólogo interesado en la especiación. El biólogo está interesado en saber los valores de las estimaciones de las probabilidades, no le alcanza con solo saber que existen. Se destaca que aún hoy en día este conjunto de datos se sigue utilizando como referencia para evaluar el desempeño de nuevos modelos y métodos de estimación. Un modelo GLMM incorpora de forma natural las probabilidades de apareamiento, ya que contempla el aporte personal de cada salamandra (efectos individuales) y permite trabajar con variables de respuesta no todas independientes entre sí. Por otra parte, un GLMM presenta el desafío de obtener las estimaciones paramétricas por máxima verosimilitud.

Palabras claves: GLMM, Monte Carlo, muestreo de importancia, aproximación de la verosimilitud, salamandras

Índice general

Índice general	v
Índice de tablas	vii
1. Introducción	1
1.1. Objetivos	2
1.1.1. Objetivo general	2
1.1.2. Objetivos específicos	3
2. Antecedentes	5
2.1. Primer acercamiento a lo mixto	6
2.2. Más mixto	7
2.3. ¿Por qué agregar una parte aleatoria?	8
2.4. Problema de tener una parte aleatoria en el modelo	10
3. Metodología	13
3.1. El modelo lineal generalizado mixto (GLMM)	13
3.1.1. Verosimilitud en un GLMM	14
3.2. Método MCLA de Knudson-Geyer para GLMM	16
3.2.1. Algunos cálculos y consideraciones previas	18
3.2.2. La densidad \tilde{f} usada para el muestreo de importancia	21
3.2.3. Fuentes de errores	22
3.3. Algunos resultados importantes	23

ÍNDICE GENERAL

3.3.1. TLC para el gradiente de MCLA	24
4. Datos utilizados para la aplicación	27
4.1. Diseño muestral	28
4.2. Resultados del experimento para analizar	31
5. Resultados	35
5.1. Modelo propuesto	35
5.1.1. Verosimilitud para el modelo propuesto	38
5.2. Resultados	39
6. Conclusiones	41
Referencias bibliográficas	43
Apéndices	47
Apéndice A. Más fundamento teórico	49
A.1. Familia exponencial de distribuciones	49
A.1.1. $\theta \in \mathbb{R}$	49
A.1.2. $\theta \in \mathbb{R}^n$	51
A.1.3. Propiedades importantes	52
A.2. Cuasi-verosimilitud penalizada (PQL)	55
A.2.1. Derivación de PQL	56
A.3. Distribución t multivariada	58
A.4. Ley de los grandes números	59
A.5. Integración Monte Carlo	59
A.6. Muestreo de importancia	61
A.7. Teorema de Slutsky	68
A.8. Test de Wald e intervalos de confianza	69
Apéndice B. Código y resultados	71

Índice de tablas

4.1. Diseño usado en el experimento de apareamiento de salamandras (McCullagh y Nelder, 1989, p. 440)	30
4.2. Apareamientos observados: verano '86 - Junio (McCullagh y Nelder, 1989, p. 442)	32
4.3. Apareamientos observados: otoño '86 - Setiembre (McCullagh y Nelder, 1989, p. 443)	33
4.4. Apareamientos observados: otoño '86 - Octubre (McCullagh y Nelder, 1989, p. 444)	34
5.1. Intervalos de confianza calculados en B	39
A.1. Parámetros de las principales distribuciones exponenciales	51

ÍNDICE DE TABLAS

Lista de Abreviaturas

- **GLMM**: Generalized linear mixed model
- **MCLA**: Monte Carlo likelihood approximation
- **MCLME**: Monte Carlo likelihood maximum estimator
- **Monte Carlo EM**: Monte Carlo expectation-maximization
- **PQL**: Penalized quasi-likelihood
- **RBM 1 ~ 5**: subconjunto de machos Rought Butt numerados de 1 a 5 (M de male, palabra usada en inglés para macho)
- **RBM 6 ~ 10**: subconjunto de machos Rought Butt numerados de 6 a 10 (M de male, palabra usada en inglés para macho)
- **RBF 1 ~ 5**: subconjunto de hembras Rought Butt numerados de 1 a 5 (F de female, palabra usada en inglés para hembra)
- **RBF 6 ~ 10**: subconjunto de hembras Rought Butt numerados de 6 a 10 (F de female, palabra usada en inglés para hembra)
- **WSM 1 ~ 5**: subconjunto de machos Whiteside numerados de 1 a 5 (M de male, palabra usada en inglés para macho)
- **WSM 6 ~ 10**: subconjunto de machos Whiteside numerados de 6 a 10 (M de male, palabra usada en inglés para macho)
- **WSF 1 ~ 5**: subconjunto de hembras Whiteside numerados de 1 a 5 (F de female,

ÍNDICE DE TABLAS

palabra usada en inglés para hembra)

- **WSF 6 ~ 10**: subconjunto de hembras Whiteside numerados de 6 a 10 (F de female, palabra usada en inglés para hembra)

Capítulo 1

Introducción

La ciencia estadística requiere el cálculo y estimación de cantidades de interés a valores escalares, vectoriales o funcionales, que dependen de la evaluación de funciones expresadas en la forma de integrales. A modo de ejemplo, se destacan las funciones de utilidad como se puede observar en [Ryan et al. \(2015\)](#), el factor de Bayes en el diseño de experimentos desde un enfoque bayesiano como se muestra en [Lewis y Raftery \(1997\)](#) o, sin más, la función de verosimilitud como paso intermedio para hallar estimaciones máximo verosímiles en el ámbito de los modelos lineales generalizados mixtos (ver los trabajos de [Kuk \(1999\)](#) y [Ruli et al. \(2016\)](#)).

El problema de la evaluación de funciones expresadas en forma integral puede resolverse por cálculo analítico directo, por medio de técnicas de simulación de tipo Monte Carlo o de integración numérica, o desarrollando métodos que combinan estas técnicas.

Visto la creciente complejidad de los modelos estadísticos propuestos para abordar problemas del mundo real sin introducir supuestos demasiado simplificadores, se observa que la aplicación del cálculo analítico directo es restringida a casos que pueden tener poca relevancia práctica.

Los métodos basados en la simulación Monte Carlo tienen amplia aplicabilidad, aunque en la mayoría de los casos reales pueden requerir de un gran número de valores simula-

dos para garantizar una estimación adecuadamente precisa, comprendida en los márgenes asignados de tolerancia del error. Veremos que las estimaciones en modelos mixtos lineales generalizados, objeto de este trabajo, y cuya formulación rigurosa será introducida en el capítulo 3, pueden obtenerse a partir de la aplicación de métodos Monte Carlo en los cuales la verosimilitud del modelo es aproximada refinando las técnicas de muestreo para reducir el número de simulaciones.

Las técnicas de aproximación analítica, como por ejemplo la aproximación de Laplace (Evans y Swartz, 2000, §4.3.3; Wong, 2001, cap. II), son adecuadas en el caso de evaluación de integrales de la forma $\int e^{L(\theta_1, \dots, \theta_n)} d\theta_1 \dots d\theta_n$. Los métodos numéricos basados en cuadraturas, por ejemplo la de Gauss-Hermite, son útiles para integrales de la forma $L(\theta_1, \dots, \theta_n) = \int e^{-x^2} f(x, \theta_1, \dots, \theta_n) dx$. En todo caso, el desempeño de estas técnicas se relaciona con el tamaño de la muestra obtenida a través del experimento, y se recomienda combinar su uso con técnicas de simulación.

En virtud de las consideraciones anteriores, este trabajo se dedicará a la presentación y aplicación de técnicas inferenciales computacionales basadas en variantes del método Monte Carlo básicas diseñadas para lograr aproximar la función de verosimilitud de dos tipos de modelos mixtos lineales generalizados, cuya verosimilitud se expresa necesariamente por medio de integrales de dimensión elevada.

1.1. Objetivos

1.1.1. Objetivo general

El objetivo general de este trabajo consiste en abordar el problema de refinar el método de simulación Monte Carlo básico para su efectiva aplicación a modelos estadísticos que no contengan supuestos demasiados simplificadores para el análisis de casos de estudio reales.

El empleo del método Monte Carlo en forma directa no resulta práctico ya que, en presen-

cia de modelos estadísticos asociados a un diseño de experimento, tales como los modelos mixtos lineales generalizados, insumiría una cantidad de tiempo demasiado elevada para disponer del número de simulaciones necesario para observar la convergencia de las estimaciones computacionales de las cantidades de interés. Este problema de convergencia puede ser atenuado logrando reducir el número de simulaciones que no aportan a la estimación de los parámetros de interés del modelo.

Una estrategia computacional que permite mejorar la convergencia del algoritmo de estimación basado en el método Monte Carlo básico consiste en el muestreo de importancia.

En este trabajo nos centraremos en la metodología introducida por Knudson (2016) en su tesis doctoral, donde la autora propone construir una densidad para el muestreo de importancia, en el caso de dos modelos mixtos lineales generalizados, la cual no depende de los datos disponibles. Se observa que la técnica propuesta por Knudson modifica una propuesta metodológica anterior (Sung y Geyer, 2007) en la cual se considera una densidad para el muestreo de importancia que depende de los datos observados.

1.1.2. Objetivos específicos

El primer objetivo específico consiste en presentar las características metodológicas de un método Monte Carlo que ha sido modificado a través de muestreo de importancia para lograr aproximar la función de verosimilitud de modelos mixtos lineales generalizados. Se destaca que la función de verosimilitud de estos modelos se expresa a través de integrales de dimensión elevada. Esta familia de modelos estadísticos es muy útil en las aplicaciones reales, dado que su formulación no presupone la independencia de las variables de respuesta entre todos los individuos.

El segundo objetivo específico consiste en mostrar cómo implementar el método de estimación computacional basado en una variante del muestreo de importancia introducida por Knudson (2016), actualmente disponible en la librería **glmm** (Knudson, 2020) de R (R Core Team, 2021), a través del célebre caso de estudio de especiación de las salaman-

CAPÍTULO 1. INTRODUCCIÓN

dras, presentado en [McCullagh y Nelder \(1989\)](#) y ampliamente utilizado en la literatura estadística como caso de prueba para medir el rendimiento de técnicas computacionales de estimación de probabilidades de apareamiento en modelos con efectos individuales.

Capítulo 2

Antecedentes

Siguiendo a [McCulloch \(2003, cap. 4\)](#), para motivar el cambio de un modelo lineal generalizado a un modelo mixto lineal generalizado, se toma un ejemplo surgido en 1990 donde se tiene una muestra de datos para estudiar efectos de un tratamiento por selenio contra el riesgo de cáncer de piel en las personas. Para eso se mide Y_{it} , el número de nuevos carcinomas de células basales (la mayoría de los casos de cáncer de piel son de células basales). Esta medición de nuevos carcinomas corresponde a la persona i en el momento t , donde los individuos se comprueban anualmente en un estudio que dura cinco años.

Las variables relevadas son:

- SEX = el sexo de la persona
- SEL = si o no, según si el individuo estuvo o no en el grupo de tratamiento por selenio
- SUN = una medida de la exposición al sol durante el año anterior

A su vez se considera el supuesto de que Y_{it} sigue una distribución Poisson, aunque se espera que algunas personas sean más susceptibles que otros. Esto es, que el individuo tiene un aporte intrínseco a la medida realizada. Esta consideración es lo que motiva introducir un efecto aleatorio (por el individuo), al modelo lineal generalizado (solo con efectos fijos), pasando a un modelo mixto lineal generalizado (con efecto fijo y aleato-

rio). Si bien, es un modelo más flexible, tiene la dificultad de que es necesario especificar correctamente el efecto aleatorio y aún si, genera una complejidad mayor al momento de hacer estimaciones por máxima verosimilitud.

2.1. Primer acercamiento a lo mixto

Retomando el modelo para el riesgo de cáncer de piel, se puede resumir lo anterior en

$$\begin{cases} Y_{it}|\lambda \sim \text{Poisson}(\lambda) \\ \log(\lambda_{it}) = \beta_{0i} + \beta_1\text{SEX}_i + \beta_2\text{SEL}_i + \beta_3\text{SUN}_{it} \end{cases} \quad (2.1)$$

Siendo la variable SUN_{it} la que provee de información sobre lo sucedido al momento t . Además, y ésta es la parte que lo convierte en un modelo mixto, el parámetro β_{0i} es una variable aleatoria y no una constante fija a determinar como lo son β_1 , β_2 y β_3 . Esta variable aleatoria β_{0i} , ahora pasa a ser el término independiente específico del sujeto, donde se puede asumir que β_{0i} sigue una distribución normal entre la población de todos los individuos, centrada en el valor β_0 . La especificación completa del modelo ahora queda:

$$\begin{cases} Y_{it}|\lambda \sim \text{Poisson}(\lambda) \\ \log(\lambda_{it}) = \beta_{0i} + \beta_1\text{SEX}_i + \beta_2\text{SEL}_i + \beta_3\text{SUN}_{it} \\ \beta_{0i} \sim \mathcal{N}(\beta_0, \sigma_{s,0}^2) \end{cases} \quad (2.2)$$

Aquí se puede mostrar un hecho muy importante debido a la nueva estructura del modelo anterior. Asumiendo como antes, que el término independiente es β_{0i} es una variable aleatoria con distribución $\mathcal{N}(\beta_0, \sigma_{s,0}^2)$, se puede observar como el mismo induce una correlación entre las medidas Y_{it} e $Y_{it'}$, esto es, entre las medidas de un mismo individuo en dos momentos diferentes. Para ver esto solo basta calcular $\text{Cov}(Y_{it}, Y_{it'})$, obteniendo lo siguiente:

$$\text{Cov}(Y_{it}, Y_{it'}) = \exp(2\beta_1\text{SEX}_i + 2\beta_2\text{SEL}_i + \beta_3\text{SUN}_{it} + \beta_3\text{SUN}_{it'}) \exp(\sigma_{s,0}^2)(\exp(\sigma_{s,0}^2) - 1)$$

Como se puede ver, siempre que $\sigma_{s,0}^2 > 0$, se tiene $\text{Cov}(Y_{it}, Y_{it'}) > 0$ lo cual significa que las medidas están correlacionadas. Esto no es de sorprender ya que provienen de la misma persona.

2.2. Más mixto

La hipótesis para β_0 hecha anteriormente se puede hacer para cualquier cantidad de parámetros. Justificar dicha asignación corresponde a un conocimiento experto del experimento en cuestión en donde la asignación es exacta o es una aproximación razonable.

Por ejemplo, siguiendo en la línea anterior, se puede suponer que además el parámetro β_3 pasa a ser una variable aleatoria β_{3i} con distribución normal centrada en el antiguo parámetro β_3 y con varianza σ_{3i}^2 . De nuevo, cada individuo tiene su aporte personal para los coeficientes β_{0i} y β_{3i} , siendo los parámetros no modificados β_1 y β_2 , parámetros con el mismo valor para todos los individuos.

En resumen, el modelo anterior queda expresado como:

$$\begin{cases} Y_{it} | \lambda \sim \text{Poisson}(\lambda) \\ \log(\lambda_{it}) = \beta_{0i} + \beta_1 \text{SEX}_i + \beta_2 \text{SEL}_i + \beta_{3i} \text{SUN}_i \\ \beta_{0i} \sim \mathcal{N}(\beta_0, \sigma_{s,0}^2) \\ \beta_{3i} \sim \mathcal{N}(\beta_3, \sigma_{s,0}^2) \end{cases} \quad (2.3)$$

El modelo en este caso se puede agrupar como sigue:

$$\begin{aligned} \log(\lambda_{it}) &= \underbrace{\beta_1 \text{SEX}_i + \beta_2 \text{SEL}_i}_{\text{efectos fijos}} + \underbrace{\beta_{0i} + \beta_{3i} \text{SUN}_i}_{\text{efectos aleatorios}} \\ &= \begin{bmatrix} \text{SEX}_i & \text{SEL}_i \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} 1 & \text{SUN}_i \end{bmatrix} \begin{bmatrix} \beta_{0i} \\ \beta_{3i} \end{bmatrix} \\ &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} \end{aligned}$$

Esta última es la forma matricial habitual del modelo con \mathbf{X} e \mathbf{Z} las matrices de diseño de la parte fija y de la parte aleatoria respectivamente, β el vector de parámetros fijos y \mathbf{u} el vector de efectos aleatorios.

2.3. ¿Por qué agregar una parte aleatoria?

Tener un efecto aleatorio genera una diferencia importante respecto de los modelos lineales y modelos lineales generalizados (sin efectos aleatorios). Como se puede ver, por ejemplo, en [McCullagh y Nelder \(1989, p. 26\)](#), [McCulloch y Searle \(2001, §5.2.a\)](#) y [Faraway \(2016, §8.1\)](#), uno de los supuestos muy utilizado para amplias clases de modelos lineales y modelos lineales generalizados es que las variables respuestas Y_1, \dots, Y_n deben ser independientes.

Pero, como se evidencia en el ejemplo anterior sobre el riesgo de cáncer en la piel, trabajar con un efecto aleatorio permite incorporar de forma natural la correlación entre las variables de respuesta observadas, siendo por lo tanto no todas mutuamente independientes entre sí. Un modelo mixto con efectos aleatorios tiene la gran ventaja de poder incluir componentes tales como los efectos individuales o de grupo.

Como dicen [Faraway \(2016, p. 195\)](#) y [Demidenko \(2013, p. 1\)](#), los datos agrupados surgen de forma muy natural y recurrente en la mayor parte de las áreas donde se usa a la estadística como ayuda para la comprensión de los fenómenos observados.

Estos tipos de datos tienen una estructura jerárquica, anidada, e incluyen datos longitudinales y/o espaciales en presencia de más de un grupo o factor de agrupación. El agrupamiento de observaciones puede producirse por el simple hecho de que se dispone de medidas repetidas para un mismo individuo o, en presencia de mediciones únicas para cada individuo, se observa que hay varios individuos que manifiestan alguna estructura grupal. Estas consideraciones ponen de manifiesto que en muchas aplicaciones no es realista suponer que las observaciones sean independientes, mientras que puede ser plausible asumir independencia entre las observaciones de grupos distintos, y que las observacio-

2.3. ¿Por qué agregar una parte aleatoria?

nes dentro de cada grupo sean dependientes dado pertenecen a la misma subpoblación. En consecuencia, todos los análisis que se fundamentan en el supuesto de independencia completa entre las observaciones individuales pueden resultar inapropiados. Sin embargo, el uso de efectos aleatorios es una forma común y conveniente de modelizar formas de dependencia, mostrándose satisfactorio para modelar la estructura de agrupación jerárquica de individuos.

A modo de ejemplo, como se puede ver en Demidenko (2013, p. 3), un estudio de precios puede tener dos resultados muy distintos entre sí, si se tienen en cuenta o no los efectos aleatorios.

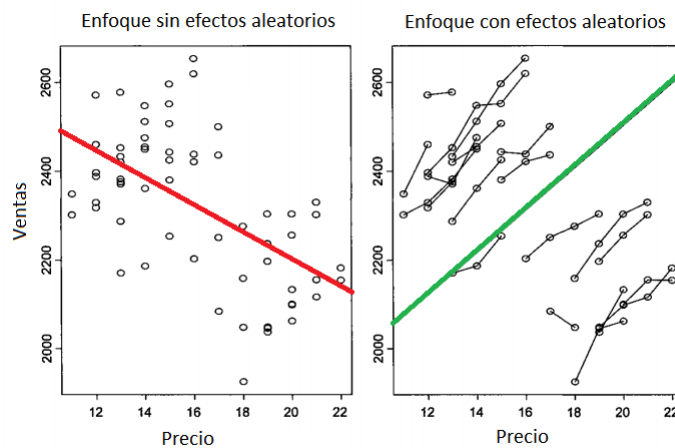


Figura 2.1: Comparación de resultados con y sin efectos aleatorios

Si se supone que todos los valores de precios son independientes, un modelo sin efectos aleatorios (2.1, gráfico a la izquierda) permite deducir que a un aumento en el precio queda asociada una disminución en las ventas. En cambio, un modelo con efectos aleatorios en el cual se consideren los datos agrupados (2.1, gráfico a la derecha), permite observar que en la mayoría de los grupos se manifiesta la tendencia que, a una suba del precio, queda asociado un aumento de las ventas.

