



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MATEMÁTICAS Y  
DE LA ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA.

UN MODELO DE REGRESIÓN SEMI-PARAMÉTRICO Y ALGUNAS  
APLICACIONES

TESIS  
QUE PARA OPTAR POR EL GRADO DE:  
MAESTRA EN CIENCIAS

PRESENTA:  
IRMA ROCÍO ZAVALA SIERRA

DIRECTOR DE LA TESIS  
DR. EDUARDO GUTIÉRREZ PEÑA  
INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS  
Y EN SISTEMAS

MÉXICO, D. F. FEBRERO 2015.



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



# Agradecimientos

El estudiar una maestría comenzó como una inquietud, una idea que con el paso del tiempo se volvió en un sueño. El hacerlo realidad se veía complicado, dada las circunstancias que vivía en esos momentos. Pero sola no estaba, estuve rodeada de amor y tuve el apoyo para salir adelante en cada prueba que se ponía en mi camino.

El principio tuvo sus complicaciones, sin embargo, Dios se encargó de tranquilizar mi mente. Ya dentro del programa Dios puso en mi corazón: *esfuérzate y sé valiente, yo estoy contigo*. Estas palabras me dieron la fuerza para seguir, confiar en mi misma y sobre todo confiar en Él.

Y así fue, durante el caminar hubieron momentos de estrés, pero afortunadamente los compartí con mis compañeros y amigos. Les agradezco a ellos su amistad y apoyo en cada semestre, que no sólo estuvo lleno de momentos de arrastrar el lápiz, también hubo tiempo para divertirnos.

En cuanto al aprendizaje obtenido durante esta etapa, sin duda alguna es invaluable. Al estudiar la maestría obtuve las herramientas necesarias para mi formación académica, profesional y personal. Se me abrió un mundo de oportunidades de conocimiento, y esto gracias a mis profesores, a ellos les agradezco el tiempo que brindaron para compartir su experiencia y conocimientos. Sobre todo le agradezco a la Dra. Ruth Fuentes quién fue mi tutora en el programa y al M. en C. José Pablo Castañeda quién fue un gran apoyo en este proyecto.

El momento de la elaboración del trabajo de tesis, el punto culmen del sueño en donde Dios seguía manifestándose. Durante la preparación de este trabajo aprendí nuevos temas y reforcé conocimientos. Todo esto con el gran apoyo del Dr. Eduardo Gutiérrez Peña a quien le agradezco el tiempo que se tomó para asesorarme y la paciencia que tuvo en cada momento. A quienes también agradezco su apoyo en este trabajo son a la Dra. Ruth Fuentes, Mtra. Patricia Romero, Dr. Alberto Contreras y Dra. Silvia Ruiz Velasco. Es importante mencionar que el trabajo presentado en esta tesis se realizó como parte del Proyecto IN106114 'Análisis Bayesiano Aplicado' del Programa UNAM-DGAPA-PAPIIT.

Y en cada una de las caídas y momentos de dicha estuvieron mis padres Arturo Zavala y Arcadia Sierra, mis hermanas Laura Zavala y Ángeles Zavala y mi sobrinito Erick López. Gracias familia por desvelarse conmigo, gracias por sufrir a mi lado, gracias por soportar mi mal humor un muchas ocasiones, gracias por estar al pendiente de mis trámites y mis deberes, gracias por velar por mi cada día, gracias por darme un techo y más que un techo un refugio. No tengo con que pagarles pues gracias a ustedes soy lo que soy, ustedes me dieron las herramientas para salir adelante.

Finalmente quiero dirigir unas palabras a mi Señor:

Dios, te entrego finalizado este proyecto. Fue tuyo desde el inicio, yo sólo fui un instrumento, Tú estuviste pendiente de él desde siempre. Gracias por las enseñanzas que recibí durante este tiempo y por no dejarme caer ni desertar. Pero sobre todo, gracias por el amor tan grande que me has tenido.

*“Yo te he mandado que seas fuerte y valeroso. No temas ni te acobardes, porque el Señor tu Dios estará contigo dondequiera que vayas”.*

Josué 1:9

# Índice general

<b>Agradecimientos</b>	<b>3</b>
<b>1. Introducción</b>	<b>7</b>
<b>2. Preliminares</b>	<b>9</b>
2.1. El modelo de regresión . . . . .	9
2.2. Análisis . . . . .	12
2.2.1. Caso $p(\beta)$ dependiente de $\sigma^2$ y ésta conocida . . . . .	12
2.2.2. Caso $p(\beta)$ no dependiente de $\sigma^2$ y ésta conocida . . . . .	14
2.2.3. Caso $p(\beta)$ dependiente de $\sigma^2$ y ésta desconocida . . . . .	15
2.3. El proceso gaussiano . . . . .	20
2.3.1. Simulación del proceso gaussiano . . . . .	21
2.3.2. Distribución condicional . . . . .	23
2.3.3. Otros aspectos . . . . .	24
<b>3. Un modelo de regresión semi-paramétrico</b>	<b>29</b>
3.1. Motivación . . . . .	29
3.2. Análisis del modelo . . . . .	31
3.2.1. Caso $p(\beta)$ dependiente de $\rho^2$ y ésta conocida . . . . .	32
3.2.2. Caso $p(\beta)$ no dependiente de $\rho^2$ y ésta conocida . . . . .	34
3.2.3. Caso $p(\beta)$ dependiente de $\rho^2$ y ésta desconocida . . . . .	35
3.3. Casos límite . . . . .	35
3.3.1. Regresión paramétrica ( $\rho^2 \rightarrow 0$ ) . . . . .	36
3.3.2. Interpolación Bayesiana ( $\sigma^2 \rightarrow 0$ ) . . . . .	36
3.4. Estimación de $w$ y $\lambda$ . . . . .	38
<b>4. Análisis numérico bayesiano</b>	<b>39</b>
4.1. Interpolación . . . . .	39
4.2. Cuadratura . . . . .	43
4.3. Optimización . . . . .	45
<b>5. Problemas Inversos</b>	<b>47</b>
5.1. Introducción . . . . .	47
5.2. El modelo teórico y el modelo empírico . . . . .	48
5.3. Otras complicaciones . . . . .	48

5.4. Propagación de la incertidumbre . . . . .	49
5.5. Ejemplo . . . . .	50
5.5.1. Ecuación de Antoine . . . . .	50
5.5.2. Aplicación del modelo . . . . .	51
<b>6. Discusión</b>	<b>55</b>
<b>A. Resultados auxiliares</b>	<b>57</b>
<b>B. Modelo General</b>	<b>61</b>
B.1. Modelo . . . . .	61
B.2. Predicción . . . . .	63
B.3. Caso a) . . . . .	63
B.4. Caso b) . . . . .	65
<b>C. Código en R</b>	<b>67</b>
<b>Bibliografía</b>	<b>71</b>

# Capítulo 1

## Introducción

El concepto de regresión surge a través de un estudio realizado por Galton sobre padres e hijos, en el que se comparaba la estatura de los mismos. De esto resultó que los hijos de padres muy altos seguían siendo altos pero con una tendencia a la estatura media. De manera análoga los hijos de padres muy bajos seguían una tendencia a la estatura media. Esto es, se percibía una “regresión” a la estatura promedio de la población. Este hecho dio pie al modelo de regresión que ahora conocemos.

El modelo de regresión plantea la relación entre una variable explicativa y una variable respuesta. En diversas áreas del conocimiento es de interés conocer la interacción de dos o más conceptos con algún otro de interés. La relación que se considera en este caso se supone con algún tipo de perturbación, lo que se conoce como error aleatorio.

Entonces, posiblemente en dicha relación entre variables o conceptos sea de interés (i) la magnitud del impacto de una variable (explicativa) sobre otra (respuesta), (ii) dado cierto valor de una variable, predecir el valor de otra o (iii) quizás lo importante es sólo probar estadísticamente la existencia de la relación entre dichas variables. El modelo de regresión brinda resultados en estos casos; es por esta razón que se utiliza como herramienta estadística en áreas como medicina, psicología, ingeniería, ciencias sociales, etc.

En la mayoría de las veces suele utilizarse el modelo bajo un supuesto de distribución sobre el error aleatorio, comúnmente se le asocia una distribución normal. Sin embargo en la práctica no siempre los datos cumplen con dicho supuesto o quizás (en el caso de ser costoso conseguir muestras grandes) sea difícil probar estadísticamente que se cumple con el mismo. En estos casos los modelos no paramétricos son una buena opción debido a que no se hacen hipótesis paramétricas sobre la distribución de los datos, lo que brinda flexibilidad y robustez.

Otro supuesto común en el modelo de regresión usual, es que los datos tienen homocedasticidad, es decir, la varianza de los errores es constante, aunque esto no siempre se da. En relación con esto se ha desarrollado el proceso gaussiano. Éste sirve para modelar datos en donde existe una correlación de las observaciones dependiendo de la cercanía, en el espacio o en el tiempo, de las mismas. Esto permite introducir el supuesto de que entre observaciones cercanas haya un mayor parecido que entre las que se encuentran más alejadas. El grado de similitud puede controlarse a través de la forma que se asocia a la matriz de correlación del proceso. Esto es importante en investigaciones espaciales, por ejemplo en estudios



geográficos, o quizás de series de tiempo.

En este trabajo se aborda el tema de un modelo de regresión semiparamétrico en donde se consideran dos tipos de errores, por un lado el error aleatorio usual y por el otro un error “determinista” que es modelado como un proceso gaussiano. Esto bajo la perspectiva de la estadística bayesiana. El modelo resultante es muy flexible y su análisis es relativamente simple. Por otro lado, tiene aplicaciones interesantes, entre las que se destacan el tratamiento bayesiano de problemas clásicos de análisis numérico.

En el siguiente capítulo se da una breve introducción al modelo de regresión usual con un enfoque bayesiano, exhibiendo diferentes casos dependiendo del supuesto en la distribución inicial de los parámetros. En este mismo capítulo se da una introducción al proceso gaussiano. En el capítulo tres se desarrolla el modelo de regresión semiparamétrico basado en el proceso gaussiano, discutiendo distintos casos de distribuciones iniciales sobre los parámetros así como lo que ocurre en casos límite. En el capítulo cuatro y cinco se dan ejemplos donde se utiliza este modelo, incluyendo interpolación bayesiana y una aplicación a problemas inversos.

# Capítulo 2

## Preliminares

En Estadística se busca describir la variable objetivo ( $Y$ ), por ejemplo a través de su media (como una característica básica) pero más generalmente se busca estimar la distribución de dicha variable. Para hacer inferencia sobre la variable es necesario recabar información sobre ella. En ocasiones esa información se obtiene a través de algunos otros datos de variables conocidas  $z$ . Se pretende encontrar cómo afectan estos valores a la distribución de la variable  $Y$ . Es decir, se busca conocer la relación para saber cómo modifican estos valores la distribución de  $Y$ .

En este capítulo se abordan conceptos importantes para poder tener un mejor panorama del tema general de este trabajo.

### 2.1. El modelo de regresión

Al analizar algún fenómeno es común encontrar una relación inherente entre variables. Es conveniente analizar este tipo de comportamientos puesto que en ocasiones con sólo conocer ciertas variables de los individuos, se podría obtener algún modelo que ayude en la toma de decisiones. Por ejemplo, suponga que se observa una relación entre el tiempo que destina la televisión abierta a los anuncios publicitarios y el número de usuarios de televisión restringida (paga). Si la meta es llegar a un cierto número de usuarios, con base en esta relación se podría encontrar el tiempo óptimo para los anuncios publicitarios. También sería útil si se deseara saber el número de usuarios dado que se ha elegido un tiempo para los anuncios publicitarios. El análisis de regresión es una técnica estadística que sirve para modelar este tipo de situaciones.

Se habla de un modelo de regresión cuando existe una función matemática que intenta ajustar una variable respuesta  $Y$  en relación a un conjunto de  $m$  variables explicativas representadas en el vector  $z = (z_1, \dots, z_m)^t$ . Debido a que  $Y$  es variable aleatoria, el objetivo principal es ver cómo incide en la distribución de ésta el valor de  $z$ . Es decir, se trata de tener información sobre dicha distribución a través de la variable  $z$ . Conociendo ésta se puede predecir la variable respuesta y se podrán explicar mejor los mecanismos de esa relación. Esto último es útil si se busca encontrar cómo afectan los valores de las variables  $z$  a la distribución de  $Y$ .

Una forma de modelar la relación es pensando en el valor esperado de  $Y$  dadas las variables explicativas

$$E(Y | z) = \gamma(z),$$

donde  $\gamma$  es una función desconocida. Esta función puede ser en principio de cualquier forma, sin embargo se suele usar una forma paramétrica simple para aproximarla

$$\gamma(z) = \eta(z; \beta),$$

donde el parámetro  $\beta = (\beta_0, \dots, \beta_k)^t$  es desconocido. Por simplicidad comúnmente se supone que  $\eta$  es una función lineal de  $\beta$  en alguna escala, es decir  $\eta(z; \beta) = g(\beta_0 + \beta_1\kappa_1(z) + \dots + \beta_k\kappa_k(z))$ . La función  $g$  (que define la escala) es uno a uno y conocida, también son conocidas las funciones  $\kappa_i : \mathbb{R}^m \rightarrow \mathbb{R}$ ,  $i = 1, \dots, k$ .

Como se puede observar, de una suposición general de la relación funcional entre la variable respuesta y las explicativas, se hace una simplificación. Lo que se hace es buscar una manera de hacer más fácil la solución al problema, de tal manera que el manejo de la relación tenga una mejor interpretación y los cálculos no sean tan complicados.

Después de las restricciones antes mencionadas se reduce el problema a encontrar los valores de  $\beta$  de la ecuación (2.1).

Cuando la distribución de  $Y$  se supone dentro de la familia exponencial, esta ecuación da lugar al desarrollo de los *modelos lineales generalizados* (McCullagh y Nelder, 1989). En este caso  $E(Y | z) = g(\beta_0 + \beta_1\kappa_1(z) + \dots + \beta_k\kappa_k(z))$ ,

$$g^{-1}(E(Y | z)) = \beta_0 + \beta_1\kappa_1(z) + \dots + \beta_k\kappa_k(z),$$

se denomina la función liga y  $\beta_0 + \beta_1\kappa_1(z) + \dots + \beta_k\kappa_k(z)$  se conoce como el predictor lineal.

Cuando se habla del modelo usual de regresión lineal, la distribución de  $Y$  se supone normal y se hace un supuesto más, la función  $g$  es la identidad. Entonces se puede reescribir el modelo como:

$$E(Y | z) = \beta_0 + \beta_1x_1 + \dots + \beta_kx_k,$$

donde  $\kappa_j(z) = x_j$ . Suponiendo un experimento de  $n$  observaciones independientes, es decir, ahora contemplando no un vector  $z$  de variables explicativas sino  $n$  vectores de  $m$  entradas cada uno, sea  $\{z_i, y_i\}_{i=1}^n$  una muestra de tamaño  $n$  de la variable explicativa y la variable respuesta. Entonces el modelo de regresión se escribe en forma matricial como en la ecuación (2.1), donde  $\mathbf{Y} = (Y_1, \dots, Y_n)$ ,  $x_{ij} = \kappa_j(z_i)$  y

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & \cdots & \ddots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix},$$

$$E(\mathbf{Y} | X) = \mathbf{X}\beta. \tag{2.1}$$

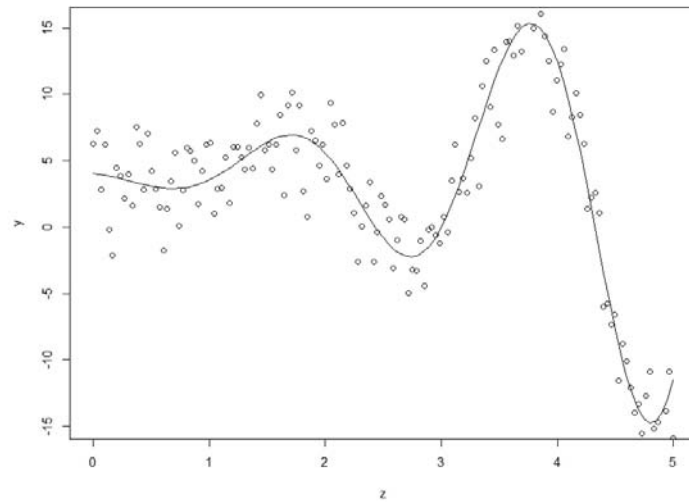


Figura 2.1: Simulación de un problema de regresión

En la Figura 2.1 se muestra una simulación de un problema de regresión no lineal en la variable  $z$ . La curva continua que se observa corresponde a la verdadera relación<sup>1</sup> entre la variable explicativa y la variable respuesta. En particular se trata de un modelo de regresión polinomial, el cual supone que  $z \in \mathbb{R}$  y además  $\kappa_j(z) = z^j$ . Es decir,  $E(Y | z) = \beta_0 + \beta_1 z + \dots + \beta_k z^k$ .

### Normalidad

Como ya se mencionó, el modelo de regresión usual supone que la variable  $Y$  tiene una distribución normal. Este puede ser un supuesto restrictivo; sin embargo, es muy útil pues simplifica los cálculos y la interpretación.

$$Y_i \sim N(y_i | \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}, \sigma^2).$$

Entonces  $y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \epsilon_i$  con  $\epsilon_i \sim N(0, \sigma^2)$  para  $i = 1, \dots, n$ . Escrito lo anterior de forma matricial,

$$\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_n(0, \sigma^2 I_n), \quad (2.2)$$

o, equivalentemente,

$$\mathbf{Y} \sim N_n(\mathbf{X}\beta, \sigma^2 I_n),$$

donde  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^t$ .

<sup>1</sup>La función de la línea continua es  $y = f(z) = 0.06 * (z + 2)^2 * \sin(3z + 3) + 4$

## 2.2. Análisis

En esta sección se analiza el modelo de regresión, bajo un modelo Bayesiano para diferentes casos, éstos dependiendo de la distribución inicial del parámetro  $\beta$  y del supuesto de desconocer o no el parámetro  $\sigma^2$ . Estos resultados son conocidos en la literatura del tema, como por ejemplo en Broemeling (1985), Zellner (1996) y Gutiérrez-Peña (1998).

