



Modelos aditivos semiparamétricos con respuesta binomial negativa

TESIS

Para optar al grado de Magíster en Estadística

Instituto de Estadística

Universidad de Valparaíso

Chile

Presentada por:

Claudia Andrea Tassara Manríquez

Profesor guía:

Dr. Germán Ibacache Pulgar

Valparaíso, Noviembre, 2017

Agradecimientos

Deseo expresar mi más profundo y sincero agradecimiento a todas aquellas personas que con su ayuda han colaborado en la realización de la presente tesis, en especial a mi profesor guía al Dr. Germán Ibacache Pulgar, por la orientación y sobre todo por la motivación y el apoyo incondicional recibido desde que comencé este nuevo camino hace ya algunos unos años y logré culminar con esfuerzo este arduo trabajo.

Agradezco también a Dios, por darme la fuerza y perseverancia para culminar esta nueva etapa.

Quisiera hacer extensiva mi gratitud a mi jefatura y compañeras de trabajo, de la Universidad de Playa Ancha, por todo el ánimo, energía, apoyo y comprensión que siempre me brindaron en cada momento que lo necesité.

Un agradecimiento especial a mi familia y amigas por la enorme paciencia, motivación y aliento en los momentos más difíciles que tuve que enfrentar en este proceso, lo cual sin duda ha facilitado este nuevo camino.

A todos ustedes, infinitas gracias.

Índice general

Resumen	4
Introducción	5
1. Modelos semiparamétricos con respuesta binomial negativa	9
1.1. Especificación del modelo	10
1.2. Función de verosimilitud penalizada	11
1.3. Función escore penalizada	12
1.3.1. Componente paramétrica	12
1.3.2. Componente no-paramétrica	13
1.3.3. Componente dispersión	13
1.4. Matriz hessiana penalizada	14
1.4.1. $\mathbf{L}_p^{\beta_j \beta_l}$	14
1.4.2. $\mathbf{L}_p^{f_j f_l}$	15
1.4.3. $\mathbf{L}_p^{\phi \phi}$	15
1.4.4. $\mathbf{L}_p^{\beta_j \mathbf{f}_l}$	15
1.4.5. $\mathbf{L}_p^{\beta \phi}$	16
1.4.6. $\mathbf{L}_p^{\mathbf{f}_j \phi}$	16
1.5. Matriz de información de Fisher penalizada	16
1.6. Proceso Iterativo	17

1.7. Grados de libertad efectivos	19
2. Modelos aditivos semiparamétricos con respuesta binomial negativa	21
2.1. Especificación del modelo	22
2.2. Función de penalización	23
2.3. Función escore penalizada	24
2.3.1. Componente paramétrica	24
2.3.2. Componente no-paramétrica	25
2.3.3. Componente dispersión	25
2.4. Matriz hessiana penalizada	25
2.4.1. $\mathbf{L}_p^{\beta_j \beta_l}$	26
2.4.2. $\mathbf{L}_p^{f_j f_l}$	26
2.4.3. $\mathbf{L}_p^{\phi \phi}$	27
2.4.4. $\mathbf{L}_p^{\beta_j \mathbf{f}_l}$	27
2.4.5. $\mathbf{L}_p^{\beta \phi}$	27
2.4.6. $\mathbf{L}_p^{\mathbf{f}_j \phi}$	28
2.5. Matriz de información de Fisher penalizada	28
2.6. Estimación de los parámetros	29
2.7. Derivación del algoritmo backfitting ponderado	30
2.8. Proceso iterativo conjunto	32
3. Aplicación	34
3.1. Descripción de los datos	34
Conclusiones	41
Bibliografía	42

Índice de figuras

3.1. Diagrama de dispersión entre la variable respuesta O ₃ y las 7 variables explicativas meteorológicas observadas en Bacia de Los Ángeles en 1976. . . .	35
3.2. Gráfico normal de probabilidades con banda de confianza generada al 95 % para el componente de desvío y los datos de ozono ajustados al modelo Poisson.	38
3.3. Gráfico normal de probabilidades con banda de confianza generada al 95 % para el componente de desvío y los datos ajustados de ozono al modelo aditivo semiparamétrico con respuesta binomial negativa.	38
3.4. Gráfico de residuos parciales bajo el modelo aditivo semiparamétrico con respuesta binomial negativa.	39

Índice de cuadros

2.1. Cantidades w_i y f_i para algunas transformaciones utilizadas en modelos con respuesta binomial negativa.	23
3.1. Estimaciones de máxima verosimilitud con sus respectivos errores estándar aproximados, con respecto al modelo Poisson.	36
3.2. Estimaciones de máxima verosimilitud con sus respectivos errores estándar aproximados, con respecto al modelo aditivo semiparamétrico con respuesta binomial negativa.	37

Resumen

En este trabajo estudiamos los modelos aditivos semiparamétricos con respuesta binomial negativa, los cuales consideran en su componente sistemática una componente aditiva no paramétrica y una componente paramétrica. Los parámetros del modelo serán estimados usando el método de máxima verosimilitud penalizada, y los errores estándar de los estimadores se estimarán a partir de la inversa de la matriz de información de Fisher penalizada. La estimación del parámetro de suavizamiento se basará en el criterio de validación cruzada o AIC. Se ilustrará la metodología a través de un conjunto de datos reales, para lo cual se utilizará el software R-Project.

Introducción

Los modelos estadísticos forman un conjunto de herramientas que permiten estudiar relaciones entre variables aleatorias. Las formas más simples de esas relaciones son lineales, y por muchos años la clase de modelos lineales normales ha sido utilizada para describir tales relaciones. Sin embargo, muchos estudios han mostrado que las suposiciones de normalidad, impuesta sobre la variable respuesta, y la de linealidad entre la variable respuesta y las variables explicativas, no siempre son adecuadas. Para lidiar con esa dificultad, algunos autores han propuesto una clase de modelos más general, denominada modelos lineales generalizados semiparamétricos. Esta clase de modelos admite diferentes opciones para la distribución de la variable respuesta, permitiendo que ésta pertenezca a la familia exponencial de distribuciones (Modelos lineales generalizados semiparamétricos, Quiroga P, 2015). Por otro lado, el modelo de regresión de respuesta binomial negativa, es utilizado para datos de conteo, generalmente cuando hay sobredispersión de los datos. Por lo general esto sucede cuando es pensada una distribución de Poisson para la respuesta, la que no impone igualdad entre la media y la varianza. Una causa probable de éste fenómeno es la heterogeneidad de la muestra, que puede ser debido a la variabilidad entre los experimentos.

Primero, estudiamos los modelos lineales semiparamétricos con respuesta binomial negativa, base fundamental para el modelo propuesto, en los cuales introducimos la estimación por máxima verosimilitud penalizada, proponemos estimar el parámetro de suavizamiento por el criterio de validación cruzada generalizada, calculamos la matriz Hessiana

y la matriz de información de Fisher. En este modelo se asume que los errores aleatorios son independientes con distribución normal.

Luego, presentamos los modelos aditivos semiparamétricos con respuesta binomial negativa, cuya principal característica es poseer una estructura paramétrica $\mathbf{X}\beta$ y una no-paramétrica (aditiva), que incorpora el efecto no lineal de varias covariables. Además se utiliza la función enlace, la cual permite que la variable respuesta pertenezca a cualquier distribución de la familia exponencial. La estimación de parámetros es a través de la función de verosimilitud penalizada, el término de suavizamiento se estima por validación cruzada generalizada, y calculamos la función score, la matriz Hessiana y la matriz de información de Fisher. En la aplicación metodológica del modelo utilizamos residuos de desviación en la bondad del ajuste.

La investigación se realiza por el interés de aprender distintas alternativas de modelación estadística en dificultades, tales como, la distribución de la variable respuesta, específicamente cuando no posee una distribución de probabilidad normal, y también casos en los cuales algunas variables explicativas contribuyan de una forma no-paramétrica con la variable de respuesta, esta situación se observa frecuentemente en variables explicativas como la edad, el tiempo, el espacio o zona en un punto geográfico, como también en situaciones económicas, por ejemplo algún ingreso familiar.

Para finalizar, ajustamos el modelo a un conjunto de datos reales. Se presenta una aplicación, donde se desarrolla la teoría respecto al modelo de regresión semiparamétrico aditivo con respuesta binomial negativa en la situación en que se considera la variable explicativa de comportamiento no paramétrico y otras de comportamiento paramétrico. Las metodologías para el desarrollo del modelo conjuntamente se realiza con la implementación computacional en Rproyect, de modelos ajustados a través del paquete GAMLSS en R.

Familia exponencial

Sean Y_1, \dots, Y_n variables aleatorias independientes. Se dice que las variables $Y_i (i = 1, \dots, n)$ pertenecen a la familia de distribución exponencial si su función de densidad o de probabilidad se puede escribir de la siguiente manera:

$$f_Y(y_i; \theta_i; \phi) = \exp\{(y_i \theta_i - b(\theta_i))/\phi + c(y_i, \phi)\}, \theta \in \Theta \quad (1)$$

donde θ_i denota el parámetro natural (o canónico) de la familia exponencial, el que contiene la información de las variables explicativas, y ϕ el parámetro de escala o de dispersión. La forma específica de la distribución es determinada por las funciones $b(\cdot)$ y $c(\cdot)$. El espacio paramétrico, Θ , es un subconjunto de la recta real, $E(Y) = \mu_i = b'(\theta_i)$ y $Var(Y_i) = (\mu_i)$, donde $V_i = V(\mu_i) = \frac{d\mu_i}{d\theta_i}$ es la función de varianza.

Distribución binomial negativa

La distribución binomial negativa puede ser construida por una distribución de dos pasos para la respuesta de conteo Y , suponiendo que $Y|Z = z \sim P(z)$ con función de probabilidades denotada por $f(y|z)$ y con $Z \sim \text{Gama}(\mu, \phi)$, siendo $\mu > 0, \phi > 0$, donde ϕ no depende de μ . Para este caso se tiene que $E(Z) = \mu$ y la $Var(Z) = \mu + \frac{\mu^2}{\phi}$. Tenemos entonces que:

$$f(y|z) = \frac{e^{-z} z^y}{y!}, y$$

$$g(z; \mu, \phi) = \frac{1}{\Gamma(\phi)} \left(\frac{z\phi}{\mu}\right)^\phi e^{-\frac{\phi z}{\mu}} \frac{1}{z}$$

A continuación:

$$E(Y) = E[E(Y|Z)] = \mu,$$

$$Var(Y) = E[Var(Y|Z)] + Var[E(Y|Z)]$$

$$= \mu + \frac{\mu^2}{\phi}.$$

INTRODUCCIÓN

La densidad de la distribución binomial negativa es dada por:

$$f(y_i) = \frac{\Gamma(\phi + y_i)}{\Gamma(y_i + 1)\Gamma(\phi)} (1 - \mu_i)^{y_i} \mu_i^\phi,$$

Propiedades de la distribución binomial negativa:

$$E(Y) = \phi \frac{1 - \mu}{\mu},$$

$$\text{y } Var(Y) = \phi \frac{1 - \mu}{\mu^2}$$

Capítulo 1

Modelos semiparamétricos con respuesta binomial negativa

En este capítulo estudiamos los modelos semiparamétricos con respuesta binomial negativa, los cuales consideran funciones no paramétricas para modelar la relación no lineal entre la variable de respuesta y las variables explicativas. Dentro de los principales trabajos relacionados con los modelos semiparamétricos para datos de conteo fue desarrollado por Green y Silverman (1994), en el contexto de los Modelos Lineales Generalizados (MLGs), Rigby y Stasinopoulos (2005) desarrollando los GAMLSS(Modelos Aditivos Generalizados de localización, escala y forma), en que flexibilizan algunas suposiciones de MLGs para los modelos aditivos generalizados con una posibilidad de modelar parámetros de asimetría y curtosis, de los parámetros de posición y dispersión. Desarrollando también un paquete de GAMLSS en R proyect, el cual es utilizado para una aplicación en este trabajo. Li (2010) también estudia la parte inferencial del modelo semiparamétrico con respuesta binomial negativa (MSRBN).