En términos generales, en la literatura se hace referencia a modelos para medidas repetidas o incluso modelos jerárquicos, cuyo uso frecuente involucra:

- modelar datos agrupados o longitudinales;

- modelar datos con múltiples fuentes de variación;
- modelar la variedad biológica y la heterogeneidad;
- favorecer un compromiso entre el enfoque frecuentista y bayesiano;
- estimar los parámetros mediante criterios de log-verosimilitud penalizada;
- proveer una base teórica para el bien conocido criterio de información de Akaike (AIC por sus siglas en inglés);
- el tratamiento de parámetros multidimensionales;
- resolver problemas mal planteados, incluidos problemas de reconstrucción de imágenes y modelar formas e imágenes.

Como se puede ver, la utilidad de la amplia familia de modelos mixtos es muy variada. Una breve reseña histórica sobre los modelos mixtos se puede consultar en [Gbur et al. \(2012, §1.3\)](#) y una reseña más detallada en [Searle et al. \(2006, Cap. 2\)](#).

2.4. Problema de tener una parte aleatoria en el modelo

Las estimaciones por máxima verosimilitud, al depender de integrales multivariadas en altas dimensiones, sin posibilidad de evaluarlas en forma exacta se vuelven un desafío importante. Gracias al avance en potencia de cálculo computacional derivado de la revolución tecnológica de las últimas décadas y a mejoras constantes en algoritmos, este muro hoy en día se puede saltar con cierta facilidad en muchos casos.

En [Zipunnikov y Booth \(2006, §2.2\)](#) y [Zipunnikov \(2009, §1.3.2\)](#) se trabaja con datos longitudinales correspondiente a 121 personas mayores inscriptos en un plan de salud en Minnesota. Los datos consisten en la cantidad de veces que cada paciente visitó o llamó a la clínica médica en cada uno de los cuatro períodos de seis meses.

- Y_{ikl} denota la cantidad para el individuo i , evento k (visita o llamada), período l .

2.4. Problema de tener una parte aleatoria en el modelo

- Es natural considerar al individuo como un efecto aleatorio. En cambio al evento y el período es más natural considerarlos como efectos fijos.
- Se considera un modelo loglineal Poisson con $Y_{ikl} | \mathbf{u}_i^\Sigma \sim \text{Poisson}(\mu_{ikl})$.
- $\log(\mu_{ikl}) = a_0 + a_k + b_l + c_{kl} + \gamma_i + v_{ik} + w_{il}$ con $i = 1, \dots, 121$, $k = 1, 2$ y $l = 1, 2, 3, 4$.
- En el modelo anterior se tiene que a_0 es el término independiente, a_k es el efecto fijo del evento k , b_l es el efecto fijo del período l , c_{kl} es el efecto fijo de la interacción **evento** \times **período**, γ_i es el efecto aleatorio asociado al individuo i , v_{ik} es el efecto aleatorio de la interacción **individuo** \times **evento** y w_{il} es el efecto aleatorio de la interacción **individuo** \times **período**. Por lo tanto, el modelo involucra un vector de efectos aleatorios \mathbf{u}_i^Σ de longitud 7 siendo $\mathbf{u}_i^\Sigma = (\gamma_i, v_{i1}, v_{i2}, w_{i1}, w_{i2}, w_{i3}, w_{i4})$ con $i = 1, \dots, 121$. Cada individuo tiene asociado un vector aleatorio \mathbf{u}_i^Σ .
- Se supone que $\mathbf{u}_i^\Sigma \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \Sigma)$ donde $\mathbf{0} = (0, 0, 0, 0, 0, 0, 0)^T$ y Σ es la matriz diagonal $\Sigma = \text{diag}(\sigma_\gamma^2, \sigma_v^2, \sigma_v^2, \sigma_w^2, \sigma_w^2, \sigma_w^2, \sigma_w^2)$.
- A los efectos de estimar los parámetros del modelo se considera la restricción $a_2 = b_4 = c_{14} = c_{21} = c_{22} = c_{23} = c_{24} = 0$ en los efectos fijos para que el modelo sea identificable. Esto implica que la matriz de información de Fisher es no singular y por lo tanto su inversa exista, que es la hessiana. Luego se tiene que el vector de parámetros de los efectos fijos es $\beta = (a_0, a_1, b_1, b_2, b_3, c_{11}, c_{12}, c_{13})^T$.
- Para adaptar el modelo a la forma $g(\mu_{ij}) = \mathbf{x}_i^T + \mathbf{z}_i^T \mathbf{u}$, se elimina el doble índice kl construyendo un nuevo índice j como $j = 4(k - 1) + l$. Esto hace que $(Y_{i1}, \dots, Y_{i4}, Y_{i5}, \dots, Y_{i8})^T = (Y_{i11}, \dots, Y_{i14}, Y_{i21}, \dots, Y_{i24})^T$ y $(\mu_{i1}, \dots, \mu_{i4}, \mu_{i5}, \dots, \mu_{i8})^T = (\mu_{i11}, \dots, \mu_{i14}, \mu_{i21}, \dots, \mu_{i24})^T$. Como se puede comprobar el nuevo índice j no es más que concatenar para cada individuo los datos en un único vector donde las primeras cuatro componentes son las referidas a $k = 1$ y las restantes cuatro a $k = 2$.
- Se define $\mathbf{x}_{ij} = (1, \mathbb{1}_{\{1,2,3,4,5\}}(j), \mathbb{1}_{\{1,5\}}(j), \mathbb{1}_{\{2,6\}}(j), \mathbb{1}_{\{3,7\}}(j), \mathbb{1}_{\{1\}}(j), \mathbb{1}_{\{2\}}(j), \mathbb{1}_{\{3\}}(j))$
- Se define $\mathbf{z}_{ij} = (1, \mathbb{1}_{\{1,2,3,4\}}(j), \mathbb{1}_{\{5,6,7,8\}}(j), \mathbb{1}_{\{1,5\}}(j), \mathbb{1}_{\{2,6\}}(j), \mathbb{1}_{\{3,7\}}(j), \mathbb{1}_{\{4,8\}}(j))$

Con lo anterior se tiene el modelo completo en su forma $g(\mu_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{u}_i^\Sigma$.

En base a esto y una reparametrización que se puede ver Demidenko (2013, p. 414) se obtiene una verosimilitud levemente diferente pero más conveniente para trabajar como muestra Zipunnikov y Booth (2006, p. 4) o Zipunnikov (2009, p. 14):

$$L(\boldsymbol{\psi} | \mathbf{Y}) = \prod_{i=1}^{121} \left(\left(\frac{1}{2\pi} \right)^{\frac{7}{2}} \left(\prod_{j=1}^8 \frac{1}{y_{ij}!} \right) \int_{\mathbb{R}^7} \exp \left(- \sum_{j=1}^8 \exp(\mathbf{x}_{ij}^T \boldsymbol{\psi}) + \sum_{j=1}^8 y_{ij} \mathbf{x}_{ij}^T \boldsymbol{\psi} - \frac{1}{2} \mathbf{u}_i^T \mathbf{u}_i \right) d\mathbf{u}_i \right)$$

Donde $\boldsymbol{\psi} = (\boldsymbol{\beta}^T, \boldsymbol{\sigma}^T) = (a_0, a_1, b_1, b_2, b_3, c_{11}, c_{12}, c_{13}, \sigma_\gamma, \sigma_v, \sigma_w)^T$ y $\mathbf{x}_{ij} = (\boldsymbol{\mathbf{x}}_{ij}^T, \boldsymbol{\xi}_{ij}^T)^T$ con $\boldsymbol{\xi}_{ij}^T = (z_{ij1}u_{i1}, z_{ij2}u_{i2} + z_{ij3}u_{i3}, z_{ij4}u_{i4} + z_{ij5}u_{i5} + z_{ij6}u_{i6} + z_{ij7}u_{i7})^T$.

Como se puede apreciar, luego de desarrollos de rutina y reparametrizaciones se llega a que

$$\int_{\mathbb{R}^7} \exp \left(- \sum_{j=1}^8 \exp(\mathbf{x}_{ij}^T \boldsymbol{\psi}) + \sum_{j=1}^8 y_{ij} \mathbf{x}_{ij}^T \boldsymbol{\psi} - \frac{1}{2} \mathbf{u}_i^T \mathbf{u}_i \right) d\mathbf{u}_i$$

no es posible evaluarla analíticamente de forma exacta y esto conlleva a buscar métodos de aproximación para las mismas.

Capítulo 3

Metodología

3.1. El modelo lineal generalizado mixto (GLMM)

Siguiendo a [McCulloch \(2003, p. 31\)](#), un GLMM en su versión más general se especifica de la siguiente forma:

$$\left\{ \begin{array}{l} \mathbf{Y}_i | \mathbf{u} \sim f_{\mathbf{Y}_i | \mathbf{u}}(\mathbf{y}_i | \mathbf{u}, \boldsymbol{\beta}, \boldsymbol{\phi}), \text{ independientes, } \forall i \in \{1, \dots, n\} \\ \mathbb{E}(\mathbf{Y}_i | \mathbf{u}) = \boldsymbol{\mu}_i \\ g(\boldsymbol{\mu}_i) = \boldsymbol{\eta}_i = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{u} \\ \mathbf{u} \sim f_{\mathbf{U}}(\mathbf{u} | \mathbf{D}) \end{array} \right.$$

\mathbf{Y} consiste en el vector de datos observados, donde además se asume la existencia de un vector de efectos aleatorios \mathbf{u} . Se asume que la distribución condicional de \mathbf{Y}_i dado el \mathbf{u} es un miembro de la familia exponencial (ver [A.1](#)).

Luego, de forma análoga al modelo lineal generalizado, se transforma $\mathbb{E}(\mathbf{Y}_i | \mathbf{u})$ mediante la función g , llamada función link. Como se puede ver en [McCullagh y Nelder \(1989, §2.2.4\)](#), las distribuciones de la tabla [A.1](#) tienen una función link particular llamada link canónico y se da cuando se toma $g(\boldsymbol{\mu}_i) = \boldsymbol{\theta}$, donde $\boldsymbol{\theta}$ es el parámetro canónico que aparece

en la tabla A.1. Este link canónico es el que se utiliza habitualmente cuando se trabaja con alguna de esas distribuciones.

Finalmente y a diferencia de la especificación de un modelo GLM, aquí se asume que el predictor lineal consiste de dos componentes. Una parte formada por los efectos fijos, modelada por $\mathbf{x}_i^T \boldsymbol{\beta}$ y otra parte formada por los efectos aleatorios, modelada por $\mathbf{z}_i^T \mathbf{u}$, para la cual se asigna una distribución a \mathbf{u} . Esto indica que \mathbf{u} es una variable aleatoria y además no hay restricción para su distribución, ya que la misma depende del experimento a modelar y de la información que se tenga del mismo.

A modo general, la forma de especificar un GLMM depende de cuatro cuestiones:

1. ¿Cuál es la distribución de los datos para valores fijos de los predictores?
2. ¿Qué aspecto del problema se desea modelar?
3. ¿Cuáles son los predictores que serán incluidos en el modelo?
4. ¿A cuáles predictores se les asignará una distribución?

Como se recalcó anteriormente, el ítem 4 es el agregado que hace la diferencia entre un GLM y un GLMM.

3.1.1. Verosimilitud en un GLMM

Siguiendo a McCulloch (2003, §7.3), para la especificación que se muestra en 3.1 se obtiene que la verosimilitud es:

$$L(\boldsymbol{\beta}, \boldsymbol{\phi}, \mathbf{D}) = \int \left[\prod_{i=1}^n f_{Y_i|\mathbf{u}}(y_i|\mathbf{u}, \boldsymbol{\beta}, \boldsymbol{\phi}) \right] f_{\mathbf{U}}(\mathbf{u}|\mathbf{D}) d\mathbf{u}$$

Debido a que las especificaciones en 3.1 son las más generales a considerar, la forma de la verosimilitud que se mostró anteriormente es la más general posible. En este caso se deben encontrar estimaciones para los parámetros $\boldsymbol{\beta}$, $\boldsymbol{\phi}$ y \mathbf{D} .

En el presente trabajo y siguiendo a Knudson (2016), se expone un método para aproximar las estimaciones por máxima verosimilitud. Siendo un método basado en simulaciones del

tipo Monte Carlo es conveniente poder asegurar que el error ocasionado por el mecanismo de muestreo sea finito (estos detalles se verán más adelante). En consecuencia, en Knudson (2016, §1.1) se dan hipótesis de trabajo que restringen la especificación de los GLMM y se puede asegurar que el error ocasionado por el mecanismo de muestreo siempre es finito. Además, estas hipótesis simplifican la estructura de la verosimilitud teniendo menos parámetros para estimar. También se ajusta la notación de las densidades tanto de los efectos aleatorios como la respuesta dados los efectos aleatorios. Esta reescritura deja como subíndice al parámetro que aporta cada densidad y que luego debe ser estimado. Por último, el nombre de la variable de la función indica a que variable aleatoria hace referencia sin necesidad de especificarla como subíndice.

Entonces, por un lado solo se va a trabajar con GLMM donde la respuesta, dado los efectos aleatorios, tiene distribución Bernoulli o Poisson. Debido a esta restricción la verosimilitud L depende de menos variables. Como se puede ver en A.1, $a(\phi) = 1$ y por lo tanto se trabaja sin el parámetro de dispersión ϕ ya que no es relevante. La densidad conjunta del vector respuesta dado los efectos aleatoria, $\mathbf{Y}|\mathbf{u}$, se reescribe como $f_{\beta}(\mathbf{y}|\mathbf{u})$.

Por otro lado, se considera un vector K -dimensional ν cuyas componentes son no negativas y ν se llama componentes de varianza. En la descripción 3.1, la matriz \mathbf{D} es la matriz de varianzas y no se restringe la forma de la misma. Aunque, como dice Knudson (2016, § 1.1), ahora se considera que la matriz \mathbf{D} es diagonal cuyas entradas son las entradas del vector ν . Considerar a la matriz \mathbf{D} como diagonal tiene la consecuencia importante de que los efectos aleatorios individuales son independientes unos de otros. También se asume que los efectos aleatorios siempre siguen una distribución normal de media $\mathbf{0}$ y de varianza \mathbf{D} y se reescribe como $f_{\nu}(\mathbf{u})$. Si bien el parámetro original es la matriz \mathbf{D} , al ser una matriz diagonal los únicos valores desconocidos están en la diagonal. Por lo tanto, el parámetro ν no contiene información redundante y se elige para representar lo desconocido a estimar.

Con lo expuesto anteriormente, el parámetro ϕ no se considera, por lo tanto no necesita ser estimado. Los parámetros β y ν son los parámetros que se buscan estimar. Por lo tanto,

haciendo $\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\nu} \end{bmatrix}$, la densidad conjunta del vector respuesta y los efectos aleatorios se puede expresar como

$$f_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{y}) = f_{\boldsymbol{\beta}}(\mathbf{y}|\mathbf{u})f_{\boldsymbol{\nu}}(\mathbf{u})$$

y la log-verosimilitud es $l(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta})$. Otra notación aceptada para la verosimilitud es $L(\boldsymbol{\theta}|\mathbf{y})$ donde aquí se especifica que es una función, dado los datos \mathbf{y} observados. Por lo tanto, la log-verosimilitud en este caso hereda esa notación y se escribe $l(\boldsymbol{\theta}|\mathbf{y}) = \log L(\boldsymbol{\theta}|\mathbf{y})$. Esta notación no es necesaria aunque es más explícita en notar que es lo asumido como dado.

Con esta notación, [Knudson \(2016\)](#) desarrolla un método eficiente para trabajar con la verosimilitud escrita de la siguiente forma:

$$L(\boldsymbol{\theta}|\mathbf{y}) = \int f_{\boldsymbol{\beta}}(\mathbf{y}|\mathbf{u})f_{\boldsymbol{\nu}}(\mathbf{u})d\mathbf{u} = \int \left[\prod_{i=1}^n f_{\boldsymbol{\beta}}(y_i|\mathbf{u}) \right] f_{\boldsymbol{\nu}}(\mathbf{u})d\mathbf{u}$$

3.2. Método MCLA de Knudson-Geyer para GLMM

En lo que sigue y como se expuso en [3.1](#), se trabaja únicamente con distribuciones condicionales $\mathbf{Y}_i|\mathbf{u}$ Bernoulli o Poisson, donde además se supone que los efectos aleatorios \mathbf{u} se consideran normales e independientes. Estas restricciones de trabajo derivan directamente de las condiciones que necesita para las propiedades asintóticas trabajadas en [Knudson \(2016\)](#).

Como se puede ver en [Knudson \(2016, §2.4\)](#), MCLA es un método basado en simulaciones del tipo Monte Carlo para aproximar la función de verosimilitud. Para más detalles sobre integración Monte Carlo se puede consultar [A.5](#). En particular, usa un muestreo de importancia con una distribución adecuada al problema que quiere resolver. Los detalles de muestreo de importancia se pueden consultar en [A.6](#). El método MCLA tiene un amplio respaldo bibliográfico, siendo la mayoría de los artículos disponibles en coautoría con

Charles J. Geyer comenzando con su tesis doctoral (Geyer, 1990).

El método MCLA es muy apreciado porque permite hacer todo tipo de inferencia basada en la verosimilitud. En particular, la máxima verosimilitud es de especial interés para encontrar el máximo de la aproximación de la verosimilitud por Monte Carlo. Este se llama el estimador por máxima verosimilitud en Monte Carlo, MCMLE por sus siglas en inglés.

Dentro de los trabajos más recientes, el método MCLA para modelos GLMM han sido estudiados en Sung y Geyer (2007) donde los autores se han enfocado en una implementación de MCLA con una distribución para el muestreo de importancia construida independientemente de los datos observados. En este contexto, en Sung y Geyer (2007) se dan condiciones bajo las cuales el MCMLE es asintóticamente normal y calculan la varianza del MCMLE, teniendo en cuenta tanto el error de muestreo habitual (relacionado al tamaño de la muestra observada) y el error Monte Carlo (relacionado al tamaño de la muestra Monte Carlo). También dan una distribución aproximada para el MLE, que puede ser usada para el MCMLE cuando el tamaño de muestra Monte Carlo es grande. Y también discuten el rol de la distribución del muestreo de importancia en el cálculo de la varianza del MCMLE. Además, acompañando el trabajo Sung y Geyer (2007), los autores implementaron la librería **bernor** (Sung y Geyer, 2018) para R (R Core Team, 2021). Su librería usa un muestreo de importancia con la mencionada distribución que no depende de los datos. Esta librería permite calcular estimaciones por máxima verosimilitud para modelos GLMM con respuesta Bernoulli. Aunque solo era funcional con modelos simples. Por ejemplo, con un modelo muy conocido como el de las salamandras, disponible en Knudson (2016, §1.3) y Karim y Zeger (1992), ya no era capaz de encontrar con precisión la estimación por máxima verosimilitud de los parámetros en cuestión como se puede ver en Sung y Geyer (2007, p. 1001).

Esto expuso la necesidad de encontrar una distribución para el muestreo de importancia que funcione bien tanto teóricamente como computacionalmente en la práctica, no sólo en modelos sencillos. Dado que los teoremas y cálculos mencionados de Sung y Geyer (2007) son únicamente válidos cuando la distribución es construida de forma independiente al

tamaño de la muestra observada, todos esos temas fueron revisados en Knudson (2016) y adaptados a una distribución que depende de los datos observados. Además, la autora basándose en sus resultados obtenidos en Knudson (2016) construye la librería **glimm** (Knudson, 2020) para R (R Core Team, 2021) con la cual es posible obtener estimaciones por máxima verosimilitud para problemas no triviales y donde la respuesta dado los efectos aleatorios puede tener distribución Bernoulli o Poisson.

