

Índice

1. Introducción al QGIS	15
1.1. Introducción	15
1.2. Sistemas de Información Geográfica	15
1.2.1. Utilidad de los Sistemas de Información Geográfica . . .	18
1.2.2. Aplicaciones de los Sistemas de Información Geográfica	18
1.2.3. Sistemas de Información Geográfica más utilizados . . .	19
1.3. Instalación de QGIS	20
1.3.1. Descripción del área de trabajo	20
1.4. Tipos de datos GIS	21
1.4.1. GIS vectorial	22
1.4.2. Ejemplo de QGIS vectorial	25
1.4.3. GIS raster	28
1.4.4. Ejemplo de QGIS raster	32
2. Utilización y Manejo de QGIS	45
2.1. Introducción	45
2.2. Incorporación de Tablas de Datos	45
2.3. Selección Espacial	51
2.4. Análisis Espacial de Proximidad	60
2.5. Presentación e Impresión	66
3. Interacción entre QGIS y R	71
3.1. Introducción	71
3.2. Configuración de QGIS	72
3.3. Ejecución de programas de R a través de QGIS	75
4. Análisis de Datos Espaciales de tipo discreto. Procesos Puntuales	81
4.1. Introducción	81
4.2. Datos espaciales y su representación	82
4.3. Procesos Puntuales Espaciales	86

4.3.1.	Análisis de la distribución espacial	87
4.3.2.	Aleatoriedad Espacial Completa (<i>CSR</i>)	90
4.3.3.	Ajuste de Modelos Espaciales Puntuales	95
4.3.4.	Análisis de la densidad espacial	104
5.	Análisis de Datos Espaciales de tipo continuo. Geoes-	
	tadística	107
5.1.	Introducción	107
5.2.	Variograma	107
5.2.1.	Utilización de covariables	108
5.2.2.	Análisis exploratorio del Variograma	108
5.3.	Interpolación espacial	111
6.	Análisis de Datos Espaciales agregados o regionales	115
6.1.	Introducción	115
6.2.	Entornos y pesos de Áreas	115
6.3.	Contraste global de autocorrelación espacial: Estadístico <i>I</i> de Moran	116
6.4.	Contraste local de autocorrelación espacial: Gráfico de disper- sión de Moran	119
6.5.	Ajuste de Modelos	121
7.	Modelos Lineales Generalizados GLM	127
7.1.	Introducción	127
7.2.	Definición de Modelo Lineal Generalizado univariante	128
7.2.1.	Dispersión excesiva (<i>Overdispersion</i>)	133
7.3.	Estimación y Contrastes basados en la verosimilitud	134
7.3.1.	Estimador de máxima verosimilitud de los β_i	134
7.3.2.	Estimador del parámetro de escala ξ	137
7.3.3.	Contrastes de hipótesis sobre los parámetros	137
7.3.4.	Contraste de bondad de ajuste del modelo	138
7.3.5.	Diagnóstico del Modelo	139
7.4.	Cálculo con R	140
7.4.1.	Regresión Logística y Regresión Binomial	140
	Interpretación de los coeficientes del Modelo de Regre- sión Logística ajustado	146
	Dispersión excesiva (<i>Overdispersion</i>)	152
7.4.2.	Regresión Logística Multinomial	154
7.4.3.	Regresión Poisson	154
7.5.	Métodos basados en la cuasi-verosimilitud	158
7.6.	Métodos Bayesianos	159
7.7.	Métodos robustos	160

7.7.1.	M -estimadores basados en la cuasi-verosimilitud	160
7.7.2.	Contraste robusto de bondad de ajuste del modelo . . .	162
7.7.3.	Cálculo con R^{mo}	163
7.8.	Ajuste de modelos GLM para datos espaciales	169
8.	Modelos Aditivos Generalizados GAM	171
8.1.	Introducción	171
8.2.	Modelos GAM clásicos	173
8.2.1.	Estimación	173
8.2.2.	Validación Cruzada (<i>Cross validation</i>)	175
8.2.3.	Cálculo con R	177
8.3.	Modelos GAM robustos	180
9.	Bibliografía	189

Capítulo 7

Modelos Lineales Generalizados GLM

7.1. Introducción

Modelos Lineales Generalizados es una denominación genérica que engloba algunos métodos ya estudiados anteriormente, tales como la Regresión Lineal Simple (CB-capítulo 9), la Regresión Lineal Múltiple (CB-capítulo 10), la Regresión Logística (TA-capítulo 9) o la Regresión Poisson (TA-capítulo 10), así como otros Métodos de Regresión aún no estudiados y que serán analizados en este capítulo.