1.1. Especificación del modelo

Suponga que Y_i , con $i = 1, \dots, n$, son variables aleatorias independientes tal que Y_i sigue una distribución binomial negativa con parámetros $\mu_i > \phi$; esto es, $Y_i \sim BN(\mu_i, \phi)$, cuya función de densidad está dada por

$$f_{Y_i}(y_i; \mu_i, \phi) = \frac{\Gamma(\phi + y_i)}{\Gamma(y_i + 1)\Gamma(\phi)} \left(\frac{\mu_i}{\mu_i + \phi}\right)^{y_i} \left(\frac{\phi}{\mu_i + \phi}\right)^\phi \quad y_i = 1, 2, \dots, \quad (1.1)$$

donde $\Gamma(\cdot)$ denota la función gama. Análogamente a los modelos lineales generalizados, asumiremos que la parte sistemática del modelo es dada por

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} + f(t_i),$$

en que $g(\cdot)$ representa la función de enlace, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ es un vector de variables explicativas, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ es un vector de parámetros desconocidos, $f(\cdot)$ es una función univariada arbitraria no especificada y t es una variable explicativa que contribuye de forma no paramétrica en el modelo.

Alternativamente, la componente sistemática se puede escribir de la siguiente forma:

$$\mathbf{g}(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{f},$$

en que \mathbf{X} es la matriz de diseño de dimensión $(n \times p)$ con filas x_i^T , $\mathbf{f} = (f_1^0, \dots, f_h^0)^T$ es un vector de dimensión $(h \times 1)$ que depende de los valores ordenados y distintos de la variable explicativa t, t_1^0, \dots, t_h^0 y \mathbf{N} es una matriz de incidencia de dimensión $(n \times h)$ cuyos elementos están dados por la función indicatriz $I(t_i = t_j^0), j = 1, \dots, h$. Cuando no hay empates (observaciones repetidas), $\mathbf{N} = \mathbf{I}_n$, con \mathbf{I}_n denotando una matriz identidad de dimensión $(n \times n)$.

1.2. Función de verosimilitud penalizada

Sea $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \mathbf{f}^T, \phi)^T$ el vector de parámetros que debe ser estimado bajo el modelo MSRBN. El logaritmo de la función de verosimilitud penalizada (función de log-verosimilitud) asociada a $\boldsymbol{\theta}$ está dada por

$$l(\boldsymbol{\theta}) = n\phi \log \phi + \sum_{i=1}^n \log \left\{ \frac{\Gamma(\phi + y_i)}{\Gamma(y_i + 1)\Gamma(\phi)} + \sum_{i=1}^n y_i \log \mu_i \right\} - \sum_{i=1}^n (\phi + y_i) \log(\mu_i + \phi).$$

Es bastante conocido el hecho de que maximizar de manera directa la función $l(\boldsymbol{\theta})$ con respecto a $\boldsymbol{\theta}$, sin imponer restricciones sobre la función f , puede generar problemas de super-ajuste e identificabilidad del parámetro $\boldsymbol{\beta}$. Algunos autores, entre ellos Green (1987), propusieron maximizar $l(\boldsymbol{\theta})$ sujeto a alguna restricción sobre la función f . Específicamente, propusieron penalizar el logaritmo de la función de verosimilitud de la siguiente manera:

$$l_p(\boldsymbol{\theta}) = l(\boldsymbol{\theta}) + \lambda^* \int_{t_a}^{t_b} [f''(t)]^2 dt,$$

en que $f''(t) = \frac{d^2}{dt^2} f(t)$ con $t \in [t_a, t_b]$, y la función f perteneciente al espacio de funciones de Sobolev definido como

$$W_2''[t_a, t_b] = \left\{ f : f'' \in L^2[t_a, t_b] \text{ tal que } f' \text{ es absolutamente continua} \right\},$$

con

$$L^2[t_a, t_b] = \left\{ f : \int_a^b f''(t), dt < \infty \right\}$$

Bajo este esquema de penalización, el estimador de la función f conduce a un spline cúbico natural con nodos dados por los puntos t_j^0 , $j = 1, \dots, h$. Green y Silverman (1994) demostraron que la función de penalización se puede expresar de la siguiente manera:

$$\lambda^* \int_a^b [f''(t)]^2 dt = \lambda^* \mathbf{f}^T \mathbf{K} \mathbf{f},$$

en que \mathbf{K} es una matriz de dimensión $(h \times h)$ definida no-negativa que depende sólo de los nodos t_j^0 's, y λ^* es un escalar que depende del parámetro de suavizamiento λ . Para

$\lambda^* = -\frac{\lambda}{2}$, la función de verosimilitud penalizada asume la siguiente forma:

$$l_p(\boldsymbol{\theta}, \lambda) = l(\boldsymbol{\theta}) - \frac{\lambda}{2} \mathbf{f}^T \mathbf{K} \mathbf{f},$$

en que $\lambda > 0$ es el parámetro de suavizamiento que controla el equilibrio entre la bondad del ajuste y la suavidad de la curva ajustada.

Es importante mencionar que si $\lambda \rightarrow 0$, estamos priorizando un ajuste perfecto, mientras que, si $\lambda \rightarrow \infty$, damos más importancia a la suavidad de la curva.

A seguir presentamos las funciones score asociadas a los elementos del vector de parámetros $\boldsymbol{\theta}$.

1.3. Función escore penalizada

Asumiendo que la función $l_p(\boldsymbol{\theta})$ es regular con respecto a los elementos de $\boldsymbol{\theta}$, la función escore viene dada por

$$\mathbf{U}_p(\boldsymbol{\theta}) = \frac{\partial l_p(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \mathbf{U}_p^\beta(\boldsymbol{\theta}) \\ \mathbf{U}_p^f(\boldsymbol{\theta}) \\ \mathbf{U}_p^\phi(\boldsymbol{\theta}) \end{pmatrix}.$$

1.3.1. Componente paramétrica

Para obtener la función escore asociada al vector de parámetros $\boldsymbol{\beta}$, debemos calcular la derivada de la función $l_p(\boldsymbol{\theta}, \lambda)$ con respecto a cada elemento de $\boldsymbol{\beta}$; esto es,

$$\begin{aligned} \frac{\partial l_p(\boldsymbol{\theta}; \lambda)}{\partial \beta_j} &= \sum_{i=1}^n \left\{ \frac{y_i}{\mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} - \frac{(\phi + y_i)}{(\phi + \mu_i)} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} \right\} \\ &= \sum_{i=1}^n \left\{ \frac{y_i}{\mu_i} \frac{\partial \mu_i}{\partial \eta_i} x_{ij} - \frac{(\phi + y_i)}{(\phi + \mu_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij} \right\} \\ &= \sum_{i=1}^n \left\{ \frac{\phi(\partial \mu_i / \partial \eta_i)}{\mu_i(\phi + \mu_i)} (y_i - \mu_i) x_{ij} \right\} \\ &= \sum_{i=1}^n w_i f_i^{*-1} (y_i - \mu_i) x_{ij}, \end{aligned}$$

en que $w_i = (\partial\mu_i/\partial\eta_i)^2/(\mu_i^2\phi^{-1} + \mu_i)$ y $f_i^* = \frac{\partial\mu_i}{\partial\eta_i}$. Luego, la función score para $\boldsymbol{\beta}$ se puede escribir en forma matricial de la siguiente manera:

$$\mathbf{U}_p^\beta(\boldsymbol{\theta}) = \mathbf{X}^T \mathbf{W} \mathbf{F}^{-1}(\mathbf{Y} - \boldsymbol{\mu}), \quad (1.2)$$

en que \mathbf{X} es una matriz con filas $\mathbf{x}_i^T, i = 1, \dots, n$, $\mathbf{W} = \text{diag}\{w_1, \dots, w_n\}$, $\mathbf{F} = \text{diag}\{f_1^*, \dots, f_n^*\}$, $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ y $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$.

1.3.2. Componente no-paramétrica

Para obtener la función score asociada al vector de parámetros $\mathbf{f} = (\psi_1, \dots, \psi_n)^T$, con $\psi_j = f(t_j^0)$ para $j = 1, \dots, h$, debemos derivar la función $l_p(\boldsymbol{\theta}, \lambda)$ con respecto a cada uno de los elementos de \mathbf{f} ; es decir,

$$\begin{aligned} \frac{\partial \ell_p(\boldsymbol{\theta}; \lambda)}{\partial \psi_j} &= \sum_{i=1}^n \left\{ \frac{y_i}{\mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \psi_j} - \frac{(\phi + y_i)}{(\phi + \mu_i)} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \psi_j} \right\} - [\lambda K f]_j \\ &= \sum_{i=1}^n \left\{ \frac{y_i}{\mu_i} \frac{\partial \mu_i}{\partial \eta_i} n_{ij} - \frac{(\phi + y_i)}{(\phi + \mu_i)} \frac{\partial \mu_i}{\partial \eta_i} \eta_{ij} \right\} - [\lambda K f]_j \\ &= \sum_{i=1}^n \left\{ \frac{\phi(\partial \mu_i / \partial \eta_i)}{\mu_i(\phi + \mu_i)} (y_i - \mu_i) \eta_{ij} \right\} - [\lambda K f]_j \\ &= \sum_{i=1}^n w_i f_i^{*-1} (y_i - \mu_i) n_{ij} - [\lambda K f]_j . \end{aligned}$$

En términos matriciales, la función score queda escrita de la siguiente manera:

$$\mathbf{U}_p^f(\boldsymbol{\theta}) = \mathbf{N}^T \mathbf{W} \mathbf{F}^{-1}(\mathbf{Y} - \boldsymbol{\mu}) - \lambda \mathbf{K} \mathbf{f}, \quad (1.3)$$

en que \mathbf{N} es una matriz de incidencia ($n \times h$) con filas \mathbf{n}_i^T .