3.2.1. Algunos cálculos y consideraciones previas

Para los cálculos relacionados con MCLA se necesita la densidad conjunta de los datos y los efectos aleatorios $f_{\theta}(\mathbf{u}, \mathbf{y}) = f_{\beta}(\mathbf{y}|\mathbf{u})f_{\nu}(\mathbf{u})$. El objetivo es construir una densidad \tilde{f} para usar como densidad de un muestreo de importancia. Esta densidad \tilde{f} a diferencia del trabajo anterior Sung y Geyer (2007), depende de los datos observados para su mejor desempeño.

Entonces, como se especificó en A.6, la forma de hacer el muestreo de importancia con \tilde{f} se reduce a simular vectores q -dimensionales $\mathbf{u}_1, \dots, \mathbf{u}_m$ de \tilde{f} donde \tilde{f} no depende de θ . Entonces la aproximación de la log-verosimilitud por Monte Carlo es:

$$l_m(\theta|\mathbf{y}) = \log \left(\frac{1}{m} \sum_{i=1}^m \frac{f_{\theta}(\mathbf{u}_k, \mathbf{y})}{\tilde{f}(\mathbf{u}_k)} \right)$$

Esta función l_m puede ser usada para inferencia basada en la verosimilitud y el valor que la maximiza es el MCMLE.

Como se puede ver en Knudson (2016, §3.1), de la función l_m se obtiene el gradiente ∇l_m y la hessiana $\nabla^2 l_m$:

$$\nabla l_m(\theta|\mathbf{y}) = \frac{\sum_{k=1}^m [\nabla \log f_{\theta}(\mathbf{u}_k, \mathbf{y})] \frac{f_{\theta}(\mathbf{u}_k, \mathbf{y})}{\tilde{f}(\mathbf{u}_k)}}{\sum_{k=1}^m \frac{f_{\theta}(\mathbf{u}_k, \mathbf{y})}{\tilde{f}(\mathbf{u}_k)}}$$

$$\begin{aligned} \blacksquare \nabla^2 l_m(\boldsymbol{\theta}|\mathbf{y}) &= \frac{\sum_{k=1}^m [\nabla^2 \log f_{\boldsymbol{\theta}}(\mathbf{u}_k, \mathbf{y})] \frac{f_{\boldsymbol{\theta}}(\mathbf{u}_k, \mathbf{y})}{\tilde{f}(\mathbf{u}_k)}}{\sum_{k=1}^m \frac{f_{\boldsymbol{\theta}}(\mathbf{u}_k, \mathbf{y})}{\tilde{f}(\mathbf{u}_k)}} + \\ &+ \frac{\sum_{k=1}^m [\nabla \log f_{\boldsymbol{\theta}}(\mathbf{u}_k, \mathbf{y}) - \nabla l_m(\boldsymbol{\theta}|\mathbf{y})][\nabla \log f_{\boldsymbol{\theta}}(\mathbf{u}_k, \mathbf{y}) - \nabla l_m(\boldsymbol{\theta}|\mathbf{y})]^T \frac{f_{\boldsymbol{\theta}}(\mathbf{u}_k, \mathbf{y})}{\tilde{f}(\mathbf{u}_k)}}{\sum_{k=1}^m \frac{f_{\boldsymbol{\theta}}(\mathbf{u}_k, \mathbf{y})}{\tilde{f}(\mathbf{u}_k)}} \end{aligned}$$

Como $\log f_{\boldsymbol{\theta}}(\mathbf{u}_k, \mathbf{y}) = \log f_{\boldsymbol{\beta}}(\mathbf{y}|\mathbf{u}_k) + \log f_{\nu}(\mathbf{u}_k)$, entonces se requieren expresiones para $\log f_{\boldsymbol{\beta}}(\mathbf{y}|\mathbf{u})$, $\log f_{\nu}(\mathbf{u})$ y las primeras dos derivadas de ambas funciones. Estos cálculos son necesarios para poder expresar tanto el gradiente ∇l_m y la hessiana $\nabla^2 l_m$.

Densidad de los efectos aleatorios: Siguiendo Knudson (2016, §3.3), para $t \in \{1, 2, \dots, K\}$ se considera \mathbf{E}_t una matriz diagonal con unos y ceros en la diagonal, siendo $\mathbf{I}_q = \sum_{t=1}^K \mathbf{E}_t$, donde \mathbf{I}_q es la matriz identidad de dimensión $q \times q$. También se debe cumplir que $\mathbf{D} = \sum_{t=1}^K \nu_t \mathbf{E}_t$. La diagonal de las matrices \mathbf{E}_t indica que efectos aleatorios tienen varianza ν_t .

Luego se definen las cantidades q_t como $q_t = \text{traza}(\mathbf{E}_t)$. Por lo tanto q_t indica la cantidad de efectos aleatorios asociados con la componente de varianza ν_t . Como en el modelo GLMM a tratar el vector de efectos aleatorios \mathbf{U} es q -dimensional, entonces se tiene que $q = \sum_{t=1}^K q_t$.

Con la matriz de varianza \mathbf{D} de los efectos aleatorios, las componentes de varianzas ν_t y los valores q_t por construcción, se tiene la siguiente igualdad: $\det(\mathbf{D}) = \nu_1^{q_1} \cdots \nu_K^{q_K}$.

La log-densidad de los efectos aleatorios es entonces:

$$\log f_{\nu}(\mathbf{u}) - \frac{q}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^K q_t \log(\nu_t) - \frac{1}{2} \sum_{t=1}^K \frac{1}{\nu_t} \mathbf{u}^T \mathbf{E}_t \mathbf{u}$$

Además

$$\begin{aligned} \partial_{\nu_t} \log f_{\nu}(\mathbf{u}) &= -\frac{q_t}{2\nu_t} + \frac{1}{2\nu_t^2} \mathbf{u}^T \mathbf{E}_t \mathbf{u} \\ \partial_{\nu_t \nu_s}^2 \log f_{\nu} &= \begin{cases} -\frac{q_t}{2\nu_t^2} - \frac{1}{\nu_t^3} \mathbf{u}^T \mathbf{E}_t \mathbf{u}, & \text{si } s = t \\ 0, & \text{si } s \neq t \end{cases} \end{aligned}$$

Densidad de los datos dados los efectos aleatorios: Como se puede ver en [Knudson \(2016, §3.4\)](#), esta expresión sale directamente de la especificación general del GLMM. La distribución condicional $f(\mathbf{y}|\mathbf{u})$ es de la familia exponencial, usando la notación que se encuentra en [A.1](#). En particular interesa los casos Bernoulli y Poisson.

El parámetro canónico que en [A.1](#) se anota como θ , a partir de ahora será η . Y la función $b(\theta)$ a partir de ahora será $c(\eta)$. Esto último es importante para no confundir con la función $c(y, \phi)$ que aparece en [A.1](#). Además, para los casos Bernoulli y Poisson se cumple que $a(\phi) = 1$. Con estas consideraciones queda que:

$$\log f_{\beta}(\mathbf{y}|\mathbf{u}) = \mathbf{y}^T \boldsymbol{\eta} - c(\boldsymbol{\eta})$$

$$\partial_{\beta} \log f_{\beta}(\mathbf{y}|\mathbf{u}) = \mathbf{X}^T [\mathbf{y} - c'(\boldsymbol{\eta})]$$

$$\partial_{\beta}^2 \log f_{\beta}(\mathbf{y}|\mathbf{u}) = -\mathbf{X}^T c''(\boldsymbol{\eta}) \mathbf{X}$$

Siguiendo a [Knudson \(2016, §3.5\)](#), la función $c(\boldsymbol{\eta})$ se expresa como $c(\boldsymbol{\eta}) = \sum_{i=1}^n c_i(\eta_i)$ de donde $c'(\boldsymbol{\eta})$ es el gradiente de $c(\boldsymbol{\eta})$ respecto a la variable $\boldsymbol{\eta}$ y $c''(\boldsymbol{\eta})$ es la hessiana. Por lo tanto $c'(\boldsymbol{\eta})$ es un vector cuya entrada i es $c'_i(\eta_i)$ y $c''(\boldsymbol{\eta})$ es una matriz diagonal cuya entrada i en la diagonal es $c''_i(\eta_i)$.

Las expresiones de $c_i(\eta_i)$, $c'_i(\eta_i)$ y $c''_i(\eta_i)$ son:

Caso Bernoulli

$$\begin{cases} c_i(\eta_i) = \log(1 + e^{\eta_i}) \\ c'_i(\eta_i) = \frac{e^{\eta_i}}{1 + e^{\eta_i}} \\ c''_i(\eta_i) = \frac{e^{\eta_i}}{1 + e^{\eta_i}} - \frac{e^{2\eta_i}}{(1 + e^{\eta_i})^2} \end{cases}$$

Caso Poisson

$$\begin{cases} c_i(\eta_i) = e^{\eta_i} \\ c'_i(\eta_i) = e^{\eta_i} \\ c''_i(\eta_i) = e^{\eta_i} \end{cases}$$

3.2.2. La densidad \tilde{f} usada para el muestreo de importancia

Para el método MCLA se toma lo siguiente:

- \mathbf{s} un vector de longitud q , el cual representa los efectos aleatorios en la escala normal estándar. O sea, \mathbf{s} es definido de tal forma que $\mathbf{u} = \mathbf{D}^{\frac{1}{2}}\mathbf{s}$.
- $\boldsymbol{\sigma}$ un vector de longitud K , con componentes $\sqrt{v_t}$ para $t = 1, \dots, K$.
- $\boldsymbol{\beta}^*$, \mathbf{s}^* y $\boldsymbol{\sigma}^*$, siendo estas las estimaciones PQL para $\boldsymbol{\beta}$, \mathbf{s} y $\boldsymbol{\sigma}$ respectivamente. Los detalles sobre como se obtienen estimaciones PQL se pueden consultar en A.2.
- $\mathbf{A}^* = \sum_{i=1}^K \mathbf{E}_t \boldsymbol{\sigma}_t^*$ y $\mathbf{D}^* = \mathbf{A}^* \mathbf{A}^*$. Ahora se desestandariza la estimación PQL \mathbf{s}^* de los efectos aleatorios y se obtiene $\mathbf{u}^* = \mathbf{A}^* \mathbf{s}^*$.
- p_1, p_2 y p_3 proporciones tales que $p_1 + p_2 + p_3 = 1$.
- $f(\mathbf{u}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, la densidad de una normal multivariada con media $\boldsymbol{\mu}$ y matriz de varianza $\boldsymbol{\Sigma}$.
- $\tilde{f}(\mathbf{u}|\mathbf{0}, \mathbf{D}^*)$ la densidad t multivariada a valores en \mathbb{R}^q con media $\mathbf{0}$, matriz de escala \mathbf{D}^* y 5 grados de libertad. Por más información sobre esta distribución se puede consultar A.3.

Ahora se construye la densidad para el muestreo de importancia como:

$$\tilde{f}(\mathbf{u}) = p_1 \tilde{f}(\mathbf{u}|\mathbf{0}, \mathbf{D}^*) + p_2 f(\mathbf{u}|\mathbf{u}^*, \mathbf{D}^*) + p_3 f(\mathbf{u}|\mathbf{u}^*, (\mathbf{Z}^T c''(\mathbf{X}\boldsymbol{\beta}^* + \mathbf{Z}\mathbf{u}^*)\mathbf{Z} + (\mathbf{D}^*)^{-1})^{-1})$$

La primer componente $\tilde{f}(\cdot|\mathbf{0}, \mathbf{D}^*)$ se toma así para asegurar que $\nabla l_m(\boldsymbol{\theta}|\mathbf{y})$, el gradiente de MCLA, tenga un teorema del límite central (ver 3.3.1).

La segunda componente se toma porque esta centrada en \mathbf{u}^* y tiene varianza \mathbf{D}^* .

La última componente esta centrada en \mathbf{u}^* tiene varianza basada en la hessiana de la verosimilitud penalizada de PQL (ver A.2). Más específicamente, la hessiana del logaritmo de la densidad de la última distribución coincide con la hessiana del logaritmo de la densidad de la distribución objetivo $f_{\theta}(\mathbf{u}, \mathbf{y})$.

3.2.3. Fuentes de errores

Como se puede ver en Knudson (2016, §3.2), el método MCLA para GLMM tiene dos tipos de errores.

El primero es debido a la variabilidad del MLE entorno al verdadero valor del parámetro y se lo denomina “**error de muestreo**”. Si $\hat{\theta}$ es el MLE calculado con una muestra de tamaño n y θ^* es el verdadero valor del parámetro, entonces esta fuente de variabilidad es la varianza de la distribución asintótica del $\sqrt{n}(\hat{\theta} - \theta^*)$. Es importante resaltar que al calcular el error de muestreo, los datos son tratados como aleatorios, siendo el MLE también aleatorio.

El segundo es debido a la variabilidad del MCMLE entorno al MLE y este error se lo denomina “**error Monte Carlo**”. Si $\hat{\theta}_m$ representa el MCMLE calculado con una muestra de tamaño m , entonces esta fuente de variabilidad es la varianza de la distribución asintótica de $\sqrt{m}(\hat{\theta}_m - \hat{\theta})$. Aquí es importante resaltar que para calcular el error Monte Carlo, a diferencia del error de muestreo, los datos son tratados como fijos. Esto implica que para el error Monte Carlo el MLE es fijo.

Para medir el error de muestreo se usa la información de Fisher. Si $J(\hat{\theta}_m)$ es la estimación de la información de Fisher observada, entonces

$$J(\hat{\theta}_m) = -\nabla^2 l_m(\hat{\theta}_m | \mathbf{y})$$

y un estimador consistente de la varianza asintótica del MLE es $(J(\hat{\theta}_m))^{-1}$, siempre y cuando se cumplan las condiciones de regularidad probadas en Geyer (2013).

Para medir el error Monte Carlo se usa que el gradiente de la log-verosimilitud en el MLE $\hat{\theta}$, el cual se muestra en la ecuación 3.1 y es

$$V = \frac{1}{\gamma_1^2} \int [\nabla \log f_{\hat{\theta}}(\mathbf{u}, \mathbf{y})][\nabla \log f_{\hat{\theta}}(\mathbf{u}, \mathbf{y})]^T \frac{(f_{\hat{\theta}}(\mathbf{u}, \mathbf{y}))^2}{\tilde{f}(\mathbf{u})} d\mathbf{u}$$

donde γ_1 se define en 3.3. Entonces se puede estimar V a partir del MCMLE $\hat{\theta}_m$ de la

siguiente manera:

$$\hat{V} = \frac{\frac{1}{m} \sum_{k=1}^m [\nabla \log f_{\hat{\theta}_m}(\mathbf{u}_k, \mathbf{y})][\nabla \log f_{\hat{\theta}_m}(\mathbf{u}_k, \mathbf{y})]^T \frac{(f_{\hat{\theta}_m}(\mathbf{u}_k, \mathbf{y}))^2}{\tilde{f}(\mathbf{u}_k)}}{\left(\frac{1}{m} \sum_{k=1}^m \frac{f_{\hat{\theta}_m}(\mathbf{u}_k, \mathbf{y})}{\tilde{f}(\mathbf{u}_k)} \right)^2}$$

Sean \hat{U} es la matriz hessiana del MCLA evaluada en el MCMLE y U es el límite de la matriz hessiana cuando el tamaño de la muestra Monte Carlo tiende a infinito. Entonces se tiene que el error Monte Carlo es $U^{-1}VU^{-1}$ y bajo las condiciones de [Geyer \(1994\)](#) el estimador plug-in es $\hat{U}^{-1}\hat{V}\hat{U}^{-1}$, donde $\hat{U} = \nabla^2 l_m(\hat{\theta}|\mathbf{y})$.

Dependiendo del muestreo de importancia que se proponga para el MCLA, es el error Monte Carlo que se va a tener. Como se puede ver en [A.5](#), se necesitan condiciones de regularidad para tener un error Monte Carlo finito. Como lo demuestra el teorema de la sección [3.3.1](#), la densidad \tilde{f} resulta de suma importancia ya que el error Monte Carlo resultante con esta densidad para la implementación de MCLA resulta siempre finito.

3.3. Algunos resultados importantes

Se fija \mathbf{y} en el valor observado y se define $\gamma_1 = f_{\theta}(\mathbf{y}) = \int f_{\theta}(\mathbf{u}, \mathbf{y})d\mathbf{u}$. Esta cantidad γ_1 es constante, finita y útil para algunos resultados.

1. Dados $\mathbf{a} \in \mathbb{R}^d$ y $b \in \mathbb{R}$. Si \mathbf{D} es una matriz de varianzas (esto es, simétrica y definida positiva), entonces la función $f : \mathbb{R}^d \rightarrow \mathbb{R}/\tilde{f}(\mathbf{u}) = \exp(-\mathbf{u}^T \mathbf{D}^{-1} \mathbf{u} + \mathbf{a}^T \mathbf{u} + b)$ es proporcional a una densidad normal con media $\frac{1}{2} \mathbf{D} \mathbf{a}$ y matriz de varianza $\frac{1}{2} \mathbf{D}$.
2. $E_{f_{\theta}}(\nabla \log f_{\theta}(\mathbf{u}, \mathbf{y})|\mathbf{y}) = \nabla l(\theta|\mathbf{y})$
3. **Ley de los grandes números para el gradiente de MCLA:**

$$\forall \theta : \nabla l_m(\theta|\mathbf{y}) \xrightarrow[m \rightarrow \infty]{\text{c.s.}} \nabla l(\theta|\mathbf{y})$$

4. Teorema del límite central para el numerador del gradiente de MCLA

La variable aleatoria

$$\frac{1}{m} \sum_{k=1}^m \nabla \log f_{\boldsymbol{\theta}}(\mathbf{u}_k, \mathbf{y}) \frac{f_{\boldsymbol{\theta}}(\mathbf{u}_k, \mathbf{y})}{\tilde{f}(\mathbf{u}_k)}$$

tiene teorema del límite central. Esto es

$$\sqrt{m} \left(\frac{1}{m} \sum_{k=1}^m \nabla \log f_{\boldsymbol{\theta}}(\mathbf{u}_k, \mathbf{y}) \frac{f_{\boldsymbol{\theta}}(\mathbf{u}_k, \mathbf{y})}{\tilde{f}(\mathbf{u}_k)} - \gamma_1 \nabla l(\boldsymbol{\theta}|\mathbf{y}) \right)$$

converge en distribución a una distribución normal con media $\mathbf{0}$ y varianza finita.

5. Teorema del límite central para el denominador del gradiente de MCLA

La variable aleatoria

$$\frac{1}{m} \sum_{k=1}^m \frac{f_{\boldsymbol{\theta}}(\mathbf{u}_k, \mathbf{y})}{\tilde{f}(\mathbf{u}_k)}$$

tiene teorema del límite central. Esto es

$$\sqrt{m} \left(\frac{1}{m} \sum_{k=1}^m \frac{f_{\boldsymbol{\theta}}(\mathbf{u}_k, \mathbf{y})}{\tilde{f}(\mathbf{u}_k)} - \gamma_1 \right)$$

converge en distribución a una distribución normal con media $\mathbf{0}$ y varianza finita.

3.3.1. TLC para el gradiente de MCLA

Todos los resultados de la sección anterior permiten probar el resultado más importante:

$\nabla l_m(\boldsymbol{\theta}|\mathbf{y})$ tiene teorema del límite central. Esto es,

$$\sqrt{m}(\nabla l_m(\boldsymbol{\theta}|\mathbf{y}) - \nabla l(\boldsymbol{\theta}|\mathbf{y}))$$

converge en distribución a una distribución normal con media $\mathbf{0}$ y varianza finita.