La razón de realizar un estudio global de estos métodos es la de obtener, de una sola vez, resultados aplicables a todos ellos. En particular en lo referente a los Métodos Robustos utilizados en dichos modelos. Esta generalización se consigue, en un primer momento, con un mayor nivel de abstracción por lo que el capítulo puede resultar, en ocasiones, demasiado técnico. Si el lector está interesado principalmente por las aplicaciones de estos métodos, encontrará más interesantes la Sección 7.4, si desea un enfoque clásico, y la Sección 7.7.3 cuando busque un análisis robusto.

A continuación aparecen tres ejemplos que serán resueltos en dichas secciones.

Ejemplo 7.1

Consideraremos el experimento realizado por Phelps (1982) en el que se anotó, para cada uno de los $i = 24$ grupos, el número de zanahorias dañadas por insectos de entre todas las del grupo. Las zanahorias fueron plantadas en tres bloques, por lo que al ser ésta una covariable de tipo cualitativo, debieron considerarse en el modelo dos covariables indicadoras, `bloque1` y `bloque2`. Además, se fumigó según ocho dosis de un determinado insecticida, considerándose la covariable cuantitativa `log(dosis)` en el modelo.

Se pretende ajustar a estos datos un Modelo de Regresión Binomial clásico y otro robusto.

Ejemplo 7.2

Feigl y Zelen (1965) analizaron datos de 33 pacientes con leucemia para los que se anotó si su tiempo de supervivencia era superior a 52 semanas (de hecho, ellos anotaron el tiempo de supervivencia y no sólo si era o no mayor a 52 semanas), asociando en ese caso un valor igual a 1, *éxito*, a una variable dependiente Y , valor que tomaría con probabilidad p . Por otro lado, asignaron el valor $Y = 0$ si ese tiempo de supervivencia era inferior o igual a 52 semanas, suceso considerado como *fracaso*, el cual tendría probabilidad $1 - p$.

Como covariables independientes que se piensa pueden explicar a Y se consideraron dos: la covariable WBC , *número de glóbulos blancos por milímetro cúbico de sangre*, (o leucocitos, o en inglés White Blood Cell Count) indicando un valor alto de esta covariable la existencia de infección, y la covariable AG , *presencia* ($AG = 1$) o *ausencia* ($AG = 0$) *de cierta característica morfológica de los glóbulos blancos*. A estos datos se ajustará en Modelo de Regresión Logística clásico y otro robusto.

Ejemplo 7.3

Los artículos de Lindenmayer y sus colaboradores (en la bibliografía damos dos de estos artículos) proporcionan multitud de datos sobre las Montañas Centrales de Victoria en Australia. Aquí trabajaremos con datos sobre diferentes especies de marsupiales arborícolas de Bosques Montano tipo Ash (*Montane Ash Forest*).

En este estudio se analizaron 151 lugares diferentes de 3ha con vegetación uniforme, observándose en cada uno de éstos la variable dependiente de respuesta, número de especies de marsupiales en el lugar (Diversidad), y las covariables siguientes: el número de arbustos (Arbustos); si había, 1, o no, 0, tocones de pasadas operaciones forestales (Tocones) que es una variable cualitativa con dos niveles; el número de árboles de porte hueco (Stags); un índice de cortezas extraídas (Cortezas); un índice de habitabilidad para marsupiales (Habitat); el área de acacias (Acacias); el tipo de Eucalipto que es una variable cualitativa con tres niveles: Eucalipto regnans (Regnans), Eucalipto delegatensis (Delegatensis) y Eucaliptus nitens (Nitens); y, por último, el aspecto del lugar que es una variable de tipo cualitativo con cuatro niveles, (NWN), (NWSE), (SESW) y (SWNW).

Se pretende ajustar un Modelo de Regresión Poisson a estos datos, clásico y robusto.

Aunque el Modelo de Regresión Lineal Simple o Múltiple es un caso particular de Modelo Lineal General y, por tanto, también puede ser considerado como otro caso más en este capítulo, no lo haremos porque ya en el texto CB lo estudiamos con detalle desde un punto de vista clásico, y en el texto MR y en el capítulo anterior, desde un punto de vista robusto. Eso sí, utilizaremos estos modelos como punto de partida.