1.3.3. Componente dispersión

Finalmente, si derivamos la función $l_p(\boldsymbol{\theta}, \lambda)$ con respecto a ϕ , la función score asociada al parámetro de escala ϕ viene dada por

$$\mathbf{U}_p^\phi(\boldsymbol{\theta}) = \sum_{i=1}^n \left[\psi(\phi + y_i) - \psi(\phi) - \frac{(y_i + \phi)}{(\phi + \mu_i)} + \log \left\{ \frac{\phi}{(\phi + \mu_i)} \right\} + 1 \right] \quad (1.4)$$

en que $\psi(\cdot)$ es la función digama.

1.4. Matriz hessiana penalizada

A continuación se determina la matriz hessiana penalizada (matriz de segundas derivadas) bajo el modelo semiparamétrico con respuesta binomial negativa. En general, esta matriz es definida de la siguiente manera:

$$\mathbf{L}_p^{\boldsymbol{\theta}\boldsymbol{\theta}} = \frac{\partial^2 \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = \begin{pmatrix} \mathbf{L}_p^{\beta\beta} & \mathbf{L}_p^{\beta f} & \mathbf{L}_p^{\beta\phi} \\ \mathbf{L}_p^{f\beta} & \mathbf{L}_p^{ff} & \mathbf{L}_p^{f\phi} \\ \mathbf{L}_p^{\phi\beta} & \mathbf{L}_p^{\phi f} & \mathbf{L}_p^{\phi\phi} \end{pmatrix}, \quad (1.5)$$

1.4.1. $\mathbf{L}_p^{\beta_j\beta_l}$

En este caso,

$$\begin{aligned} \mathbf{L}_p^{\beta_j\beta_l} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \beta_j \partial \beta_l} \\ &= \frac{\partial}{\partial \beta_l} \sum_{i=1}^n \frac{f_i^*}{(\mu_i^2 \phi^{-1} + \mu_i)} (y_i - \mu_i) x_{ij} \\ &= \sum_{i=1}^n f_i^* x_{ij} \frac{\partial}{\partial \beta_l} \frac{(y_i - \mu_i)}{(\mu_i^2 \phi^{-1} + \mu_i)} \\ &= - \sum_{i=1}^n \frac{f_i^{*2} x_{ij} x_{il}}{(\mu_i^2 \phi^{-1} + \mu_i)} - \sum_{i=1}^n \frac{f_i^{*2} x_{ij} x_{il} (y_i - \mu_i) (2\phi^{-1} \mu_i + 1)}{(\mu_i^2 \phi^{-1} + \mu_i)^2}. \end{aligned}$$

Matricialmente, esta matriz se puede escribir de la siguiente forma

$$\mathbf{L}_p^{\beta\beta} = -\mathbf{X}^T \mathbf{W} \mathbf{X} - \mathbf{X}^T \mathbf{A}(\mathbf{a}) \mathbf{X} (\mathbf{y} - \boldsymbol{\mu}),$$

donde $\mathbf{A}(\mathbf{a}) = \text{diag}\{a_1, \dots, a_n\}$ y $a_i = f_i^{*2} \frac{(2\phi^{-1} \mu_i + 1)}{(\mu_i^2 \phi^{-1} + \mu_i)^2}$, para $i = 1, \dots, n$.

1.4.2. $\mathbf{L}_p^{f_j f_l}$

Aquí,

$$\begin{aligned}
 \mathbf{L}_p^{f_j f_l} &= \frac{\partial^2 L_p(\theta, \lambda)}{\partial f_j \partial f_l} \\
 &= \frac{\partial}{\partial f_l} \sum_{i=1}^n \frac{f_i^*}{(\mu_i^2 \phi^{-1} + \mu_i)} (y_i - \mu_i) n_{ij} - \frac{\partial}{\partial f_l} [\lambda \mathbf{K} \mathbf{f}]_j \\
 &= \sum_{i=1}^n f_i^* n_{ij} \frac{\partial}{\partial f_l} \frac{(y_i - \mu_i)}{(\mu_i^2 \phi^{-1} + \mu_i)} - \lambda k_{jl} \\
 &= - \sum_{i=1}^n \frac{f_i^{*2} n_{ij} n_{il}}{(\mu_i^2 \phi^{-1} + \mu_i)} - \sum_{i=1}^n \frac{f_i^{*2} n_{ij} n_{il} (y_i - \mu_i) (2\phi^{-1} \mu_i + 1)}{(\mu_i^2 \phi^{-1} + \mu_i)^2} - \lambda k_{ij}.
 \end{aligned}$$

Matricialmente, esta matriz queda escrita como

$$\mathbf{L}_p^{\mathbf{f} \mathbf{f}} = \frac{\partial^2 L_p(\theta, \lambda)}{\partial \mathbf{f} \partial \mathbf{f}^T} = -\mathbf{N}^T \mathbf{W} \mathbf{N} - \mathbf{N}^T \mathbf{A}(\mathbf{a}) \mathbf{N} (\mathbf{y} - \boldsymbol{\mu}) - \lambda \mathbf{K}.$$

1.4.3. $\mathbf{L}_p^{\phi \phi}$

En relación al parámetro ϕ , la matriz hessiana viene dada por

$$\begin{aligned}
 \mathbf{L}_p^{\phi \phi} &= \frac{\partial^2 L_p(\theta, \lambda)}{\partial \phi \partial \phi} \\
 &= \sum_{i=1}^n \left\{ \psi'(\phi + y_i) + \frac{(y_i - 2\mu_i - \phi)}{(\phi + \mu_i)^2} \right\} + \frac{n}{\phi} [1 - \phi \psi'(\phi)].
 \end{aligned}$$

1.4.4. $\mathbf{L}_p^{\beta_j \mathbf{f}_l}$

En este caso,

$$\begin{aligned}
 \mathbf{L}_p^{\beta_j \mathbf{f}_l} &= \frac{\partial^2 \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \beta_j \partial \mathbf{f}_l} \\
 &= \sum_{i=1}^n f_i^* x_{ij} \frac{\partial}{\partial f_l} \frac{(y_i - \mu_i)}{(\mu_i^2 \phi^{-1} + \mu_i)}. \\
 \mathbf{L}_p^{\beta_j \mathbf{f}_l} &= - \sum_{i=1}^n \frac{f_i^{*2} x_{ij} n_{il}}{(\mu_i^2 \phi^{-1} + \mu_i)} - \sum_{i=1}^n \frac{f_i^{*2} x_{ij} n_{il} (y_i - \mu_i) (2\phi^{-1} \mu_i + 1)}{(\mu_i^2 \phi^{-1} + \mu_i)^2}
 \end{aligned}$$

Matricialmente, esta matriz queda escrita de la siguiente forma

$$\mathbf{L}_p^{\beta\mathbf{f}} = \frac{\partial^2 \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta} \partial \mathbf{f}^T} = -\mathbf{X}^T \mathbf{W} \mathbf{N} - \mathbf{X}^T \mathbf{A}(\mathbf{a}) \mathbf{N}(\mathbf{y} - \boldsymbol{\mu}).$$

1.4.5. $\mathbf{L}_p^{\beta\phi}$

Aquí,

$$\begin{aligned} \mathbf{L}_p^{\beta\phi} &= \frac{\partial^2 \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta} \partial \phi} \\ &= \sum_{i=1}^n \frac{f_i^* \phi^{-2}}{(\mu_i \phi^{-1} + 1)^2} (y_i - \mu_i) x_{ij} \end{aligned}$$

Matricialmente, esta matriz queda escrita como

$$\mathbf{L}_p^{\beta\phi} = \frac{\partial^2 \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta} \partial \phi} = \mathbf{X}^T \mathbf{A}(\mathbf{b})(\mathbf{y} - \boldsymbol{\mu})$$

donde $\mathbf{A}(\mathbf{b}) = \text{diag}\{b_1, \dots, b_n\}$ con $b_i = \frac{f_i^* \phi^{-2}}{(\mu_i \phi^{-1} + 1)^2}$, para $i = 1, \dots, n$.

1.4.6. $\mathbf{L}_p^{\mathbf{f}_j\phi}$

Finalmente,

$$\begin{aligned} \mathbf{L}_p^{\mathbf{f}_j\phi} &= \frac{\partial^2 \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \mathbf{f}_j \partial \phi} \\ &= \sum_{i=1}^n \frac{f_i^* \phi^{-2}}{(\mu_i \phi^{-1} + 1)^2} (y_i - \mu_i) \eta_{ij}, \end{aligned}$$

En forma matricial, se puede escribir como,

$$\mathbf{L}_p^{\mathbf{f}_j\phi} = \frac{\partial^2 \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \mathbf{f} \partial \phi} = \mathbf{N}^T \mathbf{A}(\mathbf{b})(\mathbf{y} - \boldsymbol{\mu}).$$

1.5. Matriz de información de Fisher penalizada

En general, la matriz de información de Fisher penalizada es definida de la siguiente forma:

$$\mathbf{K}_p^{\boldsymbol{\theta}\boldsymbol{\theta}} = -\mathbf{E} \left\{ \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right\}.$$

En este caso, la matriz $\mathbf{K}_p^{\theta\theta}$ asume la siguiente estructura diagonal en bloque:

$$\mathbf{K}_p^{\theta\theta} = \begin{pmatrix} \mathbf{K}_p^{\beta\beta} & \mathbf{K}_p^{\beta\mathbf{f}} & \mathbf{0} \\ \mathbf{K}_p^{\mathbf{f}\beta} & \mathbf{K}_p^{\mathbf{f}\mathbf{f}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{K}_p^{\phi\phi} \end{pmatrix},$$

en que,

$$\mathbf{K}_p^{\beta\beta} = \mathbf{X}^T \mathbf{W} \mathbf{X},$$

$$\mathbf{K}_p^{\mathbf{f}\mathbf{f}} = \mathbf{N}^T \mathbf{W} \mathbf{N} + \lambda \mathbf{K},$$

$$\mathbf{K}_{\beta\mathbf{f}} = \mathbf{X}^T \mathbf{W} \mathbf{N}$$

y

$$\mathbf{K}_p^{\phi\phi} = \sum_{i=1}^n \left\{ \sum_{j=0}^{\infty} \frac{P_r(Y_i \geq j)}{(\phi + j)^2} - \frac{\mu_i}{\phi(\mu_i + \phi)} \right\}.$$

$$\mathbf{K}_{\beta\phi} = 0,$$

$$\mathbf{K}_{\mathbf{f}\phi} = 0.$$

1.6. Proceso Iterativo

Considerando λ y ϕ fijos, utilizamos el algoritmo backfitting para encontrar los estimadores $\hat{\boldsymbol{\beta}}$ y $\hat{\mathbf{f}}$. El algoritmo backfitting es un método iterativo utilizado con frecuencia en el contexto semiparamétrico para estimar los parámetros del modelo. Las ecuaciones del algoritmo se obtienen a partir del algoritmo *scoring* de Fisher.