Demostración. Primero se reescribe la expresión

$$\begin{aligned} \sqrt{m}(\nabla l_m(\boldsymbol{\theta}|\mathbf{y}) - \nabla l(\boldsymbol{\theta}|\mathbf{y})) &= \sqrt{m} \left[\frac{\frac{1}{m} \sum_{k=1}^m [\nabla \log f_{\boldsymbol{\theta}}(\mathbf{u}_k, \mathbf{y})] \frac{f_{\boldsymbol{\theta}}(\mathbf{u}_k, \mathbf{y})}{\tilde{f}(\mathbf{u}_k)}}{\frac{1}{m} \sum_{k=1}^m \frac{f_{\boldsymbol{\theta}}(\mathbf{u}_k, \mathbf{y})}{\tilde{f}(\mathbf{u}_k)}} - \nabla l(\boldsymbol{\theta}|\mathbf{y}) \right] \\ &= \sqrt{m} \left[\frac{\frac{1}{m} \sum_{k=1}^m [\nabla \log f_{\boldsymbol{\theta}}(\mathbf{u}_k, \mathbf{y})] \frac{f_{\boldsymbol{\theta}}(\mathbf{u}_k, \mathbf{y})}{\tilde{f}(\mathbf{u}_k)} - \nabla l(\boldsymbol{\theta}|\mathbf{y}) \frac{1}{m} \sum_{k=1}^m \frac{f_{\boldsymbol{\theta}}(\mathbf{u}_k, \mathbf{y})}{\tilde{f}(\mathbf{u}_k)}}{\frac{1}{m} \sum_{k=1}^m \frac{f_{\boldsymbol{\theta}}(\mathbf{u}_k, \mathbf{y})}{\tilde{f}(\mathbf{u}_k)}} \right] \end{aligned}$$

Usando el ítem 4 de los resultados importantes, se puede afirmar que el numerador de la ecuación anterior converge a una distribución normal con media $\mathbf{0}$ y varianza finita, siendo

$$\text{Var} \left([\nabla \log f_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{y})] \frac{f_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{y})}{\tilde{f}(\mathbf{u})} \right)$$

su valor. Por la ley de los grandes números, el denominador cumple

$$\frac{1}{m} \sum_{k=1}^m \frac{f_{\boldsymbol{\theta}}(\mathbf{u}_k, \mathbf{y})}{\tilde{f}(\mathbf{u}_k)} \xrightarrow{\text{c.s.}} \mathbb{E}_{\tilde{f}} \left(\frac{f_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{y})}{\tilde{f}(\mathbf{u})} \right)$$

Con un cálculo directo se obtiene que

$$\mathbb{E}_{\tilde{f}} \left(\frac{f_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{y})}{\tilde{f}(\mathbf{u})} \right) = \int \frac{f_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{y})}{\tilde{f}(\mathbf{u})} \tilde{f}(\mathbf{u}) d\mathbf{u} = \int f_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{y}) d\mathbf{u} = \gamma_1$$

Por lo tanto, de lo anterior se deduce que

$$\frac{1}{m} \sum_{k=1}^m \frac{f_{\boldsymbol{\theta}}(\mathbf{u}_k, \mathbf{y})}{\tilde{f}(\mathbf{u}_k)} \xrightarrow{\text{c.s.}} \gamma_1$$

Por lo tanto, usando el teorema de Slutsky, se puede afirmar que la última ecuación converge en distribución a una distribución con media $\mathbf{0}$ y varianza

$$\frac{1}{\gamma_1^2} \text{Var} \left([\nabla \log f_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{y})] \frac{f_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{y})}{\tilde{f}(\mathbf{u})} \right)$$

Por último, como $\gamma_1 > 0$, entonces se puede afirmar que la varianza de la distribución normal a la que converge es finita. \square

En caso de ser necesario, se puede expresar la varianza del gradiente en términos de la siguiente integral

$$\frac{1}{\gamma_1} \int [\nabla \log f_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{y})] [\nabla \log f_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{y})]^T \frac{(f_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{y}))^2}{\tilde{f}(\mathbf{u})} d\mathbf{u} \quad (3.1)$$

Sabiendo que $\forall \boldsymbol{\theta} \forall \mathbf{y} : \sqrt{m}(\nabla l_m(\boldsymbol{\theta}|\mathbf{y}) - \nabla l(\boldsymbol{\theta}|\mathbf{y})) \xrightarrow[m \rightarrow \infty]{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, donde $\boldsymbol{\Sigma}$ es finita. Esto es análogo a probar que el gradiente de MCLA es consistente y asintóticamente normal. Estos resultados aseguran que los errores Monte Carlo detallados en 3.2.3 son finitos.

Capítulo 4

Datos utilizados para la aplicación

Los datos elegidos para trabajar son un interesante y particularmente desafiante conjunto de datos sobre apareamientos de salamandras. De acuerdo con [Karim y Zeger \(1992\)](#), son datos retadores porque la variable respuesta es binaria y el diseño del estudio es cruzado en vez de anidado. En un diseño cruzado como es el caso, todos los niveles de todos los factores se combinan entre sí, por lo tanto se observan todos los cruces posibles. Mientras que en un anidado solo se observan algunos de todos los cruces posibles. Eso hace que la verosimilitud sea particularmente difícil de evaluar.

Este conjunto de datos es ampliamente usado en la literatura hasta hoy en día como una referencia para comprobar la eficiencia de métodos computacionales en modelos GLMM. Prueba de ello son las referencias de [McCullagh y Nelder \(1989, §14.5\)](#), [Knudson \(2016, §1.3, §5.2\)](#), [Karim y Zeger \(1992\)](#) y [Shun \(1997\)](#) entre otras.

Los datos refieren a dos poblaciones de salamandras llamadas Rough Butt (RB) y White-side (WS). Cuarenta animales fueron usados en cada uno de los experimentos. El primer experimento fue realizado en el verano boreal de 1986 y los otros dos en el otoño boreal del mismo año. Las cuarenta salamandras para los experimentos se componen de diez hembras RB, diez hembras WS, diez machos RB y diez machos WS. En los dos primeros experimentos se usaron las mismas cuarenta salamandras mientras que para el último experimento se utilizaron otras salamandras. Como se puede ver en [McCullagh y Nelder](#)

(1989, p. 441), la práctica normal es usar nuevos animales para cada experimento porque repetir experimentos con los mismos animales es potencialmente importante. Pero, como es la expectativa de la práctica habitual, ese aspecto fue ignorado en el análisis de McCullagh y Nelder (1989, §4.5.3). De las diferentes preguntas y modelos que se puede abordar con estos datos, en el presente trabajo se sigue a Knudson (2016) y Karim y Zeger (1992) para estimar las probabilidades del apareamiento exitoso de cada una de las cuatro combinaciones de género y población {Población hembra · Población macho} que serán denotadas por $\{\pi_{RR}, \pi_{RW}, \pi_{WR}, \pi_{WW}\}$. El modelo usado es el tratado en Knudson (2016) que a su vez es el modelo A propuesto en Karim y Zeger (1992).

4.1. Diseño muestral

Como se explica en McCullagh y Nelder (1989, §14.5), el experimento fue realizado en 1986 por Stevan J. Arnold y Paul A. Verrell del Departamento de Ecología en la Universidad de Chicago en Estados Unidos. El experimento consiste en trabajar con dos poblaciones de salamandras: Rought Butt y Whiteside. En su hábitat natural a gran altura al sur de los montes Apalaches al este de Estados Unidos, estas dos poblaciones diferentes de salamandras están geográficamente aisladas una de otra. Por lo tanto, las salamandras emparejadas en el experimento nunca lo harían en su entorno natural, no al menos hasta 1986. Para el biólogo que estudia especiación es de gran interés saber como evoluciona el mecanismo que impide el mestizaje. Para poblaciones como las del experimento, el interés está en saber si esas barreras pueden evolucionar en poblaciones que están aisladas unas de otras.

Las 40 salamandras disponibles en cada experimento comprenden:

- 10 machos Rought Butt numerados de 1 a 10. El subconjunto de machos Rought Butt numerados de 1 a 5 se anotará RBM 1 ~ 5, mientras que los numerados de 6 a 10 se anotará RBM 6 ~ 10.
- 10 hembras Rought Butt numerados de 1 a 10. De forma análoga a los machos

Rought Butt, el subconjunto de hembras numeradas de 1 a 5 se anotará RBF 1 ~ 5 y las de 6 a 10 serán RBF 6 ~ 10.

- 10 machos Whiteside numerados numerados de 1 a 10. Análogamente a la población Rought Butt, el subconjunto de machos Whiteside numerados de 1 a 5 se anotará como WSM 1 ~ 5, mientras que los numerados de 6 a 10 se hará como WSM 6 ~ 10.
- 10 hembras Whiteside numerados de 1 a 10. Al igual que con los machos, el subconjunto de hembras numeradas de 1 a 5 se anotará WSF 1 ~ 5 y las numeradas de 6 a 10 se harán como WSF 6 ~ 10.

El diseño cruzado que se usa en cada uno de los tres experimentos para el emparejamiento de las salamandras es el siguiente:

Las salamandras RBF 1 ~ 5 fueron tomadas como parejas heterosexuales con salamandras RBM 1 ~ 5 en 3 ocasiones (día A, día A + 8, día A + 16) y con salamandras WSM 1 ~ 5 también en tres ocasiones (día A + 4, día A + 12, día A + 20). A las salamandras RBF 1 ~ 5 nunca se les permitió ver a las salamandras RBM 6 ~ 10 o las salamandras WSM 6 ~ 10. Por lo tanto, sólo 8 de los 16 posibles cruces entre subpoblaciones son permitidos.

Además, de los 25 cruces posibles entre las salamandras RBF 1 ~ 5 y por ejemplo, WSM 1 ~ 5, sólo se permiten 15 de esos cruces. Por ejemplo, la salamandra RBF 1 nunca ve a la salamandra WSM 2 o a la WSM 3. De la misma manera, por ejemplo, la salamandra WSM 1 nunca ve a la salamandra RBF 2 o a la RBF 4. Para que el diseño sea completo, hubiese sido necesario extender a 10 días y no a 6 como se hizo. Si el experimento se extiende por 10 días, cada subpoblación de 5 individuos se puede ver 5 veces con cada una de sus dos subpoblaciones asignadas para el apareamiento. Esto permite obtener, por ejemplo, que la salamandra RBF 1 se vea una vez con cada individuo de las subpoblaciones RBM 1 ~ 5 y WSM 1 ~ 5. De ahí, cada salamandra tiene 10 encuentros. En vez de eso, se optó para que lo hagan con 3 individuos de cada subpoblación y cada salamandra tiene sólo 6 encuentros por experimento. Por lo tanto, el experimento en su conjunto comprende 8 réplicas de un

CAPÍTULO 4. DATOS UTILIZADOS PARA LA APLICACIÓN

Hembras	Machos											
	Día A	Día A + 4		Día A + 8		Día A + 12		Día A + 16		Día A + 20		
1	1	4		5		1		4		5		
2	5	5		3		3		1		2		
RBF 3	RB 2	WS 1	RB 1	WS 4	RB 3	WS 3						
4	4	2		2		5		5		4		
5	3	3		4		2		2		1		
6	9	9		10		7		6		8		
7	8	8		9		9		7		6		
RBF 8	WS 6	RB 6	WS 7	RB 10	WS 10	RB 9						
9	10	7		8		6		9		10		
10	7	10		6		8		8		7		
1	9	9		7		10		10		8		
2	7	6		9		7		6		10		
WSF 3	RB 8	WS 7	RB 6	WS 9	RB 7	WS 6						
4	10	10		8		8		9		9		
5	6	8		10		6		8		7		
6	5	2		3		4		2		1		
7	4	1		5		2		1		5		
WSF 8	WS 1	RB 4	WS 2	RB 5	WS 5	RB 3						
9	3	3		1		1		4		4		
10	2	5		4		3		3		2		

Tabla 4.1: Diseño usado en el experimento de apareamiento de salamandras (McCullagh y Nelder, 1989, p. 440)

cuadrado Latino incompleto. Aunque, si bien el diseño es importante, se hace poco uso del diseño en sí para el análisis de los resultados observados. Lo importante es que este diseño experimental permite una comparación de las probabilidades de apareamiento exitoso para los cuatro cruces posibles de poblaciones y género.

4.2. Resultados del experimento para analizar

Los resultados observados se anotaron como 0 o 1, donde 0 indica que la pareja de salamandras no se apareó exitosamente en ese día, mientras que el 1 indica que sí lo hicieron. Las siguientes tres tablas reúnen esas observaciones:

CAPÍTULO 4. DATOS UTILIZADOS PARA LA APLICACIÓN

		Machos											
Hembras		Junio 4	Junio 8	Junio 12	Junio 16	Junio 20	Junio 24						
	1	1	1	1	0	1	1						
	2	1	1	1	1	1	1						
RBF	3	RB	1	WS	0	RB	1	WS	1	RB	1	WS	1
	4	1	1	1	0	1	1						
	5	1	1	1	1	1	1						
	6	1	1	1	0	1	1						
	7	0	0	0	1	0	0						
RBF	8	WS	0	RB	1	WS	0	RB	0	WS	1	RB	1
	9	0	0	1	1	1	1						
	10	0	0	1	0	1	0						
	1	0	1	1	1	0	1						
	2	0	0	0	1	0	0						
WSF	3	RB	0	WS	0	RB	0	WS	0	RB	0	WS	1
	4	0	1	1	1	0	1						
	5	0	1	0	0	0	0						
	6	0	0	1	0	0	0						
	7	1	1	1	0	1	1						
WSF	8	WS	1	RB	0	WS	1	RB	0	WS	1	RB	0
	9	1	1	1	1	1	1						
	10	1	0	0	1	1	0						

Tabla 4.2: Apareamientos observados: verano '86 - Junio (McCullagh y Nelder, 1989, p. 442)

4.2. Resultados del experimento para analizar

		Machos											
Hembras		Set. 4	Set. 8	Set. 12	Set. 16	Set. 20	Set. 24						
	1	1	0	1	0	1	0						
	2	1	1	1	0	0	1						
RBF	3	RB	0	WS	1	RB	1	WS	1	RB	1	WS	1
	4	1	0	0	0	0	0						
	5	1	0	1	0	0	0						
	6	1	1	1	0	0	0						
	7	1	1	0	1	0	1						
RBF	8	WS	1	RB	0	WS	0	RB	0	WS	1	RB	0
	9	1	1	1	1	0	1						
	10	0	0	1	1	1	1						
	1	0	1	0	0	0	0						
	2	0	1	0	0	0	0						
WSF	3	RB	0	WS	1	RB	0	WS	1	RB	0	WS	1
	4	0	0	1	1	0	1						
	5	0	0	0	0	0	0						
	6	1	0	1	1	1	1						
	7	1	0	1	0	0	0						
WSF	8	WS	0	RB	0	WS	0	RB	0	WS	0	RB	0
	9	1	1	1	1	1	1						
	10	1	0	1	1	1	1						

Tabla 4.3: Apareamientos observados: otoño '86 - Setiembre (McCullagh y Nelder, 1989, p. 443)

CAPÍTULO 4. DATOS UTILIZADOS PARA LA APLICACIÓN

Hembras		Machos									
		Oct. 4	Oct. 8	Oct. 12	Oct. 16	Oct. 20	Oct. 24				
	1	1	1	1	0	1	1				
	2	0	0	0	1	0	0				
RBF	3	RB 1	WS 1	RB 1	WS 0	RB 1	WS 1				
	4	1	0	1	1	0	0				
	5	0	1	1	0	1	0				
	6	0	1	1	1	1	1				
	7	0	1	0	0	0	0				
RBF	8	WS 1	RB 1	WS 1	RB 1	WS 1	RB 0				
	9	0	0	1	0	1	1				
	10	0	1	1	1	1	1				
	1	0	0	0	1	0	1				
	2	0	1	0	1	0	1				
WSF	3	RB 1	WS 0	RB 1	WS 0	RB 0	WS 1				
	4	1	1	1	0	0	0				
	5	0	1	1	1	0	1				
	6	0	0	1	0	0	0				
	7	1	0	1	0	0	0				
WSF	8	WS 0	RB 0	WS 1	RB 0	WS 1	RB 0				
	9	1	0	0	0	1	0				
	10	0	0	1	0	1	0				

Tabla 4.4: Apareamientos observados: otoño '86 - Octubre (McCullagh y Nelder, 1989, p. 444)

Capítulo 5

Resultados

5.1. Modelo propuesto

Como se puede ver en [McCullagh y Nelder \(1989, §14.5.1\)](#), el experimento que se describe en el capítulo anterior involucra varias componentes de dispersión asociadas a diferentes subgrupos y eso no puede ser considerado como una simple sobredispersión en donde las observaciones son independientes y sólo hay una componente de dispersión.

Como se puede ver en [Karim y Zeger \(1992, p. 633\)](#), la pregunta importante es ¿cómo hacer inferencia correctamente con el conjunto de datos disponible para poder estimar las cuatro probabilidades de apareamiento exitoso? Por cada experimento se obtienen 120 datos binarios de resultados. Esos 120 datos binarios corresponden a 30 datos por experimento para cada uno de los cruces RR, RW, WR y WW. Como primer observación se puede decir que es poco creíble que el total de respuestas de un experimento sean independientes. Por cada una de las hembras, es poco creíble que las 6 observaciones de los cruces hechos lo sean. Y por otro, observaciones de diferentes hembras pero con el mismo macho también pueden estar correlacionadas. En síntesis, si dos observaciones tienen al menos un individuo repetido, es posible que no sean independientes. Lo que se puede decir, es que en cada experimento, todas las observaciones que corresponden a un mismo día sí lo son. En la tabla [4.1](#) eso equivale a decir que las observaciones son independientes por

columnas. Otra observación es preguntarse que tanta heterogeneidad hay entre hembras y machos. El grado relativo de heterogeneidad entre machos y hembras y entre verano y otoño reflejan posibles aspectos importantes del mecanismo de reproducción.

Los aspectos mencionados anteriormente determinan que un modelo del estilo GLM no sea suficiente para tratar el problema de la estimación de las probabilidades deseadas. En referencia a lo dicho en el capítulo 2, la mejor alternativa es un modelo del estilo GLMM donde la fórmula del modelo mixto condicional es por lo tanto:

efecto cruzado + efecto hembra individual + efecto macho individual

A partir de ahora se trabaja con las especificaciones del modelo A propuesto en [Karim y Zeger \(1992\)](#) y que además se trabaja en [Knudson \(2016, §1.3, §5.2\)](#). El modelo A propone un efecto aleatorio para cada hembra, un efecto aleatorio para cada macho, un predictor de efecto fijo para cada tipo de cruce y una respuesta Bernoulli de si el par de salamandras tuvieron apareamiento exitoso que depende del tipo de cruce, del efecto aleatorio de la hembra y del efecto aleatorio del macho. Además cada efecto aleatorio de las salamandras se asume independiente de los demás. Los efectos aleatorios de las salamandras machos comparten la misma varianza mientras que los efectos aleatorios de las salamandras hembras también comparten la misma varianza, pero diferente a la varianza ν_M de los machos. Estas varianzas son las componentes de varianza y se denotarán a partir de ahora como ν_M y ν_F . Aunque las salamandras del primer y segundo experimento son las mismas, los datos se trabajan como se hace tradicionalmente y como se sugiere en [McCullagh y Nelder \(1989, p. 441\)](#), considerando que las salamandras son diferentes. Por lo tanto se considera que en total hubieron 120 salamandras diferentes. Según el modelo presentado en 3.1, el mismo en este caso queda:

$$\left\{ \begin{array}{l} Y_i | \mathbf{u} \sim f_{\beta}(y_i | \mathbf{u}) = \exp(y_i(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u}) - \log(1 + \exp(\mathbf{x}_i \boldsymbol{\beta} + \mathbf{z}_i \mathbf{u}))) \\ \mathbb{E}(Y_i | \mathbf{u}) = \text{P}(Y_i = 1) \\ \text{logit } \text{P}(Y_i = 1) = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{u} \\ \mathbf{u} \sim f_{\nu}(\mathbf{u}) \end{array} \right.$$

El modelo asume que $Y_i|\mathbf{u}$ es Bernoulli, por lo tanto la distribución condicional es Bernoulli, expresada según la notación que se puede ver en A.1 tomando $\theta = \log\left(\frac{p}{1-p}\right) = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}$ y de forma equivalente, $p = \frac{\exp(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u})}{1 + \exp(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u})}$.