### 2.2.1. Caso $p(\beta)$ dependiente de $\sigma^2$ y ésta conocida

En esta sección se analizará el modelo anterior suponiendo  $\sigma^2$  conocida. Es ilustrativo desarrollar esta situación, pero además será útil al analizar otros casos que se desarrollan más adelante.

#### Distribución inicial

La distribución inicial del parámetro  $\beta$  se expresa en (2.3). Esta distribución está basada en la familia conjugada (ver Bernardo y Smith, 2000)

$$[\beta \mid \sigma^2] \sim N(b_0, \sigma^2 B_0^{-1}), \quad (2.3)$$

donde  $b_0 \in \mathbb{R}^{k+1}$  y  $B_0$   $(k+1) \times (k+1)$  es una matriz simétrica y positiva definida.

#### Distribución final

Para los casos particulares de distribución inicial antes mencionados, se obtienen los siguientes resultados

$$[\beta \mid \sigma^2, \mathbf{y}] \sim N(b_1, \sigma^2 B_1^{-1}), \quad (2.4)$$

donde

$$\begin{aligned} \mathbf{B}_1 &= \mathbf{X}^t \mathbf{X} + \mathbf{B}_0, \\ \mathbf{b}_1 &= \mathbf{B}_1^{-1} (\mathbf{X} \mathbf{y} + \mathbf{B}_0 \mathbf{b}_0). \end{aligned}$$

Este resultado se obtiene de usar el teorema de Bayes,

$$p(\beta \mid \sigma^2, \mathbf{y}) \propto p(\beta \mid \sigma^2) L(\beta, \sigma^2; \mathbf{y}),$$

donde

$$\begin{aligned} L(\beta, \sigma^2; \mathbf{y}) &\propto (2\pi\sigma^2)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta)^t(\mathbf{y} - \mathbf{X}\beta)\right\} \\ &\propto (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}[(\beta - \hat{\beta})^t \mathbf{X}^t \mathbf{X}(\beta - \hat{\beta}) + n\hat{\sigma}^2]\right\}, \\ \hat{\beta} &= (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{y}, \\ \hat{\sigma}^2 &= \frac{1}{n}(\mathbf{y} - \mathbf{X}^t \beta)^t(\mathbf{y} - \mathbf{X}^t \beta). \end{aligned} \quad (2.5)$$

### Inferencia

De acuerdo al resultado anterior se hará inferencia sobre los parámetros e inclusive intervalos de máxima densidad sin perder de vista que se está en el caso donde  $\sigma^2$  es conocida.

En el caso de  $\beta$ , con base en (2.4) se conoce su esperanza y varianza final. El estimador para el parámetro puede ser su esperanza final y con base en la matriz de varianzas construir intervalos de probabilidad para cada uno de los coeficientes de regresión.

### Predicción

Un problema común que se desea resolver con base en regresión, es el predecir el valor de una variable dados valores conocidos. Es decir, suponga que se tiene el vector de variables explicativas  $\mathbf{x}_*^t = (1, x_{1*}, \dots, x_{k*})$  y se desea predecir el valor de la variable respuesta  $Y_*$ . Bajo el supuesto de regresión antes descrito, el modelo es

$$Y_* = \mathbf{x}_*^t \beta + \epsilon_*, \quad (2.6)$$

donde  $\epsilon_* \sim N(0, \sigma^2)$  e independiente de  $\epsilon$ .

Definiendo a  $\mu_*$  como el valor esperado de la observación futura  $Y_*$  y bajo el modelo anterior se tiene

$$\begin{aligned} \mu_* &= E(Y_* | \beta, \sigma^2) \\ &= \mathbf{x}_*^t \beta, \\ \text{Var}(Y_* | \beta, \sigma^2) &= \sigma^2. \end{aligned} \quad (2.7)$$

Este problema se puede resolver desde dos perspectivas, una de ella es con base en la distribución de  $\mu_*$  y la otra bajo la distribución de  $Y_*$ . A continuación se explican ambos enfoques.

#### *Inferencia sobre $\mu_*$*

Desde este punto de vista se busca la distribución final de  $\mu_*$ ; para ello hay que notar que éste es una combinación lineal del  $\beta$ . Se sabe que la distribución de  $\beta$  dados  $\mathbf{y}$  y  $\sigma^2$  tiene una distribución normal, entonces la distribución  $\mu_*$  dados  $\mathbf{y}$  y  $\sigma^2$  también es normal. Por lo tanto, con base en lo que se hizo en (2.4), la distribución final de dicho parámetro es

$$p(\mu_* | \mathbf{y}) = N(\mu_* | \mathbf{x}_*^t \mathbf{b}_1, \sigma^2 \mathbf{x}_*^t B_1^{-1} \mathbf{x}_*).$$

De esto se sigue que el intervalo de máxima densidad de  $(1 - \alpha) \times 100\%$  es

$$\mathbf{x}_*^t \mathbf{b}_1 \pm z^{1-\alpha/2} \sigma \sqrt{\mathbf{x}_*^t B_1^{-1} \mathbf{x}_*},$$

donde  $z^{1-\alpha/2}$  es el cuantil de orden  $(1 - \frac{\alpha}{2})$  de la distribución normal estándar.

*Inferencia sobre  $Y_*$* 

En este caso se busca la distribución predictiva final de  $Y_*$ , es decir,  $p(y_* | \mathbf{y})$ . Con base en el modelo (2.6) y la distribución de  $\beta$  y  $\epsilon_*$  dados  $\mathbf{y}$  y  $\sigma^2$ , se tiene

$$p(y_* | \sigma^2, \mathbf{y}) = N(y_* | \mathbf{x}_*^t \mathbf{b}_1, \sigma^2(1 + \mathbf{x}_*^t \mathbf{B}_1^{-1} \mathbf{x}_*)). \quad (2.8)$$

Un intervalo predictivo de máxima densidad del  $(1 - \alpha) \times 100\%$  es

$$\mathbf{x}_*^t \mathbf{b}_1 \pm z^{1-\alpha/2} \sigma \sqrt{1 + \mathbf{x}_*^t \mathbf{B}_1^{-1} \mathbf{x}_*}. \quad (2.9)$$

**2.2.2. Caso  $p(\beta)$  no dependiente de  $\sigma^2$  y ésta conocida**

Este caso es de utilidad en situaciones en las que la relación entre  $\beta$  y  $\sigma^2$  que establece la distribución conjugada se considera demasiado restrictiva *a priori*.

**Distribución inicial**

La distribución inicial del parámetro  $\beta$  se expresa como

$$[\beta | \sigma^2] \sim N(b_0, B_0^{-1}), \quad (2.10)$$

donde  $b_0 \in \mathbb{R}^{k+1}$  y  $B_0$   $_{(k+1) \times (k+1)}$  es una matriz simétrica y positiva definida.

**Distribución final**

Para los casos particulares de distribución inicial antes mencionados, se obtienen los siguientes resultados

$$[\beta | \sigma^2, \mathbf{y}] \sim N(b_1, B_1^{-1}), \quad (2.11)$$

donde

$$\begin{aligned} \mathbf{B}_1 &= \sigma^{-2} \mathbf{X}^t \mathbf{X} + \mathbf{B}_0, \\ \mathbf{b}_1 &= \mathbf{B}_1^{-1} \left( \frac{1}{\sigma^2} \mathbf{X}^t \mathbf{y} + \mathbf{B}_0 \mathbf{b}_0 \right). \end{aligned} \quad (2.12)$$

Como en el caso anterior, este resultado se obtiene de usar el teorema de Bayes,

$$p(\beta | \sigma^2, \mathbf{y}) \propto p(\beta | \sigma^2) L(\beta, \sigma^2; \mathbf{y}),$$

donde  $L(\beta, \sigma^2; \mathbf{y})$  está dada por (2.5).

### Inferencia

Como en el caso anterior, se puede hacer inferencia sobre los coeficientes de regresión, por ejemplo a través de intervalos de máxima densidad.

### Predicción

Suponga que se tiene el vector de variables explicativas  $\mathbf{x}_*^t = (1, x_{1*}, \dots, x_{k*})$  y se desea predecir el valor de la variable respuesta  $Y_*$ . Considerando  $Y_*$  como en (2.6) y  $\mu_*$  como en la (2.7).

#### *Inferencia sobre $\mu_*$*

Desde este punto de vista se busca la distribución final de  $\mu_*$ , para ello hay que notar que ésta es una combinación lineal del  $\beta$ , al igual que en el caso anterior. Por lo tanto, con base en (2.11), la distribución final de dicha variable es

$$p(\mu_* | \mathbf{y}) = N(\mu_* | \mathbf{x}_*^t \mathbf{b}_1, \mathbf{x}_*^t B_1^{-1} \mathbf{x}_*).$$

De esto se sigue el intervalo de máxima densidad de  $(1 - \alpha) \times 100\%$  es

$$\mathbf{x}_*^t \mathbf{b}_1 \pm z^{1-\alpha/2} \sqrt{\mathbf{x}_*^t B_1^{-1} \mathbf{x}_*},$$

donde  $z^{1-\alpha/2}$  es el cuantil de orden  $(1 - \frac{\alpha}{2})$  de la distribución normal estándar.

Es importante recalcar que los valores de  $b_1$  y  $B_1$  que se utilizan en esta sección son los expuestos en (2.12).

#### *Inferencia sobre $Y_*$*

En este caso se busca la distribución predictiva final de  $Y_*$ , es decir,  $p(y_* | \mathbf{y})$ . Con base en el modelo (2.6) y la distribución de  $\beta$  y  $\epsilon_*$  dados  $\mathbf{y}$  y  $\sigma^2$  se tiene

$$p(y_* | \sigma^2, \mathbf{y}) = N(y_* | \mathbf{x}_*^t \mathbf{b}_1, \sigma^2 + \mathbf{x}_*^t \mathbf{B}_1^{-1} \mathbf{x}_*).$$

Un intervalo de máxima densidad del  $(1 - \alpha) \times 100\%$  es

$$\mathbf{x}_*^t b_1 \pm z^{1-\alpha/2} \sqrt{\sigma^2 + \mathbf{x}_*^t B_1^{-1} \mathbf{x}_*}, \quad (2.13)$$

de nuevo se hace hincapié en que las expresiones para  $b_1$  y  $B_1$  en este caso están en (2.12).

### 2.2.3. Caso $p(\beta)$ dependiente de $\sigma^2$ y ésta desconocida

En el caso anterior, al suponer  $\sigma^2$  conocida los cálculos se hacen sencillos. Sin embargo a continuación se analizará el modelo suponiéndola desconocida debido a que en la mayoría de las situaciones ese será el caso. Para esto se hará uso de algunos de los resultados de la Sección 2.2.1.



En algunos libros y artículos, sobre todo los que manejan una perspectiva Bayesiana, se hace un cambio de notación para facilitar cálculos. Este cambio consisten en usar la precisión en lugar de la varianza como parámetro  $\tau = \frac{1}{\sigma^2}$ , de manera que  $\mathbf{Y} \sim N_n(\mathbf{X}\beta, \tau^{-1}I_n)$ .

### Distribución inicial

Pensando en la familia conjugada, la distribución inicial de los parámetros  $\beta$  y  $\tau$  debe ser tal que su función de densidad sea de la forma:

$$p(\beta, \tau) \propto \tau^{n_0/2} \exp\left\{-\frac{\tau}{2}[(\beta - b_0)^t B_0(\beta - b_0) + t_0]\right\} \quad (2.14)$$

donde  $n_0, t_0 \in \mathbb{R}$ ,  $b_0 \in \mathbb{R}^{k+1}$  y  $B_0$  es una matriz simétrica y positiva definida.

Notemos que la situación anterior se sigue si

$$\begin{aligned} [\beta \mid \tau] &\sim N_{k+1}(b_0, \tau^{-1}B_0^{-1}), \\ \tau &\sim \text{Gamma}(\tau \mid a_0/2, t_0/2). \end{aligned}$$

donde  $a_0 = n_0 - k + 1$  (ver, por ejemplo, Bromeling, 1985). Este esquema corresponde a la distribución *Normal-Gamma*.

En particular la distribución de  $\tau$  se puede reescribir como una Ji-cuadrada con los parámetros adecuados. Esto recordando que si  $G \sim \text{Gamma}(g/2, 1/2)$  entonces  $G \sim \chi_g^2$ , para efectos del parámetro  $\tau$  se tiene que

$$\begin{aligned} \tau &\sim \text{Gamma}(\tau \mid a_0/2, t_0/2), \\ &\sim t_0 \chi_{a_0}^2. \end{aligned}$$

Por lo tanto la distribución inicial para la varianza  $\sigma^2$ , bajo este esquema, es  $\sigma^2 \sim s_0 \chi_{a_0}^{-2}$ , con  $t_0^{-1} = s_0$ . Aquí,  $\chi_{a_0}^{-2}$  denoto una distribución Ji-cuadrada inversa con  $a_0$  grados de libertad.

Cuando no se tiene información sobre la distribución inicial de los parámetros se utilizan reglas para encontrar distribuciones no informativas. Utilizando la *Regla de Jeffreys* dicha distribución es  $\pi(\beta, \tau) \propto \tau^{(k-1)/2}$ . Sin embargo, la más aceptada es la distribución inicial de referencia que se puede obtener a partir del método de Bernardo (1979), dada por  $\pi(\beta, \tau) \propto \tau^{-1}$ .

### Distribución final

Establecida la distribución inicial que refleje la información inicial que se tiene sobre el parámetro, se obtiene la distribución final haciendo uso del teorema de Bayes. Para los casos particulares de distribución inicial antes mencionados, se obtienen los siguientes resultados.

**Proposición 2.1.** *La distribución final para el modelo (2.2) si se utiliza una distribución inicial conjugada de la forma (2.14) es*

$$p(\beta, \tau | \mathbf{y}) = N_{k+1}(\beta | \mathbf{b}_1, \tau^{-1} \mathbf{B}_1^{-1}) \text{Ga}(\tau | \frac{a_1}{2}, \frac{d_1}{2}), \quad (2.15)$$

donde

$$\begin{aligned} \mathbf{B}_1 &= \mathbf{X}^t \mathbf{X} + \mathbf{B}_0, \\ \mathbf{b}_1 &= \mathbf{B}_1^{-1}(\mathbf{X} \mathbf{y} + \mathbf{B}_0 \mathbf{b}_0) \\ a_1 &= n + a_0, \\ d_1 &= (\mathbf{y} - \mathbf{X} \mathbf{b}_1)^t (\mathbf{y} - \mathbf{X} \mathbf{b}_1) + (\mathbf{b}_1 - \mathbf{b}_0)^t \mathbf{B}_0 (\mathbf{b}_1 - \mathbf{b}_0) + t_0. \end{aligned}$$

Con base en este resultado, las distribuciones finales marginales son:

$$p(\tau | \mathbf{y}) = \text{Ga}(\tau | \frac{a_1}{2}, \frac{d_1}{2}), \quad (2.16)$$

$$p(\beta | \mathbf{y}) = \text{St}_{k+1}(\beta | \mathbf{b}_1, \mathbf{T}_1^{-1}, a_1), \quad (2.17)$$

donde  $\text{St}_{k+1}$  corresponde a la distribución  $t$ -Student en el caso multivariado con  $a_1$  grados de libertad, parámetro de localización  $\mathbf{b}_1$  y parámetro de escala  $\mathbf{T}_1^{-1}$ , con  $\mathbf{T}_1 = \frac{a_1}{d_1} \mathbf{B}_1$ .

Dado que  $\tau$  tiene una distribución Gamma, entonces la distribución de  $\sigma^2$  es una Gamma Inversa, es decir, con base en (2.16)

$$p(\sigma^2 | \mathbf{y}) = \text{IGa}(\sigma^2 | \frac{a_1}{2}, \frac{d_1}{2}). \quad (2.18)$$

Cuando se utiliza una distribución inicial de referencia,  $\pi(\beta, \tau) \propto \tau^{-1}$ , la distribución final es

$$\pi(\beta, \tau | \mathbf{y}) = N_{k+1}(\beta | \hat{\beta}, \tau^{-1} (\mathbf{X}^t \mathbf{X})^{-1}) \text{Ga}(\tau | \frac{n-k-1}{2}, \frac{n-k-1}{2} \hat{\sigma}^2),$$

donde

$$\hat{\beta} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{y}$$

y

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X} \hat{\beta})^t (\mathbf{y} - \mathbf{X} \hat{\beta})}{n - k - 1},$$

son los estimadores insesgados usuales para  $\beta$  y  $\sigma^2$  respectivamente. Ésta es un caso límite del modelo (2.14) con  $t_0 = 0$ ,  $a_0 = -(k+1)$  y  $\mathbf{B}_0 = \mathbf{0}$ .

Con base en la distribución conjunta anterior, las distribuciones marginales son:

$$\begin{aligned} \pi(\tau | \mathbf{y}) &= \text{Ga}(\tau | \frac{n-k-1}{2}, \frac{n-k-1}{2} \hat{\sigma}^2), \\ \pi(\beta | \mathbf{y}) &= \text{St}_{k+1}(\beta | \hat{\beta}, \hat{\sigma}^2 (\mathbf{X}^t \mathbf{X})^{-1}, n - k - 1). \end{aligned}$$

### Inferencia

De acuerdo a estos resultados, se puede hacer inferencia sobre los parámetros, por ejemplo, a través de intervalos de máxima densidad, como en los casos anteriores.

En el caso de  $\beta$ , con base en (2.17), su esperanza y varianza final se expresan como

$$\begin{aligned} E(\beta | \mathbf{y}) &= \mathbf{b}_1 \quad \text{si } a_1 > 1, \\ \text{Var}(\beta | \mathbf{y}) &= \frac{a_1}{a_1 - 2} \mathbf{T}_1^{-1} \quad \text{si } a_1 > 2. \end{aligned} \quad (2.19)$$

La esperanza puede ser el estimador de dicho parámetro y se pueden construir intervalo de máxima densidad usando la varianza y tomando en cuenta su distribución normal.