Es posible mostrar que la etapa $(u+1)$ -ésima del proceso iterativo *scoring* de Fisher para maximizar la función de verosimilitud penalizada con respecto a $\boldsymbol{\beta}$ y \mathbf{f} para ϕ fijo es igual a

$$\begin{pmatrix} \mathbf{K}_p^{\beta\beta} & \mathbf{K}_p^{\beta\mathbf{f}} \\ \mathbf{K}_p^{\mathbf{f}\beta} & \mathbf{K}_p^{\mathbf{f}\mathbf{f}} \end{pmatrix}^{(u)} \begin{pmatrix} \boldsymbol{\beta}^{(u+1)} & - \boldsymbol{\beta}^{(u)} \\ \mathbf{f}^{(u+1)} & - \mathbf{f}^{(u)} \end{pmatrix} = \begin{pmatrix} \mathbf{U}_p^{\beta} \\ \mathbf{U}_p^{\mathbf{f}} \end{pmatrix}^{(u)}$$

o, equivalentemente

$$\begin{pmatrix} \boldsymbol{\beta}^{(u+1)} \\ \mathbf{f}^{(u+1)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}^{(u)} \\ \mathbf{f}^{(u)} \end{pmatrix} + \begin{pmatrix} \mathbf{X}^T \mathbf{W} \mathbf{X} & \mathbf{X}^T \mathbf{W} \mathbf{N} \\ \mathbf{N}^T \mathbf{W} \mathbf{X} & \mathbf{N}^T \mathbf{W} \mathbf{N} + \lambda \mathbf{K} \end{pmatrix}^{(u)} \begin{pmatrix} \mathbf{X}^T \mathbf{W} \mathbf{z} \\ \mathbf{N}^T \mathbf{W} \mathbf{z} - \lambda \mathbf{K} \mathbf{f} \end{pmatrix}^{(u)}$$

En consecuencia las ecuaciones backfitting están dadas por

$$\mathbf{f}^{(u+1)} = (\mathbf{N}^T \mathbf{W}^{(u)} \mathbf{N} + \lambda \mathbf{K})^{-1} \mathbf{N}^T \mathbf{W}^{(u)} (\mathbf{z}^{(u)} - \mathbf{X} \boldsymbol{\beta}^{(u)}) \quad (1.6)$$

y

$$\boldsymbol{\beta}^{(u+1)} = (\mathbf{X}^T \mathbf{W}^{(u)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(u)} (\mathbf{z}^{(u)} - \mathbf{N} \mathbf{f}^{(u)}), \quad (1.7)$$

donde $\mathbf{z} = \mathbf{F}^{-1}(\mathbf{y} - \boldsymbol{\mu}) + \mathbf{X} \boldsymbol{\beta} + \mathbf{N} \mathbf{f}$. Se puede observar que, si $\boldsymbol{\beta}$ es conocido y constante en $(\mathbf{z} - \mathbf{X} \boldsymbol{\beta})$, (1.6) nos da cuenta sobre el ajuste del spline cúbico vía máxima verosimilitud penalizada de las diferencias $(\mathbf{z} - \mathbf{X} \boldsymbol{\beta})$. Al igual sucede si f es conocido, (1.7) es una estimación de $\boldsymbol{\beta}$ mediante máxima verosimilitud penalizada de las diferencias $(\mathbf{z} - \mathbf{N} \mathbf{f})$. Esta interpretación puede realizarse, en el sentido que se puede hacer repetidamente una alternancia entre (1.6) y (1.7) para así obtener $\hat{\boldsymbol{\beta}}$ y $\hat{\mathbf{f}}$ vía máxima verosimilitud penalizada hasta lograr la convergencia deseada. Ese algoritmo es conocido como backfitting, y fue introducido por Breiman y Friedman (1985). Como los parámetros $(\boldsymbol{\beta}, \mathbf{f})$ son ortogonales con respecto a ϕ , podemos desarrollar un proceso iterativo para ϕ dadas las estimaciones de $\boldsymbol{\beta}$ y \mathbf{f} , como sigue:

$$\phi^{(s+1)} = \phi^{(s)} + \{U_p^\phi / L_p^{\phi\phi}\}^{(s)}, \quad (1.8)$$

para $s = 0, 1, 2, \dots$, donde U_p^ϕ y $L_p^{\phi\phi}$ denotan, respectivamente, la función score y la información de Fisher para ϕ . El proceso iterativo (1.10) y (1.11) se debe alternar con el proceso iterativo (1.12).

Green y Silverman (1994) probaron que, bajo determinadas condiciones sobre las matrices \mathbf{N} , \mathbf{W} y \mathbf{K} , el algoritmo backfitting siempre converge y puede terminar en poco

más de 5 iteraciones.

1.7. Grados de libertad efectivos

Un tópico importante en los modelos semiparamétricos es estimar el número de grados de libertad efectivos, esto es, cuantos grados de libertad están siendo efectivamente utilizados para estimar los parámetros de componente no paramétrico. En el caso de los modelos semiparamétricos con respuesta binomial negativa el número total de parámetros para ser estimados es dado por $q = 1 + p + df(\lambda)$, donde $p + 1$ se refiere a los parámetros β y ϕ , mientras que $df(\lambda)$ a los parámetros de la componente no paramétrica. En Green y Silverman (1994), existe una expresión para $df(\lambda)$ para los MLGs que puede ser extendida para los modelos semiparamétricos con respuesta binomial negativa en virtud de la semejanza entre las ecuaciones de estimación. Esa expresión puede ser escrita de la siguiente forma:

$$df(\lambda) = tr\{\mathbf{H}(\lambda)\} - tr\{\mathbf{H}_1(\lambda)\},$$

en que

$$\mathbf{H}(\lambda) = \mathbf{N}(\mathbf{N}^T \mathbf{W} \mathbf{N} + \lambda \mathbf{K})^{-1} \mathbf{N}^T \mathbf{W}$$

y

$$\mathbf{H}_1(\lambda) = [\mathbf{X}^T \mathbf{W} \{\mathbf{I}_n - \mathbf{H}(\lambda)\} \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W} \{\mathbf{I}_n - \mathbf{H}(\lambda)\}^2 \mathbf{X},$$

siendo \mathbf{I}_n la matriz de identidad de dimensión $(n \times n)$. Por lo tanto, la expresión para $df(\lambda)$ asume la forma

$$df(\lambda) = tr\{\mathbf{H}(\lambda)\} - tr\{\mathbf{H}(\lambda) \tilde{\mathbf{H}}(\lambda)\},$$

con

$$\tilde{\mathbf{H}}(\lambda) = \{\mathbf{I}_n - \mathbf{H}(\lambda)\}^{-1} \tilde{\mathbf{X}} (\tilde{\mathbf{X}}^T \mathbf{W} \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \mathbf{W} \{\mathbf{I}_n - \mathbf{H}(\lambda)\},$$

y

$$\tilde{\mathbf{X}} = \{\mathbf{I}_n - \mathbf{H}(\lambda)\} \mathbf{X}.$$

Capítulo 2

Modelos aditivos semiparamétricos con respuesta binomial negativa

En este capítulo estudiamos los modelos aditivos generalizados semiparamétricos (MAGS's). Esta clase de modelos permite modelar fenómenos en los cuales la variable de respuesta pertenece a la familia exponencial y su componente sistemática depende de una parte paramétrica, $\mathbf{X}\boldsymbol{\beta}$, y de una parte no-paramétrica (aditiva) que incorpora el efecto no lineal de varias covariables. Análogamente a los MLGS's, el proceso de estimación de los parámetros y el análisis inferencial está basado en la función de log-verosimilitud penalizada. En la primera sección se presenta una descripción del modelo aditivo generalizado semiparamétrico. En la segunda sección, se define la función de log-verosimilitud y la función de penalización basada en el método de suavizamiento spline. En la tercera y cuarta sección, se obtienen la función score y la matriz de segundas derivadas (Hessiana) asociadas al vector de parámetros del MAGS. Finalmente, en la quinta y sexta sección, se determinan la matriz de información de Fisher y los errores estándar de los estimadores, respectivamente.

2.1. Especificación del modelo

Suponga que Y_1, \dots, Y_n son variables aleatorias independientes tales que $Y_i \sim BN(\mu_i, \phi)$ cuya función de probabilidad de Y_i está dada por :

$$f(y_i; \mu_i, \phi) = \frac{\Gamma(\phi + y_i)}{\Gamma(y_i + 1)\Gamma(\phi)} \left(\frac{\mu_i}{\mu_i + \phi} \right)^{y_i} \left(\frac{\phi}{\mu_i + \phi} \right)^\phi, y_i = 1, 2, \dots \quad (2.1)$$

donde $E(Y_i) = \mu_i$ y $Var(Y_i) = \mu_i + \mu_i^2/\phi$.

Análogamente a lo observado en el Capítulo 1, asumiremos que la parte sistemática del modelo toma la forma

$$\begin{aligned} g(\mu_i) &= \eta_i \\ &= \mathbf{x}_i^T \boldsymbol{\beta} + f_1(t_{1i}) + \dots + f_s(t_{si}), \end{aligned} \quad (2.2)$$

donde $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ representa los valores de las variables explicativas, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$, $p < n$, es un vector de parámetros desconocidos, $g(\cdot)$ es una función monótona diferenciable, denominada función de enlace, y $f_k(\cdot)$ es una función suave que depende de la variable explicativa $t_k (k = 1, \dots, s)$, y que debe ser estimada. Alternativamente, la componente sistemática (2.2) se puede escribir de la forma

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{n}_{1i}^T \mathbf{f}_1 + \dots + \mathbf{n}_{si}^T \mathbf{f}_s \quad (2.3)$$

ó, de manera matricial, como:

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}_1\mathbf{f}_1 + \dots + \mathbf{N}_s\mathbf{f}_s, \quad (2.4)$$

donde $g(\boldsymbol{\mu})$ es un vector $(n \times 1)$ tal que $g(\boldsymbol{\mu}) = (g(\mu_1), \dots, g(\mu_n))^T$, \mathbf{X} es una matriz de diseño $(n \times p)$ con filas \mathbf{x}_i^T , \mathbf{N}_k es una matriz de incidencia $(n \times r_k)$ con filas \mathbf{n}_{ki}^T cuyo (i, j) -ésimo elemento es igual a la indicatriz $(t_{ki} = t_{kj}^0) = 1$, $(j = 1, \dots, q)$ con $t_{kj}^0 (j = 1, \dots, r_k)$ denotando los valores distintos y ordenados de la variable explicativa t_k , $\mathbf{f}_k = (f_k(t_{k1}^0), \dots, f_k(t_{kr_k}^0))^T$ es un vector de parámetros $(r_k \times 1)$.

Note que si todos los valores de la variable t_k son distintos, la matriz $\mathbf{N}_k = \mathbf{I}_n$. En lo que sigue, denotaremos los elementos de la matriz \mathbf{N}_k de la forma n_{kil} ($i = 1, \dots, n; l = 1, \dots, q$).