Las observaciones son además condicionalmente independientes dado los efectos aleatorios, por lo tanto tienen distribución conjunta dada por:

$$f_{\boldsymbol{\beta}}(\mathbf{y}|\mathbf{u}) = \prod_{i=1}^{360} f_{\boldsymbol{\beta}}(y_i|\mathbf{u}) = \exp\left(\mathbf{y}^T(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) - \sum_{i=1}^{360} \log(1 + \exp(\mathbf{x}_i\boldsymbol{\beta} + \mathbf{z}_i\mathbf{u}))\right)$$

donde \mathbf{X}_i es la fila i de la matriz para los efectos fijos \mathbf{X} cuyo tamaño es 360×4 . La cantidad de filas es la cantidad de ensayos hechos en total para los cuatro cruces heterosexuales posibles (RR, RW, WR y WW).

Luego, \mathbf{Z}_i es la fila i de la matriz para los efectos aleatorios \mathbf{Z} . Esta matriz es de tamaño 360×120 ya que el experimento consta de 360 ensayos y de 120 salamandras diferentes (bajo el supuesto de que en cada experimento se tomaron 40 salamandras diferentes a todas las anteriores).

El vector de efectos fijos es $\boldsymbol{\beta} = (\beta_{RR}, \beta_{RW}, \beta_{WR}, \beta_{WW})^T$, donde β_{ij} denota el logaritmo de las probabilidades (log odds por su nombre en inglés) del apareamiento entre una hembra de la población i con un macho de la población j (con $i, j \in \{R, W\}$).

El vector de efectos aleatorios es $\mathbf{u} = \begin{bmatrix} \mathbf{u}^F \\ \mathbf{u}^M \end{bmatrix}$, donde \mathbf{u}^F y \mathbf{u}^M son vectores de largo 60 cada uno, correspondientes a los efectos aleatorios de las 60 hembras y los 60 machos respectivamente. Las componentes u_i^F de \mathbf{u}^F se asumen con distribución normal de media 0 y varianza ν_F y de forma análoga las componentes u_i^M de \mathbf{u}^M se asumen con distribución normal de media 0 y varianza ν_M . Además, como se dijo anteriormente, todos los efectos aleatorios son independientes unos de otros. Se considera el vector de componentes de varianza $\boldsymbol{\nu} = \begin{bmatrix} \nu_F \\ \nu_M \end{bmatrix}$ y por lo tanto:

$$f_{\boldsymbol{\nu}}(\mathbf{u}) = \exp\left(-60 \log(2\pi) - 30 \log(\nu_F) - 30 \log(\nu_M) - \frac{(\mathbf{u}^F)^T \mathbf{u}^F}{2\nu_F} - \frac{(\mathbf{u}^M)^T \mathbf{u}^M}{2\nu_M}\right)$$

5.1.1. Verosimilitud para el modelo propuesto

Como se expuso en 3.1.1 y siguiendo los cálculos de la sección anterior, se obtiene que:

$$L(\boldsymbol{\beta}, \boldsymbol{\nu} | \mathbf{y}) = \int \exp \left(\mathbf{y}^T (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) - \sum_{i=1}^{360} \log(1 + \exp(\mathbf{x}_i\boldsymbol{\beta} + \mathbf{z}_i\mathbf{u})) \right. \\ \left. - 60 \log(2\pi) - 30 \log(\nu_F) - 30 \log(\nu_M) - \frac{(\mathbf{u}^F)^T \mathbf{u}^F}{2\nu_F} - \frac{(\mathbf{u}^M)^T \mathbf{u}^M}{2\nu_M} \right) d\mathbf{u}$$

donde $d\mathbf{u} = du_1^F \cdots du_{60}^F du_1^M \cdots du_{60}^M$. Por lo tanto, la verosimilitud depende de una integral de dimensión 120 (esto es, evaluada sobre un subconjunto de \mathbb{R}^{120}). Esto representa un obstáculo importante y desafiante para encontrar los estimadores por máxima verosimilitud.

Como se puede ver en Knudson (2016, pp.8-9), la evaluación directa de la integral es impensable. El método clásico de integración numérica sólo es aplicable cuando se tiene una integral de poca dimensión. No se puede pensar que esa integral se puede factorizar como producto de 120 integrales univariadas ya que los efectos aleatorios están cruzados. A lo sumo, se puede reducir a un producto de 6 integrales de dimensión 20 cada una. Esto es debido a que en cada experimento, hay dos subgrupos de interacción hembra-macho que son cerrados. Un grupo son RBF 1 ~ 5, WSF 5 ~ 10, RBM 1 ~ 5, WSM 1 ~ 5 y otro son RBF 6 ~ 10, WSF 1 ~ 5, RBM 6 ~ 10 y WSM 6 ~ 10. Cada experimento produce una integral de dimensión 20 por cada uno de esos subgrupos. Aún así, resolver una integral de dimensión 20 sigue siendo un problema intratable por integración numérica.

Como alternativas algunos investigadores han usado un enfoque bayesiano basado en cadenas de Markov Monte Carlo, como en Karim y Zeger (1992). Otros han basado su análisis en la pseudo verosimilitud o en el método Monte Carlo EM.

Más recientemente en Sung y Geyer (2007) se implementó un muestreo de importancia cuya distribución fue construida de forma independiente de los datos para analizar problemas con integrales pero no mostró un buen desempeño con la integral anterior como se dice en Sung y Geyer (2007, p. 1001). Esta es la implementación de MCLA, antecesora

Parámetro	Intervalo de confianza al 99 %
β_{RR}	[0,119; 1,492]
β_{RW}	[-0,092; 1,240]
β_{WR}	[-3,140; 1,521]
β_{WW}	[0,390; 1,804]
ν_F	[0,690; 1,928]
ν_M	[0,930; 2,594]

Tabla 5.1: Intervalos de confianza calculados en B

de la implementación actual de Knudson (2016) donde la distribución para el muestreo de importancia se tomó independiente de los datos y sobre este método que se presentan las estimaciones por máxima verosimilitud (el MCMLE).

5.2. Resultados

Como se dice en B, de las ejecuciones hechas se puede concluir que un modelo para los datos es:

$$\begin{aligned} \text{logit } P(Y_i = 1) = & 0,8503 \cdot \mathbb{1}_{RR}(\text{Cross}) + 0,5740 \cdot \mathbb{1}_{RW}(\text{Cross}) - 2,3300 \cdot \mathbb{1}_{WR}(\text{Cross}) \\ & + 1,0965 \cdot \mathbb{1}_{WW}(\text{Cross}) + u_i^F + u_i^M \end{aligned}$$

donde

- $\mathbb{1}_{ij}(\text{Cross}) = \begin{cases} 1, & \text{si Cross} = ij \\ 0, & \text{si Cross} \neq ij \end{cases}$
- $u_i^F \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0; 1, 3086)$
- $u_i^M \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0; 1, 7620)$
- las u_i^F son independientes de las u_i^M

Para cada uno de los parámetros se tienen los siguientes intervalos de confianza:

CAPÍTULO 5. RESULTADOS

Con los parámetros del efecto fijo finalmente se puede obtener las probabilidades de apareamiento exitoso buscadas mediante

$$P(\text{Apareamiento entre especie } i \text{ y } j) = \frac{e^{\beta_{ij}}}{1 + e^{\beta_{ij}}}$$

De lo cual se obtiene los siguientes resultados:

Cruce	RR	RW	WR	WW
Prob. de apareamiento	0,7006	0,6397	0,0887	0,7496

Para finalizar, se comparan estos valores con los que se muestran en Knudson (2016, §5.4.1):

Autor	$\hat{\beta}_{RR}$	$\hat{\beta}_{RW}$	$\hat{\beta}_{WR}$	$\hat{\beta}_{WW}$	$\hat{\nu}_F$	$\hat{\nu}_M$
Presente trabajo, librería glmm	0,8503	0,574	-2,33	1,0965	1,3086	1,762
Knudson (2015), librería glmm	1,03	0,34	-1,94	1,00	1,36	1,23
Booth and Hobert (1999) MCMC	1,03	0,32	-1,95	0,99	1,4	1,25
Bates et al. (2014), librería lme4	1,01	0,31	-1,89	0,99	1,17	1,04

Como se puede observar, los resultados obtenidos en el presente trabajo concuerdan con la precisión que hay en la bibliografía al respecto sobre el conjunto de datos de las salamandras.

Capítulo 6

Conclusiones

En este trabajo se ha presentado una técnica estadística computacional que representa el estado del arte sobre los métodos de aproximación de la función de verosimilitud de modelos mixtos lineales generalizados (GLMM) con componentes del vector de variables respuesta que, dado los efectos aleatorios, tienen distribuciones condicionales todas Bernoulli o todas Poisson. Como se mostró en las secciones 2.4 y 3.1, la optimización de la función de verosimilitud que corresponde a GLMM representa un desafío, dado que la función objetivo se expresa usualmente en forma de integrales de dimensión elevada.

La característica principal de la implementación de este método de aproximación Monte Carlo para los GLMM es proponer una distribución de muestreo de importancia para generar los efectos aleatorios la cual, a diferencia de trabajos anteriores (Sung y Geyer, 2007; Geyer, 1994), se construye utilizando los datos observados.

Siguiendo el trabajo de Knudson (2016), se ha mostrado mediante técnicas de estadística para muestras grandes, que la varianza del error debido a la discrepancia entre el estimador basado en el método Monte Carlo de aproximación de la verosimilitud (MCMLE) y el estimador de máxima verosimilitud (MLE), es finita. Asimismo, en el capítulo 3 se ha tratado el error ocasionado por el mecanismo de muestreo de los datos observados con respecto a los verdaderos valores de los parámetros.

CAPÍTULO 6. CONCLUSIONES

La implementación del trabajo teórico de Knudson (2016) y los resultados de su aplicación al famoso conjunto de datos procedente de un experimento de especiación realizado con salamandras, han sido obtenidos empleando la librería **glmm** (Knudson, 2020), actualmente disponible en el CRAN de R (R Core Team, 2021). La librería **glmm** resulta ser una herramienta de trabajo efectiva para llevar a cabo inferencia estadística basada en la verosimilitud para GLMM con respuestas de Bernoulli y de Poisson, condicionadas a los efectos aleatorios. El apéndice B contiene el código en el lenguaje de programación R (R Core Team, 2021) y los resultados correspondientes, comentados e interpretados en el capítulo 5.

Como propuesta metodológica para un futuro trabajo se puede extender el enfoque al área de la estadística espacial.

Referencias bibliográficas

- Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New Jersey: John Wiley & Sons.
- Breslow, N. E., y Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421), 9–25. doi: 10.2307/2290687
- Casella, G., y Berger, R. L. (2001). *Statistical inference* (2nd ed.). California: Duxbury Press.
- Demidenko, E. (2013). *Mixed models: theory and applications with R* (2nd ed.). New Jersey: John Wiley & Sons.
- Evans, M., y Swartz, T. (2000). *Approximating integrals via Monte Carlo and deterministic methods*. Oxford: Oxford University Press.
- Faraway, J. J. (2016). *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models* (2nd ed.). Florida: CRC Press.
- Gbur, E. E., Stroup, W. W., McCarter, K. S., Durham, S., Young, L. J., Christman, M., West, M., y Kramer, M. (2012). *Analysis of generalized linear mixed models in the agricultural and natural resources sciences*. American Society of Agronomy.
- Geyer, C. J. (1990). *Likelihood and exponential families* (Tesis Doctoral, University of Washington, United States). Descargado de <http://purl.umn.edu/56330>
- Geyer, C. J. (1994). On the convergence of Monte Carlo maximum likelihood calculations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(1), 261-274. doi:

REFERENCIAS BIBLIOGRÁFICAS

10.1111/j.2517-6161.1994.tb01976.x

Geyer, C. J. (2013). Asymptotics of maximum likelihood without the LLN or CLT or sample size going to infinity. En *Advances in modern statistical theory and applications: A Festschrift in honor of Morris L. Eaton* (Vol. 10, pp. 1–24). Institute of Mathematical Statistics. doi: 10.1214/12-IMSCOLL1001

Gut, A. (2005). *Probability: A graduate course*. New York: Springer.

Jacod, J., y Protter, P. (2003). *Probability essentials* (2nd ed.). New York: Springer.

Jiang, J. (2007). *Linear and generalized linear mixed models and their applications*. New York: Springer.

Karim, M. R., y Zeger, S. L. (1992). Generalized linear models with random effects; salamander mating revisited. *Biometrics*, 48(2), 631–644. doi: 10.2307/2532317

Kendall, M. G., y Stuart, A. (1961). *The advanced theory of statistics* (3rd ed., Vol. II). New York: Hafner Publishing Company.

Knudson, C. (2016). *Monte Carlo likelihood approximation for generalized linear mixed models* (Tesis Doctoral, University of Minnesota, United States). Descargado de <http://hdl.handle.net/11299/178948>

Knudson, C. (2020). glmm: Generalized linear mixed models via Monte Carlo likelihood approximation [Manual de software informático]. Descargado de <https://CRAN.R-project.org/package=glmm> (R package version 1.4.2)

Kotz, S., y Nadarajah, S. (2004). *Multivariate t distributions and their applications*. Cambridge: Cambridge University Press.

Kuk, A. Y. C. (1999). Laplace importance sampling for generalized linear mixed models. *Journal of Statistical Computation and Simulation*, 63(2), 143-158. doi: 10.1080/00949659908548522

- Lewis, S. M., y Raftery, A. E. (1997). Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator. *Journal of the American Statistical Association*, 92(438), 648-655. doi: 10.1080/01621459.1997.10474016
- McCullagh, P., y Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman and Hall.
- McCulloch, C. E. (2003). Generalized linear mixed models. *NSF-CBMS Regional Conference Series in Probability and Statistics*, 7, i-84. Descargado de <http://www.jstor.org/stable/4153190>
- McCulloch, C. E., y Searle, S. R. (2001). *Generalized, linear and mixed models* (1st ed., Vol. 1). New York: John Wiley & Sons.
- Metropolis, N. (1987). The beginning of the Monte Carlo method. *Los Alamos Science*, 15, 125-130. (Disponible en <https://fas.org/sgp/othergov/doe/lanl/pubs/LaScience.htm>. Visto por última vez el 12 de diciembre de 2020.)
- Metropolis, N., y Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association*, 44(247), 335-341. doi: 10.2307/2280232
- R Core Team. (2021). R: A language and environment for statistical computing [Manual de software informático]. Vienna, Austria. Descargado de <https://www.R-project.org/>
- Ripley, B. D. (1987). *Stochastic simulation*. John Wiley & Sons.
- Robert, C., y Casella, G. (2004). *Monte Carlo statistical methods* (2nd ed.). New York: Springer.
- Ruli, E., Sartori, N., y Ventura, L. (2016). Improved Laplace approximation for marginal likelihoods. *Electron. J. Statist.*, 10(2), 3986-4009. doi: 10.1214/16-EJS1218
- Ryan, E., Drovandi, C., y Pettitt, A. (2015, Mar). Fully Bayesian experimental design for pharmacokinetic studies. *Entropy*, 17(3), 1063-1089. doi: 10.3390/e17031063

REFERENCIAS BIBLIOGRÁFICAS

- Searle, S. R., Casella, G., y McCulloch, C. E. (2006). *Variance components*. New Jersey: John Wiley & Sons.
- Shun, Z. (1997). Another look at the salamander mating data: a modified Laplace approximation approach. *Journal of the American Statistical Association*, 92(437), 341-349. doi: 10.1080/01621459.1997.10473632
- Stroup, W. W. (2012). *Generalized linear mixed models: modern concepts, methods and applications*. Florida: CRC Press.
- Sung, Y. J., y Geyer, C. J. (2007). Monte Carlo likelihood inference for missing data models. *Annals of Statistics*, 35(3), 990–1011. doi: 10.1214/009053606000001389
- Sung, Y. J., y Geyer, C. J. (2018). Bernoulli regression with normal random effects [Manual de software informático]. Descargado de <https://github.com/cjgeyer/bernor> (R package version 0.3-12)
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54(3), 426–482. doi: 10.2307/1990256
- Wasserman, L. (2003). *All of statistics: a concise course in statistical inference*. New York: Springer.
- Wong, R. (2001). *Asymptotic approximations of integrals*. Boston: Society for Industrial and Applied Mathematics.
- Zipunnikov, V. V. (2009). *Topics generalized linear mixed models* (Tesis Doctoral no publicada). Cornell University, United States.
- Zipunnikov, V. V., y Booth, J. G. (2006). *Monte Carlo EM for generalized linear mixed models using randomized spherical radial integration*. Descargado de <http://faculty.bsccb.cornell.edu/~booth/papers/mcem-sr.pdf> (Trabajo no publicado, consultado por última vez el 12 de diciembre de 2020)

Apéndices

Apéndice A

Más fundamento teórico

A.1. Familia exponencial de distribuciones

En general se dice que una densidad f es un miembro de la familia exponencial cuando su densidad puede expresarse como se muestra en [Casella y Berger \(2001, §3.4\)](#)

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{i=1}^n w_i(\boldsymbol{\theta})t_i(x)\right)$$

donde el parámetro de la distribución es $\boldsymbol{\theta}$ un vector de largo n ; h y c son funciones a valores reales no negativos mientras las t_i y w_i son funciones a valores reales cualesquiera. En vez de usar esta notación se opta por una equivalente y más cómoda para trabajar con la especificación de este modelo.

A.1.1. $\theta \in \mathbb{R}$

Siguiendo a [Stroup \(2012, §4.2.1\)](#), la notación para familia exponencial es

$$f(y|\theta) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right]$$

o de forma equivalente

$$l(y, \theta, \phi) = \log(f(y|\theta)) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)$$

donde ϕ denota un parámetro de escala. Esta forma tiene la ventaja de contabilizar explícitamente el parámetro de escala y presentarse con naturalidad para la estimación por máxima verosimilitud. Esta notación también se presenta en [Jiang \(2007, p. 121\)](#) y con más detalle en [McCullagh y Nelder \(1989, §2.2.2\)](#).