7.2. Definición de Modelo Lineal Generalizado univariante

Para definir los Modelos Lineales Generalizados, partiremos del Modelo de Regresión Lineal estudiado en el capítulo anterior. Vimos allí que modelizar nuestros datos con un Modelo de Regresión Lineal Múltiple supone considerar una variable *dependiente* o de *respuesta* Y sobre la que pensamos influyen

linealmente k variables *independientes* o *covariables* X_1, \dots, X_k de la forma

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + e \quad [7.1]$$

siendo e una variable de error con distribución normal $N(0, \sigma)$.

En el Modelo de Regresión Lineal [7.1] se persigue, entre otras cosas, estimar los parámetros $\beta_0, \beta_1, \dots, \beta_k$ en base a una muestra aleatoria de tamaño $n(> k + 1)$ de las variables independientes y de la dependiente, dando origen a los datos

$$\begin{array}{cccc} y_1 & x_{11} & \dots & x_{1k} \\ \vdots & & & \\ y_i & x_{i1} & \dots & x_{ik} \\ \vdots & & & \\ y_n & x_{n1} & \dots & x_{nk} \end{array}$$

Si denominamos $\mathbf{y} = (y_1, \dots, y_n)^t$ al vector de las observaciones de la variable dependiente, englobamos los parámetros en un vector de parámetros $\boldsymbol{\beta} = (\beta_0, \dots, \beta_k)^t$, y llamamos *matriz del diseño*

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ \vdots & & & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{pmatrix}$$

a la matriz $n \times (k + 1)$ de las observaciones de las variables independientes, el Modelo de Regresión Lineal se suele expresar de la forma

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

en donde $\mathbf{e} = (e_1, \dots, e_n)^t$ es el vector de errores.

Con este modelo estamos interesados en estimar los parámetros de $\boldsymbol{\beta}$ en base a los datos

$$(y_i, \mathbf{x}_i^t) = (y_i, 1, x_{i1}, \dots, x_{ik}) \quad , \quad i = 1, \dots, n.$$

En este Modelo de Regresión Lineal, la variable de respuesta Y es de tipo cuantitativo. Las covariables suelen ser de tipo cuantitativo (aunque también podrían considerarse de tipo cualitativo), y pueden ser *determinísticas*, es decir valores conocidos o condiciones experimentales, o pueden ser *estocásticas*, es decir valores de un vector aleatorio.

Si suponemos que las covariables son de tipo determinístico, el modelo lineal [7.1] puede reformularse diciendo que tenemos n observaciones independientes y_1, \dots, y_n procedentes de distribuciones $N(\mu_i, \sigma)$ en donde la media μ_i es de la forma

$$\mu_i = \mathbf{x}_i^t \boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} \quad i = 1, \dots, n.$$

Si, como habitualmente sucede, las covariables se consideran estocásticas, el esquema sería el mismo aunque, ahora, condicional; en concreto, los n pares (y_i, \mathbf{x}_i^t) se suponen observaciones independientes y, dadas las \mathbf{x}_i , las Y_i serán (condicionalmente) independientes con distribución

$$Y_i | \mathbf{x}_i \rightsquigarrow N(\mu_i, \sigma) \quad i = 1, \dots, n$$

con

$$E[Y_i | \mathbf{x}_i] = \mu_i = \mathbf{x}_i^t \boldsymbol{\beta} \quad i = 1, \dots, n.$$

En un Modelo Lineal Generalizado (univariante) ampliamos un poco la situación anterior. De nuevo suponemos que, dadas las \mathbf{x}_i , las n variables Y_i son (condicionalmente) independientes aunque ahora, la variable de respuesta Y_i puede ser de tipo continuo, puede ser de recuentos de observaciones, o puede ser de tipo binario.

Las dos condiciones antes recuadradas ahora también se generalizan. En este tipo de modelos suponemos que la distribución de las Y_i (condicionada por las \mathbf{x}_i) no es necesariamente normal, sino una familia de tipo exponencial con esperanza (condicional) $E[Y_i | \mathbf{x}_i] = \mu_i$ y, posiblemente, con un parámetro de escala (común para todas las Y_i) denominado ξ . Más en concreto, se supone que la distribución de las $Y_i | \mathbf{x}_i$ tiene por función de densidad una *familia de tipo exponencial* (Vélez y García Pérez, 1993, Ejemplo 5.17) de la forma

$$f(y_i | \theta_i, \xi) = \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{\xi} + c(y_i, \xi) \right\} \quad [7.2]$$

en donde θ_i se denomina *parámetro natural*, ξ es el *parámetro de escala* o *dispersión*, y b y c dos funciones que determinan el tipo de familia exponencial.