Cuadro 2.1: Cantidades w_i y f_i^* para algunas transformaciones utilizadas en modelos con respuesta binomial negativa.

transformación	w_i	f_i^*
$\log \mu_i = \eta_i$	$\mu_i(\mu_i\phi_{-1} + 1)^{-1}$	μ_i
$\sqrt{\mu_i} = \eta_i$	$4(\mu_i\phi_{-1} + 1)^{-1}$	$2\sqrt{\mu_i}$
$\mu_i = \eta_i$	$(\mu_i^2\phi_{-1} + \mu_i)^{-1}$	1

2.2. Función de penalización

Asumiendo que las variables aleatorias Y_1, \dots, Y_n son independientes y que cada una de ellas tiene función de densidad de probabilidad dada por la ecuación (2.1), el logaritmo de la función de verosimilitud para $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \mathbf{f}_1^T, \dots, \mathbf{f}_s^T, \phi)^T$ está dado por

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \phi^{-1} \{y_i \theta_i - b(\theta_i)\} + \sum_{i=1}^n c(y_i, \phi).$$

Análogamente al caso de los MLGS's, la maximización directa de la función $\ell(\boldsymbol{\theta})$ sin imponer una restricción sobre la función f_k , conduce a problemas de súper ajuste e identificabilidad del vector de parámetros $\boldsymbol{\beta}$. Siguiendo la idea de Green y Silverman (1994), asumiremos que las funciones f_k 's pertenecen al espacio de Sobolev y penalizamos el logaritmo de la función de verosimilitud $\ell(\boldsymbol{\theta})$ de la siguiente manera:

$$\begin{aligned} \ell_p(\boldsymbol{\theta}; \boldsymbol{\lambda}) &= \ell(\boldsymbol{\theta}, \phi) - \frac{1}{2} \left\{ \int \lambda_1 f_1''(t_1)^2 dt + \dots + \int \lambda_s f_s''(t_s)^2 dt \right\} \\ &= \ell(\boldsymbol{\theta}, \phi) - \frac{1}{2} \sum_{k=1}^s \int \lambda_k f_k''(t_k)^2 dt \\ &= \ell(\boldsymbol{\theta}, \phi) - \frac{1}{2} \sum_{k=1}^s \lambda_k \mathbf{f}_k^T \mathbf{K}_k \mathbf{f}_k, \end{aligned}$$

donde $\mathbf{f}_k = (f_{k1}(t_{k1}^0), \dots, f_{kq}(t_{kq}^0))^T$ es un vector $(q \times 1)$, \mathbf{K}_k es una matriz de suavizamiento definida no-negativa de dimensión $(q \times q)$, y $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_s)^T$ es un vector $(s \times 1)$.

El parámetro de suavizamiento, λ_k , al igual que en el caso de los MLGS's, tiene un rol fundamental en el proceso de modelación. Bajo el contexto de los MAGS's, este parámetro es seleccionado mediante los criterios de validación cruzada (VC), validación cruzada generalizada (VCG) y Akaike (AIC), entre otros.

2.3. Función score penalizada

En esta sección obtenemos la función score penalizada asociada al vector de parámetros $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \mathbf{f}_1^T, \dots, \mathbf{f}_s^T, \phi)^T$, asumiendo que la función de log-verosimilitud penalizada $\ell_p(\boldsymbol{\theta}; \boldsymbol{\lambda})$ es regular con respecto a los elementos de $\boldsymbol{\theta}$.

2.3.1. Componente paramétrica

Para obtener la función score para el parámetro $\boldsymbol{\beta}$ debemos calcular

$$\begin{aligned} \frac{\partial \ell_p(\boldsymbol{\theta}; \boldsymbol{\lambda})}{\partial \beta_j} &= \sum_{i=1}^n \left\{ \frac{y_i}{\mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} - \frac{(\phi + y_i)}{(\phi + \mu_i)} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} \right\} \\ &= \sum_{i=1}^n \left\{ \frac{y_i}{\mu_i} \frac{\partial \mu_i}{\partial \eta_i} x_{ij} - \frac{(\phi + y_i)}{(\phi + \mu_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij} \right\} \\ &= \sum_{i=1}^n w_i f_i^{*-1} (y_i - \mu_i) x_{ij} . \end{aligned}$$

donde $w_i = (\partial \mu_i / \partial \eta_i)^2 / V_i$. Luego, la función score para $\boldsymbol{\beta}$ se puede escribir en forma matricial de la siguiente manera:

$$\begin{aligned} U_p^\beta(\boldsymbol{\theta}) &= \frac{\partial \ell_p(\boldsymbol{\theta}; \boldsymbol{\lambda})}{\partial \boldsymbol{\beta}} \\ &= \phi^{-1} \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{V}^{-1/2} (\mathbf{Y} - \boldsymbol{\mu}), \end{aligned}$$

donde \mathbf{X} es una matriz $(n \times p)$ de rango completo cuyas filas son denotadas por \mathbf{x}_i^T , $i = 1, \dots, n$, $\mathbf{W} = \text{diag}\{w_1, \dots, w_n\}$ es una matriz de pesos, $\mathbf{V} = \text{diag}\{V_1, \dots, V_n\}$, $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ y $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$.

2.3.2. Componente no-paramétrica

Para obtener la función escora asociada al vector del parámetro $\mathbf{f}_k = (\psi_{k1}, \dots, \psi_{kq})^T$, con $\psi_{kl} = f_k(t_{kl}^0)$ para $l = 1, \dots, q$, debemos calcular

$$\begin{aligned} \frac{\partial \ell_p(\boldsymbol{\theta}; \boldsymbol{\lambda})}{\partial f_j} &= \sum_{i=1}^n \left\{ \frac{y_i}{\mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial f_j} - \frac{(\phi + y_i)}{(\phi + \mu_i)} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial f_j} \right\} - [\lambda K f]_j \\ &= \sum_{i=1}^n \left\{ \frac{y_i}{\mu_i} \frac{\partial \mu_i}{\partial \eta_i} n_{ij} - \frac{(\phi + y_i)}{(\phi + \mu_i)} \frac{\partial \mu_i}{\partial \eta_i} n_{ij} \right\} - [\lambda K f]_j \\ &= \sum_{i=1}^n w_i f_i^{*-1} (y_i - \mu_i) n_{ij} - [\lambda K f]_j. \end{aligned}$$

Luego, la función escora para \mathbf{f}_k en términos matriciales está dada por

$$\begin{aligned} U_p^f(\boldsymbol{\theta}) &= \frac{\partial \ell_p(\boldsymbol{\theta}; \boldsymbol{\lambda})}{\partial \mathbf{f}_k} \\ &= \phi^{-1} \mathbf{N}_k^T \mathbf{W}^{1/2} \mathbf{V}^{-1/2} (\mathbf{Y} - \boldsymbol{\mu}) - \lambda_k \mathbf{K}_k \mathbf{f}_k \quad (k = 1, \dots, s), \end{aligned}$$

donde \mathbf{N}_k es una matriz de incidencia ($n \times q$) cuyo (i, j) -ésimo elemento es igual a la función indicatriz $I(t_{ki} = t_{kj}^0) = 1$, $j = 1, \dots, q$. Note que las filas de esta matriz son denotadas por \mathbf{n}_{ki}^T y sus elementos por n_{kil} , $i = 1, \dots, n$ y $l = 1, \dots, q$.

2.3.3. Componente dispersión

Finalmente, la función escora para ϕ está dada por

$$\frac{\partial \ell_p(\boldsymbol{\theta}; \boldsymbol{\lambda})}{\partial \phi} = \sum_{i=1}^n -\phi^{-2} \{y_i \theta_i - b(\theta_i)\} + \sum_{i=1}^n c'(y_i, \phi),$$

donde $c'(y_i, \phi) = \partial c(y_i, \phi) / \partial \phi$.

2.4. Matriz hessiana penalizada

Usando resultados de derivadas de matrices tenemos que los elementos de la matriz de segundas derivadas con respecto a $\boldsymbol{\theta}$ se pueden calcular como sigue.

2.4.1. $\mathbf{L}_p^{\beta_j \beta_l}$

En este caso,

$$\begin{aligned}
 \mathbf{L}_p^{\beta_j \beta_l} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \beta_j \partial \beta_l} \\
 &= \frac{\partial}{\partial \beta_l} \sum_{i=1}^n \frac{f_i^*}{(\mu_i^2 \phi^{-1} + \mu_i)} (y_i - \mu_i) x_{ij} \\
 &= \sum_{i=1}^n f_i^* x_{ij} \frac{\partial}{\partial \beta_l} \frac{(y_i - \mu_i)}{(\mu_i^2 \phi^{-1} + \mu_i)} \\
 &= - \sum_{i=1}^n \frac{f_i^{*2} x_{ij} x_{il}}{(\mu_i^2 \phi^{-1} + \mu_i)} - \sum_{i=1}^n \frac{f_i^{*2} x_{ij} x_{il} (y_i - \mu_i) (2\phi^{-1} \mu_i + 1)}{(\mu_i^2 \phi^{-1} + \mu_i)^2}.
 \end{aligned}$$

Matricialmente, esta matriz se puede escribir de la siguiente forma:

$$\mathbf{L}_p^{\beta\beta} = \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = -\mathbf{X}^T \mathbf{W} \mathbf{X} - \mathbf{X}^T \mathbf{A}(\mathbf{a}) \mathbf{X} (\mathbf{y} - \boldsymbol{\mu}),$$

donde $\mathbf{A}(\mathbf{a}) = \text{diag}\{a_1, \dots, a_n\}$ y $a_i = f_i^{*2} \frac{(2\phi^{-1} \mu_i + 1)}{(\mu_i^2 \phi^{-1} + \mu_i)^2}$, para $i = 1, \dots, n$.