Para  $\sigma^2$ , con base en 2.18, el estimador considerando su esperanza final, se expresa como

$$E(\sigma^2 | \mathbf{y}) = \frac{d_1}{a_1 - 2}.$$

Y su varianza y moda finales son

$$\begin{aligned} \text{Var}(\sigma^2 | \mathbf{y}) &= \frac{2d_1^2}{(a_1 - 2)^2(a_1 - 4)}, \\ \text{Moda}(\sigma^2 | \mathbf{y}) &= \frac{d_1}{(a_1 + 2)}. \end{aligned}$$

Si se utiliza la distribución de referencia las expresiones se reducen y quedan en función de las estimadores insesgados antes mencionados. Para el caso de  $\sigma^2$  las expresiones son

$$\begin{aligned} E(\sigma^2 | \mathbf{y}) &= \frac{n - k - 1}{n - k - 3} \hat{\sigma}^2, \\ \text{Var}(\sigma^2 | \mathbf{y}) &= \frac{2(n - k - 1)^2}{(n - k - 3)^2(n - k - 5)} \hat{\sigma}^4, \\ \text{Moda}(\sigma^2 | \mathbf{y}) &= \frac{n - k - 1}{(n - k + 1)} \hat{\sigma}^2; \end{aligned}$$

y para  $\beta$

$$\begin{aligned} E(\beta | \mathbf{y}) &= \hat{\beta} \quad \text{si } n > k + 2, \\ \text{Var}(\beta | \mathbf{y}) &= \frac{n - k - 1}{n - k - 3} \hat{\sigma}^2 (\mathbf{X}^t \mathbf{X})^{-1} \quad \text{si } n > k + 3. \end{aligned}$$

### Predicción

Como en los casos ya mencionados anteriormente, suponga que se tiene el vector de variables explicativas  $x_*^t = (1, x_{1*}, \dots, x_{k*})$  y se desea predecir el valor de la variable respuesta

$Y_*$ . Este problema se abordará desde las dos perspectivas antes mencionadas. A continuación se explican ambos enfoques.

*Inferencia sobre  $\mu_*$*

Desde este punto de vista se busca la distribución final de  $\mu_*$ . Para ello hay que notar que ésta es una combinación lineal del  $\beta$ . Se sabe que la distribución de  $\beta$  dados  $\mathbf{y}$  y  $\sigma^2$  tiene una distribución normal, entonces la distribución  $\mu_*$  dados  $\mathbf{y}$  y  $\sigma^2$  también es normal. Por lo tanto, con base en lo que se hizo en (2.17), la distribución final de dicho parámetro es

$$p(\mu_* | \mathbf{y}) = St(\mu_* | \mathbf{x}_*^t \mathbf{b}_1, \frac{d_1}{a_1} \mathbf{x}_*^t \mathbf{B}_1^{-1} \mathbf{x}_*, a_1).$$

Cuando se utiliza una distribución inicial de referencia la distribución final es

$$\pi(\mu_* | \mathbf{y}) = St(\mu_* | \mathbf{x}_*^t \hat{\beta}, \hat{\sigma}^2 \mathbf{x}_*^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_*, n - k - 1). \quad (2.20)$$

Un intervalo de máxima densidad del  $(1 - \alpha) \times 100\%$  usando la distribución (2.20) es

$$\mathbf{x}_*^t \hat{\beta} \pm t_{n-k-1}^{1-\alpha/2} \hat{\sigma} \sqrt{\mathbf{x}_*^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_*} \quad (2.21)$$

donde  $t_{n-p}^{1-\alpha/2}$  es el cuantil de orden  $(1 - \alpha/2)$  de una distribución  $t$ -Student estandarizada con  $(n - k - 1)$  grados de libertad.

*Inferencia sobre  $Y_*$*

En este caso se busca la distribución predictiva final de  $Y_*$ , es decir,  $p(y_* | \mathbf{y})$ . Con base en el modelo (2.6) y la distribución de  $\beta$  y  $\epsilon_*$  dados  $\mathbf{y}$  y  $\sigma^2$ , se tiene

$$p(y_* | \sigma^2, \mathbf{y}) = N(y_* | \mathbf{x}_*^t \mathbf{b}_1, \sigma^2 (1 + \mathbf{x}_*^t \mathbf{B}_1^{-1} \mathbf{x}_*), a_1).$$

Para obtener la distribución deseada, sólo se integra con respecto a la distribución final de  $\sigma^2$ , lo que da como resultado

$$p(y_* | \mathbf{y}) = St(y_* | \mathbf{x}_*^t \mathbf{b}_1, \frac{d_1}{a_1} (1 + \mathbf{x}_*^t \mathbf{B}_1^{-1} \mathbf{x}_*), a_1).$$

Cuando se utiliza una distribución inicial de referencia la distribución final es

$$\pi(y_* | \mathbf{y}) = St(y_* | \mathbf{x}_*^t \hat{\beta}, \hat{\sigma}^2 (1 + \mathbf{x}_*^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_*), n - k - 1). \quad (2.22)$$

Un intervalo predictivo de máxima densidad del  $(1 - \alpha) \times 100\%$  usando la distribución (2.22) es

$$\mathbf{x}_*^t \hat{\beta} \pm t_{n-k-1}^{1-\alpha/2} \hat{\sigma} \sqrt{1 + \mathbf{x}_*^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_*}. \quad (2.23)$$

## 2.3. El proceso gaussiano

Existen diferentes situaciones donde se estudia una secuencia de observaciones o puntos entre los cuales existe una correlación. Esta correlación depende de la cercanía entre las observaciones, es decir, la medición en un punto estará relacionada de manera diferente a puntos cercanos que a los lejanos. El proceso gaussiano sirve para modelar este tipo de datos.

**Definición 2.1.** *Un **proceso Gaussiano** es un proceso estocástico continuo en el tiempo, el cual para cualquier vector de dimensión finita  $Y = (Y(t_1), \dots, Y(t_n))^t$  tiene una distribución normal multivariada  $N_n(\boldsymbol{\mu}(t), \mathbf{V}(t))$  para algún vector de medias  $\boldsymbol{\mu}$  y matriz de covarianzas  $\mathbf{V}$  que depende de  $t = (t_1, \dots, t_n)^t$ . Un **proceso Gaussiano** también puede indexarse por medio de un vector  $\mathbf{x} \in \mathbb{R}^m$  que se interpreta como una posición en el espacio.*

Sea  $\{x_i\}_{i=1}^n$  un conjunto de puntos ubicados en el espacio ( $\mathbb{R}^m$ ), el conjunto  $\{y(x_i)\}_{i=1}^n$  representa una realización del proceso gaussiano. La característica de este proceso es la distribución normal multivariada de  $y = (y(x_1), \dots, y(x_n))^t$ , es decir:

$$y \sim N_n(\boldsymbol{\mu}, \Sigma) \quad (2.24)$$

donde  $\Sigma$  es el *kernel* del proceso. Existen varias funciones para representar cada elemento de esta matriz  $\Sigma_{ij}$ , suelen dividirse según sea el proceso estacionario o no.

Para procesos estacionarios<sup>2</sup>:

### Exponencial Cuadrado

$$\Sigma_{ij} = \exp(-\|x_i - x_j\|^2)$$

### $\gamma$ -Exponencial

$$\Sigma_{ij} = \exp(-\|x_i - x_j\|^\gamma) \quad 0 < \gamma < 2$$

### Matérn

$$\Sigma_{ij} = \|x_i - x_j\|^v K_v(\|x_i - x_j\|),$$

donde  $K_v$  es modificación de la función de Bessel  $v > 0$  **Rotación cuadrática**

$$\Sigma_{ij} = (1 + \|x_i - x_j\|^2)^{-\alpha} \quad \alpha > 0$$

---

<sup>2</sup>Un proceso es estrictamente estacionario si todas sus distribuciones finito dimensionales son invariantes ante transformaciones, es decir,  $F(Y(x_1), \dots, Y(x_n)) = F(Y(x_1+s), \dots, Y(x_n+s))$  (Rasmussen & Williams, 2006)

Para procesos no estacionarios:

### Lineal

$$\Sigma_{ij} = x_i^t x_j$$

### Red neuronal

$$\Sigma_{ij} = \arcsin \left( \frac{2x_i^t x_j}{\sqrt{1 + 2x_i^t x_i} \sqrt{1 + 2x_j^t x_j}} \right)$$

Otro kernel usado comúnmente es

$$\Sigma_{ij}(\sigma^2, r) = \sigma^2 \exp(-(\|x_i - x_j\|/r)^2) \quad (2.25)$$

donde  $\sigma^2$  es la varianza marginal de  $Y$ .

Distintas muestras de un proceso Gaussiano pueden tener diferentes propiedades dependiendo de la función de covarianza. Un ejemplo de esto es cuando se utiliza la función tal que  $\Sigma_{ij} = \exp(-|x_i - x_j|)$ , cuando las variables explicativas son de dimensión uno. Este proceso es similar al Ornstein-Uhlenbeck.

Información más a detalle del proceso gaussiano y del kernel puede encontrarse en Geoffrey y Stirzaker (2001) sección 9.6, Bishop (2006) capítulo 6 y Barber (2012) capítulo 19.

### 2.3.1. Simulación del proceso gaussiano

Los paquetes de computación, como 'gptk' del sistema R (R Core Team, 2013) y 'gpml' del lenguaje Matlab, suelen generar una realización del proceso gaussiano con media cero y matriz de covarianzas  $\Sigma$ , basados en las propiedades de la distribución normal multivariada.

Si  $z \sim N_n(m, S)$  entonces  $Az \sim N_n(Am, ASA^t)$ , donde  $A$  es una matriz de  $n \times m$ . Por lo tanto, una forma de simular el proceso gaussiano es siguiendo el siguiente algoritmo.

1. Calcular una matriz  $A$  tal que  $AA^t = \Sigma$ .
2. Generar  $u \sim N_n(0, I_n)$ .
3. Calcular  $y = Au$ .

Con base en la definición anterior y los diferentes kernels, se muestran algunas simulaciones. Estas fueron realizadas usando la función *gaussSamp* de la librería 'gptk' del sistema R (R Core Team, 2013).

En las Figuras 2.2, 2.3 y 2.4 se muestra una realización del proceso gaussiano para diferentes tipos de kernel; en las primeras dos figuras la ubicación espacial está sobre  $x = \{0, 1, \dots, 7\}$  y en la tercera  $x = \{-5, \dots, 20\}$ . Como se puede observar, al usarse una matriz de covarianza diferente el comportamiento de  $Y$  cambia. En el caso de la Figura 2.3 hay un comportamiento más errático, a diferencia de las otras dos. Sin embargo en la Figura 2.4 existe una mayor suavidad en la curva.

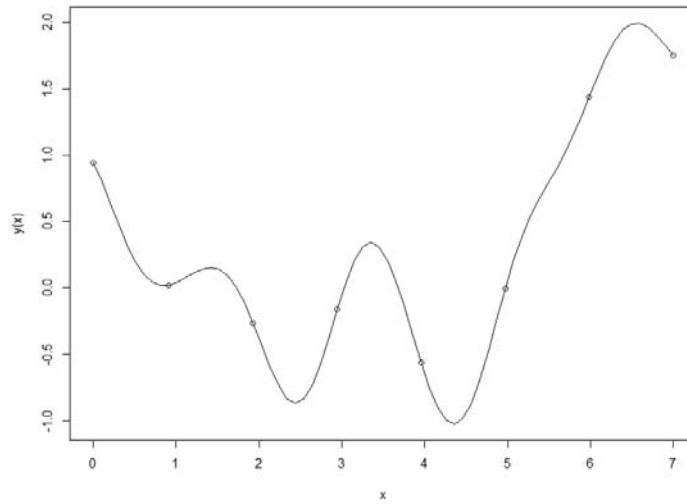


Figura 2.2: Proceso gaussiano con  $\Sigma_{ij} = \exp(-(x_i - x_j)^2)$

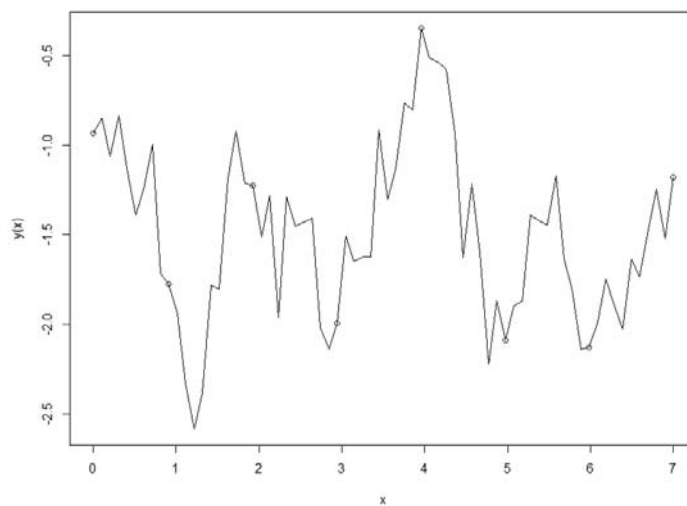


Figura 2.3: Proceso gaussiano con  $\Sigma_{ij} = \exp(-0.5\|x_i - x_j\|)$

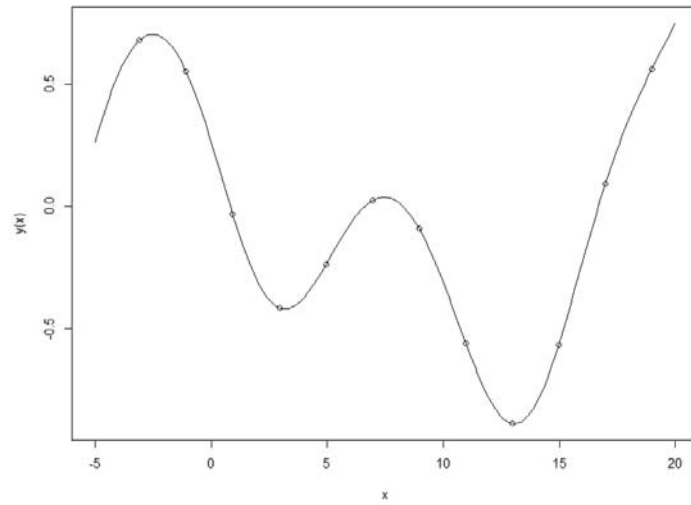


Figura 2.4: Proceso gaussiano con  $\Sigma_{ij} = \exp(-(\frac{\|x_i - x_j\|}{5})^2)$

En la Figura 2.5 se muestran cuatro realizaciones del proceso con el mismo kernel. En esta figura, a diferencia de las otras, se utiliza un kernel especial

$$\Sigma_{ij} = \rho^2 \exp\{-\lambda(x_i - x_j)^t(x_i - x_j)\},$$

el cual será mencionado en los Capítulos 3 y 4 de este trabajo (ver ecuaciones (3.5) y (3.6)). El valor  $\rho^2 = 0.5$  representa la varianza marginal del proceso y  $\lambda = 0.1$  es el valor del parámetro relacionado con lo errático del proceso.

### 2.3.2. Distribución condicional

En ocasiones se suele hacer inferencia sobre un conjunto de valores en el dominio de  $x$ , dado que son conocidos los valores del proceso en algunos puntos particulares. Sea  $y(x)$  un proceso gaussiano con covarianza  $\Sigma$  sobre  $x = (x_1, \dots, x_n)^t$ . Suponga que se observaron  $(x_{p_1}, \dots, x_{p_k})^t$  donde  $p_j \in \{1, \dots, n\}$  para  $0 < j \leq k < n$ . Es decir algunos puntos de  $x$  son conocidos. Llamando  $x_1^* = (x_{p_1}, \dots, x_{p_k})^t$  y  $x_2^*$  el conjunto de puntos no contemplados en  $x_1^*$ , se desea encontrar la distribución de las variables no observadas dadas las conocidas.

Para encontrar dicha distribución se utilizan las propiedades de la distribución normal multivariada descritas en el Apéndice A. Sea  $x^* = (x_1^{*t}, x_2^{*t})^t$ , al tener  $Y(x)$  una distribución normal multivariada,

$$\begin{pmatrix} Y(x_1^*) \\ Y(x_2^*) \end{pmatrix} \sim N \left( \begin{pmatrix} \mu_1^* \\ \mu_2^* \end{pmatrix}, \begin{pmatrix} \Sigma_{11}^* & \Sigma_{12}^* \\ \Sigma_{21}^* & \Sigma_{22}^* \end{pmatrix} \right), \quad (2.26)$$

al reordenarla seguirá siendo normal multivariada pero con matriz de covarianza según el reordenamiento.



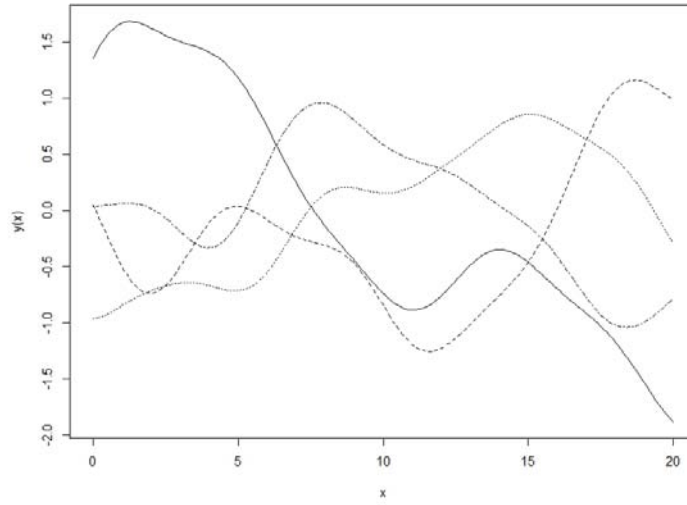


Figura 2.5: Cuatro simulaciones de proceso gaussiano con  $\Sigma_{ij} = 0.5 \exp(-0.1\|x_i - x_j\|^2)$

La ecuación (2.27) expresa la distribución condicional, con la cual se puede hacer inferencia o simulaciones sobre el conjunto de datos que no se observaron,

$$[Y(x_2^*) | Y(x_1^*)] \sim N_{n-k}(\mu_{2.1}^*(x_1^*), \Sigma_{22.1}^*), \quad (2.27)$$

donde  $\mu_{2.1}^* = \mu_2^* + \Sigma_{21}^* \Sigma_{11}^{*-1} (x_1^* - \mu_1^*)$  y  $\Sigma_{22.1}^* = \Sigma_{22}^* - \Sigma_{21}^* \Sigma_{11}^{*-1} \Sigma_{12}^*$ .