Además, en un Modelo Lineal Generalizado, la forma en que las covariables suministran información sobre la media μ_i de la variable dependiente ya no es necesariamente lineal mediante el *predictor lineal* $\eta_i = \mathbf{x}_i^t \boldsymbol{\beta}$, sino que lo hacen mediante una función de respuesta h con inversa $h^{-1} = g$, denominada esta última función *link*, es decir, de la forma

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

o bien,

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

Por tanto, un Modelo Lineal Generalizado vendrá especificado cuando fijemos el tipo de familia exponencial para las distribuciones condicionadas $Y_i|\mathbf{x}_i$, la función *link* g y el vector (o matriz) del diseño \mathbf{x}_i .

En estas distribuciones de $Y_i|\mathbf{x}_i$, se supone que el parámetro natural es una función w_1 de la media; es decir, $\theta_i = w_1(\mu_i)$ siendo $\mu_i = b'(\theta_i) = \partial b(\theta_i)/\partial \theta_i$.

Además, la varianza en estas distribuciones también es de una forma peculiar, $Var(Y_i|\mathbf{x}_i) = \xi w_2(\mu_i)$, en donde la función w_2 también se determina a partir de la función b de la forma $w_2(\mu_i) = b''(\theta_i) = \partial^2 b(\theta_i)/\partial \theta_i^2$. Es decir, suponemos que es $E(Y_i|\mathbf{x}_i) = b'(\theta_i)$ y $Var(Y_i|\mathbf{x}_i) = \xi b''(\theta_i)$.

Para cada familia exponencial existe una función *link natural* o *canónica* que es la que iguala al parámetro natural con el predictor lineal; es decir, $\theta_i = w_1(\mu_i) = g(\mu_i) = \eta_i = \mathbf{x}_i^t \boldsymbol{\beta}$; es decir, la obtenida a partir de la ecuación

$$g(\mu) \equiv w_1(\mu).$$

Ejemplo 7.4

Si las $Y_i|\mathbf{x}_i$ se distribuyen como normales $N(\mu_i, \sigma)$, su función de densidad será

$$\frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^2}(y_i - \mu_i)^2 \right\} = \frac{1}{\sigma\sqrt{2\pi}} e^{-y_i^2/(2\sigma^2)} \exp \left\{ \frac{y_i\mu_i - \mu_i^2/2}{\sigma^2} \right\}.$$

Si comparamos la expresión anterior con [7.2], podemos identificar, observando el término clave (el que involucra a las y_i y las μ_i), que es $\theta_i = \mu_i = w_1(\mu_i)$ (con lo que será $w_1(\mu) = \mu$), $b(\theta_i) = \mu_i^2/2$ y $\xi = \sigma^2$.

El término restante deberá ser $\exp\{c(y_i, \xi)\} = 1/(\sigma\sqrt{2\pi}) e^{-y_i^2/(2\sigma^2)}$ aunque éste es irrelevante a la hora de identificar los elementos de la distribución modelo.

Como se observa, es $b'(\theta_i) = \partial b(\theta_i)/\partial \theta_i = \mu_i$ y $w_2(\mu_i) = b''(\theta_i) = \partial^2 b(\theta_i)/\partial \theta_i^2 = 1$, con lo que $Var(Y_i|\mathbf{x}_i) = \xi w_2(\mu_i) = \xi$.

Finalmente, de la ecuación clave

$$g(\mu) \equiv w_1(\mu) = \mu$$

se deduce que, en el caso de ser f una distribución normal (caso de Regresión Lineal), debe de ser $g(\mu) = \mu$, lo que implica una función *link* canónica igual a la identidad.

En el caso de ser f una distribución Poisson, $\mathcal{P}(\lambda_i)$ la distribución de probabilidad se puede expresar como

$$f(y_i|\theta_i, \xi) = \frac{1}{y_i!} \exp\{y_i \log \lambda_i - \lambda_i\}$$

con lo que, observando [7.2], deberá ser

$$\theta_i = \log \lambda_i \quad \text{y} \quad b(\theta_i) = \lambda_i$$

de la primera de estas igualdades se deduce que debe ser $\lambda_i = e^{\theta_i}$, obteniendo de la segunda, en consecuencia, que es $b(\theta_i) = \lambda_i = e^{\theta_i}$.

Por otro lado, al ser λ_i la media de Y_i , deberá ser $\theta_i = w_1(\mu_i)$, es decir, $\log \lambda_i = w_1(\lambda_i)$, por lo que la función w_1 es $w_1(\lambda) = \log \lambda$. Finalmente, de la ecuación $g(\mu) \equiv w_1(\mu)$ obtenemos $g(\lambda) = \log \lambda$, que indica a la función logaritmo como la función *link* canónica en este tipo de modelos de Regresión Poisson.