2.4.2. $\mathbf{L}_p^{f_j f_l}$

Aquí,

$$\begin{aligned}
 \mathbf{L}_p^{f_j f_l} &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial f_j \partial f_l} \\
 &= \frac{\partial}{\partial f_l} \sum_{i=1}^n \frac{f_i^*}{(\mu_i^2 \phi^{-1} + \mu_i)} (y_i - \mu_i) n_{ij} - \frac{\partial}{\partial f_l} [\lambda \mathbf{K} \mathbf{f}]_j \\
 &= \sum_{i=1}^n f_i^* n_{ij} \frac{\partial}{\partial f_l} \frac{(y_i - \mu_i)}{(\mu_i^2 \phi^{-1} + \mu_i)} - \lambda k_{jl} \\
 &= - \sum_{i=1}^n \frac{f_i^{*2} n_{ij} n_{il}}{(\mu_i^2 \phi^{-1} + \mu_i)} - \sum_{i=1}^n \frac{f_i^{*2} n_{ij} n_{il} (y_i - \mu_i) (2\phi^{-1} \mu_i + 1)}{(\mu_i^2 \phi^{-1} + \mu_i)^2} - \lambda k_{ij}.
 \end{aligned}$$

Matricialmente, esta matriz queda escrita como

$$\mathbf{L}_p^{\mathbf{f}\mathbf{f}} = -\mathbf{N}^T \mathbf{W} \mathbf{N} - \mathbf{N}^T \mathbf{A}(\mathbf{a}) \mathbf{N} (\mathbf{y} - \boldsymbol{\mu}) - \lambda \mathbf{K}.$$

2.4.3. $\mathbf{L}_p^{\phi\phi}$

En relación al parámetro ϕ , la matriz hessiana viene dada por

$$\begin{aligned}\mathbf{L}_p^{\phi\phi} &= \frac{\partial^2 \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \phi \partial \phi} \\ &= \sum_{i=1}^n \left\{ \psi'(\phi + y_i) + \frac{(y_i - 2\mu_i - \phi)}{(\phi + \mu_i)^2} \right\} + \frac{n}{\phi} [1 - \phi \psi'(\phi)].\end{aligned}$$

2.4.4. $\mathbf{L}_p^{\beta_j \mathbf{f}_l}$

En este caso,

$$\begin{aligned}\mathbf{L}_p^{\beta_j \mathbf{f}_l} &= \frac{\partial^2 \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \beta_j \partial \mathbf{f}_l} \\ &= \sum_{i=1}^n f_i^* x_{ij} \frac{\partial}{\partial f_l} \frac{(y_i - \mu_i)}{(\mu_i^2 \phi^{-1} + \mu_i)} \\ \mathbf{L}_p^{\beta_j \mathbf{f}_l} &= - \sum_{i=1}^n \frac{f_i^{*2} x_{ij} n_{il}}{(\mu_i^2 \phi^{-1} + \mu_i)} - \sum_{i=1}^n \frac{f_i^{*2} x_{ij} n_{il} (y_i - \mu_i) (2\phi^{-1} \mu_i + 1)}{(\mu_i^2 \phi^{-1} + \mu_i)^2}\end{aligned}$$

$$\mathbf{L}_p^{\beta \mathbf{f}} = \frac{\partial^2 \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta} \partial \mathbf{f}^T} = -\mathbf{X}^T \mathbf{W} \mathbf{N} - \mathbf{X}^T \mathbf{A}(\mathbf{a}) \mathbf{N} (\mathbf{y} - \boldsymbol{\mu}).$$

2.4.5. $\mathbf{L}_p^{\beta\phi}$

Aquí,

$$\begin{aligned}\mathbf{L}_p^{\beta\phi} &= \frac{\partial^2 \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta} \partial \phi} \\ &= \sum_{i=1}^n \frac{f_i^* \phi^{-2}}{(\mu_i \phi^{-1} + 1)^2} (y_i - \mu_i) x_{ij}\end{aligned}$$

Matricialmente, esta matriz queda escrita como

$$\mathbf{L}_p^{\beta\phi} = \frac{\partial^2 \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta} \partial \phi} = \mathbf{X}^T \mathbf{A}(\mathbf{b})(\mathbf{y} - \boldsymbol{\mu})$$

donde $\mathbf{A}(\mathbf{b}) = \text{diag}\{b_1, \dots, b_n\}$ con $b_i = \frac{f_i^* \phi^{-2}}{(\mu_i \phi^{-1} + 1)^2}$, para $i = 1, \dots, n$.

2.4.6. $\mathbf{L}_p^{\mathbf{f}_j\phi}$

Finalmente,

$$\begin{aligned} \mathbf{L}_p^{\mathbf{f}_j\phi} &= \frac{\partial^2 \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \mathbf{f}_j \partial \phi} \\ &= \sum_{i=1}^n \frac{f_i^* \phi^{-2}}{(\mu_i \phi^{-1} + 1)^2} (y_i - \mu_i) \eta_{ij}, \end{aligned}$$

En forma matricial, se puede escribir como,

$$\mathbf{L}_p^{\mathbf{f}\phi} = \frac{\partial^2 \mathbf{L}_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \mathbf{f} \partial \phi} = \mathbf{N}^T \mathbf{A}(\mathbf{b})(\mathbf{y} - \boldsymbol{\mu}).$$

2.5. Matriz de información de Fisher penalizada

Los cálculos para la obtención de la matriz observada de Fisher (matriz Hessiana), para la cual basta con tomar la segunda derivada del vector escore en relación a $\boldsymbol{\theta}$. Se calcula nuevamente la derivada de Información de Fisher en relación al vector $\boldsymbol{\beta}$ y se tiene la segunda derivada del logaritmo de función de verosimilitud penalizada. Por lo tanto, las submatrices diagonal en bloque de la matriz de Información de Fisher son dadas por:

$$\begin{aligned} \mathbf{K}_p^{\beta\beta} &= -\mathbf{E}(L_p^{\beta\beta}) \\ &= -\mathbf{E}(-\mathbf{X}^T \mathbf{W} \mathbf{X} - \mathbf{X}^T \mathbf{A}(\mathbf{a}) \mathbf{X} (y - \mu)) \\ &= \mathbf{X}^T \mathbf{W} \mathbf{X} \end{aligned}$$

$$\begin{aligned}
 \mathbf{K}_p^{\mathbf{f}_k} &= -\mathbf{E}(L_p^{\mathbf{f}_k}) \\
 &= -\mathbf{E}(-\mathbf{N}_k^T \mathbf{W} \mathbf{N}_k - \mathbf{N}_k^T \mathbf{A}(\mathbf{a}) \mathbf{N}_k (y - \mu) - \lambda_k \mathbf{K}_k) \\
 &= \mathbf{N}_k^T \mathbf{W} \mathbf{N}_k - \lambda_k \mathbf{K}_k, \quad k = 1, \dots, s.
 \end{aligned}$$

La matriz de Información de Fisher para ϕ , (vea Lawless,1987), dada por

$$\mathbf{K}_p^{\phi\phi} = \sum_{i=1}^n \left\{ \sum_{j=0}^{\infty} \frac{P_r(Y_i \geq j)}{(\phi + j)^2} - \frac{\mu_i}{\phi(\mu_i + \phi)} \right\}$$

Note que

$$\begin{aligned}
 \mathbf{K}_p^{\beta\mathbf{f}} &= \mathbf{X}^T \mathbf{W} \mathbf{X}, \\
 \mathbf{K}_p^{\beta\phi} &= 0, \\
 \mathbf{K}_p^{\mathbf{f}\phi} &= 0.
 \end{aligned}$$

Los vectores de β y \mathbf{f} son ortogonales a ϕ .

2.6. Estimación de los parámetros

El problema de estimación de los parámetros en el contexto de los MAGS's ha sido estudiado por diferentes autores. Por ejemplo, Hastie y Tibshirani (1986) propusieron estimar los parámetros del modelo a través del algoritmo de score local usando suavizadores lineales, y observaron que otros suavizadores, como los splines cúbicos, también pueden ser utilizados. Berhane y Tibshirani (1988) introdujeron los modelos aditivos generalizados para datos longitudinales y derivaron el algoritmo de score local a partir de la casi-log-verosimilitud penalizada, asumiendo que las funciones no-paramétricas pertenecen al espacio de funciones de Sobolev. En este contexto, propusieron algunas condiciones sobre la matriz de pesos para garantizar la consistencia del sistema de ecuaciones y la convergencia del algoritmo backfitting. Lin y Zhang (1999) propusieron los modelos mixtos aditivos generalizados (MMAG's), los cuales son una extensión de los modelos mixtos lineales generalizados, y estimaron las funciones no-paramétricas mediante suavizamiento spline cúbico.

Además, estimaron de manera conjunta los parámetros de suavizamiento y los componentes de varianza usando la casi-verosimilitud marginal. Finalmente Rigby y Stasinopoulos (2005) introdujeron los modelos aditivos generalizados para posición, escala y forma (del inglés, GAMLSS), y ajustaron el modelo usando el método de máxima verosimilitud penalizada. En este contexto, la función de verosimilitud penalizada es maximizada a través del algoritmo score de Fisher y la componente aditiva del modelo es ajustada a través del algoritmo backfitting.

2.7. Derivación del algoritmo backfitting ponderado

Sea $\boldsymbol{\theta} = (\mathbf{f}_0^T, \mathbf{f}_1^T, \dots, \mathbf{f}_s^T, \phi)^T$, con $\mathbf{f}_0^T = \boldsymbol{\beta}$ y $\mathbf{N}_0 = \mathbf{X}$. Considere $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_s)$, donde $\alpha_k = \lambda_k \phi$, y ϕ fijos. El proceso iterativo de score de Fisher que permite maximizar la función de log-verosimilitud penalizada $\ell_p(\boldsymbol{\theta}; \boldsymbol{\lambda})$, con respecto a \mathbf{f}_0 y $(\mathbf{f}_1, \dots, \mathbf{f}_s)$, está dada por la solución de la siguiente ecuación matricial:

$$\begin{pmatrix} \mathbf{N}_0^T \mathbf{W} \mathbf{N}_0 & \mathbf{N}_0^T \mathbf{W} \mathbf{N}_1 & \dots & \mathbf{N}_0^T \mathbf{W} \mathbf{N}_s \\ \mathbf{N}_1^T \mathbf{W} \mathbf{N}_0 & \mathbf{N}_1^T \mathbf{W} \mathbf{N}_1 + \alpha_1 \mathbf{K}_1 & \dots & \mathbf{N}_1^T \mathbf{W} \mathbf{N}_s \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{N}_s^T \mathbf{W} \mathbf{N}_0 & \mathbf{N}_s^T \mathbf{W} \mathbf{N}_1 & \dots & \mathbf{N}_s^T \mathbf{W} \mathbf{N}_s + \alpha_s \mathbf{K}_s \end{pmatrix} \begin{pmatrix} \mathbf{f}_0^{(r)} \\ \mathbf{f}_1^{(r)} \\ \vdots \\ \mathbf{f}_s^{(r)} \end{pmatrix} = \begin{pmatrix} \mathbf{N}_0^T \mathbf{W} \mathbf{z} \\ \mathbf{N}_1^T \mathbf{W} \mathbf{z} \\ \vdots \\ \mathbf{N}_s^T \mathbf{W} \mathbf{z} \end{pmatrix}, \quad (2.5)$$

donde el vector de respuestas \mathbf{z} contiene las variables modificadas cuyo i -ésimo elemento está dado por

$$z_i = (Y_i - \mu_i)g'(\mu_i) + \left(\sum_{k=0}^s \mathbf{N}_k \mathbf{f}_k \right)_i,$$

$(\mathbf{f}_0^{(r)}, \mathbf{f}_1^{(r)}, \dots, \mathbf{f}_s^{(r)})$ corresponden a los valores actualizados de los parámetros estimados, y la matriz diagonal \mathbf{W} está definida en la Subsección 1.4.1. Después de algunas manipulaciones algebraicas, tenemos que las ecuaciones de score de Fisher para actualizar de manera

simultánea $(\mathbf{f}_0^{(r)}, \mathbf{f}_1^{(r)}, \dots, \mathbf{f}_s^{(r)})$ están dadas por la siguiente ecuación matricial:

$$\begin{pmatrix} \mathbf{I}_p & \mathbf{S}_0\mathbf{N}_1 & \dots & \mathbf{S}_0\mathbf{N}_s \\ \mathbf{S}_1\mathbf{N}_0 & \mathbf{I}_q & \dots & \mathbf{S}_1\mathbf{N}_s \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_s\mathbf{N}_0 & \mathbf{S}_s\mathbf{N}_1 & \dots & \mathbf{I}_q \end{pmatrix} \begin{pmatrix} \mathbf{f}_0^{(r)} \\ \mathbf{f}_1^{(r)} \\ \vdots \\ \mathbf{f}_s^{(r)} \end{pmatrix} = \begin{pmatrix} \mathbf{S}_0\mathbf{z} \\ \mathbf{S}_1\mathbf{z} \\ \vdots \\ \mathbf{S}_s\mathbf{z} \end{pmatrix}, \quad (2.6)$$

donde

$$\begin{aligned} \mathbf{S}_0 &= (\mathbf{N}_0^T \mathbf{W} \mathbf{N}_0)^{-1} \mathbf{N}_0^T \mathbf{W}, \\ \mathbf{S}_k &= (\mathbf{N}_k^T \mathbf{W} \mathbf{N}_k + \alpha_k \mathbf{K}_k)^{-1} \mathbf{N}_k^T \mathbf{W} \quad (k = 1, \dots, s), \end{aligned}$$

corresponden a las matrices de suavizamiento. Hastie y Tibshirani (1990) probaron que la solución a la ecuación anterior es única y que el siguiente algoritmo backfitting converge a esta solución:

$$\mathbf{f}_k = \mathbf{S}_k \left(\mathbf{z} - \sum_{l=0, l \neq k}^s \mathbf{N}_l \mathbf{f}_l \right), \quad (k = 0, \dots, s). \quad (2.7)$$

La convergencia del algoritmo backfitting se obtiene iterando repetidamente entre las ecuaciones anteriores hasta alcanzar el criterio de convergencia establecido.