En las Figuras 2.6 y 2.7 se simularon procesos gaussianos dados tres valores conocidos,  $x_1^* = (4.924, 9.947, 14.974)$ . En ambas figuras se grafican las simulaciones de cuatro procesos condicionados con líneas discontinuas, la línea continua corresponde al proceso original de donde fueron extraídas las observaciones para el condicionamiento.

### 2.3.3. Otros aspectos

En esta sección se mencionan algunos aspectos del proceso gaussiano. Esta información es una breve recopilación de algunos artículos y manuscritos de algunos autores que han desarrollado más a fondo los temas.

#### Covarianzas en el modelo de regresión

Sin considerar un proceso gaussiano inicialmente, se ha analizado la matriz de covarianzas del modelo de regresión como en Mackay(1998). En dicho artículo se explica más a detalle lo que a continuación se expone.

Suponga un modelo de regresión lineal en  $\beta$ , de la forma

$$y(\mathbf{x}, \beta) = \sum_{h=1}^H \beta_h \phi_h(\mathbf{x}), \quad (2.28)$$

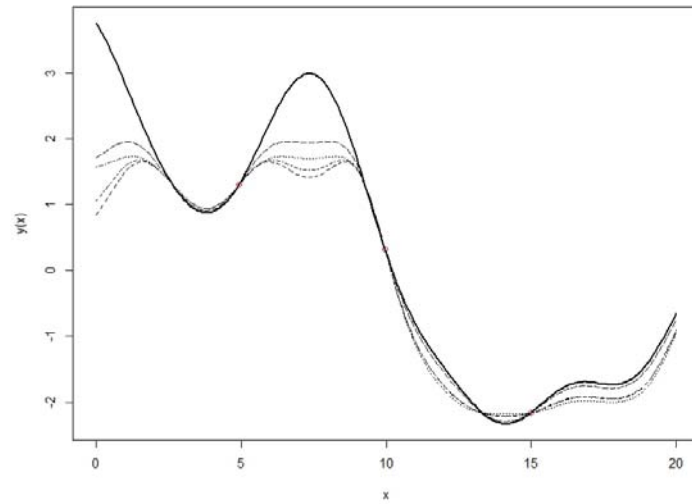


Figura 2.6: Simulaciones del proceso gaussiano condicionado a tres puntos con  $\Sigma_{ij} = 0.5 \exp(-0.1\|x_i - x_j\|^2)$

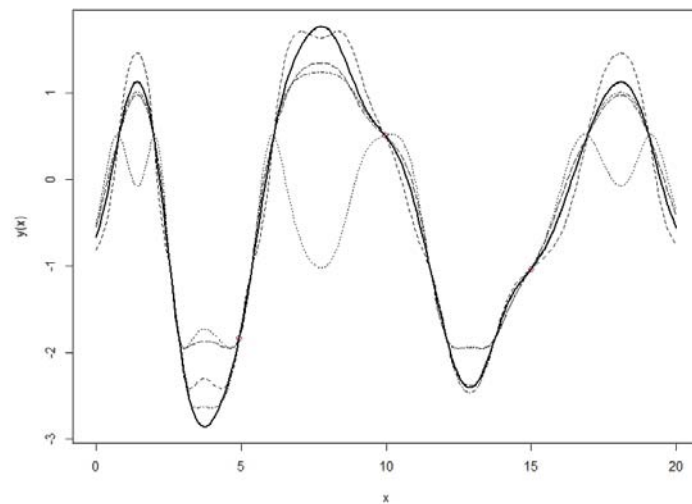


Figura 2.7: Simulaciones del proceso gaussiano condicionado a tres puntos con  $\Sigma_{ij} = 2 \exp(-0.3\|x_i - x_j\|^2)$

donde  $\mathbf{x} = (x_1, x_2, \dots, x_n)^t$  es un vector,  $\phi_h(z)$  una función base, como por ejemplo  $\phi_h(z) = \exp -\frac{(z-c_h)^2}{2r^2}$  con  $c_h \in \mathbb{R}^n$  un punto fijo en el espacio, para  $h = 1, \dots, H$ .

Definimos el vector  $\mathbf{y}$  como el vector de valores de  $y(\mathbf{x}, \beta)$  en  $n$  puntos  $\{\mathbf{x}^{(n)}\}$ . Es decir, cada entrada del vector  $\mathbf{y}$  es  $y_j = \sum_{h=1}^H \beta_h \phi_h(\mathbf{x}^{(j)})$  para  $j = 1, \dots, n$ .

Desde el punto de vista bayesiano  $\beta = (\beta_1, \beta_2, \dots, \beta_H)^t$  es una variable aleatoria a la cual se le asignará distribución inicial normal con media cero y matriz de covarianzas  $\sigma_\beta^2 I$ . Entonces  $\mathbf{y}$  es también distribuida normal con media cero y matriz de covarianzas dado por

$$\begin{aligned} \Sigma_{i,j} &= E(y_i y_j) \\ &= E\left(\sum_{h_1=1}^H \beta_{h_1} \phi_{h_1}(\mathbf{x}^{(i)}) \sum_{h_2=1}^H \beta_{h_2} \phi_{h_2}(\mathbf{x}^{(j)})\right) \\ &= \sigma_\beta^2 \sum_{h=1}^H \phi_h(\mathbf{x}^{(i)}) \phi_h(\mathbf{x}^{(j)}). \end{aligned}$$

Para el caso en donde las variables explicativas son unidimensionales, asumiendo que las funciones  $\phi_h$  son uniformemente espaciadas y tomando el límite cuando  $H \rightarrow \infty$ ; la expresión para  $\Sigma_{i,j}$  se simplifica,

$$\Sigma_{i,j} = \sqrt{\pi r^2} S \exp\left(\frac{(\mathbf{x}^{(i)} - \mathbf{x}^{(j)})^2}{4r^2}\right) \quad (2.29)$$

donde  $S$  es un valor de cambio de escala para  $\sigma_\beta^2$ .

### Estimación de parámetros para la matriz de covarianzas

Para hacer inferencia, predicciones, simulaciones, etc., con base en procesos gaussianos es necesaria una función de covarianzas dada. Sin embargo, esta función no siempre está completamente especificada, por lo que se ha hecho investigación que permita inferirla. En Williams (1998), Mackay (1998) y Williams y Rasmussen (1996) se pueden encontrar algunos desarrollos.

En los artículos antes mencionados se escoge una familia paramétrica de funciones de covarianzas. Se denota  $\theta$  como el vector de parámetros de dicha función, se pretende estimarlos ya sea por máxima verosimilitud o usando técnicas Bayesianas.

Al ser los datos tratados como un proceso gaussiano, se sabe que la distribución conjunta de los mismos es normal multivariada. Sin perder generalidad se supone con media cero y matriz de covarianza denotada por  $\Sigma$ , a quien se le asignará una familia paramétrica de funciones de covarianza.

Entonces la log-verosimilitud de los datos tiene una forma conocida expresada como

$$\log(P(\mathbf{y} | \Sigma)) = -\frac{1}{2} \log |\Sigma| - \frac{1}{2} \mathbf{y}^t \Sigma^{-1} \mathbf{y} - \frac{n}{2} \log 2\pi. \quad (2.30)$$

De igual manera es fácil expresar las derivadas parciales de la log-verosimilitud con respecto a cada uno de los parámetros de la función de covarianza. Dichas derivadas se expresan

$$\frac{\partial \log(P(\mathbf{y} | \Sigma))}{\partial \theta_i} = \frac{1}{2} \text{tr}(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_i}) + \frac{1}{2} \mathbf{y}^t \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_i} \Sigma^{-1} \mathbf{y}. \quad (2.31)$$

El obtener dichas derivadas es costoso por el número de cálculos que hay que realizar, sin embargo son necesarias para obtener los estimadores máximos verosímiles de  $\Sigma$ .

También se puede hacer inferencia desde un punto de vista bayesiano, esto es definiendo una distribución inicial sobre  $\theta$  y obteniendo la distribución final considerando los datos observados que se denotan por  $D$ . Para hacer predicción para un nuevo punto  $x_*$  se puede utilizar la ecuación:

$$P(y_* | D) = \int P(y_* | \theta, D) P(\theta | D) d\theta. \quad (2.32)$$

No obstante resolver la anterior integral no se puede realizar analíticamente, es por eso que se han implementado algunos métodos numéricos, sobre todo cuando el número de parámetros es pequeño. Cuando el número de parámetros a estimar es grande, resolver este problema es complicado. Para ello se han utilizado métodos de Monte Carlo con cadenas de Markov (MCMC).



# Capítulo 3

## Un modelo de regresión semi-paramétrico

En este capítulo se desarrolla el modelo de regresión semi-paramétrico haciendo uso del proceso Gaussiano, comenzando con un planteamiento sobre la relación entre las variables explicativas y la variable respuesta. Finalmente, se desarrollan casos límite de este modelo que llevan a problemas ya conocidos como son el caso de interpolación y el de regresión lineal paramétrica.

### 3.1. Motivación

Suponiendo un experimento de  $n$  observaciones, sea  $\{x_i, y_i\}_{i=1}^n$  una muestra de tamaño  $n$  de la variable explicativa y la variable respuesta. Entonces el modelo de regresión se escribe como

$$y_i = f(x_i) + \epsilon_i. \quad (3.1)$$

La función  $f(\cdot)$  denota la relación entre la variable explicativa y la variable respuesta. Las variables aleatorias  $\{\epsilon_i\}_{i=1}^n$  representan errores en el modelo. En el caso general  $x_i \in \mathbb{R}^m$ .

Se supondrá que  $\{\epsilon_i\}_{i=1}^n$  son independientes e idénticamente distribuidas. Al ser ésta la parte aleatoria del modelo, se le asigna una distribución, la cual se expresa en

$$[\epsilon_i \mid \sigma^2] \sim N(0, \sigma^2). \quad (3.2)$$

Un supuesto más que se observa en dicha distribución es la misma varianza para cada observación, lo que se llama homoscedasticidad.

La función  $f(\cdot)$  es determinista pero desconocida, por lo que se trata de aproximar por otra función propuesta por el experimentador. En términos generales podría expresarse como  $f(x) = h(x)^t \beta + \delta(x)$ , donde  $h(x) : \mathbb{R}^m \rightarrow \mathbb{R}^q$  y  $\beta = (\beta_1, \dots, \beta_k)^t$ . De esta expresión se observa un nuevo error, el cual se denotará como  $\delta_i = \delta(x_i)$ . Este error corresponde al error de aproximación, que al ser desconocido se ve como una variable aleatoria y por ende se le asigna una distribución. En este trabajo se supondrá que

$$[\delta_i \mid \rho^2] \sim N(0, \rho^2). \quad (3.3)$$

Notemos que existen dos tipos de errores dentro del modelo: los  $\{\epsilon_i\}$  que representan la variabilidad de los datos y los  $\{\delta_i\}$  que representan el error “determinista” pero desconocido. Estos errores son deterministas en el sentido de que  $\delta_i = \delta_j$  si  $x_i = x_j$ . Esto no ocurre en el caso de los errores  $\{\epsilon_i\}$ , debido a que si se muestrea dos veces en un mismo punto  $x$ , el valor de la variable respuesta en ambas observaciones puede ser distinto.

Entre más cercanas entre sí sean las observaciones  $\{x_i\}$ , existirá una mayor correlación entre los valores correspondientes de la función  $f$ . Es decir, debe existir una función de correlación entre dichas variables, la cual se ve reflejada en los errores  $\{\delta_i\}_{i=1}^n$

$$\begin{aligned} \text{Cor}(f(x_i), f(x_j)) &= \frac{\text{Cov}(\delta_i, \delta_j)}{\rho^2} \\ &= v_0(x_i, x_j). \end{aligned} \quad (3.4)$$

La función  $v_0(\cdot, \cdot)$  debe ser tal que refleje una mayor correlación para  $x_i$ 's cercanas y una correlación menor entre las lejanas. Esta función también es sugerida por el experimentador. Una propuesta ha sido:

$$v_0(x, x^*) = \exp\{-\lambda(x - x^*)^t V^{-1}(x - x^*)\}. \quad (3.5)$$

Notemos que  $\lambda$  puede considerarse un parámetro de suavizamiento, mientras que  $\rho^2$  determina la flexibilidad del modelo para tomar en cuenta desviaciones de  $f(\cdot)$  respecto a  $h(\cdot)^t \beta$ .

Con todo lo anterior  $\delta = (\delta_1, \dots, \delta_n)^t$  se puede ver como un proceso gaussiano con función de correlación dada por (3.4). Es decir

$$[\delta(\cdot) \mid \rho^2] \sim \mathcal{N}(0, \rho^2 v_0(\cdot, \cdot)). \quad (3.6)$$

Análogamente  $\epsilon = (\epsilon_1, \dots, \epsilon_n)^t$  se puede ver como un proceso gaussiano con una función de correlación especial, debido a que se supone que no existe correlación entre dichos errores

$$[\epsilon(\cdot) \mid \sigma^2] \sim \mathcal{N}(0, \sigma^2 I(\cdot, \cdot)), \quad (3.7)$$

donde

$$I(x, x^*) = \begin{cases} 1 & \text{si } x = x^*; \\ 0 & \text{si } x \neq x^* \end{cases}$$

Entonces el modelo (3.1) se puede reescribir como  $y_i = h(x_i)^t \beta + \delta_i + \epsilon_i$  y en términos de un proceso gaussiano se escribe como (3.8). Este modelo es semi-paramétrico, la parte paramétrica corresponde a la estimación del parámetro  $\beta$  y la parte no paramétrica a la estimación del error  $\delta$ . Finalmente, el modelo puede escribirse como

$$[y(\cdot) \mid \beta, \sigma^2, \rho^2] \sim \mathcal{N}(h(\cdot)^t \beta, \rho^2 v_0(\cdot, \cdot) + \sigma^2 I(\cdot, \cdot)). \quad (3.8)$$

El objetivo es encontrar un estimador para  $f(\cdot)$  así como predecir los valores de la respuesta  $y$  para distintos valores de  $x$ .

*Comentario:* Lo más común en regresión es usar una relación lineal entre las variables explicativas y la variable respuesta, conservando la linealidad en el parámetro. Pero en ocasiones la relación que existe entre dichas variables es bajo una curva no necesariamente lineal. Para este tipo de casos usar un polinomio para ajustar la relación es aceptable, lo cual se puede justificar través del Teorema de Stone-Weierstrass, Stone (1948). Una función continua se puede aproximar uniformemente con un polinomio en un intervalo cerrado.

Este caso particular fue explorado por Blight y Ott (1975), cuyos resultados son utilizados en este capítulo.

En el caso particular del modelo de regresión polinomial, la función  $h(x)$  puede ser escrita como:

$$h(x) = \begin{pmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^k \end{pmatrix},$$

donde  $k = q - 1$ .

## 3.2. Análisis del modelo

En esta sección analizaremos el modelo de regresión (3.1) suponiendo que  $\rho^2$ ,  $\lambda$  y  $\sigma^2$  son conocidas. Posteriormente, consideraremos el caso en el que  $\rho^2$  es desconocida. En ambos casos será posible obtener expresiones cerradas para las correspondientes distribuciones finales y predictivas. Al final de esta sección discutiremos la situación en la que  $\lambda$  y  $\sigma^2$  son desconocidas.

El modelo de regresión (3.1) se puede reescribir como:

$$[y_i | f(x_i)] \sim N(f(x_i), \sigma^2 I) \quad i = 1, \dots, n, \quad (3.9)$$

donde  $\sigma^2$  se supondrá conocida.

En vista de la discusión de la sección anterior, supondremos la siguiente distribución inicial para  $f$

$$f(\cdot) \sim \mathcal{N}(m_0(\cdot), \rho^2 v_0(\cdot, \cdot)), \quad (3.10)$$

donde  $v_0(\cdot, \cdot)$  es la función de correlación (3.5) y, por el momento,  $m_0(\cdot)$  es una función conocida que describe la forma que el experimentador espera tenga la función  $f(\cdot)$ .

La distribución final de  $f(\cdot)$  es en este caso:

$$[f(\cdot) | y_1, \dots, y_n] \sim \mathcal{N}(m_1(\cdot), \rho^2 v_1(\cdot, \cdot)),$$



donde

$$\begin{aligned} m_1(x) &= m_0(x) + \rho^2 T_0(x)^t [\rho^2 A + \sigma^2 I]^{-1} (y - m_0(x)) \\ &= m_0(x) + T_0(x)^t [A + wI]^{-1} (y - m_0(x)), \\ v_1(x, x^*) &= v_0(x, x^*) - \rho^2 T_0(x)^t [\rho^2 A + \sigma^2 I]^{-1} T_0(x^*) \\ &= v_0(x, x^*) - T_0(x)^t [A + wI]^{-1} T_0(x^*); \end{aligned}$$

con

$$A = \begin{pmatrix} v_0(x_1, x_1) & \cdots & v_0(x_1, x_n) \\ \vdots & \ddots & \vdots \\ v_0(x_n, x_1) & \cdots & v_0(x_n, x_n) \end{pmatrix}$$

y

$$T_0(x) = \begin{pmatrix} v_0(x, x_1) \\ \vdots \\ v_0(x, x_n) \end{pmatrix},$$

tomando  $w = \frac{\sigma^2}{\rho^2}$  (ver O'Hagan, 1978).

Con base en lo anterior la distribución predictiva de  $y(\cdot)$  se expresa como

$$[y(\cdot) \mid y_1, \dots, y_n] \sim \mathcal{N}(m_1(\cdot), \rho^2(w + v_1(\cdot, \cdot))), \quad (3.11)$$

Consideremos ahora una estructura jerárquica suponiendo que en realidad  $m_0(x) = h(x)^t \beta$  con  $\beta$  desconocida. En este caso (3.9) corresponde al primer nivel de la jerarquía (observaciones), mientras que (3.10) corresponde al segundo nivel (parámetros). Respecto al tercer nivel (hiperparámetros) para el parámetro  $\beta$  supondremos una distribución inicial normal. Para ello consideremos los siguientes dos casos.