En el caso de seguir las $Y_i|\mathbf{x}_i$ una distribución binomial $B(n_i, p_i)$, será

$$f(y_i|\theta_i, \xi) = \binom{n_i}{y_i} p_i^{y_i} (1-p_i)^{n_i-y_i} = \binom{n_i}{y_i} \exp \left\{ y_i \log \frac{p_i}{1-p_i} + n_i \log(1-p_i) \right\}$$

con lo que, observando [7.2], deberá ser

$$\theta_i = \log \frac{p_i}{1-p_i}, \quad b(\theta_i) = -n_i \log(1-p_i) \quad \text{y} \quad \xi = 1.$$

Como la media de la distribución binomial, $B(n_i, p_i)$, es $\mu_i = n_i p_i$, de la ecuación $\theta_i = w_1(\mu_i)$ obtenemos

$$w_1(\mu_i) = w_1(n_i p_i) = \log \frac{p_i}{1-p_i} = \log \frac{n_i p_i}{n_i - n_i p_i} = \log \frac{\mu_i}{n_i - \mu_i}$$

y, finalmente, de la ecuación $g(\mu) \equiv w_1(\mu)$, la función *link* canónica $g(\mu) = \log(\mu/(n - \mu))$. Por tanto, la ecuación que relaciona la media de la variable de respuesta con las covariables $g(\mu_i) = \mathbf{x}_i^t \boldsymbol{\beta}$, será

$$\log \left(\frac{\mu_i}{n_i - \mu_i} \right) = \log \left(\frac{n_i p_i}{n_i - n_i p_i} \right) = \log \left(\frac{p_i}{1-p_i} \right) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k.$$

Observemos que, en el caso de que la variable respuesta sea Bernoulli, $Y_i|\mathbf{x}_i \rightsquigarrow B(1, p_i)$ en donde ésta sólo toma los valores *éxito* y *fracaso*, tendremos un caso particular del anterior (correspondiente a la Regresión Logística) en donde la función *link* será $g(\mu) = \log(\mu/(1-\mu))$ o lo que es lo mismo, $g(p) = \log(p/(1-p))$ por ser para esta distribución $\mu = p$. La ecuación que relaciona la media de la variable de respuesta con las covariables es, en este caso, la misma de antes,

$$\log \left(\frac{\mu_i}{1-\mu_i} \right) = \log \left(\frac{p_i}{1-p_i} \right) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

por lo que no se suele hacer distinción entre estos dos últimos casos y se habla de la función *link* canónica $g(\mu) = \log(\mu/(1-\mu))$, denominada *logit*.

En resumen, prescindiendo de la nomenclatura dada a la variable de la función considerada, hemos obtenido tres funciones *link*, la función *link identidad*, $g(\mu) = \mu$, la función *link logaritmo* o simplemente *log*, $g(\mu) = \log \mu$ y la función *link logit*, $g(\mu) = \log(\mu/(1-\mu))$, funciones *link* naturales o canónicas de los modelos, respectivamente, normal, Poisson y binomial (Bernoulli).

Se utilizan también otras funciones *link*, la función *link inversa*, $g(\mu) = -1/\mu$ y la función *link gaussiana-inversa*, $g(\mu) = -2/\mu^2$, funciones *link* canónicas de los modelos, respectivamente, gamma y gaussiano-inverso.

Otras funciones *link* no canónicas, pero que se pueden utilizar en algún modelo son, la función *link probit*, $g(\mu) = \Phi^{-1}(\mu)$, es decir, la inversa de la función de distribución de una normal estándar $N(0, 1)$, la función *link*

complementaria log-log, $g(\mu) = \log(-\log(1-\mu))$ y la función *link raíz cuadrada*, $g(\mu) = \sqrt{\mu}$.

Con el paquete R podemos trabajar con los cinco modelos antes mencionados, formando la Tabla 7.1 en la que aparece una **C** indicando la función *link* canónica. Las opciones marcadas con una **p** indican que también pueden elegirse como funciones *link*, pero que no son las canónicas.