Análogamente a los MLGS's, si igualamos la función escora U_p^ϕ a cero obtenemos la siguiente solución:

$$\sum c'(y_i, \phi) = \frac{1}{2} D(\mathbf{Y}; \hat{\boldsymbol{\mu}}) - \sum \left\{ y_i \tilde{\theta}_i - b(\tilde{\theta}_i) \right\},$$

donde $D(\mathbf{Y}; \hat{\boldsymbol{\mu}})$ denota la función de desvío del modelo aditivo generalizado semiparamétrico. En este caso podemos considerar el estimador consistente de ϕ que no requiere de un proceso iterativo para encontrarlo, y que está dado por

$$\hat{\phi} = \frac{n - p^*}{\sum_{i=1}^n (y_i - \hat{\mu}_i)^2 / V(\hat{\mu}_i)},$$

donde $p^* = p + \sum_{k=1}^s \text{tr}(\mathbf{N}_k \mathbf{S}_k)$ denota el número de parámetros efectivos considerados en el modelo. Note que la $\text{tr}(\mathbf{N}_k \mathbf{S}_k)$ representa el número de parámetros asociados a la componente aditiva del modelo.

2.8. Proceso iterativo conjunto

La solución al sistema de ecuaciones (2.7) conduce al estimador de máxima verosimilitud penalizado de $\boldsymbol{\theta}$, el cual se obtiene iterando entre un algoritmo backfitting ponderado, con matriz de peso \mathbf{W} , y una estimación de máxima verosimilitud del parámetro ϕ . Esto es equivalente a resolver el siguiente proceso iterativo:

(i) Inicio

(a) Ajuste un MAGS normal para obtener $\mathbf{f}_l^{(0)}$ ($l = 0, \dots, s$).

(b) Obtenga un valor inicial para ϕ usando los valores ajustados obtenidos en (a) y la ecuación

$$\phi = \frac{1}{n - p^*} \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{V(\mu_i)}, \quad p^* = p + \sum_{l=1}^s \text{tr}(\mathbf{N}_l \mathbf{S}_l).$$

(c) A partir de los valores actualizados $\boldsymbol{\theta}^{(0)} = (\mathbf{f}_0^{(0)T}, \mathbf{f}_1^{(0)T}, \dots, \mathbf{f}_s^{(0)T}, \phi^{(0)})^T$ obtenga la matriz de pesos $\mathbf{W}^{(0)}$ y las matrices de suavizamiento

$$\begin{aligned} \mathbf{S}_0^{(0)} &= (\mathbf{N}_0^T \mathbf{W}^{(0)} \mathbf{N}_0)^{-1} \mathbf{N}_0^T \mathbf{W}^{(0)}, \\ \mathbf{S}_k^{(0)} &= (\mathbf{N}_k^T \mathbf{W}^{(0)} \mathbf{N}_k + \alpha_k^{(0)} \mathbf{K}_k)^{-1} \mathbf{N}_k^T \mathbf{W}^{(0)} \quad (k = 1, \dots, s). \end{aligned}$$

(ii) Etapa 1: Itere repetidamente entre las siguientes ecuaciones:

$$\begin{aligned} \mathbf{f}_0^{(r+1)} &= \mathbf{S}_0^{(r)} \left(\mathbf{z} - \sum_{l=1}^s \mathbf{N}_l \mathbf{f}_l^{(r)} \right), \\ \mathbf{f}_1^{(r+1)} &= \mathbf{S}_1^{(r)} \left(\mathbf{z} - \mathbf{N}_0 \mathbf{f}_0^{(r+1)} - \sum_{l=2}^s \mathbf{N}_l \mathbf{f}_l^{(r)} \right), \\ &\vdots \\ \mathbf{f}_s^{(r+1)} &= \mathbf{S}_s^{(r)} \left(\mathbf{z} - \sum_{l=0}^{s-1} \mathbf{N}_l \mathbf{f}_l^{(r+1)} \right), \end{aligned}$$

para $r = 0, 1, \dots$. Repita (ii) reemplazando $\mathbf{f}_l^{(r)}$ por $\mathbf{f}_l^{(r+1)}$ por $\mathbf{f}_l^{(r+1)}$ ($l = 0, 1, \dots, s$) hasta que el criterio de convergencia

$$\Delta_r(\mathbf{f}_l^{(r+1)}, \mathbf{f}_l^{(r)}) = \frac{\sum_{l=0}^s \left\| \mathbf{f}_l^{(r+1)} - \mathbf{f}_l^{(r)} \right\|}{\sum_{l=0}^s \left\| \mathbf{f}_l^{(r)} \right\|},$$

sea menor que un valor de tolerancia pequeño.

- (iii) Etapa 2: Para los valores actualizados $\mathbf{f}_l^{(r+1)}$ ($l = 0, 1, \dots, s$), obtenga $\phi^{(r+1)}$ usando la ecuación

$$\phi^{(r+1)} = \frac{1}{n - p^*} \sum_{i=1}^n \frac{(y_i - \mu_i^{(r)})^2}{V(\mu_i^{(r)})}.$$

- (iv) Itere entre las etapas (ii) y (iii) reemplazando $\mathbf{f}_l^{(r)}$ ($l = 0, 1, \dots, s$) y $\phi^{(r)}$ por $\mathbf{f}_l^{(r+1)}$ y $\phi^{(r+1)}$, respectivamente, hasta alcanzar la convergencia.

Capítulo 3

Aplicación

3.1. Descripción de los datos

Los datos que utilizamos en esta aplicación fueron obtenidos del paquete Faraway (2006, Cap.12) y se refieren a un estudio de la relación entre la concentración de ozono en la atmósfera en Bacia de Los Ángeles durante el año 1976. Específicamente se consideran las siguientes variables:

O3: máxima diaria de concentración media de ozono(ppm).

vh: Altura de presión 500 milibares numérica continua.

wind: Velocidad del viento en mph en aeropuerto de LAX.

humidity: Unidades en porcentaje en LAX.

temp: Temperatura en la base aérea de Sandburg en grados Fahrenheit.

ibh: Altura de inversión de temperatura base en peso.

dpg: Gradiente de presión a partir de Los Angeles para Daggert en mmHg.

ibt: Inversión de temperatura base de LAX en grados Fahrenheit.

vis: Visibilidad de LAX en millas.

doy: Día del año.

La variable de respuesta O3 y las 9 variables explicativas, fueron observadas en el año 1976 durante 330 días. Para ilustrar la aplicabilidad de los modelos aditivos, se considerarán sólo 7 variables explicativas; humidity, temp, ibh, vis, doy, vh, ibt. Las Figuras 3.1 (a), (b), (c) y (d) muestran una tendencia lineal en los datos, lo que sugiere incorporar el efecto de las variables humidity, temp, ibh, y vis, de manera paramétrica en el modelo, mientras que las Figuras 3.1 (e) - (f) - (g), muestran una tendencia no-lineal en los datos que puede ser cuantificada a través de funciones no-paramétricas.

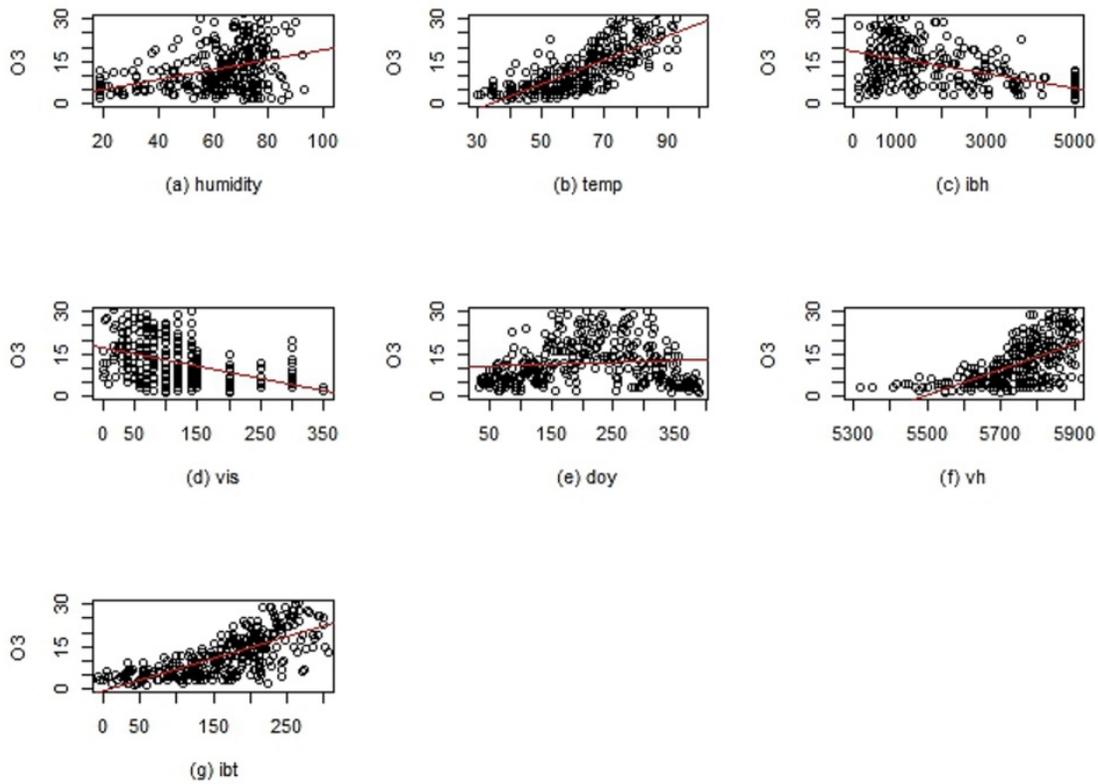


Figura 3.1: Diagrama de dispersión entre la variable respuesta O3 y las 7 variables explicativas meteorológicas observadas en Bacia de Los Ángeles en 1976.