### 3.2.1. Caso $p(\beta)$ dependiente de $\rho^2$ y ésta conocida

Una razón para suponer a  $\beta$  dependiente de  $\rho^2$  es que  $\beta$  se utiliza para aproximar  $f$  y la varianza del error de dicha aproximación es  $\rho^2$ . Este caso será de interés al considerar qué pasa cuando  $\sigma^2 \rightarrow 0$  (ver Sección 3.3.2). Por lo tanto es lógico que  $\beta$  debiera heredar dicha variabilidad. Entonces

$$[\beta \mid \rho^2] \sim N(b_0, \rho^2 B_0^{-1}). \quad (3.12)$$

La distribución final de  $\beta$  en este caso es

$$[\beta \mid y_1, \dots, y_n] \sim N(b_1, \rho^2 B_1^{-1}),$$

con

$$\begin{aligned} b_1 &= [H^t(A + wI)^{-1}H + B_0]^{-1}(H^t(A + wI)^{-1}y + B_0b_0) \\ B_1 &= H^t(A + wI)^{-1}H + B_0. \end{aligned}$$

y

$$H = \begin{pmatrix} h(x_1)^t \\ \vdots \\ h(x_n)^t \end{pmatrix},$$

(ver O'Hagan, 1992).

Es decir, en términos de la función  $m_0$  la distribución inicial es

$$m_0(\cdot) \sim \mathcal{N}(h(\cdot)^tb_0, \rho^2h(\cdot)^tB_0^{-1}h(\cdot))$$

y la final es

$$[m_0(\cdot) \mid y_1, \dots, y_n] \sim \mathcal{N}(h(\cdot)^tb_1, \rho^2h(\cdot)^tB_1^{-1}h(\cdot)).$$

Puesto que estamos en el caso donde  $f(x_i) = h(x_i)^t\beta + \delta(x_i)$ , la distribución inicial de  $f(\cdot)$  dado el parámetro  $\beta$  es

$$[f(\cdot) \mid \beta] \sim \mathcal{N}(h(\cdot)^t\beta, \rho^2v_0(\cdot, \cdot)) \quad (3.13)$$

y la distribución final es

$$[f(\cdot) \mid \beta, y_1, \dots, y_n] \sim \mathcal{N}(m_1(\cdot \mid \beta), \rho^2v_1(\cdot, \cdot)) \quad (3.14)$$

con

$$m_1(x \mid \beta) = h(x)^t\beta + T_0(x)^t[A + wI]^{-1}(y - H\beta)$$

Entonces la distribución final de  $f(\cdot)$ , la cual se obtiene al integrar (3.14) con respecto a  $\beta$  se expresa como

$$[f(\cdot) \mid y_1, \dots, y_n] \sim \mathcal{N}(m_1^*(\cdot), \rho^2v_1^*(\cdot, \cdot)), \quad (3.15)$$

donde

$$\begin{aligned} m_1^*(x) &= h(x)^tb_1 + T_0(x)^t[A + wI]^{-1}(y - Hb_1), \\ v_1^*(x, x^*) &= v_1(x, x^*) + h(x)^tB_1^{-1}h(x^*) - T_0(x)^t[A + wI]^{-1}HB_1^{-1}H^tT_0(x^*) \\ &= v_1(x, x^*) + \{h(x) - H^t[A + wI]^{-1}T_0(x)\}^tB_1^{-1}\{h(x^*) - H^t[A + wI]^{-1}T_0(x^*)\}. \end{aligned}$$

Finalmente, la distribución predictiva está dada por

$$[y(\cdot) \mid y_1, \dots, y_n] \sim \mathcal{N}(m_1^*(\cdot), \rho^2[w + v_1^*(\cdot, \cdot)]). \quad (3.16)$$

### 3.2.2. Caso $p(\beta)$ no dependiente de $\rho^2$ y ésta conocida

En este apartado se expresan los resultados bajo el supuesto de que la distribución inicial del parámetro  $\beta$  no depende de  $\rho^2$ . Este caso será de interés al considerar qué pasa cuando  $\rho^2 \rightarrow 0$  (ver Sección 3.3.1). Esta distribución inicial se expresa como:

$$\beta \sim N(b_0, B_0^{-1}). \quad (3.17)$$

Con base en lo anterior distribución final de  $\beta$  es

$$[\beta \mid y_1, \cdot, y_n] \sim N(b_1, B_1^{-1}),$$

con

$$\begin{aligned} b_1 &= [H^t(\rho^2 A + \sigma^2 I)^{-1}H + B_0]^{-1}(H^t(\rho^2 A + \sigma^2 I)^{-1}y + B_0 b_0) \\ B_1 &= H^t(\rho^2 A + \sigma^2 I)^{-1}H + B_0. \end{aligned} \quad (3.18)$$

(ver O'Hagan, 1992).

Escribiendo lo anterior en términos de la función  $m_0$  la distribución inicial es

$$m_0(\cdot) \sim \mathcal{N}(h(\cdot)^t b_0, h(\cdot)^t B_0^{-1} h(\cdot))$$

y la final es

$$[m_0(\cdot) \mid y_1, \dots, y_n] \sim \mathcal{N}(h(\cdot)^t b_1, h(\cdot)^t B_1^{-1} h(\cdot)).$$

La distribución inicial de  $f(\cdot)$  dado el parámetro  $\beta$  es

$$[f(\cdot) \mid \beta] \sim \mathcal{N}(h(\cdot)^t \beta, v_0(\cdot, \cdot))$$

y la distribución final es

$$[f(\cdot) \mid \beta, y_1, \dots, y_n] \sim \mathcal{N}(m_1(\cdot \mid \beta), v_1(\cdot, \cdot)). \quad (3.19)$$

con

$$m_1(x \mid \beta) = h(x)^t \beta + \rho^2 T_0(x)^t [\rho^2 A + \sigma^2 I]^{-1} (y - H\beta).$$

Entonces la distribución final de  $f(\cdot)$ , la cual se obtiene al integrar (3.19) con respecto a  $\beta$  se expresa como

$$[f(\cdot) \mid y_1, \dots, y_n] \sim \mathcal{N}(m_1^*(\cdot), v_1^*(\cdot, \cdot)), \quad (3.20)$$

donde

$$\begin{aligned} m_1^*(x) &= h(x)^t b_1 + \rho^2 T_0(x)^t [\rho^2 A + \sigma^2 I]^{-1} (y - H b_1), \\ v_1^*(x, x^*) &= \rho^2 v_1(x, x^*) + h(x)^t B_1^{-1} h(x^*) - \rho^4 T_0(x)^t [\rho^2 A + \sigma^2 I]^{-1} H B_1^{-1} H^t T_0(x^*) \\ &= \rho^2 v_1(x, x^*) + \{h(x) - \rho^2 H^t [\rho^2 A + \sigma^2 I]^{-1} T_0(x)\}^t B_1^{-1} \{h(x^*) - \rho^2 H^t [\rho^2 A + \sigma^2 I]^{-1} T_0(x^*)\}. \end{aligned} \quad (3.21)$$

Finalmente, la distribución predictiva está dada por

$$[y(\cdot) \mid y_1, \dots, y_n] \sim \mathcal{N}(m_1^*(\cdot), [\sigma^2 + v_1^*(\cdot, \cdot)]). \quad (3.22)$$

### 3.2.3. Caso $p(\beta)$ dependiente de $\rho^2$ y ésta desconocida

En esta sección retomaremos el caso discutido en la Sección 3.2.1 pero relajaremos el supuesto de que  $\rho^2$  es conocida. Específicamente, supondremos que  $\rho^2$  es desconocida y que se le asigna una distribución inicial conjugada de la forma

$$\rho^2 \sim s_0 \chi_{a_0}^{-2}. \quad (3.23)$$

Una vez establecida dicha distribución inicial, se actualiza la distribución inicial del parámetro  $\beta$ , con base también en los datos observados  $\mathbf{y} = (y_1, \dots, y_n)^t$ . Dicho resultado se observa en la siguiente ecuación

$$[\beta \mid \rho^2, y_1, \dots, y_n] \sim \mathcal{N}(b_1, \rho^2 B_1^{-1}). \quad (3.24)$$

Utilizado cálculos análogos a los de O'Hagan (1992) la distribución final del parámetro  $\rho^2$  es:

$$[\rho^2 \mid y_1, \dots, y_n] \sim s_1 \chi_{a_1}^{-2}, \quad (3.25)$$

donde  $a_1 = a_0 + n - q$  y

$$s_1 = s_0 + y^t \{ [A + wI]^{-1} - [A + wI]^{-1} (H^t [A + wI]^{-1} H)^{-1} [A + wI]^{-1} \} y.$$

Los cálculos bajo este supuesto son análogos a los establecidos en la Sección 3.2.1. Lo que restaría es integrar la expresión (3.16) con respecto a  $\rho^2$ , dicho resultado se expresa a continuación,

$$[y(\cdot) \mid y_1, \dots, y_n] \sim t_{a; m_1^*(\cdot), v_2(\cdot)}, \quad (3.26)$$

donde  $t_{a; m_1^*(\cdot), v_2(\cdot)}$  denota la distribución  $t$ -Student no estandarizada con  $a$  grados de libertad, media  $m_1^*(\cdot)$  y varianza  $v_2(\cdot)$ , expresadas como:

$$\begin{aligned} E(y(x) \mid y) &= m_1^*(x) \\ &= h(x)^t b_1 + T_0(x)^t [A + wI]^{-1} (y - Hb_1), \\ \text{Var}(y(x) \mid y) &= v_2(x) \\ &= \frac{s_1 [w + v_1^*(x, x)]}{a - 2}, \end{aligned}$$

donde

$$v_1^*(x, x^*) = v_1(x, x^*) + h(x)^t B_1^{-1} h(x^*) - T_0(x)^t [A + wI]^{-1} H B_1^{-1} H^t T_0(x^*).$$

Entonces el estimador de  $y(\cdot)$  corresponde al valor esperado de la distribución (3.26).

## 3.3. Casos límite

En esta sección se consideran dos casos de interés, por un lado se analiza el modelo cuando no existe error determinista y por el otro cuando usando este modelo se supone que no existe error aleatorio.

### 3.3.1. Regresión paramétrica ( $\rho^2 \rightarrow 0$ )

A continuación se explora el caso límite cuando el parámetro  $\rho^2$ , de la distribución inicial de la función  $f$ , tiende a cero (ver (3.10)). El hacer tender a cero la varianza de dicha variable aleatoria se interpreta que cada punto en el espacio  $x$ , tiene probabilidad uno de que al ser evaluado en la función desconocida  $f$  su valor sea  $m_0(x)$ . Por lo tanto la distribución predictiva de  $f(\cdot)$  al hacer tender  $\rho^2$  a cero de la expresión (3.11) es

$$[y(\cdot) \mid y_1, \dots, y_n] \sim \mathcal{N}(m_0(\cdot), \sigma^2).$$

Bajo una estructura jerárquica de  $m_0(x) = h(x)^t \beta$  donde a  $\beta$  se le asigna una distribución inicial (3.17), los resultados en este caso límite son análogos a los obtenidos en la regresión paramétrica usual.

La distribución predictiva final de  $y(\cdot)$ , obtenida al llevar al límite (3.22) y sus respectivas funciones asociadas (ver (3.18) y (3.21)), es entonces

$$[f(\cdot) \mid y_1, \dots, y_n] \sim \mathcal{N}(m_1^*(\cdot), \sigma^2 + v_1^*(\cdot, \cdot)),$$

donde

$$\begin{aligned} m_1^*(x) &= h(x)^t b_1, \\ v_1^*(x, x^*) &= h(x)^t B_1^{-1} h(x^*), \end{aligned}$$

con

$$\begin{aligned} b_1 &= [\sigma^{-2} H^t H + B_0]^{-1} (\sigma^{-2} H^t y + B_0 b_0) \quad y \\ B_1 &= \sigma^{-2} H^t H + B_0. \end{aligned}$$

Cabe señalar que estos resultados son los mismos que se obtienen en el problema de regresión paramétrica de la Sección 2.2.2.

### 3.3.2. Interpolación Bayesiana ( $\sigma^2 \rightarrow 0$ )

En esta sección se hace tender a cero el parámetro de dispersión ( $\sigma^2$ ) correspondiente a la parte aleatoria del modelo (3.1). Al hacer cero dicho parámetro se está forzando al modelo a ser exactamente  $f$ , es decir que para cada valor  $x$  en el espacio  $y(x) = f(x)$ . En este sentido no existe algún tipo de perturbación sobre el valor que pudiera tomar  $y(\cdot)$ , sin embargo al ser desconocida la función  $f(\cdot)$  hay un problema que en ocasiones suele resolverse con técnicas usuales de interpolación.

La interpolación es una técnica que consiste en, dada una serie de valores que toma una función (usualmente difícil de evaluar), usar una nueva función que se aproxime lo mejor posible a la función original. Esta nueva función es más simple que la original e idéntica a ella en los puntos observados. En general esto suele utilizarse para aproximar procesos complicados y hacer predicciones.

Entonces el problema de interpolación se puede ver como un problema de regresión, debido a que existe una función matemática que ajusta una variable respuesta en relación a una variable explicativa. Ese ajuste es exacto, en el sentido que no existe aleatoriedad en la muestra. El enfoque de esta sección se basa en el trabajo de O'Hagan (1992).

De nuevo, suponiendo un experimento de  $n$  observaciones, sea  $\{x_i, y_i\}_{i=1}^n$  una muestra de tamaño  $n$  de la variable explicativa y variable respuesta. Entonces el modelo se escribe como

$$y_i = f(x_i). \quad (3.27)$$

La función  $f(\cdot)$  denota la función que relaciona las variables explicativas y la variable respuesta, es decir, la función a interpolar.

La distribución inicial de  $f(\cdot)$  es la expresada en (3.10) suponiendo por el momento que  $m_0$  es una función descrita por el experimentador que se espera describa la forma de la función  $f(\cdot)$ .

Por consiguiente la distribución predictiva del modelo bajo los mencionados supuestos es

$$[y(\cdot) \mid y_1, \dots, y_n] \sim \mathcal{N}(m_1(\cdot), \rho^2 v_1(\cdot, \cdot)),$$

donde

$$\begin{aligned} m_1(x) &= m_0(x) + \rho^2 T_0(x)^t [\rho^2 A]^{-1} (y - m_0(x)) \\ &= m_0(x) + T_0(x)^t A^{-1} (y - m_0(x)), \\ v_1(x, x^*) &= v_0(x, x^*) - \rho^2 T_0(x)^t [\rho^2 A]^{-1} T_0(x^*) \\ &= v_0(x, x^*) - T_0(x)^t A^{-1} T_0(x^*), \end{aligned}$$

lo cual se deriva de (3.11) con  $\sigma^2 \rightarrow 0$ .

Ahora, retomando la idea de la Sección 3.2.1 pero bajo el supuesto que  $\sigma^2 \rightarrow 0$ , el objetivo principal de este problema es encontrar una función que se ajuste adecuadamente a los datos. Para esto es necesario encontrar un estimador para  $\beta$  y se busca el estimador de  $y(\cdot) = h(\cdot)^t \beta + \delta(\cdot)$ . Lo que se tiene ahora es un modelo paramétrico, en donde la distribución inicial para el parámetro  $\beta$  es dada por (3.12).

Dados los supuestos antes mencionados la distribución predictiva del modelo se deduce a partir de (3.16) y es

$$[y(\cdot) \mid y_1, \dots, y_n] \sim \mathcal{N}(m_1^*(\cdot), \rho^2 v_1^*(\cdot, \cdot)).$$

Donde

$$\begin{aligned} m_1^*(x) &= h(x)^t b_1 + T_0(x)^t A^{-1} (y - H b_1), \\ v_1^*(x, x^*) &= v_1(x, x^*) + h(x)^t B_1^{-1} h(x^*) - T_0(x)^t A^{-1} H B_1^{-1} H^t T_0(x^*) \\ &= v_1(x, x^*) + \{h(x) - H^t A^{-1} T_0(x)\}^t B_1^{-1} \{h(x^*) - H^t A^{-1} T_0(x^*)\}. \end{aligned}$$

Se supondrá ahora que  $\rho^2$  es desconocida para lo cual se retoma la Sección 3.2.3 asignándole una distribución inicial a dicho parámetro, la cual se expresa en (3.23) con  $s_0$  y  $a_0$  valores conocidos.

Finalmente la distribución predictiva del modelo es (3.26), pero con media y varianza:

$$\begin{aligned} E(y(x) | y) &= m_1^*(x) \\ &= h(x)^t b_1 + T_0(x)^t A^{-1}(y - Hb_1), \end{aligned} \quad (3.28)$$

$$\begin{aligned} \text{Var}(y(x) | y) &= v_2(x) \\ &= \frac{sv_1^*(x, x)}{a - 2}. \end{aligned} \quad (3.29)$$

donde

$$v_1^*(x, x^*) = v_1(x, x^*) + h(x)^t B_1^{-1} h(x^*) - T_0(x)^t A^{-1} H B_1^{-1} H^t T_0(x^*)$$

y

$$\begin{aligned} s &= s_0 + y^t \{A^{-1} - A^{-1}(H^t A^{-1} H)^{-1} A^{-1}\} y, \\ b_1 &= [H^t A^{-1} H + B_0]^{-1} (H^t A^{-1} y + B_0 b_0), \\ B_1 &= H^t A^{-1} H + B_0. \end{aligned}$$

En el caso particular,  $m^*(x_i) = f(x_i)$  y  $w^*(x_i, x_i) = 0$ . Como se puede observar, en este caso de interpolación se llega a la estimación de una función de  $x$ , no sólo a la estimación de un parámetro en específico. Con esto se puede obtener el valor estimado de la función evaluado en cualquier punto  $x$ , esto implica predicción.