	Modelos				
	Normal	Poisson	Binomial	Gamma	Gaussiano-inverso
Funciones <i>link</i>					
identidad	C	p	–	p	–
logaritmo	–	C	–	p	–
logit	–	–	C	–	–
inversa	–	–	–	C	–
gaussiana-inversa	–	–	–	–	C
probit	–	–	p	–	–
complementaria log-log	–	–	p	–	–
raíz cuadrada	–	p	–	–	–

Tabla 7.1: Modelos y funciones *link*

7.2.1. Dispersión excesiva (*Overdispersion*)

Supongamos que queremos modelizar nuestros datos mediante un Modelo de Regresión Logístico. En este caso, la distribución asociada a las Y_i en el Modelo Lineal Generalizado sería la Bernoulli $B(1, p)$, con media p y varianza $p(1 - p)$.

Si quisiéramos modelizar los datos con un Modelo de Regresión Poisson, la distribución sería Poisson, $\mathcal{P}(\lambda)$, de media λ y varianza λ .

Supongamos ahora que, al observar nuestros datos, vemos que, en uno u otro caso, su varianza es mayor de la que debería ser. En estos casos, modelizaremos los datos, para la primera situación, con un Modelo de Regresión Logística, de varianza $\varsigma p(1 - p)$ y, en el segundo caso, mediante un Modelo de Regresión Poisson, pero con varianza $\varsigma \lambda$.

En estas situaciones decimos que nuestros datos presentan una *dispersión excesiva* (*overdispersion*), con parámetro de *overdispersion* ς , problema que trataremos más adelante al hablar de cada uno de los dos modelos.

7.3. Estimación y Contrastes basados en la verosimilitud

La estimación de los parámetros del Modelo Lineal Generalizado (así como contrastes de hipótesis referentes a éstos), además de dos tests de bondad del ajuste, se pueden realizar siguiendo métodos basados en la *verosimilitud*. En posteriores secciones estudiaremos Métodos basados en la cuasi-verosimilitud y Métodos Bayesianos.

7.3.1. Estimador de máxima verosimilitud de los β_i

En esta sección determinaremos la forma en la que estimar los parámetros β_i del modelo; es posible que los diferentes parámetros y funciones que intervienen en el Modelo Lineal Generalizado puedan entorpecer la comprensión del proceso, pero hemos querido desgranar éste puesto que la ecuación de verosimilitud resultante (en realidad, sistema de ecuaciones) es clave en las posteriores generalizaciones y robustificación.

La manera en la que habitualmente estimamos los parámetros de un modelo es mediante la utilización del Método de la Máxima Verosimilitud (CB-sección 5.2). Para ello, primero debemos expresar la función de verosimilitud como función del parámetro. Si observamos [7.2] los parámetros del modelo serán θ_i y ξ ; de momento supondremos ξ conocido (aunque más abajo volveremos sobre ello). La función de verosimilitud será, por tanto,

$$L(\theta_1, \dots, \theta_n) = \prod_{i=1}^n f(y_i|\theta_i) = \exp \left\{ \sum_{i=1}^n \left(\frac{y_i\theta_i - b(\theta_i)}{\xi} - c(y_i, \xi) \right) \right\}.$$

El Método de la Máxima Verosimilitud indica asignar como estimadores de los parámetros a aquellos valores que hagan máxima dicha función de verosimilitud. Como el máximo de una función y de su logaritmo se alcanzan en el mismo punto, determinaremos el máximo del logaritmo de $L(\theta_1, \dots, \theta_n)$,

$$\log L(\theta_1, \dots, \theta_n) = \sum_{i=1}^n \left(\frac{y_i\theta_i - b(\theta_i)}{\xi} \right) - \sum_{i=1}^n c(y_i, \xi).$$

Como suponemos ξ conocido y vamos a maximizar esta función derivando respecto al parámetro e igualando a cero la derivada, el segundo sumando de la expresión anterior se anulará por lo que prescindiremos de él en lo que sigue considerándolo, simplemente, como una constante, cte.

Si reparametrizamos la función anterior (es decir, cambiamos los parámetros), al ser $\theta_i = w_1(\mu_i)$ tendremos, (la última igualdad es sólo notación)

$$\log L(\mu_1, \dots, \mu_n) = \sum_{i=1}^n \left(\frac{y_i w_1(\mu_i) - b(w_1(\mu_i))}{\xi} \right) + \text{cte} = \sum_{i=1}^n l_i(\mu_i) + \text{cte} \quad [7.3]$$

y si volvemos a reparametrizar, expresando la verosimilitud anterior en términos de las β_i y las covariables, por ser $\mu_i = h(\mathbf{x}_i^t \boldsymbol{\beta})$ tendremos

$$\log L(\boldsymbol{\beta}) = \sum_{i=1}^n \left(\frac{y_i w_1(h(\mathbf{x}_i^t \boldsymbol{\beta})) - b(w_1(h(\mathbf{x}_i^t \boldsymbol{\beta})))}{\xi} \right) + \text{cte}. \quad [7.4]$$

La derivada de esta expresión la debemos obtener teniendo en cuenta las funciones que aparecen en ella y la denominación que hemos dado a sus variables.