Inicialmente, consideraremos un modelo log lineal de Poisson , esto es,

$$O3 \sim P(\mu_i),$$

con

$$\log(\mu_i) = \beta_0 + \beta_1 humidity_i + \beta_2 temp_i + \beta_3 ibh_i + \beta_4 vis_i + f_1(doy_i) + f_2(vh_i) + f_3(ibt_i),$$

para $i = 1, \dots, 330$, donde μ_i representa la media (ppm) de ozono y $\beta = (\beta_0; \beta_1; \beta_2; \beta_3; \beta_4)^T$ representa el vector de coeficientes de regresión, y $f_k(\cdot), (k = 1, 2, 3)$ son funciones suaves desconocidas. Usando la librería GAMLSS del software R-project se ajustó el modelo aditivo semiparamétrico de Poisson obteniéndose los siguientes resultados:

Cuadro 3.1: Estimaciones de máxima verosimilitud con sus respectivos errores estándar aproximados, con respecto al modelo Poisson.

Efecto	Estimaciones	Error Est.	Valor t
Intercepto	-4.960e+00	3.425e-01	< 2e-16
humidity	6.433e-03	1.193e-03	1.36e-07
temp	1.204e-02	2.791e-03	2.14e-05
ibh	-5.194e-05	1.047e-06	< 2e-16
vis	-8.844e-04	2.833e-04	0.001969
cs(doy)	-8.906e-04	2.104e-04	3.03e-05
cs(vh)	-1.073e-03	2.737e-06	< 2e-16
cs(ibt)	2.474e-03	6.948e-04	0.000426

En el cuadro 3.1, se observan las estimaciones, errores estándar y el valor del estadístico t asociado. Se puede apreciar que todos los coeficientes son significativos, y por lo tanto cada variable explicativa contribuye de manera significativa sobre el modelo. En este caso, el valor de AIC es de 1725.3.

Por otro lado, si consideramos el ajuste de un modelo con respuesta binomial negativa, es decir,

$$O3 \sim \text{BN}(\mu_i, \phi),$$

con

$$\log(\mu_i) = \beta_0 + \beta_1 \text{humidity}_i + \beta_2 \text{temp}_i + \beta_3 \text{ibh}_i + \beta_4 \text{vis}_i + f_1(\text{doy})_i + f_2(\text{vh})_i + f_3(\text{ibt})_i,$$

$i = 1, \dots, 330$.

Usando la librería GAMLSS del software R-project, se obtuvieron los siguientes resultados:

Cuadro 3.2: Estimaciones de máxima verosimilitud con sus respectivos errores estándar aproximados, con respecto al modelo aditivo semiparamétrico con respuesta binomial negativa.

Efecto	Estimaciones	Error Est.	Valor t
Intercepto	-4.652e+00	2.447e+00	0.05819
humidity	6.292e-03	1.235e-03	6.00e-07
temp	1.226e-02	3.097e-03	9.33e-05
ibh	-5.191e-05	2.072e-05	0.01275
vis	-9.096e-04	3.007e-04	0.00269
cs(doy)	-8.675e-04	2.288e-04	0.00018
cs(vh)	1.012e-03	4.460e-04	0.02400
cs(ibt)	2.595e-03	8.843e-04	0.00358

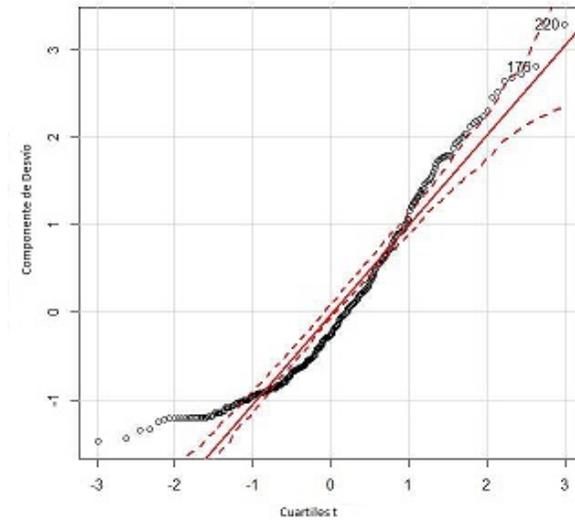


Figura 3.2: Gráfico normal de probabilidades con banda de confianza generada al 95 % para el componente de desvío y los datos de ozono ajustados al modelo Poisson.

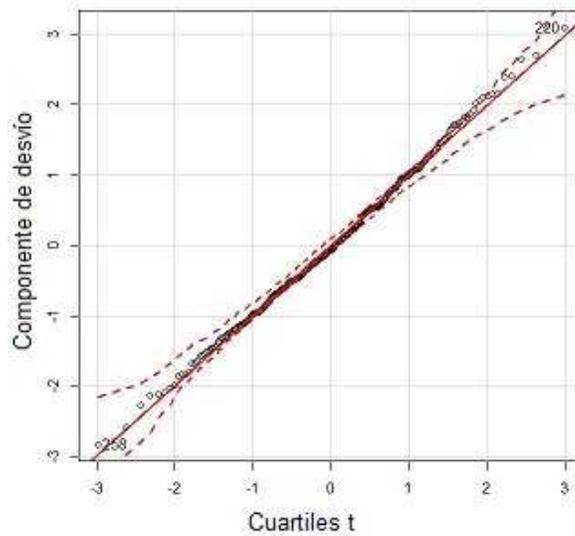


Figura 3.3: Gráfico normal de probabilidades con banda de confianza generada al 95 % para el componente de desvío y los datos ajustados de ozono al modelo aditivo semiparamétrico con respuesta binomial negativa.

Finalmente, la figura 3.4 muestra los gráficos de los residuos parciales para cada componente no paramétrica bajo el modelo binomial negativo.

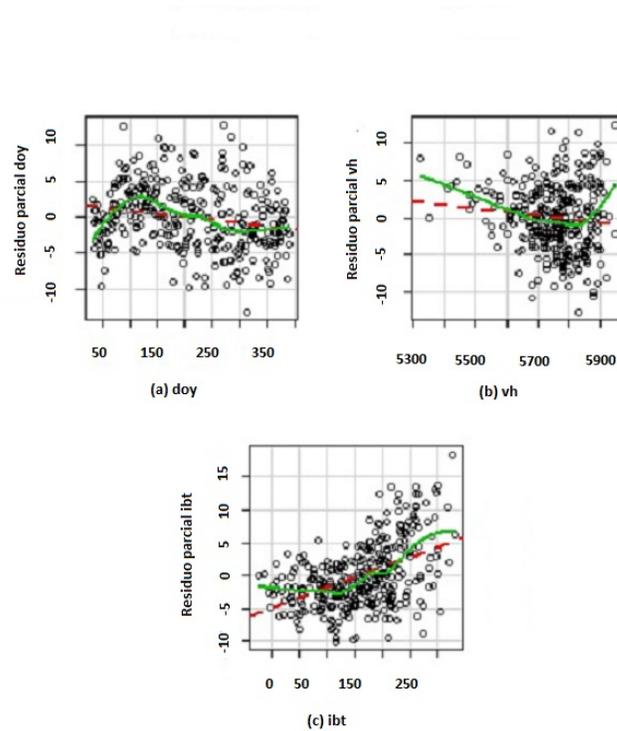


Figura 3.4: Gráfico de residuos parciales bajo el modelo aditivo semiparamétrico con respuesta binomial negativa.

Se observa que todos los coeficientes son significativos. Es decir, a medida que aumenta una unidad, la temperatura en la base aérea, se espera un aumento de la concentración de ozono. Por otro lado, a medida que aumenta la altura de inversión de temperatura, se

espera una reducción de la concentración de ozono.

El valor de AIC es de 1725, por lo cual no existe una gran diferencia con el modelo de Poisson, pero bajo el modelo binomial negativo se logra una menor dispersión de los datos, con un 17.33, mientras que bajo el modelo de Poisson, es de 18.42 la variabilidad de los datos, por lo cual se sugiere el modelo binomial negativo como el más adecuado. esto también es confirmado por el gráfico QQ-Plot de ambos modelos (ver figuras 3.2 y 3.3)

Conclusiones

En esta tesis estudiamos los modelos semiparamétricos con respuesta binomial negativa. Posteriormente, consideramos una extensión de tales modelos específicamente los modelos aditivos semiparamétricos con respuesta binomial negativa. En ambos modelos las funciones no-paramétricas son estimadas por splines cúbicos, usando la función de verosimilitud penalizada para obtener los estimadores de los parámetros. La inferencia sobre los parámetros y el cálculo de su varianza aproximada se basó en la matriz de información de Fisher penalizada.

Para ilustrar la aplicabilidad de estos modelos se consideró un conjunto de datos del área medio ambiental. Con fines comparativos, se ajustó el modelo aditivo semiparamétrico con respuesta Poisson y Binomial Negativa. Para ambos modelos el valor de AIC es el mismo, sin embargo, los gráficos QQ-Plots mostraron un ajuste más adecuado cuando se asume un modelo con respuesta Binomial Negativa. Los ajustes de ambos modelos fueron realizados usando la librería GAMLSS de R-Project.

Bibliografía

Eubank, R. L. (1984). The hat matrix for smoothing splines. *Statistics and Probability Letters*, **2**, 9-14.

Green, P. J. (1987). Penalized likelihood for general semi-parametric regression models. *International Statistical Review*, **55**, 245-259.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman and Hall, Boca Raton.

Yandell, B. S and Green, P. J. (1986). Semi-parametric generalized linear model diagnostics. *Technical Report, Dept. of Statistics, U. of Wisconsin*.

Hideto, F. (2015). Modelos semiparamétricos com resposta binomial negativa. *Instituto de Matemática e Estatística da Universidade de Sao Paulo*.

Hastie, T., Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, **1**, 297-318.

Ibacache-Pulgar, G., Paula, G. A. and J. A. Cysneiros (2013). Semiparametric additive models under symmetric distributions. *TEST*, **22**, 103-121.

Lin, X., Zhang, D. (1999). Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society. Series B*, **61**, 381-400.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall/CRC, Boca Raton.

Nelder, Wedderburn. (1972). Generalized Linear Models. *J. R. Statistical Society*, **135**, 370-384.

Paula, G. (2013). Modelos de regressao com apoio computacional. *Instituto de Matemática e Estatística*.

Quiroga P. (2014). Modelos lineales semiparamétricos. *Facultad de Ciencias-Instituto de Estadística*.

Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society B*, **50**, 413-436.

Stasinopoulos Mikis, P., Rigby Robert A. (2007). Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. *Journal of Statistical Software*. London Metropolitan University.