### 3.4. Estimación de $w$ y $\lambda$

Un análisis bayesiano completo del modelo presentado en esta sección requiere de la especificación de distribuciones iniciales para  $w$  y  $\lambda$ . Sin embargo, la distribución final  $p(\beta, \rho^2, w, \lambda | \mathbf{y})$  que resulta es complicada debido a la cantidad de parámetros a estimar y por lo general es necesario utilizar técnicas de Monte Carlo con Cadenas de Markov para analizarla. Un inconveniente al utilizar técnicas de MCMC es la necesidad de evaluar en cada iteración el determinante de la matriz asociada a la matriz de correlación del proceso gaussiano así como su inversa. Lo anterior se requiere evaluar en un gran número de veces y con valores pequeños de  $\lambda$ , por lo que suele haber problemas numéricos en este tipo de implementaciones.

En este trabajo usaremos un enfoque Bayesiano empírico, estimando a  $w$  y a  $\lambda$  con base en su función de verosimilitud bajo el modelo discutido en la Sección 3.2.3 (ver la Sección 2.3.3). Notemos que esto resulta muy conveniente porque el análisis para  $\beta$  y  $\rho^2$  cuando se conoce  $w$  y  $\lambda$  se puede obtener en forma cerrada ya que  $p(\beta, \rho^2 | w, \lambda, \mathbf{y})$  es una distribución normal-gamma inversa.

# Capítulo 4

## Análisis numérico bayesiano

El modelo presentado en el capítulo anterior (Sección 3.2.3), puede servir para el caso particular de interpolación así como para otros problemas de análisis numérico tales como optimización y cuadratura. En este capítulo discutiremos el trabajo desarrollado por O'Hagan (1992).

### 4.1. Interpolación

A continuación dos ejemplos que ilustran el uso de modelos gaussianos para resolver un problema de interpolación. En ambos, de antemano es conocida la función a interpolar, pero se supondrá que es difícil o costosa de evaluar en puntos diferentes a los proporcionados en cada ejemplo.

Para la elaboración de los ejemplos se utilizó el lenguaje estadístico *R* (R Core Team, 2013) y los códigos del Apéndice C.

**Ejemplo 4.1.** La función a interpolar es

$$f(x) = \exp\left(\frac{x + x^2}{100}\right) \sin(x)$$

, en el intervalo  $[-5, 5]$ . En la Figura 4.1 se muestra la gráfica de dicha función. Con este ejemplo se pretende mostrar cómo se modifica el resultado y la varianza en cada interpolación, conforme el número de observaciones aumenta.

La función  $h(\cdot)$  que se utiliza en este ejemplo es un polinomio de primer grado, con esto se espera poner a prueba el modelo (pues se aproximará una función con gráfica 4.1 con una línea recta). Los valores para la distribuciones iniciales de  $\beta$  y  $\rho^2$  son:

$$\begin{aligned} b_0 &= (1, 1)^t, \\ B_0 &= \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix}, \\ s_0 &= 1, \\ a_0 &= 1. \end{aligned}$$



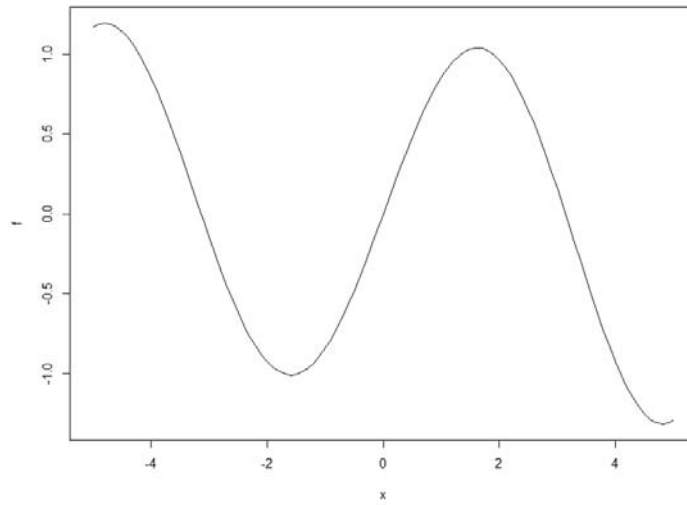


Figura 4.1: Gráfica de función a interpolar del Ejemplo 4.1 :  $f(x) = \exp\left(\frac{x^2}{100}\right) \sin(x)$

La matriz de correlación del proceso gaussiano está dada bajo la función de la ecuación (3.5), con  $\lambda = 0.69$  y  $V = 1$ .

Con los valores iniciales mencionados anteriormente y usando la ecuación (3.28) se obtienen estimadores para cincuenta puntos equidistantes, en el intervalo  $[-5, 5]$ . Con la ecuación (3.29) se calculan las bandas de confianza para los mismos puntos. Una vez que se obtiene la estimación para cada punto se le suma y resta dos veces la varianza final de ese punto para calcular las bandas.

En la Figura 4.2 se muestran dos estimaciones usando tres y cuatro observaciones, la línea negra es la estimación, las líneas azules son las bandas y la línea punteada es la función original. Las mencionadas observaciones son tomadas en el intervalo antes mencionado y equidistantes con respecto al eje horizontal. Para el caso de tres observaciones por ejemplo, se tomaron los puntos  $-5$ ,  $0$  y  $5$  con sus respectivas evaluaciones, es decir, las observaciones son :  $\{(-5, 1.1712), (0, 0), (5, -1.2944)\}$ . Para este caso no fue posible dibujar las bandas debido a que al ser muy grande la varianza computacionalmente se marca un error. Se observa un mejor ajuste para el caso de cuatro observaciones, sin embargo las bandas son considerablemente anchas.

En la Figura 4.2 también se puede apreciar que conforme crece el número de observaciones la incertidumbre es menor y el ajuste es mayor. Esto se aprecia en la amplitud de las bandas de confianza (líneas azules), las cuales se construyeron sumando y restando el cuantil al 97.5% de la distribución  $t$ -Student asociada (ver ecuaciones (3.28) y (3.29)). La Figura 4.3 es análoga a la anterior pero la escala en las cuatro figuras es la misma para que se pueda apreciar de otra forma el efecto del número de observaciones sobre el ajuste del modelo.

Ahora consideremos el caso de una dimensión mayor.

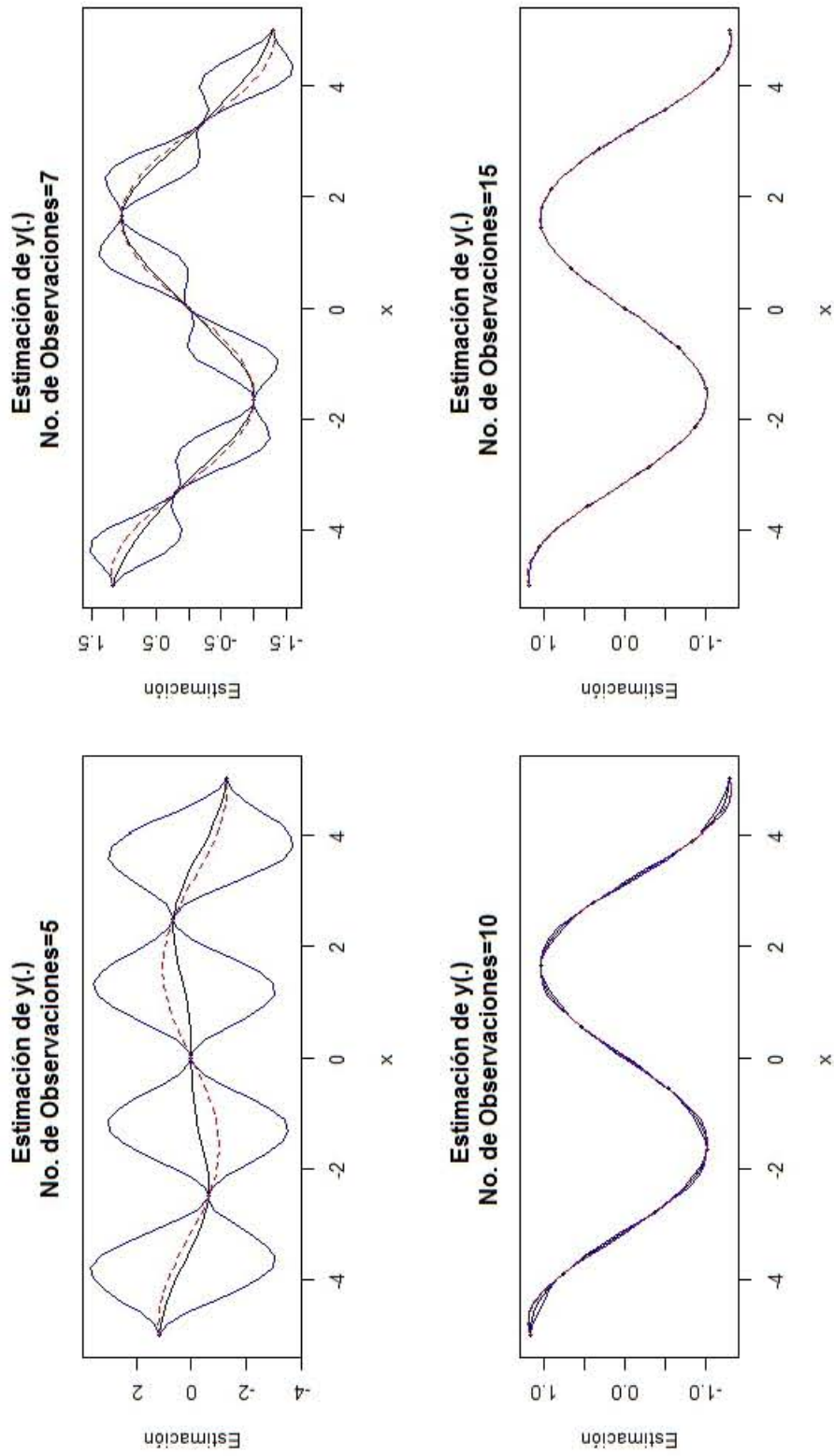


Figura 4.2: Interpolaciones y bandas de confianza.

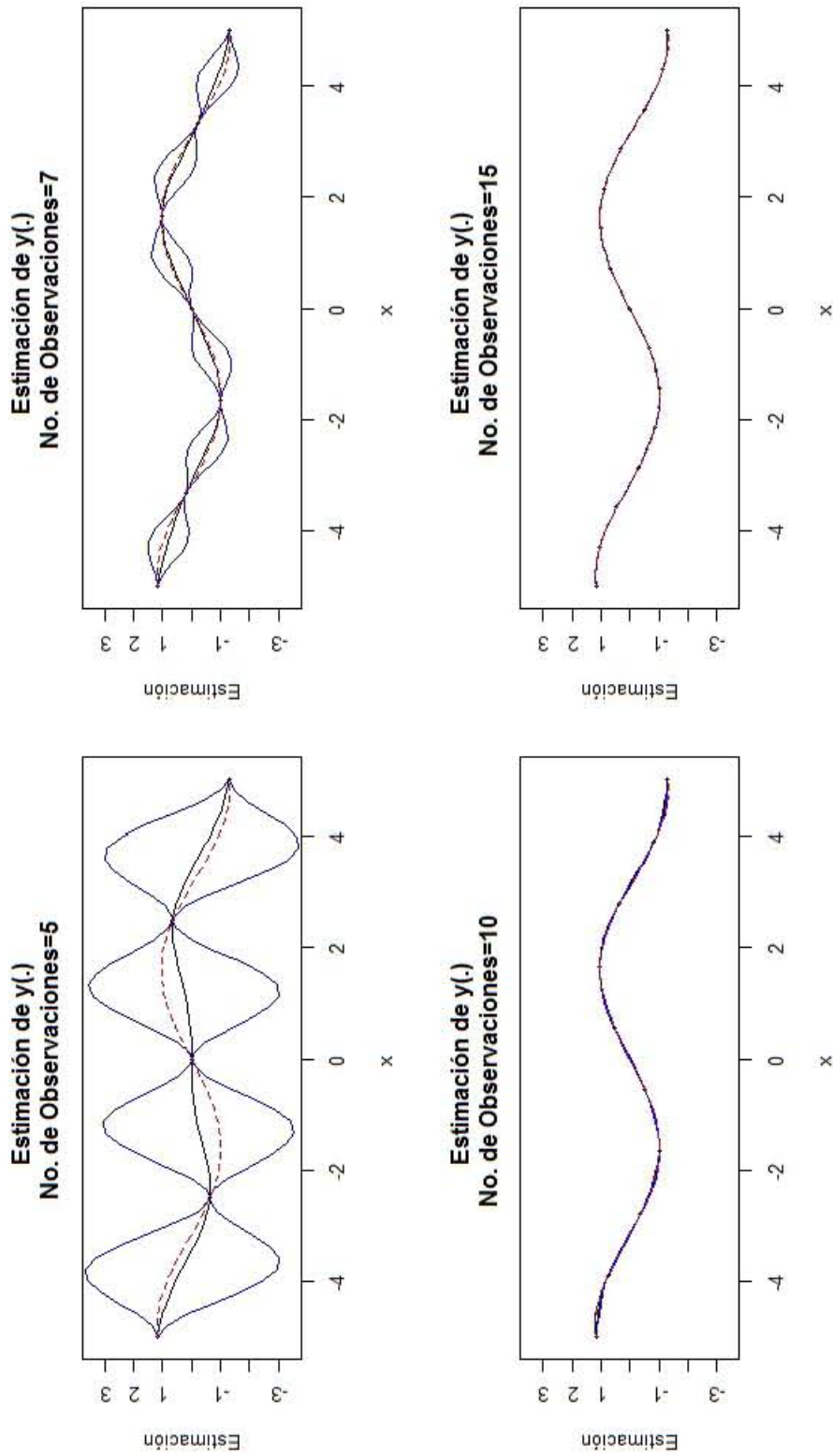


Figura 4.3: Interpolaciones y bandas de confianza usando la misma escala en el eje vertical.

**Ejemplo 4.2.** La verdadera función corresponde a la función de densidad de una mezcla de normales bivariadas. Dicha función se expresa como:

$$f(\mathbf{x}) = \frac{1}{2}(N_2(\mathbf{x} \mid \mathbf{m}_1, \mathbf{S}) + N_2(\mathbf{x} \mid \mathbf{m}_2, \mathbf{S})), \quad (4.1)$$

donde  $\mathbf{m}_1 = (2, -2)^t = -\mathbf{m}_2$  y  $\mathbf{S} = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}$ . La expresión  $N_2(\mathbf{x} \mid \mathbf{m}_1, \mathbf{S})$  denota la función de densidad de una distribución normal con media  $\mathbf{m}_1$  y matriz de covarianzas  $\mathbf{S}$ .

La función  $h(\cdot)$  que se utilizó en este ejemplo es la correspondiente a un plano en el espacio, es decir,  $h(\mathbf{x}) = (1, x_1, x_2)^t$ . La distribución inicial de los parámetros  $\beta$  y  $\rho^2$  están dadas por las ecuaciones (3.12) y (3.23) con hiperparámetros:

$$\begin{aligned} b_0 &= (1, 1, 1)^t, \\ B_0 &= \begin{pmatrix} 0.01 & 0 & 0 \\ 0 & 0.01 & 0 \\ 0 & 0 & 0.01 \end{pmatrix}, \\ s_0 &= 1, \\ a_0 &= 1. \end{aligned}$$

La matriz de correlación del proceso gaussiano está dado bajo la función de la ecuación (3.5), con  $\lambda = 0.8$  y  $V = I_2$ .

Las observaciones fueron tomadas del intervalo  $[-3, 3] \times [-3, 3]$  de tal manera que estuvieran en una retícula de treinta y seis puntos equidistantes. En la Figura 4.4 se muestran estas observaciones graficadas en el espacio. Se desea que con estas observaciones se obtenga la interpolación en el intervalo antes mencionado de tal manera que el ajuste a la verdadera función sea adecuado.

En la Figura 4.5 se muestran los resultados de la interpolación. Como se puede observar el ajuste es bueno, tanto las curvas de nivel como la gráficas son parecidas.

## 4.2. Cuadratura

La cuadratura a menudo se ve como un método numérico que trata de dar una aproximación a una integral definida dada. Esta aproximación es necesaria por varias razones, la principal sería la imposibilidad de realizar la integración de forma analítica. Es decir, integrales que requerirían de un gran conocimiento y manejo de matemática avanzada pueden ser resueltas de una manera más sencilla mediante métodos numéricos. Incluso existen funciones integrables pero cuya primitiva no puede ser calculada, siendo la integración numérica de vital importancia. La solución analítica de una integral nos arrojaría una solución exacta, mientras que la solución numérica nos daría una solución aproximada. El error de la aproximación, que depende del método que se utilice y de qué tan fino sea, puede llegar a ser tan pequeño que es posible obtener un resultado idéntico a la solución analítica en las primeras cifras decimales.