Las distribuciones más importantes para el presente trabajo son la Bernoulli y Poisson. Para estos casos particulares se tiene lo siguiente:

- **Bernoulli** $f(y) = p^y(1 - p)^{1-y}$ con $y \in \{0, 1\}$. De aquí se tiene que

$$\begin{aligned}\log(f(y)) &= y \log(p) + (1 - y) \log(1 - p) \\ &= y \log\left(\frac{p}{1 - p}\right) + \log(1 - p) \\ &= \frac{y \log\left(\frac{p}{1 - p}\right) - (-\log(1 - p))}{1} + 0\end{aligned}$$

Comparando con la forma genérica se puede ver que

- $\theta = \log\left(\frac{p}{1 - p}\right)$
 - $b(\theta) = -\log(1 - p)$
 - $a(\phi) = 1$
 - $c(y, \phi) = 0$
- **Poisson** $f(y) = \frac{e^{-\lambda}\lambda^y}{y!}$ con $y \in \{0, 1, 2, 3, \dots\}$. De aquí se tiene que

$$\begin{aligned}\log(f(y)) &= -\lambda + y \log(\lambda) - \log(y!) \\ &= \frac{y \log(\lambda) - \lambda}{1} - \log(y!)\end{aligned}$$

Comparando con la forma genérica se puede ver que

- $\theta = \log(\lambda)$
- $b(\theta) = \lambda$
- $a(\phi) = 1$
- $c(y, \phi) = -\log(y!)$

En la siguiente tabla (A.1) se resumen estos casos con algunos adicionales no deducidos aquí:

Distribución	Parámetro canónico θ	$b(\theta)$	$a(\phi)$	$c(y, \phi)$
Bernoulli	$\log\left(\frac{p}{1-p}\right)$	$-\log(1-p)$	1	0
Poisson	$\log(\lambda)$	λ	1	$-\log(y!)$
Binomial	$\log\left(\frac{p}{1-p}\right)$	$-n \log(1-p)$	1	$\log(C_y^n)$
Normal	μ	$\frac{\mu^2}{2}$	σ^2	$\frac{y^2}{2\sigma^2} + \log(\sigma\sqrt{2\pi})$

Tabla A.1: Parámetros de las principales distribuciones exponenciales

Por más ejemplos de distribuciones usuales se puede consultar la tabla disponible también en [McCullagh y Nelder \(1989, p. 30\)](#).

Una observación importante sobre el parámetro canónico, es que siempre es una función de la esperanza de la distribución en cuestión.

A.1.2. $\boldsymbol{\theta} \in \mathbb{R}^n$

Mientras la sección anterior captura la forma esencial de la log-verosimilitud, ahora se debe reflejar el hecho de que en general se observa un vector de datos y en base a esos datos estimar un vector de parámetros. Ahora tanto y como θ son vectores y hay que generalizar la expresión A.1.1. Para eso se reescribe la misma adaptándola a un lenguaje matricial para usar en la especificación de un modelo GLMM como sigue:

$$l(\boldsymbol{\theta}; \mathbf{y}, \boldsymbol{\phi}) = \sum_{i=1}^n \frac{y_i \theta_i}{a(\phi_i)} - \sum_{i=1}^n \frac{b(\theta_i)}{a(\phi_i)} + c(\mathbf{y}, \boldsymbol{\phi}) = \mathbf{y}^T \mathbf{A} \boldsymbol{\theta} - \mathbf{1}^T \mathbf{A} b(\boldsymbol{\theta}) + c(\mathbf{y}, \boldsymbol{\phi})$$

donde

- $\mathbf{A} = \text{diag}\left[\frac{1}{a(\phi_i)}\right]$ es una matriz diagonal (asumiendo que se tienen n observaciones) donde el elemento en la entrada (i, i) es el valor del parámetro de escala para la observación y_i .
- $\boldsymbol{\theta}$ es el vector de parámetros canónico.

- $\mathbf{1}$ es un vector $n \times 1$ de unos.
- $b(\boldsymbol{\theta})$ es un vector $n \times 1$ definido por la aplicación de la función b a cada elemento de $\boldsymbol{\theta}$.
- $c(\mathbf{Y}, \boldsymbol{\phi})$ es lo restante que no depende de $\boldsymbol{\theta}$.

A.1.3. Propiedades importantes

Trabajando con la función score $S(\boldsymbol{\theta}) = \nabla l(\boldsymbol{\theta}; \mathbf{y}, \boldsymbol{\phi})$, el vector cuya componente i es $\partial_{\theta_i} l(\boldsymbol{\theta}; \mathbf{y}, \boldsymbol{\phi})$. A partir de esto, se define a la información de Fisher como $I(\boldsymbol{\theta}) = \text{Var}(S(\boldsymbol{\theta}))$ de un uso muy importante en estadística.

De la función S se pueden deducir dos propiedades para usar en A.1.1 y obtener relaciones muy útiles. Estas propiedades, en su formulación matricial son:

- $\mathbb{E}(S(\boldsymbol{\theta})) = \mathbf{0}$
- $\mathbb{E}(S(\boldsymbol{\theta})S(\boldsymbol{\theta})^T) = -\nabla^2 S(\boldsymbol{\theta})$

donde $\mathbf{0}$ es el vector nulo y $\nabla^2 S(\boldsymbol{\theta})$ es la matriz hessiana de S evaluada en $\boldsymbol{\theta}$.

Expresando las igualdades en cada una de las entradas de los vectores y matrices, lo anterior se reduce a los siguientes casos unidimensionales:

- $\mathbb{E}(\partial_{\theta_j} l(\boldsymbol{\theta}; \mathbf{y}, \boldsymbol{\phi})) = 0$
- $\mathbb{E}[\partial_{\theta_k}^2 l(\boldsymbol{\theta}; \mathbf{y}, \boldsymbol{\phi})] + \mathbb{E}[\partial_{\theta_k} l(\boldsymbol{\theta}; \mathbf{y}, \boldsymbol{\phi}) \partial_{\theta_j} l(\boldsymbol{\theta}; \mathbf{y}, \boldsymbol{\phi})] = 0$

Una demostración detallada de estas últimas se puede encontrar en [Kendall y Stuart \(1961, p. 53\)](#).

Las mismas propiedades, pero para la función score en el caso unidimensional pueden encontrarse en [Kendall y Stuart \(1961, p. 9\)](#), [McCullagh y Nelder \(1989, p. 28\)](#) y [Stroup \(2012, §4.2.1.1\)](#).

En lo que sigue y usando las anteriores propiedades de la función score, se deducen expresiones para el caso vectorial.

Teniendo $l(\boldsymbol{\theta}; \mathbf{y}, \boldsymbol{\phi}) = \mathbf{y}^T \mathbf{A}\boldsymbol{\theta} - \mathbf{1}^T \mathbf{A}b(\boldsymbol{\theta}) + c(\mathbf{y}, \boldsymbol{\phi})$, se calculan las derivadas parciales:

$$\begin{aligned} \partial_{\theta_j} l(\boldsymbol{\theta}; \mathbf{y}, \boldsymbol{\phi}) &= \partial_{\theta_j} (\mathbf{y}^T \mathbf{A}\boldsymbol{\theta} - \mathbf{1}^T \mathbf{A}b(\boldsymbol{\theta}) + c(\mathbf{y}, \boldsymbol{\phi})) \\ &= \partial_{\theta_j} (\mathbf{y}^T \mathbf{A}\boldsymbol{\theta}) - \partial_{\theta_j} (\mathbf{1}^T \mathbf{A}b(\boldsymbol{\theta})) + \partial_{\theta_j} (c(\mathbf{y}, \boldsymbol{\phi})) \\ &= \frac{y_j}{a(\phi_j)} - \frac{b'(\theta_j)}{a(\phi_j)} + 0 \\ &= \frac{y_j - b'(\theta_j)}{a(\phi_j)} \end{aligned}$$

Por otro lado, usando la propiedad de la función score y teniendo en cuenta que la única variable aleatoria es y_i :

$$\begin{aligned} 0 &= \mathbb{E}(\partial_{\theta_j} l(\boldsymbol{\theta}; \mathbf{y}, \boldsymbol{\phi})) \\ &= \mathbb{E}\left(\frac{y_j - b'(\theta_j)}{a(\phi_j)}\right) \\ &= \frac{\mu_j - b'(\theta_j)}{a(\phi_j)} \end{aligned}$$

Por lo tanto, debe ser $\mu_j - b'(\theta_j) = 0$, así que $\mu_j = b'(\theta_j)$. Como se cumple para cada componente, entonces se puede deducir:

$$\boldsymbol{\mu} = b'(\boldsymbol{\theta})$$

Ahora se calculan las derivadas parciales segundas:

$$\begin{aligned} \partial_{\theta_k \theta_j}^2 l(\boldsymbol{\theta}; \mathbf{y}, \boldsymbol{\phi}) &= \partial_{\theta_k} (\partial_{\theta_j} l(\boldsymbol{\theta}; \mathbf{y}, \boldsymbol{\phi})) \\ &= \partial_{\theta_k} \left(\frac{y_j - b'(\theta_j)}{a(\phi_j)} \right) \\ &= \begin{cases} -\frac{b''(\theta_k)}{a(\phi_k)}, & \text{si } k = j \\ 0, & \text{si } k \neq j \end{cases} \end{aligned}$$

Ahora al igual que en las anteriores partes, usando la otra propiedad de la función score y teniendo en cuenta que la única variable aleatoria es y_i de nuevo se llega a expresiones

útiles. Dado que la $\partial_{\theta_k \theta_j}^2 l(\boldsymbol{\theta}; \mathbf{y}, \boldsymbol{\phi})$ tiene dos expresiones según sea $k = j$ y $k \neq j$, se presenta cada caso por separado:

▪ **k = j:**

$$\begin{aligned}
 0 &= \mathbb{E}[\partial_{\theta_k \theta_j}^2 l(\boldsymbol{\theta}; \mathbf{y}, \boldsymbol{\phi})] + \mathbb{E}[\partial_{\theta_k} l(\boldsymbol{\theta}; \mathbf{y}, \boldsymbol{\phi}) \partial_{\theta_j} l(\boldsymbol{\theta}; \mathbf{y}, \boldsymbol{\phi})] \\
 &= \mathbb{E}\left(-\frac{b''(\theta_k)}{a(\phi_k)}\right) + \mathbb{E}\left[\left(\frac{y_k - b'(\theta_k)}{a(\phi_k)}\right)^2\right] \\
 &= -\frac{b''(\theta_k)}{a(\phi_k)} + \frac{\mathbb{E}[(y_k - b'(\theta_k))^2]}{(a(\phi_k))^2} \\
 &= -\frac{b''(\theta_k)}{a(\phi_k)} + \frac{1}{(a(\phi_k))^2} \underbrace{\mathbb{E}[(y_k - \mu_k)^2]}_{\text{Var}(y_k)}, \text{ debido a que } \mathbb{E}(y_k) = \mu_k = b'(\theta_k) \\
 &= \frac{-b''(\theta_k)a(\theta_k) + \text{Var}(y_k)}{(a(\theta_k))^2}
 \end{aligned}$$

Por lo tanto debe ser $0 = -b''(\theta_k)a(\theta_k) + \text{Var}(y_k)$, así que $\text{Var}(y_k) = b''(\theta_k)a(\theta_k)$.

▪ **k ≠ j:**

$$\begin{aligned}
 0 &= \mathbb{E}[\partial_{\theta_k \theta_j}^2 l(\boldsymbol{\theta}; \mathbf{y}, \boldsymbol{\phi})] + \mathbb{E}[\partial_{\theta_k} l(\boldsymbol{\theta}; \mathbf{y}, \boldsymbol{\phi}) \partial_{\theta_j} l(\boldsymbol{\theta}; \mathbf{y}, \boldsymbol{\phi})] \\
 &= \mathbb{E}(0) + \mathbb{E}\left[\left(\frac{y_k - b'(\theta_k)}{a(\phi_k)}\right) \left(\frac{y_j - b'(\theta_j)}{a(\phi_j)}\right)\right] \\
 &= 0 + \frac{\mathbb{E}[(y_k - b'(\theta_k))(y_j - b'(\theta_j))]}{a(\phi_k)a(\phi_j)} \\
 &= \frac{1}{a(\phi_k)a(\phi_j)} \underbrace{\mathbb{E}[(y_k - \mu_k)(y_j - \mu_j)]}_{\text{Cov}(y_k, y_j)}, \text{ debido a que } \mathbb{E}(y_k) = \mu_k = b'(\theta_k) \\
 &= \frac{\text{Cov}(y_k, y_j)}{a(\theta_k)a(\theta_j)}
 \end{aligned}$$

Por lo tanto debe ser $\text{Cov}(y_k, y_j) = 0$.

En conclusión la matriz de varianzas y covarianzas del vector aleatorio y queda una matriz

diagonal de la forma

$$\text{diag}(b''(\theta_1)a(\theta_1), b''(\theta_2)a(\theta_2), \dots, b''(\theta_n)a(\theta_n))$$

$$\parallel$$

$$\begin{bmatrix} b''(\theta_1)a(\theta_1) & 0 & \cdots & 0 \\ 0 & b''(\theta_2)a(\theta_2) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & b''(\theta_n)a(\theta_n) \end{bmatrix}$$

En los anteriores cálculos, tomando $n = 1$ se obtienen los resultados para el caso unidimensional como se puede encontrar en [Stroup \(2012, §4.2.1.1\)](#):

- $\mathbb{E}(y) = \mu = b'(\theta)$
- $\text{Var}(y) = a(\phi)b''(\theta)$

La derivada segunda $b''(\theta)$ se llama *función de varianza* y se suele anotar como $V(\mu)$. Este nombre es debido a que $V(\mu) = b''(\theta)$ describe la dependencia de la varianza de y en su esperanza. Por ejemplo, en la distribución normal se tiene que $V(\mu) = 1$ y significa que la varianza no depende de la media en nada. Pero, por ejemplo, para la distribución binomial se tiene que $V(p) = p(1 - p)$ o para la distribución poisson $V(\lambda) = \lambda$; esto junto a que el parámetro de escala es una constante conocida para ambas distribuciones, implica que la varianza es completamente determinada por la esperanza. En cambio, para otras distribuciones como beta, gamma o binomial negativa, la varianza depende de ambos como se puede ver en [McCullagh y Nelder \(1989, p. 30\)](#).

A.2. Cuasi-verosimilitud penalizada (PQL)

Este método para encontrar aproximaciones de estimadores por máxima verosimilitud. En vez de trabajar con la verosimilitud exacta se trabaja con una función más adecuada para maximizar, la cual se conoce como cuasi-verosimilitud penalizada (PQL por sus siglas en

inglés). Siguiendo a [Breslow y Clayton \(1993, §1\)](#) y [Jiang \(2007, §3.5.2\)](#), se tiene que para este procedimiento computacional se trabaja con los siguientes supuestos:

- $\mathbb{E}(Y_i|\mathbf{u}) = \mu_i = h(\mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{z}_i^T\mathbf{u})$. Matricialmente, definiendo $\mathbf{Y} = (Y_1, \dots, Y_n)$ y las matrices \mathbf{X} y \mathbf{Z} de filas \mathbf{x}_i^T y \mathbf{z}_i^T respectivamente, lo anterior queda como $\mathbb{E}(\mathbf{Y}|\mathbf{u}) = h(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})$. Donde:
 - h es la función inversa de la función link del modelo especificado
- $\text{Var}(Y_i|\mathbf{u}) = a_i(\phi)v(\mu_i)$ donde:
 - ϕ es un parámetro de dispersión adicional
 - a_i es una función conocida que depende de ϕ y en general se toma $a_i(\phi) = \frac{\phi}{w_i}$ donde los w_i son pesos conocidos
 - v es una función de varianza conocida

En este contexto usando el método de aproximación de Laplace se puede encontrar estimadores de máxima verosimilitud con la verosimilitud aproximada en vez de usar la verdadera verosimilitud.

Es de destacar aquí que los supuestos para este procedimiento son más débiles respecto a los incluidos en la definición de modelos GLMM. Específicamente, la distribución condicional de Y_i dado \mathbf{u} no tiene que ser un miembro de la familia exponencial, mientras que en la definición de los GLMM sí, deben ser de una familia exponencial. Pero, a diferencia de la especificación general de los GLMM y siguiendo a [Breslow y Clayton \(1993\)](#), sólo se consideran respuestas individuales Y_i univariadas.

A.2.1. Derivación de PQL

Siguiendo a [Jiang \(2007\)](#), la deducción de su expresión se hace bajo supuestos agregados a los mencionados:

- Los Y_1, \dots, Y_n son condicionalmente independientes dado el vector \mathbf{u} .

- $\mathbf{u} \in \mathbb{R}^n$ tiene distribución normal multivariada con media $\mathbf{0} = (0, \dots, 0)^T$ y matriz de covarianza \mathbf{G} , donde \mathbf{G} depende de un vector desconocido ζ de componentes de varianzas. Formalmente, el valor de la matriz de covarianzas debe ser anotado como $\mathbf{G}(\zeta)$, pero para no saturar la notación dado que solo es importante en este párrafo, simplemente se la anota como \mathbf{G} . Con lo cual $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{G})$.

Una cuasi-verosimilitud basada en $\mathbf{y} = (y_1, \dots, y_n)^T$ puede ser expresada de la siguiente manera:

$$L_Q \propto |\mathbf{G}|^{-\frac{1}{2}} \int \exp\left(-\frac{1}{2} \sum_{i=1}^n d_i - \frac{1}{2} \mathbf{u}^T \mathbf{G}^{-1} \mathbf{u}\right) d\mathbf{u}$$

donde $d_i = -2 \int_{y_i}^{\mu_i} \frac{y_i - t}{a_i(\phi)v(t)} dt$ es conocida como la cuasi-devianza. Tiene su origen en la devianza para los modelos lineales generalizados ya que en con los supuestos de trabajo que se adoptan aquí ambas cantidades están relacionadas como se puede encontrar en [McCullagh y Nelder \(1989, §2.3\)](#) o [Jiang \(2007, p. 128\)](#). Por más detalles de la devianza se puede consultar [McCullagh y Nelder \(1989, p. 33\)](#) o [Stroup \(2012, §5.3.1\)](#).

El problema de maximizar L_Q se reduce a maximizar una nueva función l_{PQ} la cual, bajo ciertas condiciones, es una aproximación de $l_Q = \log L_Q$. Esta función que realmente es la que se va a maximizar, es

$$l_{PQ} = -1 \frac{1}{2} \left(\sum_{i=1}^n d_i + \mathbf{u}^T \mathbf{G}^{-1} \mathbf{u} \right)$$

El procedimiento computacional habitual para maximizar esa función corresponde a la solución de un sistema de ecuaciones no lineales dadas por $\partial_{\beta} l_{PQ} = 0$ y $\partial_{\mathbf{u}} l_{PQ} = 0$. O en su forma equivalente

$$\begin{cases} \sum_{i=1}^n \frac{(y_i - \mu_i) \mathbf{x}_i}{a_i(\phi)v(\mu_i)g'(\mu_i)} = 0 \\ \sum_{i=1}^n \frac{(y_i - \mu_i) \mathbf{z}_i}{a_i(\phi)v(\mu_i)g'(\mu_i)} - \mathbf{G}^{-1} \mathbf{u} = 0 \end{cases}$$

Este último sistema depende explícitamente de β y \mathbf{u} ya que de la definición del modelo GLMM se tiene que $\mu_i = h(\mathbf{x}_i^T \beta + \mathbf{z}_i^T \mathbf{u})$.

Para más detalles sobre la deducción completa de la función l_{PQ} y de como resolver el sistema se puede consultar [Breslow y Clayton \(1993\)](#) o [Jiang \(2007, p. 130\)](#).

A.3. Distribución t multivariada

Como se puede ver en [Kotz y Nadarajah \(2004, §1.1\)](#), un vector \mathbf{x} de largo q se dice que tiene distribución t multivariada a valores en \mathbb{R}^q con ν grados de libertad, vector de media $\boldsymbol{\mu}$, matriz de correlación \mathbf{R} (obtenida de la matriz de covarianzas $\boldsymbol{\Sigma}$), si su función de densidad conjunta está dada por:

$$f(\mathbf{x}) = \frac{\Gamma\left(\frac{\nu+p}{2}\right)}{(\pi\nu)^{\frac{p}{2}} \Gamma\left(\frac{\nu}{2}\right) |\mathbf{R}|^{\frac{1}{2}}} \left[1 + \frac{1}{\nu}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{R}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right]^{-\frac{\nu+p}{2}}$$

El parámetro ν , el grado de libertad, también se suele llamar parámetro de forma. Modificando el valor de este parámetro se puede modificar la forma puntiaguda de la silueta de f .