Conviene recordar también que, como $\boldsymbol{\beta}$ es un vector, al hablar de la derivada de $\log L(\boldsymbol{\beta})$ con respecto a $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)^t$, la cual representamos por $\partial \log L(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}$, nos referimos al vector de derivadas parciales

$$(\partial \log L(\boldsymbol{\beta}) / \partial \beta_0, \dots, \partial \log L(\boldsymbol{\beta}) / \partial \beta_k)^t$$

el cual igualaremos al vector de ceros, dando origen a un sistema de ecuaciones de verosimilitud, de $k+1$ ecuaciones con $k+1$ incógnitas, $\beta_0, \beta_1, \dots, \beta_k$.

Observamos también que derivar [7.4] respecto a $\boldsymbol{\beta}$ va a consistir, básicamente, en aplicar reiteradamente la derivada de una función de función por lo que expresaremos cada una de las funciones de la composición con respecto a su variable; además, como el mismo lector puede comprobar fácilmente, es

$$\frac{\partial \mathbf{x}_i^t \boldsymbol{\beta}}{\partial \boldsymbol{\beta}} = \mathbf{x}_i$$

Derivando en [7.4] será

$$\begin{aligned} \frac{\partial \log L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \frac{1}{\xi} \sum_{i=1}^n \left[y_i \cdot \left(\frac{\partial w_1(\mu_i)}{\partial \mu_i} \right) \Big|_{\mu_i=h(\mathbf{x}_i^t \boldsymbol{\beta})} \cdot \boldsymbol{\mu}_i' - b'(w_1(\mu_i)) \cdot \left(\frac{\partial w_1(\mu_i)}{\partial \mu_i} \right) \Big|_{\mu_i=h(\mathbf{x}_i^t \boldsymbol{\beta})} \cdot \boldsymbol{\mu}_i' \right] \\ &= \frac{1}{\xi} \sum_{i=1}^n \left(\frac{\partial w_1(\mu_i)}{\partial \mu_i} \right) \Big|_{\mu_i=h(\mathbf{x}_i^t \boldsymbol{\beta})} \boldsymbol{\mu}_i' (y_i - \mu_i(\boldsymbol{\beta})) \end{aligned}$$

por ser $b'(w_1(\mu_i)) = \mu_i(\boldsymbol{\beta}) = \mu_i$, y siendo

$$\boldsymbol{\mu}_i' = \frac{\partial \mu_i}{\partial \boldsymbol{\beta}} = \left(\frac{\partial h(\eta)}{\partial \eta} \right) \Big|_{\eta=h(\mathbf{x}_i^t \boldsymbol{\beta})} \cdot \frac{\partial \mathbf{x}_i^t \boldsymbol{\beta}}{\partial \boldsymbol{\beta}} = \left(\frac{\partial h(\eta)}{\partial \eta} \right) \Big|_{\eta=h(\mathbf{x}_i^t \boldsymbol{\beta})} \cdot \mathbf{x}_i = D_i(\boldsymbol{\beta}) \mathbf{x}_i$$

en donde la última igualdad sólo se ha introducido como notación para definir $D_i(\boldsymbol{\beta})$.

Como es $\mu_i = b'(\theta_i)$ será $\theta_i = (b')^{-1}(\mu_i)$ y, como era $\theta_i = w_1(\mu_i)$, será $w_1(\mu_i) = (b')^{-1}(\mu_i)$ por lo que, utilizando la fórmula para la derivada de la función inversa, será

$$\frac{\partial w_1(\mu_i)}{\partial \mu_i} = \frac{\partial (b')^{-1}(\mu_i)}{\partial \mu_i} = \frac{1}{b''((b')^{-1}(\mu_i))} = \frac{1}{b''(\theta_i)} = \frac{1}{w_2(\mu_i)} = \frac{\xi}{Var(Y_i|\mathbf{x}_i)}.$$

Por tanto, la derivada buscada se podrá expresar en cualquiera de las siguientes dos maneras,

$$\frac{\partial \log L(\beta)}{\partial \beta} = \sum_{i=1}^n \frac{z_i D_i(\beta)}{Var(Y_i|\mathbf{x}_i)} (y_i - \mu_i(\beta)) = \sum_{i=1}^n \frac{\mu_i'}{\xi w_2(\mu_i)} (y_i - \mu_i)$$

como aparece, respectivamente, en Fahrmeir y Tutz (1994, pp. 38) o en Cantoni y Ronchetti (2001, pp. 1022).