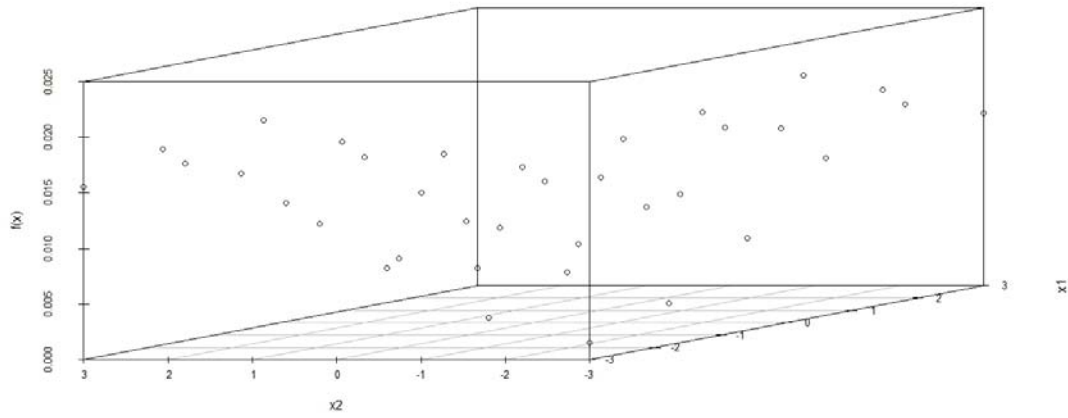


Figura 4.4: Observaciones para el Ejemplo 4.2

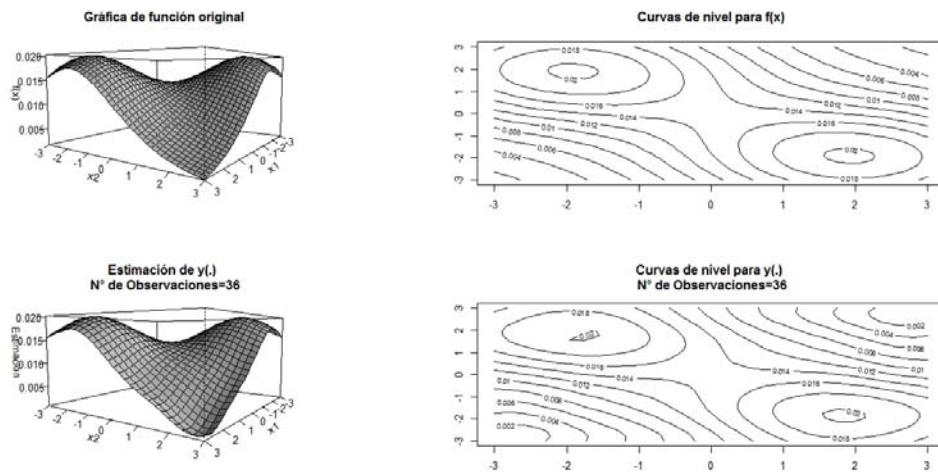


Figura 4.5: Interpolación y función original del Ejemplo 4.2

De nuevo, se puede dar una aproximación usando la teoría antes mencionada en el caso particular de  $\sigma^2 = 0$ . En este caso el problema es integrar la función  $f(\cdot)$  con respecto a una medida  $M(\cdot)$  sobre  $\mathcal{X}$ .

$$z = \int_{\mathcal{X}} f(x) dM(x).$$

Al ser  $z$  un valor desconocido se toma como una variable aleatoria cuya distribución final dados los valores  $y$  y los parámetros está dada por (4.2). La distribución final de los hiperparámetros corresponden de igual manera a las expresiones en (3.24) y (3.25).

$$[z \mid y, \beta, \rho^2] \sim \mathcal{N}(m_k, \rho^2 w_k), \quad (4.2)$$

donde

$$\begin{aligned} m_z &= \int_{\mathcal{X}} m(x) dM(x), \\ w_z &= \int_{\mathcal{X}^2} w(x, x^*) dM(x) dM(x^*). \end{aligned} \quad (4.3)$$

Finalmente, se actualiza la información para obtener la distribución final de  $z$ . El resultado de nuevo es una distribución  $t$  con  $a$  grados de libertad. La media y la varianza están dadas por las correspondientes a las distribuciones de las ecuaciones (3.24) para  $\beta$  y (3.25) para  $\rho^2$ .

$$[z \mid f] \sim t_{a_1}$$

$$m_z^* = \int_{\mathcal{X}} m^*(x) dM(x) = h^* t b + t^* A^{-1} (f - Hb), \quad (4.4)$$

donde

$$h^* = \int_{\mathcal{X}} h(x) dM(x), \quad t^* = \int_{\mathcal{X}} t(x) dM(x). \quad (4.5)$$

### 4.3. Optimización

En muchas áreas del conocimiento se habla del tema de optimización, entendido en general como la selección del mejor elemento (con respecto a algún criterio) de un conjunto de elementos disponibles. Este tema es muy importante en la práctica ya que siempre se busca aquella estrategia que, por ejemplo, maximicen las ventas, aquella decisión que minimice las pérdidas, la trayectoria en donde se hace el menor tiempo, etc. En términos un poco más matemáticos, la optimización incluye el descubrimiento de los “mejores valores” de alguna función objetivo dado un dominio definido, incluyendo una variedad de diferentes tipos de funciones objetivo y diferentes tipos de dominios. Dicha función puede ser explícita o no.

Sin embargo muchas veces no es sencillo encontrar esos valores que maximizan o minimizan una función debido a la complejidad de la misma. Es por eso que se utilizan métodos que ayuden a dar quizás no el valor exacto pero si una aproximación del máximo o mínimo. La mayoría de estos métodos utilizan una serie de iteraciones recurrentes, en donde se hace una gran cantidad de evaluaciones debido a que se obtienen nuevos puntos que oscilan alrededor del real.

De nuevo se tiene el problema del desconocimiento del valor exacto del punto en donde una función alcanza un valor mínimo o máximo. Es por eso que de nuevo se puede ver como una variable aleatoria.

Suponiendo que se tienen valores observados  $f(x_1), \dots, f(x_n)$ , se hace el análisis como en el caso de la interpolación. Es decir, se obtiene que la media de la distribución final expresada en (3.28), la cual es una función poco complicada o barata. De una función complicada, se llega a una función más sencilla de la cual es más fácil de obtener el valor óptimo.

# Capítulo 5

## Problemas Inversos

En este capítulo se habla sobre una aplicación más de los procesos gaussianos. Esta aplicación está relacionada con una variedad de problemas que existen en áreas como la física, geofísica, arqueología, etc. La idea es usar problemas de estas áreas y tratar de darles un enfoque estadístico basado en el proceso gaussiano.

### 5.1. Introducción

En varias áreas de las ciencias se trata de dar una descripción completa de un sistema, de tal manera que se pueda predecir el resultado de futuras mediciones. Cuando ocurre esto se le conoce como el *problema directo*. Por ejemplo, para problemas de caída libre desde cierta altura se puede conocer la velocidad con la que el objeto llega al suelo.

Pero no siempre se conoce del todo al sistema; es decir, existen algunos parámetros desconocidos dentro del mismo que es necesario determinar. A los problemas de este estilo se les llama *problemas inversos*. Para resolver éstos se espera utilizar mediciones conocidas para encontrar esos parámetros desconocidos del sistema. Por lo tanto, el problema inverso consiste en usar los resultados actuales de algunas medidas para inferir el valor de los parámetros que caracterizan el sistema.

Mientras que el problema directo tiene una única solución, por lo general el problema inverso no. Un ejemplo simple que se ha manejado es el producto de dos números: dados dos números encontrar su producto es un problema directo. Sin embargo, si dado un número se deseara conocer los valores que fueron multiplicados para dar ese resultado, ese sería un problema inverso. En este tipo de problemas se necesita entonces hacer explícita cualquier información inicial sobre los parámetros del modelo.

Otro ejemplo de problemas inversos esta en el diagnóstico de una enfermedad a partir de los síntomas o quizás reconstruir un suceso pasado a partir de las huellas del presente.

Las siguientes líneas dan una breve explicación y características de los problemas inversos, para mayor detalles se recomienda consultar Tarantola (2005).



## 5.2. El modelo teórico y el modelo empírico

En ocasiones el sistema puede escribirse en términos de una ecuación diferencial como la ecuación

$$F(x, x_0, y, y_0, y'; \beta) = 0. \quad (5.1)$$

La solución a dicha ecuación será denotada por  $y = f(x; \beta)$  con valores iniciales  $x_0$  y  $y_0$  tal que  $y_0 = f(x_0; \beta)$ .

Lo anterior corresponde al modelo teórico del sistema con el cual, si está completamente especificado, se puede resolver el problema directo.

Existen complicaciones debido a que no siempre es fácil determinar analíticamente cuál es la función  $f(x; \beta)$ . Por esta razón se suelen usar métodos numéricos para evaluarla en puntos determinados, los cuales pudieran ser pocos debido al costo que la evaluación conlleve. Esto último lleva a usar métodos de interpolación para poder encontrar aproximaciones de la función evaluada en algunos otros puntos.

Por otro lado, en muchas ocasiones la función  $f(\cdot, \beta)$  está especificada salvo por el parámetro  $\beta$  y se cuenta con una colección de observaciones de la forma  $\{(y_i, x_i)\}$ ,  $i = 1, \dots, n$  donde  $y_i = f(x_i) = f(x_i; \beta)$ . En este caso el problema inverso radica en inferir el valor de  $\beta$  con base en las observaciones y todo tipo de información acerca del parámetro. Con dichas observaciones se puede plantear un sistema de ecuaciones cuya solución resolvería el problema, es decir,  $y_i = g(\beta) = f(x_i; \beta)$ . A este sistema se le llamará modelo empírico del sistema físico.

Cuando no existen suficientes observaciones para resolverlo de manera única, se deben poner restricciones basadas en información adicional sobre el parámetro. Este proceso es conocido como regularización. En la literatura del tema se dice que un problema está bien definido, en el sentido de Hadamard, si las siguientes condiciones se cumplen:

1. El problema tiene solución (Existencia).
2. La solución del problema es única (Unicidad).
3. El problema depende continuamente de los datos (Estabilidad).

Un problema se dice mal planteado si alguna de las condiciones antes mencionadas no se satisfacen. En general la mayoría de los problemas inversos son mal planteados.

Los problemas inversos se formulan a menudo en espacios de dimensión infinita, con la limitación de un número finito de mediciones. Esto puede conducir a los problemas que son replanteados en forma discreta. En estos casos, los métodos de regularización pueden ser utilizados para introducir supuestos relativamente débiles en la solución y prevenir sobreajuste.

## 5.3. Otras complicaciones

Existen diferentes complicaciones en la práctica que no permiten describir con exactitud las características del modelo. Esto puede deberse a una incorrecta especificación de la función o a errores en las mediciones.

Habrán algunos casos en los que la ecuación diferencial que realmente refleja todas las características del sistema no sea (5.1) sino otra, digamos  $\tilde{F}(x, x_0, \tilde{y}, \tilde{y}_0, \tilde{y}'; \tilde{\beta}) = 0$ , cuya solución denotamos por  $\tilde{y} = \tilde{f}(x; \tilde{\beta})$  con valores iniciales  $x_0$  y  $\tilde{y}_0$  tal que  $\tilde{y}_0 = \tilde{f}(x_0; \tilde{\beta})$ .

Entonces la función  $f(x; \beta)$  es sólo una aproximación. En la práctica suele modelarse el error con una función  $\delta(\cdot)$  que aparece de manera aditiva en el modelo. El modelo verdadero se puede suponer entonces como  $\tilde{f}(x; \beta) = f(x; \beta) + \delta(x)$ , donde  $\delta(\cdot)$  es una función desconocida pero de cierto modo determinista, debido a que  $\delta(x_i) = \delta(x_j)$  si  $x_i = x_j$  con  $i \neq j$ .

Ahora, si se tiene una colección de observaciones  $\{(\tilde{y}_i, x)\}$ ,  $i = 1, \dots, n$  tales que  $\tilde{y}_i = \tilde{f}(x_i; \beta) = f(x_i; \beta) + \delta(x_i)$ , se plantea el sistema de ecuaciones que corresponde al modelo empírico. En este caso las observaciones no necesariamente serán compatibles con dicho sistema para algún  $\beta$ , lo que trae como consecuencia que no haya solución. Para mitigar esta complicación, se suele usar el criterio de mínimos cuadrados. Éste consiste en minimizar  $\|\mathbf{y} - \tilde{\mathbf{y}}\|^2$ , donde  $\tilde{\mathbf{y}} = (\tilde{y}_1, \dots, \tilde{y}_n)$  es el vector de respuestas esperadas bajo  $\tilde{f}(x, \beta)$  y  $\mathbf{y} = (y_1, \dots, y_n)$  el de respuestas postuladas bajo  $f(x, \beta)$ .

Por otra parte, dentro de la especificación del modelo pueden existir errores relacionados con la medición de las observaciones. Estos errores pueden darse tanto en las observaciones de salida como en las de entrada. Cuando la imprecisión de las mediciones se da en las observaciones de salida del sistema, el error aleatorio se representa por  $\epsilon$  y el modelo se escribe como

$$\begin{aligned}\tilde{y}_i &= y_i + \epsilon_i \\ &= f(x_i, \beta) + \epsilon_i; \quad i = 1, \dots, n.\end{aligned}\tag{5.2}$$

En el caso donde los errores son en las variables de entrada del sistema, el error se denota por  $e$  y se considera de igual forma aleatorio. La representación es

$$\begin{aligned}y_i &= f(x_i; \beta) \\ \tilde{x}_i &= x_i + e_i; \quad i = 1, \dots, n.\end{aligned}\tag{5.3}$$

Este tipo de complicaciones suelen darse simultáneamente, por lo tanto la forma de resolver el problema se complica pues se deben considerar todas las fuentes de incertidumbre.

## 5.4. Propagación de la incertidumbre

Es importante mencionar que existe una afectación de las complicaciones antes mencionadas sobre los resultados del modelo. Esto debido a que, si llega a existir incertidumbre en cualquiera de los componentes del sistema, ésta se propagaría al resultado.

Esta propagación se da tanto en el problema directo como en el inverso. En el directo se da cuando existen errores en las mediciones de entrada. En el inverso, se da sobre todo cuando los errores están en las mediciones de las observaciones de salida.

## 5.5. Ejemplo

En esta sección se aplica el modelo semiparámetro basado en el proceso gaussiano en un problema termoquímico en el que se busca describir la relación entre la temperatura y la presión del vapor de sustancias puras.

### 5.5.1. Ecuación de Antoine

Para describir dicha relación se ha utilizado la ecuación de Antoine (Antoine, 1996), la cual se deriva de la relación de Clausius-Clapeyron. Ésta es una manera de caracterizar una transición de fase de primer orden que tiene lugar en un sistema monocomponente.

La ecuación de Antoine relaciona empíricamente el logaritmo de la presión del vapor ( $p$  generalmente en mmHg) con una función que depende de la temperatura ( $T$ , generalmente en grados Kelvin) de la sustancia

$$\log p = A - \frac{B}{C + T}. \quad (5.4)$$

Los parámetros  $A$ ,  $B$  y  $C$  son desconocidos pero han sido estimados; de hecho estas estimaciones se pueden encontrar en libros para las sustancias más usadas. Por ejemplo, se ha sabido que para el agua ( $H_2O$ ) se necesitan dos conjuntos de los parámetros para describir la relación de la temperatura y presión del vapor del agua en todo el rango requerido de la temperatura. Esto se muestran en el Cuadro 5.1, como se puede observar, se necesitan dos conjuntos de parámetros  $A$ ,  $B$  y  $C$  uno para cada intervalo de la temperatura. Este mismo fenómeno ocurre con otras sustancias puras.

T mín °C	T máx °C	A	B	C
1	100	8.07131	1730.63	233.426
99	374	8.14019	1810.94	244.485

Cuadro 5.1: Valores empíricos de la ecuación de Antoine para el agua

En general, la ecuación de Antoine no es suficientemente flexible y no se puede utilizar para describir con suficiente precisión toda la curva en función de la temperatura, razón por lo cual usualmente se ajusta por pedazos. Es decir, se utilizan varios conjuntos de parámetros para una sola sustancia. Como en el caso del agua, hay parámetros  $A$ ,  $B$  y  $C$  para describir la curva de presión del vapor hasta el punto de ebullición normal y el segundo conjunto para describir desde el punto de ebullición normal a un punto crítico. Además, se puede apreciar en el Cuadro 5.1 que en el intervalo de temperatura de  $[99, 100]$  grados centígrados hay problemas pues se podría utilizar ambos conjuntos de parámetros.

Existe una simplificación de la ecuación de Antoine. Cuando se supone  $C = 0$  entonces la ecuación describe una relación lineal entre el logaritmo de la presión y el valor inverso de la temperatura. Esta nueva relación se conoce como ecuación de August. Por otro lado, también se ha utilizado una extensión a la ecuación, pero son necesarios más parámetros

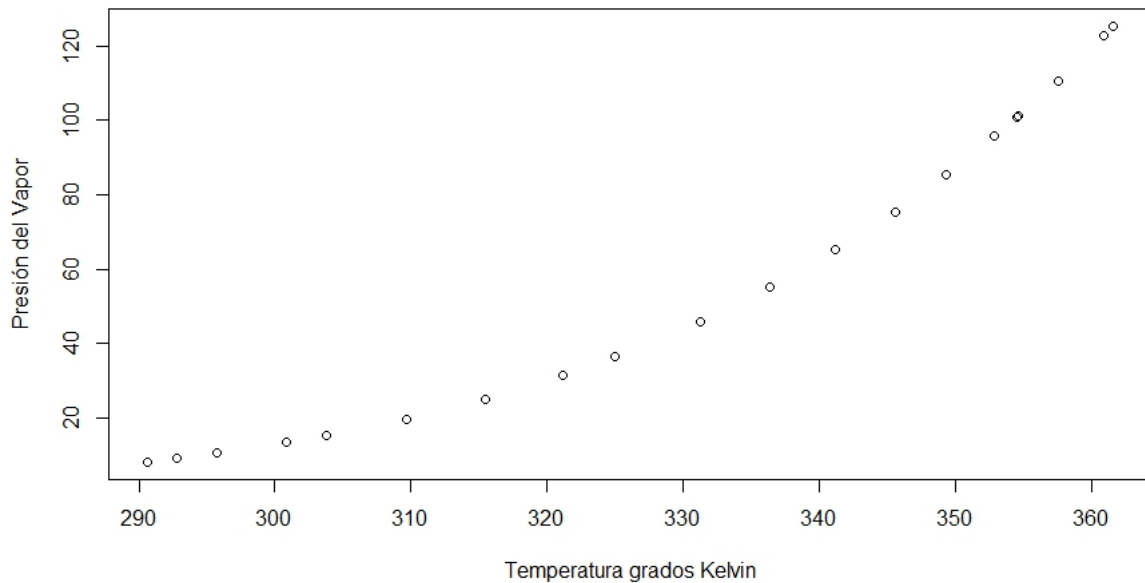


Figura 5.1: Mediciones observadas de temperatura y presión durante un experimento.

$$p = \exp\left\{A + \frac{B}{C + T} + D \cdot T + E \cdot T^2 + F \cdot \log(T)\right\}. \quad (5.5)$$

En la práctica, los parámetros se estiman por medio de método de mínimos cuadrados no lineales. Los datos se obtienen de experimentos diseñados para medir la presión del vapor bajo diferentes temperaturas. Desde el punto de vista estadístico éste es un problema que puede ser resuelto utilizando la técnica de regresión no lineal. Sin embargo, la selección de un modelo paramétrico no lineal adecuado no es una tarea sencilla. Tampoco lo es la evaluación de la incertidumbre basada en dicho modelo. Es por esta razón que proponemos el uso del modelo de regresión semiparamétrico de este trabajo para dar la una estimación predictiva de la presión dada la temperatura.