Si $\boldsymbol{\mu} = \mathbf{0}$ se dice que la distribución esta centrada o es central. En caso contrario, se dice que es no centrada.

Si $q = 1$, $\boldsymbol{\mu} = 0$ y $\mathbf{R} = 1$ entonces la densidad corresponde a la densidad de una t -Student univariada con ν grados de libertad.

De forma análoga al caso univariado, si \mathbf{x} es un vector de largo q y tiene distribución t multivariada con ν grados de libertad se puede representar a partir de un vector normal multivariado y una variable con distribución χ^2 . Si:

- \mathbf{y} es un vector de largo q con distribución normal multivariada de media $\mathbf{0}$ y matriz de covarianzas $\boldsymbol{\Sigma}$
- $\frac{\nu S^2}{\sigma^2}$ es una variable aleatoria con ν grados de libertad
- $\frac{\nu S^2}{\sigma^2}$ e \mathbf{y} son independientes

entonces se tiene que $\mathbf{x} = S^{-1}\mathbf{y} + \boldsymbol{\mu}$.

A.4. Ley de los grandes números

Para la siguiente sección se recuerda la ley de los grandes números, disponible por ejemplo en [Jacod y Protter \(2003, p. 175\)](#).

Ley fuerte de los grandes números de Kolmogorov: Dada la sucesión $\{X_i\}_{i \in \mathbb{N}}$ variables aleatorias i.i.d. y $\mu \in \mathbb{R}$. Entonces:

$$\mathbb{E}(X_1) = \mu \Leftrightarrow \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{c.s.} \mu$$

Esta es la forma minimal para la ley de los grandes números. Además, dado que la convergencia casi segura implica la convergencia en probabilidad, esta ley fuerte de los grandes números implica la análoga ley débil bajo las mismas hipótesis.

De la ley fuerte de los grandes números de Kolmogorov es inmediato el siguiente corolario:

Corolario Si además $\mathbb{E}(X_1^2) < \infty \Rightarrow s_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \xrightarrow{c.s.} \sigma^2 = \text{Var}(X_1)$.

A.5. Integración Monte Carlo

Como se ha mencionado, es muy común la necesidad de evaluar una integral de la forma $I = \int_a^b h(x) dx$ para alguna función h en el caso unidimensional. Lo ideal es tener una expresión explícita para la misma, pero es bien sabido que tal cosa no es posible en la mayoría de los casos, bien sea porque la h carece de primitiva expresable por funciones elementales conocidas o bien porque es prácticamente imposible hallarla. Ante la necesidad de tener un valor, al menos aproximado al valor real, es que se buscan métodos para obtener tales aproximaciones. Integración por el método de Monte Carlo es uno de los tantos que hay para aproximar integrales destacándose por su simplicidad y generalidad para aplicarlo. Este método forma parte del llamado método Monte Carlo desarrollado en su versión moderna a finales de la década de 1940. Se desarrolló para resolver problemas de difusión de neutrones, entre otros problemas físico-matemáticos que se presentaron en

el desarrollo de bombas termonucleares. La idea original del método vino de la mano de Stanislaw Ulam, quién tuvo intercambio de ideas con otros destacados matemáticos como John von Neumann o Nicholas Metropolis. Dos importantes desarrollos con los cuales se contaban en la época hicieron posible el éxito y uso extensivo del método. Estos inventos son los generadores de números pseudo aleatorios y ENIAC, la primer computadora electrónica de propósito general. El método original se puede consultar en [Metropolis y Ulam \(1949\)](#) y más detalles históricos en [Metropolis \(1987\)](#).

Siguiendo a [Wasserman \(2003, §24.2\)](#), el método básico de integración Monte Carlo en el caso unidimensional consiste en estimar la integral $I = \int_a^b h(x) dx$ que se reescribe como $I = \int_a^b w(x)f(x) dx$ donde $w(x) = (b - a)h(x)$ y $f(x) = \frac{1}{b-a}\mathbb{1}_{[a,b]}(x)$, donde se puede ver que f es una densidad de probabilidad para una $U[a, b]$. Con esta reformulación del problema la integral a estimar es $I = \mathbb{E}_f(w(X))$ donde $X \sim U[a, b]$. Entonces, si se genera una muestra X_1, \dots, X_n i.i.d. $U[a, b]$ y se considera como estimador \hat{I} de I a $\hat{I} = \overline{w(X)}_n = \frac{1}{n} \sum_{i=1}^n w(X_i)$, por la ley de los grandes números se tiene que

$$\hat{I} = \overline{w(X)}_n \xrightarrow{\text{c.s.}} \mathbb{E}_f(w(X)) = I$$

Además se cumple que

$$\mathbb{E}(\hat{I}) = \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n w(X_i)\right) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(w(X_i)) = \mathbb{E}(w(X_1))$$

por lo tanto se tiene que \hat{I} es un estimador insesgado para I . Por lo tanto, su error cuadrático medio es

$$\text{ECM}(\hat{I}) = \text{Var}(\hat{I}) + (\text{Sesgo}(\hat{I}))^2 = \text{Var}(\hat{I})$$

Como

$$\text{Var}(\hat{I}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n w(X_i)\right) \stackrel{\text{indep.}}{=} \frac{1}{n} \text{Var}(w(X_1))$$

entonces se tiene que

$$\text{ECM}(\hat{I}) = \frac{1}{n} \text{Var}(w(X_1))$$

Si $\mathbb{E}(w^2(X_1)) < +\infty \Rightarrow \text{Var}(w(X_1)) < +\infty$, por lo tanto se puede tomar como medida

del error cometido a

$$\sqrt{\text{ECM}(\hat{I})} = \frac{1}{\sqrt{n}} \sqrt{\text{Var}(\hat{I})} = \frac{1}{\sqrt{n}} \sigma(w(X_1))$$

Por lo tanto, como se puede ver en Ripley (1987, p. 120), bajo el supuesto de momento segundo finito de $w(X)$, la medida del error cometido $\sqrt{\text{ECM}(\hat{I})}$ es $O\left(\frac{1}{\sqrt{n}}\right)$ a medida que crece el número n de simulaciones.

Además, en este caso se puede calcular el error estándar del estimador como $s.\hat{e}. = \frac{s}{\sqrt{n}}$ donde $s^2 = \frac{1}{n-1} \sum_{i=1}^n (w(X_i) - \hat{I})^2$.

Siguiendo a Robert y Casella (2004, §3.2) la generalización del método básico en el caso unidimensional consiste en estimar cantidades de la forma $I = \mathbb{E}_f(h(X)) = \int_{\mathcal{X}} h(x)f(x) dx$ donde f es una densidad de probabilidad, siendo este tipo de integrales de uso muy frecuente en estadística. Al igual que el desarrollo del método básico de integración, si se genera una muestra X_1, \dots, X_n i.i.d. de f , es natural proponer como estimador \hat{I} de $I = \mathbb{E}_f(h(X))$ al promedio empírico $\hat{I} = \overline{h(X)}_n = \frac{1}{n} \sum_{i=1}^n h(X_i)$, donde por la ley débil de los grandes números se tiene que

$$\hat{I} = \overline{h(X)}_n \xrightarrow{c.s.} \mathbb{E}_f(h(X)) = I$$

Además, cuando $\mathbb{E}_f(h^2) < +\infty$, la velocidad de convergencia de \hat{I} se puede evaluar como la varianza

$$\text{Var}(\hat{I}) = \frac{1}{n} \text{Var}(h(X_1)) = \frac{1}{n} \int_{\mathcal{X}} (h(x) - \mathbb{E}_f(h(x)))^2 f(x) dx$$

donde además puede ser estimada a partir de la muestra como

$$v_n = \frac{1}{n^2} \sum_{i=1}^n (h(x_i) - \hat{I})^2$$

A.6. Muestreo de importancia

Considerando $I = \int_{\mathcal{X}} h(x)f(x) dx$ donde f es una densidad de probabilidad y siguiendo a Wasserman (2003, §24.3). El método de integración Monte Carlo de la sección anterior

se basa en un muestreo directo de la f sin importar cuál sea la misma. Este procedimiento, en general resulta poco efectivo por dos razones principales.

Hay casos en que es posible que no se sepa como muestrear valores de la f por más conocida que la misma sea. Por ejemplo, en inferencia bayesiana, la densidad posterior $f(\theta|x)$ se obtiene como proporcional al producto de la verosimilitud $L(\theta)$ y la densidad prior $f(\theta)$. Esto no garantiza que $f(\theta|x)$ sea una distribución de una familia conocida como la normal, la gamma o cualquier otra y por lo tanto no se puedan aplicar los métodos de simulación conocidos.

Siguiendo a Ripley (1987, p. 122), otro problema viene por la forma de la propia f ya que algunos resultados pueden ser más importantes que otros para determinar I , por lo tanto se necesita que seleccionar estos valores con más frecuencia. Un ejemplo puede ser tomar I como la probabilidad de un evento A , cuya ocurrencia sea muy poco probable según la densidad f . En este contexto se tiene que $I = P(A) = \int_{\mathcal{X}} \mathbb{1}_A(x)f(x) dx$. La única alternativa para estimar I con precisión es producir ese evento raro con más frecuencia.

El muestreo de importancia es una alternativa basada en la integración Monte Carlo generalizada, que logra superar estos problemas. El método consiste en tomar una función de densidad g de la cual es fundamental saber como simular valores. Entonces:

$$I = \mathbb{E}_f(h(X)) = \int_{\mathcal{X}} h(x)f(x) dx = \int_{\mathcal{X}} \frac{h(x)f(x)}{g(x)}g(x) dx = \mathbb{E}_g(Y)$$

donde $Y = \frac{h(X)f(X)}{g(X)}$, X es una nueva variable aleatoria con densidad g y \mathbb{E}_g es la esperanza respecto a la densidad g . Por comodidad, se suele seguir llamando X a la variable aleatoria de cual se va a hacer el muestreo de importancia, para no sobrecargar de notación. En este caso, siempre tener presente que al aplicar muestreo de importancia, se cambia el evento original X .

En este nuevo escenario, se puede simular X_1, \dots, X_n , una muestra i.i.d. de la densidad g y estimar I como en el método Monte Carlo básico con

$$\hat{I} = \hat{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i = \frac{1}{n} \sum_{i=1}^n \frac{h(X_i)f(X_i)}{g(X_i)}$$

Esto es lo que se llama muestreo de importancia. En base a esa muestra simulada por la ley de los grandes números se puede afirmar que como en el método de integración Monte Carlo se cumple que $\hat{I} \xrightarrow{c.s.} I$. Sin embargo, hay dos detalles.

En primer lugar, es posible que \hat{I} pueda tener un error estándar infinito. Esto es debido a que I es la esperanza de $w(X) = \frac{h(X)f(X)}{g(X)}$. Por lo tanto el momento segundo de $w(X)$ es

$$\mathbb{E}_g(w^2(X)) = \int_{\mathcal{X}} \left(\frac{h(x)f(x)}{g(x)} \right)^2 g(x) \, dx = \int_{\mathcal{X}} \frac{h^2(x)f^2(x)}{g(x)} \, dx$$

Entonces, si g tiene colas con probabilidades demasiado pequeñas respecto a las de f , la integral anterior puede ser infinita. Para evitar esto, una regla básica es elegir a la densidad g con colas lo suficientemente probables como para evitar ese inconveniente.

En segundo lugar, puede ocurrir que la densidad g sea pequeña en algún conjunto B donde f es grande. Nuevamente el cociente $\frac{f}{g}$ puede ser grande, haciendo que la varianza sea grande.

Estos dos inconvenientes muestran que una condición para la densidad g es que sea de una forma similar a la f pero con colas de g más probables que las de f .

En base a esto se puede demostrar que existe g^* , el cambio de densidad óptimo que minimiza $\text{Var}(\hat{I})$ siendo $g^*(x) = \frac{|h(x)f(x)|}{\int_{\mathcal{X}} |h(u)f(u)| \, du}$. Aunque, si bien es interesante este teorema, no pasa más allá de la cuenta teórica ya que si no se sabe muestrear de f va a ser poco probable que se sepa como simular valores de $\frac{|h|f}{\int_{\mathcal{X}} |h|f}$.

En resumen y en la práctica, los esfuerzos en construir la densidad g para simular valores se van a centrar en una g cuya forma sea similar a la de $f|h|$ pero con colas más pesadas.

A modo de ejemplo y para ver la ventaja de utilizar un muestreo de importancia adecuado ahora se considera una variable $Z \sim \mathcal{N}(0, 1)$. El objetivo es estimar $I = \mathbb{P}(Z > 3) = \int_3^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2} \, dz$.

- Primero se escribe de una forma más adecuada la integral de interés I . La misma queda $I = \int_3^{+\infty} 1 \cdot \frac{1}{\sqrt{2\pi}} e^{-z^2} \, dz = \int_{\mathbb{R}} \mathbb{1}_{[3, +\infty)}(z) \cdot \frac{1}{\sqrt{2\pi}} e^{-z^2} \, dz$. Con la notación usada en los párrafos anteriores, aquí se tiene que $h(z) = \mathbb{1}_{[3, +\infty)}(z)$ y $f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2}$ y

su dominio es $\mathcal{X} = \mathbb{R}$.

- Por lo tanto, una forma de estimar dicha cantidad puede ser aplicando integración Monte Carlo generalizada ya que de la densidad f se conoce como simular valores. De esta forma se estima I como $\hat{I} = \frac{1}{n} \sum_{i=1}^n h(X_i)$ donde X_1, \dots, X_n es i.i.d. de la $\mathcal{N}(0, 1)$. Por ejemplo, usando $n = 100$ y simulando varias veces se obtiene lo siguiente que $\mathbb{E}(\hat{I}) \cong 0,001$ y $\text{Var}(\hat{I}) = 10^{-5}$. El código para R (R Core Team, 2021) con las respectivas salidas se puede ver más abajo. También es fácil notar que muchas observaciones simuladas no son cercanas a la cola derecha, por lo tanto hay muchas observaciones que no son aprovechadas adecuadamente. Este es un claro ejemplo en que la I está asociada a un evento raro. Por lo tanto, integración Monte Carlo resulta poco efectivo ya que no se logra observar al evento raro lo necesario. Una solución puede ser, aumentar el n para así poder tener mejor información sobre el evento raro. Eso lleva más tiempo de ejecución, que en el presente ejemplo puede no tener relevancia, pero otro tipo de modelo algo más complejo puede llevar días de ejecución.
- Con lo observado en el item anterior y tratando de no aumentar el tamaño de la muestra simulada, se puede proponer un muestreo de una g , de una forma similar a la f pero que contemple al evento $[3, +\infty)$ como algo no tan raro. Por ejemplo, se puede considerar como g a la densidad de una $X \sim \mathcal{N}(4, 1)$. En este nuevo contexto y usando muestreo de importancia con g , se simula X_1, \dots, X_{100} i.i.d. de una $\mathcal{N}(4, 1)$ y se estima I con $\hat{I}_{\text{MI}} = \frac{1}{n} \sum_{i=1}^n \frac{f(X_i)h(X_i)}{g(X_i)}$. Para comparar, de nuevo se vuelve a usar $n = 100$ y se hace varias veces dicho experimento. En este caso se obtiene $\mathbb{E}(\hat{I}_{\text{MI}}) \cong 0,001462102$ y $\text{Var}(\hat{I}) \cong 4,126957 \times 10^{-08}$. Como bien se puede apreciar la desviación estándar se ha reducido en un factor de 242. El código en R (R Core Team, 2021) con las respectivas salidas se muestra a continuación.

```
options(width = 65)

rm(list = ls())
```

```

u <- runif(10^4)
(I <- mean(u^3))

## [1] 0.2489514

sd(u^3)/sqrt(length(u))

## [1] 0.002840099

sqrt(sum((u^3 - I)^2)/(length(u) - 1))/sqrt(length(u))

## [1] 0.002840099

## runif(1)
set.seed(583808)
n <- 2
m <- 10

## Integración Monte Carlo generalizada

hat.I.mcg <- rep(NA, m)
k <- 1
{
  t <- proc.time()[3]
  while (k <= m) {
    hat.I.mcg[k] <- mean(rnorm(10^n) > 3)
    k <- k + 1
  }
  proc.time()[3] - t
}

## elapsed
##      0

head(hat.I.mcg)

## [1] 0.01 0.00 0.00 0.00 0.00 0.00

tail(hat.I.mcg)

## [1] 0 0 0 0 0 0

hat.I.mcg

```

APÉNDICE A. MÁS FUNDAMENTO TEÓRICO

```
## [1] 0.01 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00

mean(hat.I.mcg)

## [1] 0.001

var(hat.I.mcg)

## [1] 1e-05

## Muestreo de importancia con  $g \sim N(4, 1)$ 

hat.I.mi <- rep(NA, m)
k <- 1
{
  t <- proc.time()[3]
  while (k <= m) {
    aux <- rnorm(10^n, mean = 4, sd = 1)
    hat.I.mi[k] <- mean((aux > 3) * dnorm(aux)/dnorm(aux, mean = 4,
      sd = 1))
    k <- k + 1
  }
  proc.time()[3] - t
}

## elapsed
## 0.02

head(hat.I.mi)

## [1] 0.001768874 0.001551710 0.001438460 0.001268603 0.001147765
## [6] 0.001418896

tail(hat.I.mi)

## [1] 0.001147765 0.001418896 0.001438633 0.001341236 0.001445376
## [6] 0.001801473

hat.I.mi

## [1] 0.001768874 0.001551710 0.001438460 0.001268603 0.001147765
## [6] 0.001418896 0.001438633 0.001341236 0.001445376 0.001801473
```

```

mean(hat.I.mi)
## [1] 0.001462102

var(hat.I.mi)
## [1] 4.126957e-08

var(hat.I.mcg)/var(hat.I.mi)
## [1] 242.3093

## MI con U[2.8, 4]

hat.I.mi.u <- rep(NA, m)
k <- 1
{
  t <- proc.time()[3]
  while (k <= m) {
    aux <- runif(10^n, 2.8, 4)
    hat.I.mi.u[k] <- mean((aux > 3) * dnorm(aux)/dunif(aux, 2.8,
      4))
    k <- k + 1
  }
  proc.time()[3] - t
}

## elapsed
## 0.02

head(hat.I.mi.u)
## [1] 0.001488576 0.001403190 0.001387890 0.001340843 0.001390903
## [6] 0.001141166

tail(hat.I.mi.u)
## [1] 0.001390903 0.001141166 0.001338299 0.001133618 0.001378680
## [6] 0.001083544

hat.I.mi.u
## [1] 0.001488576 0.001403190 0.001387890 0.001340843 0.001390903

```

```
## [6] 0.001141166 0.001338299 0.001133618 0.001378680 0.001083544
mean(hat.I.mi.u)
## [1] 0.001308671
var(hat.I.mi.u)
## [1] 1.894831e-08
var(hat.I.mcg)/var(hat.I.mi.u)
## [1] 527.7515
var(hat.I.mi)/var(hat.I.mi.u)
## [1] 2.178008
```

A.7. Teorema de Slutsky

Para el resultado final del teorema del límite central para el gradiente de MCLA se recuerda el teorema de Slutsky. El mismo, en una versión general se puede encontrar en [Gut \(2005, p. 249\)](#).

Teorema de Slutsky: Se consideran dos sucesiones de variables aleatorias $\{X_i\}_{i \in \mathbb{N}}$ $\{Y_i\}_{i \in \mathbb{N}}$, una variable aleatoria X y una constante $a \in \mathbb{R}$.

Si $X_n \xrightarrow{d} X$ e $Y_n \xrightarrow{p} a$, entonces:

- $X_n + Y_n \xrightarrow{d} X + a$
- $X_n - Y_n \xrightarrow{d} X - a$
- $X_n \cdot Y_n \xrightarrow{d} X \cdot a$
- si además $a \neq 0$: $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{a}$

Este teorema también se puede encontrar como teorema de Cramer, aunque es más común encontrarlo atribuido a Slutsky.