El sistema de ecuaciones de verosimilitud

$$\frac{\partial \log L(\beta)}{\partial \beta} = \sum_{i=1}^n \frac{\mu_i'}{\xi w_2(\mu_i)} (y_i - \mu_i) = \mathbf{0} \quad [7.5]$$

no va a tener habitualmente una solución analítica y debe de resolverse de forma numérica mediante un método iterativo. R utiliza el más habitual, el de mínimos cuadrados ponderados IWLS (*iteratively reweighted least squares*), también denominado de las marcas de Fisher (*Fisher scoring*). Otras alternativas son el Método de Newton-Raphson o, mejor, los Métodos Quasi-Newton.

El estimador de máxima verosimilitud $\hat{\beta}$ obtenido mediante alguno de los métodos anteriores, cuando exista y sea único, tendrá una distribución asintótica normal multivariante,

$$\hat{\beta} \rightsquigarrow N(\beta, V)$$

siendo la matriz de covarianzas V aproximadamente igual a la inversa de la matriz de información de Fisher

$$V \approx A^{-1}(\hat{\beta})$$

siendo dicha matriz de información igual a

$$A(\hat{\beta}) = \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^t D_i^2(\hat{\beta}) \frac{1}{w_2(h(\mathbf{x}_i^t \hat{\beta})) \xi}$$

7.3.2. Estimador del parámetro de escala ξ

Si el parámetro de escala ξ no fuese conocido podría estimarse, a partir del estimador $\hat{\beta}$, por la expresión,

$$\hat{\xi} = \frac{1}{n - (k + 1)} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{w_2(\hat{\mu}_i)} \quad [7.6]$$

en donde $\hat{\mu}_i = h(\mathbf{x}_i^t \hat{\beta})$, obteniéndose de esta manera un estimador consistente para ξ , el cual puede utilizarse en la expresión de $A^{-1}(\hat{\beta})$.

Obsérvese que, en un Modelo de Regresión Normal, el estimador anterior del parámetro de escala coincide con el obtenido para la varianza σ^2 mediante la suma de residuos al cuadrado.

7.3.3. Contrastes de hipótesis sobre los parámetros

Una vez obtenidos los estimadores para los β_i , podemos considerar el realizar tests de hipótesis sobre ellos de la forma $H_0 : \mathbf{C}\beta = \mathbf{c}_0$ frente a la alternativa $H_1 : \mathbf{C}\beta \neq \mathbf{c}_0$. (En esta sección supondremos que el parámetro de escala ξ es conocido o reemplazado por el valor [7.6].)

Un caso particular de estas hipótesis, muy importante, es el contraste de $H_0 : \beta_r = \mathbf{0}$ frente a $H_0 : \beta_r \neq \mathbf{0}$ siendo β_r un subvector de β ; es decir, el contraste de ser cero algunas β_i frente a la alternativa de modelo completo, en el que todas las β_i son distintas de cero.

Se consideran tres tipos de tests de hipótesis. El primero es el *test de razón de verosimilitudes* (Vélez y García Pérez, 1993, Sección 9.2) basado en el estadístico de contraste

$$\Lambda = \frac{\sup_{\beta \in \Theta_0} L(\beta)}{\sup_{\beta \in \Theta} L(\beta)} = \frac{L(\tilde{\beta})}{L(\hat{\beta})}$$

siendo Θ el espacio paramétrico y Θ_0 la parte de este espacio definido por la hipótesis nula; es decir, el cociente entre el máximo de la función de verosimilitud $L(\beta)$ alcanzado cuando las variables β varían en la región definida por la hipótesis nula, $L(\tilde{\beta})$, y el máximo alcanzado por esta función cuando los parámetros toman cualquier valor posible, $L(\hat{\beta})$, por la definición de estimador de máxima verosimilitud.

Como todo test de hipótesis, éste requiere para su ejecución de la distribución del estadístico de contraste bajo la hipótesis nula. Aunque la distribución exacta no es fácilmente calculable, no obstante, sí se sabe (Vélez y García Pérez, 1993, Teorema 9.1) que, para tamaños muestrales suficientemente grandes, se tiene aproximadamente una distribución χ^2