A.8. Test de Wald e intervalos de confianza

Como se puede ver en Agresti (2002) y gracias al trabajo de Wald (1943), se puede introducir un test de hipótesis nula $H_0)\beta = \beta_0$ y se puede discutir su relación con intervalos de confianza. Esto se puede hacer gracias a la normalidad en muestras grandes de los estimadores máximo verosímiles.

Con un error estándar no nulo $s_{\hat{\beta}}$, el estadístico $Z = \frac{\hat{\beta} - \beta_0}{s_{\hat{\beta}}}$ tiene una distribución asintótica normal estándar cuando $n \rightarrow \infty$ y asumiendo como cierta la hipótesis nula $\beta = \beta_0$.

Entonces, según sea la hipótesis alternativa que acompañe a $H_0)\beta = \beta_0$, se usa el valor z del estadístico para obtener los p-valor de los test, tanto unilateral $(H_A)\beta < \beta_0$ o $(H_A)\beta > \beta_0$ como bilateral $(H_A)\beta \neq \beta_0$.

Este tipo de estadístico, usando el error estándar no nulo, es llamado un *estadístico de Wald* (Wald, 1943).

En la práctica también aporta información relevante el intervalo de confianza. Un intervalo de confianza al $100(1 - \alpha)\%$ se puede obtener invirtiendo el test de hipótesis

$$\begin{cases} H_0)\beta = \beta_0 \\ H_A)\beta \neq \beta_0 \end{cases} \quad \text{al nivel } \alpha.$$

El intervalo de confianza de Wald es el conjunto de los β_0 tales que $\frac{|\hat{\beta} - \beta_0|}{s_{\hat{\beta}}} < z_{\frac{\alpha}{2}}$. Esto da como resultado el intervalo $\hat{\beta} \pm z_{\frac{\alpha}{2}} s_{\hat{\beta}}$, con confianza $100(1 - \alpha)\%$.

Apéndice B

Código y resultados

Para los cálculos se utiliza la función `glm` de la librería homónima **glm** (Knudson, 2020) en una computadora portátil estándar del mercado con un procesador Intel® Core™ i5-5200U y 8 GB de memoria RAM. La principal referencia usada para este apéndice es Knudson (2016, §5.1, §5.2).

El conjunto de datos de las salamandras se distribuye en la librería **glm** de Knudson (2020). A continuación una pequeña descripción de los datos:

```
options(width = 65)
if (!require(glm)) {
  install.packages("glm")
  library(glm)
}

## Loading required package: glm

## Loading required package: trust

## Loading required package: mvtnorm

## Loading required package: Matrix

## Loading required package: parallel
```

APÉNDICE B. CÓDIGO Y RESULTADOS

```
## Loading required package: doParallel

## Loading required package: foreach

## Loading required package: iterators

data(salamander)
head(salamander)

##   Mate Cross Female Male
## 1    1  R/R    10   10
## 2    1  R/R    11   14
## 3    1  R/R    12   11
## 4    1  R/R    13   13
## 5    1  R/R    14   12
## 6    1  R/W    15   28

tail(salamander)

##   Mate Cross Female Male
## 355    1  W/W    64   66
## 356    0  W/R    65   50
## 357    0  W/R    66   54
## 358    0  W/R    67   52
## 359    0  W/R    68   53
## 360    0  W/R    69   51

str(salamander)

## 'data.frame': 360 obs. of  4 variables:
##  $ Mate   : num  1 1 1 1 1 1 0 0 0 0 ...
##  $ Cross  : Factor w/ 4 levels "R/R","R/W","W/R",...: 1 1 1 1 1 2 2 2 2 2 ...
##  $ Female : Factor w/ 60 levels "10","11","12",...: 1 2 3 4 5 6 7 8 9 10 ...
##  $ Male   : Factor w/ 60 levels "10","11","12",...: 1 5 2 4 3 19 18 16 20 17 ...

table(salamander$Mate)

##
##  0  1
## 171 189

table(salamander$Cross)
```

```
##
## R/R R/W W/R W/W
## 90 90 90 90

table(table(salamander$Female))

##
## 6
## 60

table(table(salamander$Male))

##
## 6
## 60
```

Los datos muestran los resultados expuestos en [4](#).

La variable **Mate** muestra un total de 360 ensayos de los cuales 189 fueron apareamientos exitosos, mientras que los restantes 171 no lo fueron. Este es el vector que se va a considerar como respuesta y debe ser numérico.

La variable **Cross** muestra los cruces totales hechos. Como se puede ver en [4.1](#), los 360 ensayos se componen de 90 ensayos por cada cruce. Se puede comprobar en dicha tabla ya que se usó en los tres experimentos. Cada cruce se compone de 5×2 ensayos por día. Luego, de los 6 días del experimento, cada uno de esos cruces se repite 3 veces. Y por último, esto sucede en cada uno de los experimentos. Por lo tanto, el total de ensayos por cruce es $5 \times 2 \times 3 \times 3 = 90$. Esta variable debe ser factor o numérica para ser usada en `glmm`.

Tanto la variable **Female** como la variable **Male** corresponde a los identificadores de las salamandras que interviene en cada ensayo. Un ensayo corresponde a una fila, donde en la variable **Mate** está el resultado del cruce de la hembra identificada con el valor que aparece en **Female**, el macho con el identificador que aparece en **Male** y corresponde a un cruce del tipo que se indica en **Cross**. Las variables **Female** y **Male** se utilizan para definir los efectos aleatorios y deben ser factor. Cada variable contiene 60 identificadores diferentes,

APÉNDICE B. CÓDIGO Y RESULTADOS

correspondiente a las 60 salamandras diferentes usadas (20 por cada experimento y con el supuesto de que las salamandras del segundo experimento son diferentes a las del primero). Recordando que cada salamandra se usó para hacer 6 ensayos, $60 \times 6 = 360$, el total de ensayos.

Ahora se presenta el código para algunas estimaciones, mostrando en una línea de comentario el tiempo de ejecución de cada una de ellas.

```
options(width = 65)
set.seed(7784)
{
  t0 <- proc.time()[3]
  m1 <- glmm(Mate ~ 0 + Cross, random = list(~0 + Female, ~0 +
    Male), varcomps.names = c("F", "M"), data = salamander,
    family.glmm = bernoulli.glmm, m = 10^4)
  tf <- proc.time()[3]
}
# tf-t0 = 748 = total de segundos que demoró esta ejecución

{
  t0 <- proc.time()[3]
  m2 <- glmm(Mate ~ 0 + Cross, random = list(~0 + Female, ~0 +
    Male), varcomps.names = c("F", "M"), data = salamander,
    family.glmm = bernoulli.glmm, m = 10^5)
  tf <- proc.time()[3]
}
# tf-t0 = 5965

{
  t0 <- proc.time()[3]
  m3 <- glmm(Mate ~ 0 + Cross, random = list(~0 + Female, ~0 +
    Male), varcomps.names = c("F", "M"), data = salamander,
    family.glmm = bernoulli.glmm, m = 10^4, par.init = c(0.8053,
    0.574, -2.33, 1.0965, 1.3086, 1.762))
  tf <- proc.time()[3]
```

```

}
# Al usar el argumento par.int, la función no ejecuta las
# aproximaciones por PQL.

# tf-t0 = 746

set.seed(7784)
{
  t0 <- proc.time()[3]
  m4 <- glmm(Mate ~ 0 + Cross, random = list(~0 + Female, ~0 +
    Male), varcomps.names = c("F", "M"), data = salamander,
    family.glmm = bernoulli.glmm, m = 3 * 10^4, par.init = c(0.8053,
      0.574, -2.33, 1.0965, 1.3086, 1.762))
  tf <- proc.time()[3]
}
# tf-t0 = 2555

set.seed(7784)
{
  t0 <- proc.time()[3]
  m5 <- glmm(Mate ~ 0 + Cross, random = list(~0 + Female, ~0 +
    Male), varcomps.names = c("F", "M"), data = salamander,
    family.glmm = bernoulli.glmm, m = 3 * 10^4)
  tf <- proc.time()[3]
}
# tf-t0 = 1712

```

Los resultados **m1**, **m2** y **m5** se obtiene utilizan como valores iniciales las aproximaciones PQL de los parámetros. La misma se hace por defecto cuando no se especifica un vector de valores iniciales para los parámetros. Cada resultado se diferencia en el parámetro *m* que corresponde a la cantidad de iteraciones para Monte Carlo. Los resultados **m3** y **m4** utilizan como valores iniciales los valores estimados en el modelo **m1**. Estos valores se incorporan al proceso de estimación mediante el parámetro *par.init*, siguiendo el orden de parámetros según los muestra la función `summary.glmm`. El uso del parámetro

APÉNDICE B. CÓDIGO Y RESULTADOS

`par.init` hace que no se ejecute la aproximación PQL para valores iniciales. Por ejemplo para el modelo **m3**, esto se puede comprobar haciendo `m3$u.pql`. Este vector es el vector nulo, indicando que no se ejecutó la aproximación PQL para dicho elemento. A su vez, `m3$beta.pql` y `m3$nu.pql` contienen los valores que se incorporan mediante el parámetro `m3$par.init`.

Una vez calculados varios modelos en diferentes escenarios se muestran los valores estimados de β y la componente de varianza ν . Además, se calculan las fuentes de error, el error de muestreo y el error Monte Carlo expuesto en 3.2.3.

```
options(width = 65)
# Coeficientes y componentes de varianza estimadas
cc <- rbind(round(c(m1$beta, m1$nu), 3), round(c(m2$beta, m2$nu),
3), round(c(m3$beta, m3$nu), 3), round(c(m4$beta, m4$nu), 3),
round(c(m5$beta, m5$nu), 3))
row.names(cc) <- c("m1", "m2", "m3", "m4", "m5")
cc
##      CrossR/R CrossR/W CrossW/R CrossW/W      F      M
## m1      0.805      0.574     -2.330      1.097  1.309  1.762
## m2      1.016      0.320     -2.025      0.931  1.406  1.205
## m3      1.100      0.514     -1.551      1.094  0.702  0.824
## m4      0.877      0.214     -1.582      0.768  0.527  0.659
## m5      0.879      0.332     -2.153      1.065  1.486  1.423
# Error de muestreo
errm <- rbind(round(se(m1), 3), round(se(m2), 3), round(se(m3), 3),
round(se(m4), 3), round(se(m5), 3))
row.names(errm) <- c("m1", "m2", "m3", "m4", "m5")
errm
##      CrossR/R CrossR/W CrossW/R CrossW/W      F      M
## m1      0.267      0.259      0.314      0.274  0.240  0.323
## m2      0.385      0.387      0.572      0.410  0.587  0.376
## m3      0.255      0.237      0.294      0.261  0.133  0.151
## m4      0.247      0.245      0.292      0.269  0.100  0.129
## m5      0.313      0.284      0.375      0.316  0.323  0.323
```

```

# Error Monte Carlo
{
  t0 <- proc.time()[3]
  errmc <- rbind(round(mcse(m1), 4), round(mcse(m2), 4), round(mcse(m3),
    4), round(mcse(m4), 4), round(mcse(m5), 4))
  proc.time()[3] - t0
}

## elapsed
## 722.89

row.names(errmc) <- c("m1", "m2", "m3", "m4", "m5")
errmc

##      CrossR/R CrossR/W CrossW/R CrossW/W      F      M
## m1    0.0009    0.0014    0.0013    0.0004 0.0010 0.0018
## m2    0.0261    0.0383    0.1990    0.0905 0.2501 0.0275
## m3    0.0033    0.0058    0.0019    0.0088 0.0063 0.0010
## m4    0.0123    0.0206    0.0175    0.0488 0.0071 0.0108
## m5    0.0667    0.0124    0.0564    0.0920 0.0834 0.0624

```

Las estimaciones para los cinco modelos no presentan mayor diferencia para que alguno de los modelos se destaque sobre los otros.

El error de muestreo presenta valores similares para todos los modelos. Como se expone en 3.2.3, es independiente del muestreo Monte Carlo, por lo tanto no es posible reducirlo aumentando el tamaño **m** del muestreo ni haciendo otra simulación. Sólo es posible cambiar este error si se hiciera otro experimento con 120 salamandras, análogo al que se describe en 4.

Por último, la diferencia más notable se encuentra en el error Monte Carlo. Este se produce por la simulación Monte Carlo, por el tamaño de la misma y los valores simulados. El más destacable es **m1** el cual tiene los errores más pequeños en comparación con los demás. Por lo tanto, se elige el resultado **m1** para seguir trabajando.

El modelo elegido se analiza con la función `summary.glm`. A continuación el resultado:

APÉNDICE B. CÓDIGO Y RESULTADOS

```
options(width = 65)
summary.glmm(m1)

##
## Call:
## glmm(fixed = Mate ~ 0 + Cross, random = list(~0 + Female, ~0 +
##   Male), varcomps.names = c("F", "M"), data = salamander,
##   family.glmm = bernoulli.glmm,
##   m = 10^4)
##
##
## Link is: "logit (log odds)"
##
## Fixed Effects:
##           Estimate Std. Error z value Pr(>|z|)
## CrossR/R    0.8053     0.2666   3.021  0.00252 **
## CrossR/W    0.5740     0.2590   2.219  0.02646 *
## CrossW/R   -2.3300     0.3140  -7.417  1.20e-13 ***
## CrossW/W    1.0965     0.2745   3.995  6.47e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Variance Components for Random Effects (P-values are one-tailed):
##   Estimate Std. Error z value Pr(>|z|)/2
## F    1.3086     0.2404   5.444   2.60e-08 ***
## M    1.7620     0.3230   5.456   2.44e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Los efectos fijos se muestran con su correspondiente estimación y su estimación del error de muestreo. Luego se presenta información de los test de hipótesis correspondientes a

$$\begin{cases} H_0) \beta_i = 0 \\ H_A) \beta_i \neq 0 \end{cases}$$
 (para más detalles ver [A.8](#)). Se presenta tanto el valor del estadístico en z value como el p-valor en $\text{Pr}(>|z|)$. Como se puede ver, todas estimaciones de los pará-

metros del efecto fijo son significativamente diferentes de 0 para un nivel de significación $\alpha = 0,05$.

Los efectos aleatorios se muestran en forma análoga. Se muestra la estimación de las componentes de varianzas y su estimación del error de muestreo. Luego se presenta el resultado

de los test de hipótesis $\begin{cases} H_0) \nu_i = 0 \\ H_A) \nu_i > 0 \end{cases}$ (para más detalles ver [A.8](#)). Se presenta tanto el

valor del estadístico en `z` value y el p-valor de la prueba unilateral en `Pr(>|z|)/2`. Y como se puede observar, ambas estimaciones son significativamente diferente de 0 para un nivel de significación $\alpha = 0,05$.

Dadas las estimaciones de los parámetros, es posible calcular para cada una de ellas un intervalo de confianza con la función `confint` (para más detalles ver [A.8](#)). Para dichos intervalos se elige una confianza de 99%.

```
options(width = 65)
round(confint(m1, level = 0.99), 3)

##           0.005  0.995
## CrossR/R  0.119  1.492
## CrossR/W -0.092  1.240
## CrossW/R -3.140 -1.521
## CrossW/W  0.390  1.804
## F         0.690  1.928
## M         0.930  2.594
```

Luego, para los parámetros estimados se puede analizar cuales son significativamente diferentes.

Por ejemplo, dado un nivel α , si se busca hacer el test de hipótesis $\begin{cases} H_0) \beta_{RR} = \beta_{WW} \\ H_A) \beta_{RR} \neq \beta_{WW} \end{cases}$

se puede reescribir como $\begin{cases} H_0) \beta_{RR} - \beta_{WW} = 0 \\ H_A) \beta_{RR} - \beta_{WW} \neq 0 \end{cases}$ y utilizar el test de Wald ([A.8](#)).

APÉNDICE B. CÓDIGO Y RESULTADOS

El estadístico para la prueba es $E = \frac{\hat{\beta}_{RR} - \hat{\beta}_{WW} - 0}{\sqrt{\text{Var}(\hat{\beta}_{RR} - \hat{\beta}_{WW})}}$ donde $E \sim \mathcal{N}(0, 1)$ de forma asintótica si $\beta_{RR} = \beta_{WW}$. Para calcular la varianza del denominador se puede utilizar la igualdad $\text{Var}(\hat{\beta}_{RR} - \hat{\beta}_{WW}) = \text{Var}(\hat{\beta}_{RR}) + \text{Var}(\hat{\beta}_{WW}) - 2\text{Cov}(\hat{\beta}_{RR}, \hat{\beta}_{WW})$.

El p-valor para el test de hipótesis se calcula como $P(|E| \geq |e_{\text{Obs}}|) = 2\Phi(-|e_{\text{Obs}}|)$.

Los valores de varianzas para estimadores de parámetros y covarianzas entre ellos se pueden obtener de la matriz de covarianzas. Esta matriz se puede obtener del modelo con la función `vcov`. A continuación se muestra la matriz de covarianzas y los p-valor de todos los test de Wald posibles para el vector $\hat{\beta}$.

```
options(width = 65)
mcov <- vcov(m1)
round(mcov, 5)

##          CrossR/R CrossR/W CrossW/R CrossW/W          F          M
## CrossR/R  0.07106 -0.00004  0.00026 -0.00002 -0.00002 -0.00021
## CrossR/W -0.00004  0.06689 -0.00023  0.00018 -0.00005  0.00031
## CrossW/R  0.00026 -0.00023  0.09871 -0.00004 -0.00028 -0.00042
## CrossW/W -0.00002  0.00018 -0.00004  0.07534  0.00011  0.00015
## F         -0.00002 -0.00005 -0.00028  0.00011  0.05777  0.00017
## M         -0.00021  0.00031 -0.00042  0.00015  0.00017  0.10434

param <- c("R/R", "R/W", "W/R", "W/W")
salida <- as.character()
for (i in 1:3) {
  for (j in (i + 1):4) {
    varianza <- mcov[i, i] + mcov[j, j] - 2 * mcov[i, j]
    e_obs <- (m1$beta[i] - m1$beta[j])/sqrt(varianza)
    salida <- rbind(salida, paste(paste(paste("p-valor de test entre",
      paste(param[i], param[j], sep = "-")), "es"), round(2 *
      pnorm(-abs(e_obs)), 4)))
  }
}

salida
```

```
##      [,1]
## [1,] "p-valor de test entre R/R-R/W es 0.5335"
## [2,] "p-valor de test entre R/R-W/R es 0"
## [3,] "p-valor de test entre R/R-W/W es 0.4467"
## [4,] "p-valor de test entre R/W-W/R es 0"
## [5,] "p-valor de test entre R/W-W/W es 0.1654"
## [6,] "p-valor de test entre W/R-W/W es 0"
```

Por lo tanto, para un nivel de significación $\alpha = 0,05$ se puede decir que

- β_{RR} y β_{RW} no son significativamente diferentes
- β_{RR} y β_{WW} no son significativamente diferentes
- β_{RW} y β_{WW} no son significativamente diferentes

Aunque, como se puede observar, la diferencia mayor la tienen β_{RW} y β_{WW} .

También se puede hacer el test de Wald con las componentes de varianza. En este caso el resultado es el siguiente:

```
options(width = 65)
varianza <- mcov[5, 5] + mcov[6, 6] - 2 * mcov[5, 6]
e_obs <- (m1$nu[1] - m1$nu[2])/sqrt(varianza)
print(paste(paste(paste("p-valor de test entre nu_F y nu_M es"),
  round(2 * pnorm(-abs(e_obs)), 4))))
## [1] "p-valor de test entre nu_F y nu_M es 0.2593"
```

Por lo tanto, fijando también un nivel de significación $\alpha = 0,05$ se puede concluir que ν_F y ν_M no son significativamente diferentes.