### 5.5.2. Aplicación del modelo

Se realizó un experimento y se obtuvieron veinte observaciones de la temperatura y presión del vapor del acetonitrilo ( Ewing & Sanchez, 2004 ). En términos del modelo de este trabajo se tiene una muestra de tamaño 20 donde  $\{x_i\}_{i=1}^{20}$  corresponden a las mediciones de la temperatura y  $\{y_i\}_{i=1}^{20}$  corresponden a las mediciones de la presión del vapor. En la Figura 5.1 se muestran los datos observados.

Para estimar una curva que se ajuste a los datos se hace el supuesto de que  $\{y_i\}_{i=1}^{20}$  tiene una distribución inicial como en la ecuación (3.8) bajo el esquema descrito en la Sección 3.2.3. Es decir, se hace el supuesto de la existencia de un error aleatorio en las observaciones,

así como un error determinista y por lo tanto la distribución predictiva final es (3.26). Con esta distribución se hace la estimación de la curva.

En un principio se ajustó el modelo utilizando polinomios de grado cero, uno y dos. Conforme se disminuye el grado del polinomio se fuerza al proceso gaussiano a capturar la variabilidad del ajuste. La función  $h(\cdot)$  que se reporta en este ejemplo es un polinomio de primer grado debido a que con éste se obtuvieron mejores resultados en el ajuste. Se utilizaron hiperparámetros para las distribuciones iniciales de  $\beta$  y  $\rho^2$  de tal manera que las distribuciones fueran no informativas. Estos valores son:

$$\begin{aligned} b_0 &= (1, 1)^t, \\ B_0 &= \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix}, \\ s_0 &= 1, \\ a_0 &= 1. \end{aligned}$$

Para escoger los valores de  $w$  y  $\lambda$  se utiliza un enfoque Bayesiano empírico el cual consiste en obtener los valores que maximicen la log-verosimilitud de los datos dados los parámetros

$$\log(P(\mathbf{y} | \Sigma)) = -\frac{1}{2} \log |\Sigma| - \frac{1}{2} (\mathbf{y} - \mu)^t \Sigma^{-1} (\mathbf{y} - \mu) - \frac{n}{2} \log 2\pi. \quad (5.6)$$

Con base en este modelo,  $\Sigma = \hat{\rho}^2 [wI_{20} + A_1^\lambda]$  y la media  $\mu$  es  $H\hat{\beta}$  con

$$A_1 = \begin{pmatrix} \exp\{-(x_1 - x_1)^t(x_1 - x_1)\} & \cdots & \exp\{-(x_1 - x_{20})^t(x_1 - x_{20})\} \\ \vdots & \ddots & \vdots \\ \exp\{-(x_{20} - x_1)^t(x_{20} - x_1)\} & \cdots & \exp\{-(x_{20} - x_{20})^t(x_{20} - x_{20})\} \end{pmatrix}. \quad (5.7)$$

Dado lo anterior es necesario contar con valores  $\hat{\rho}^2$  y  $\hat{\beta}$ ; es por esto que se hace un primer análisis del modelo dando valores iniciales de  $w = 0.1$  y  $\lambda = 0.1$ . Al correr el modelo se obtienen estimaciones preliminares para  $\rho^2$  y  $\beta$  con base en las distribuciones (3.25) y (3.24) respectivamente. Éstas son utilizadas para maximizar la función (5.6) con respecto a  $w$  y  $\lambda$ . Sin embargo para el caso de este experimento se observa que al momento de maximizar la log-verosimilitud existen problemas numéricos al invertir y obtener el determinante de  $A_1$ . Es por ello que se buscó sumar una penalización a la función (5.6) para hacerla mas estable. La función que se maximizó en este ejemplo es

$$\begin{aligned} \log(P(\mathbf{y} | \Sigma)) &= -0.04 \left( \frac{1}{2} \log |\Sigma| - \frac{1}{2} (\mathbf{y} - \mu)^t \Sigma^{-1} (\mathbf{y} - \mu) - \frac{n}{2} \log 2\pi \right) \\ &\quad - 0.073 \frac{3.70^{1.37}}{\Gamma(1.37)} w^{0.37} e^{-3.70w} \\ &\quad - 0.073 \frac{3.70^{1.37}}{\Gamma(1.37)} \lambda^{0.37} e^{-3.70\lambda}. \end{aligned} \quad (5.8)$$

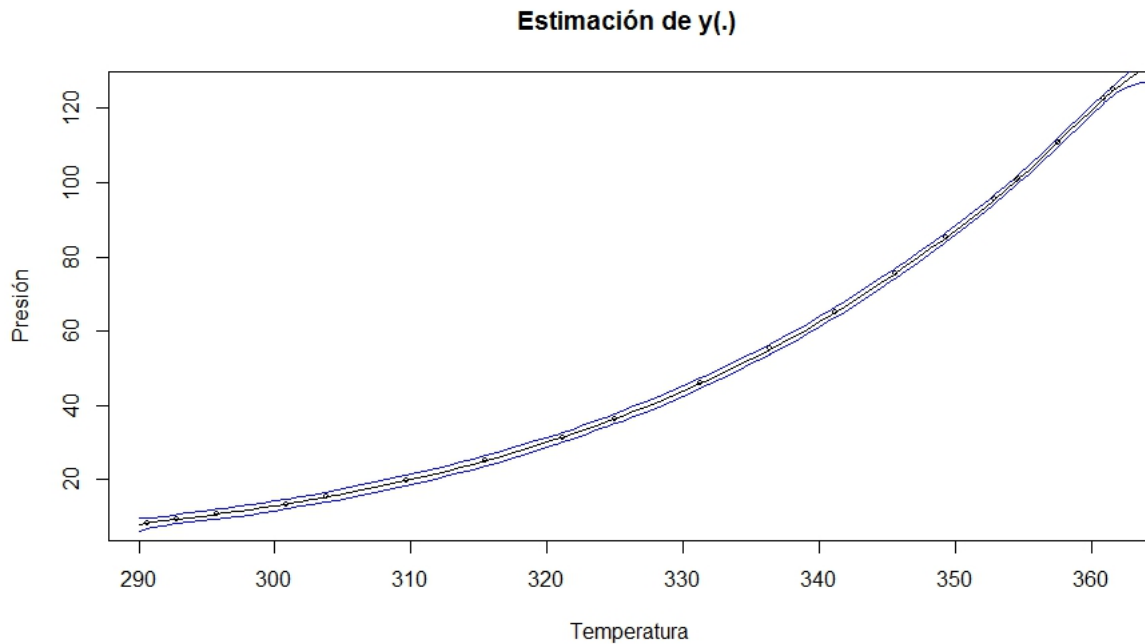


Figura 5.2: Mediciones observadas, el ajuste del modelo y bandas de confianza.

Se puede observar que los sumandos que se agregaron corresponden al logaritmo de una densidad gamma de moda 0.1 y varianza 0.1. Esto se puede ver como añadir una distribución inicial gamma para  $w$  y otra para  $\lambda$ , procurando que sean no informativas. Cada sumando de la función tiene un peso, el cual sirve para asignar un peso a la información que aporta cada sumando. Haciendo este tipo de cálculos la log-verosimilitud pudo ser maximizada sin problemas numéricos ni de inestabilidad para valores pequeños.

Una vez calculados los valores máximos de  $w$  y  $\lambda$  para la función (5.8) se usan para obtener nuevos valores de  $\hat{\rho}^2$  y  $\hat{\beta}$ . Con estas nuevas estimaciones se maximiza de nuevo la función (5.8) y así se encuentran nuevos valores de  $w$  y  $\lambda$ . Este algoritmo se detiene en el momento que los valores máximos en cada iteración se diferencian en 0.001 o menos.

En la Figura 5.2 se puede apreciar el ajuste del modelo con  $w = 0.00409$  y  $\lambda = 0.00663$  y bandas de confianza (líneas azules).

Con las observaciones de este experimento, los valores de la ecuación de Antoine (que ajusta los logaritmos de las presiones de vapor a diferentes temperaturas del acetonitrilo) por el método de mínimos cuadrados no lineales son  $A = 14.7325$ ,  $B = -3267.6599$  y  $C = -31.65307$ . En la Figura 5.3 se aprecia este ajuste en comparación con el ajuste obtenido con el modelo de este trabajo. Para poder ser comparadas se aplicó la transformación logarítmica a la estimación que se muestra en la Figura 5.2 debido a que la ecuación de Antoine ajusta el logaritmo de la presión.

Como se puede apreciar, en este caso las estimaciones son muy similares. Sin embargo, para otras sustancias menos conocidas la ecuación de Antoine podría no ser apropiada y se

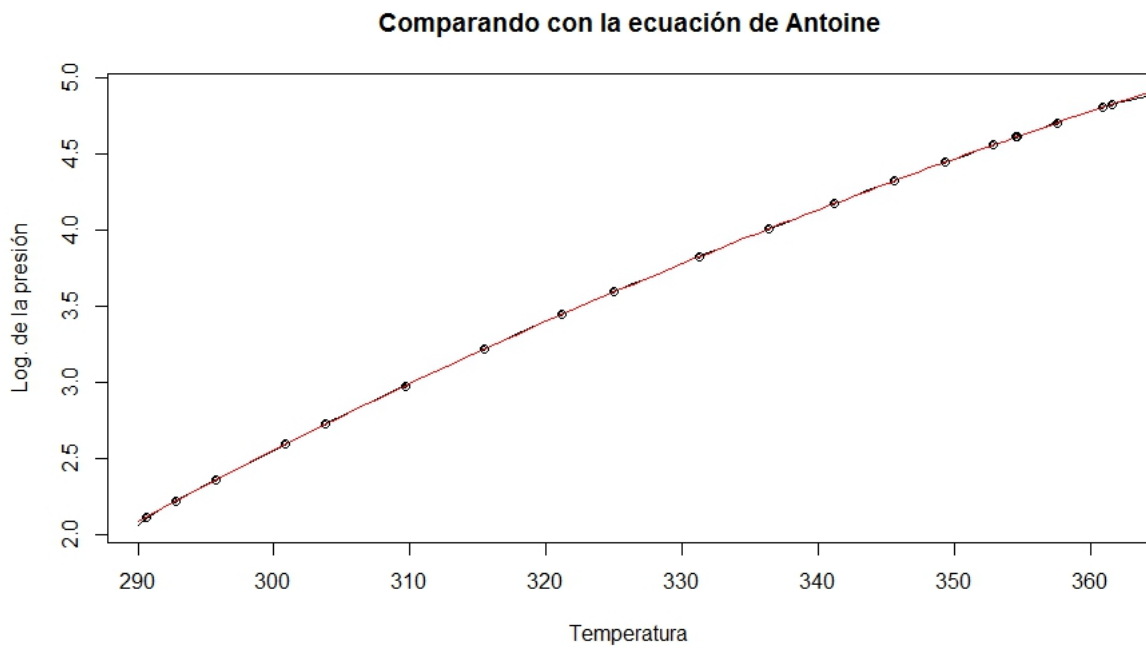


Figura 5.3: Ajuste de los datos usando la ecuación de Antoine (línea roja) y el modelo de regresión semi-parámétrico (línea negra).

tendría que buscar un modelo no lineal alternativo, mientras que el modelo semi-paramétrico propuesto en este trabajo sería básicamente el mismo.

# Capítulo 6

## Discusión

Seleccionar un modelo paramétrico suficientemente flexible para una aplicación dada no siempre es una tarea sencilla. En este trabajo se discute un modelo de regresión semi-paramétrico en el que se consideran dos tipos de errores: uno “determinista” y el otro aleatorio. En el error determinista se toma en cuenta la correlación que pudiera existir entre los valores de la respuesta para dos observaciones distintas pero cercanas. Este error está asociado a la función  $\delta(\cdot)$  que sigue un proceso gaussiano cuyos parámetros  $\rho^2$  y  $\lambda$  determinan la flexibilidad y la suavidad del modelo, respectivamente. Ésta es la parte no paramétrica del modelo. Por otro lado, el error aleatorio está asociado a los posibles errores en la observación de la respuesta  $f(\cdot)$ . Esta función se aproxima, como es usual en un modelo de regresión lineal, a través de  $h(\cdot)^t\beta$ , lo que corresponde a la parte paramétrica del modelo. La función  $\delta(\cdot)$  modela entonces las desviaciones del modelo paramétrico propuesto respecto al modelo “verdadero”.

El proceso gaussiano es tratable y tiene una interpretación relativamente simple, principalmente en este contexto de regresión. Esto nos ha permitido plantear un modelo general que tiene, como casos límite, escenarios conocidos tales como interpolación y regresión lineal. De hecho, el modelo de regresión semi-paramétrico discutido aquí tiene aplicaciones prácticas en problemas de interpolación y en otros problemas clásicos del análisis numérico, así como en la solución de problemas inversos, tal como se muestra en los Capítulos 4 y 5. Con base en los resultados de la aplicación del modelo para el caso de problemas inversos, es de destacar la flexibilidad del modelo semi-paramétrico. Por otra parte, el enfoque bayesiano permite tomar en cuenta y describir todas las posibles fuentes de incertidumbre, incluida aquella relacionada con la especificación del modelo.





# Apéndice A

## Resultados auxiliares

El conocer las propiedades y resultados sobre la normal multivariada en ocasiones puede simplificar algunas cuentas. A continuación, un teorema relacionado con la distribución normal multivariada que facilita la aplicación computacional del modelo. Estos resultados se puede ver también en Mardia, Kent y Bibby (1979).

Sea  $\mathbf{x} \in \mathbb{R}^{n+m}$  una variable aleatoria con distribución normal multivariada, tal que  $\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}$ , donde  $\mathbf{x}_1 \in \mathbb{R}^n$  y  $\mathbf{x}_2 \in \mathbb{R}^m$ . Entonces su distribución se puede escribir como:

$$\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \sim N_{n+m} \left( \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right). \quad (\text{A.1})$$

**Teorema A.1.** Si  $\mathbf{x} = (\mathbf{x}_1^t, \mathbf{x}_2^t)^t \sim N_{n+m}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , entonces  $\mathbf{x}_1$  y  $\mathbf{x}_{2.1} = \mathbf{x}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{x}_1$  tienen la siguiente distribución y son independientes:

$$\mathbf{x}_1 \sim N_n(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}) \quad y \quad \mathbf{x}_{2.1} \sim N_m(\boldsymbol{\mu}_{2.1}, \boldsymbol{\Sigma}_{22.1}), \quad (\text{A.2})$$

donde  $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1$  y  $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$ .

El teorema anterior sirve para encontrar distribuciones condicionales de  $\mathbf{x}_2$  dado que  $\mathbf{x}_1$  es conocida.

**Teorema A.2.** Usando los supuestos y notación del Teorema A.1, la distribución condicional de  $\mathbf{x}_2$  dado que  $\mathbf{x}_1$  es:

$$[\mathbf{x}_2 \mid \mathbf{x}_1] \sim N_m(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22.1}). \quad (\text{A.3})$$

Recordando que el modelo de regresión se escribe como en la ecuación (3.1). La función  $f(\cdot)$  denota la relación entre las variables explicativas y la variable respuesta,

$$y_i = f(x_i) + \epsilon_i.$$

Suponiendo que se tiene una muestra de tamaño  $m$  de la variable explicativa,  $\{x_i^*\}_{i=1}^m$ , un objetivo es encontrar un estimador para la variable respuesta correspondiente a cada

nueva variable explicativa. Dichos estimadores pueden ser calculados con base en procesos gaussianos, como se mostró anteriormente.

En el caso del modelo de este trabajo se tiene (3.8) y con esto se llega al resultado (3.14). Lo anterior corresponde al caso de una observación. Usando los resultados anteriores se pueden encontrar expresiones para toda una muestra de tamaño  $m$ .

Sea  $\mathbf{Y}$  una variable aleatoria de la forma,  $\mathbf{Y} = \begin{pmatrix} y_1 \\ y^* \end{pmatrix}$ . Donde  $y^*$  es el valor de la variable respuesta para cada  $x_i^*$   $i = 1, \dots, m$ , es decir,  $y^* = (y(x_1^*), \dots, y(x_m^*))^t$ . Análogamente el vector aleatorio  $y_1 = (y(x_1), \dots, y(x_n))^t$ .

Con base en lo mencionado al principio y por la definición de proceso gaussiano, la variable  $Y$  dados los parámetros tiene la siguiente distribución:

$$\left[ \begin{pmatrix} y_1 \\ y^* \end{pmatrix} \mid \beta, \rho^2 \right] \sim N_{n+m} \left( \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right), \quad (\text{A.4})$$

donde

$$\begin{aligned} \boldsymbol{\mu}_1 &= \begin{pmatrix} h(x_1)^t \beta \\ \vdots \\ h(x_m)^t \beta \end{pmatrix} \\ &= H\beta, \\ \boldsymbol{\mu}_2 &= \begin{pmatrix} h(x_1^*)^t \beta \\ \vdots \\ h(x_m^*)^t \beta \end{pmatrix} \\ &= H^* \beta, \end{aligned} \quad (\text{A.5})$$

$$\boldsymbol{\Sigma} = \rho^2 \begin{pmatrix} w + v(x_1, x_1) & v(x_1, x_2) & \cdots & v(x_1, x_n) & v(x_1, x_1^*) & v(x_1, x_2^*) & \cdots & v(x_1, x_m^*) \\ v(x_2, x_1) & w + v(x_2, x_2) & \cdots & v(x_2, x_n) & v(x_2, x_1^*) & v(x_2, x_2^*) & \cdots & v(x_2, x_m^*) \\ \vdots & \ddots & \cdots & \vdots & \vdots & \ddots & \cdots & \vdots \\ v(x_n, x_1) & v(x_n, x_2) & \cdots & w + v(x_n, x_n) & v(x_n, x_1^*) & v(x_n, x_2^*) & \cdots & v(x_n, x_m^*) \\ v(x_1^*, x_1) & v(x_1^*, x_2) & \cdots & v(x_1^*, x_n) & w + v(x_1^*, x_1^*) & v(x_1^*, x_2^*) & \cdots & v(x_1^*, x_m^*) \\ \vdots & \ddots & \cdots & \vdots & \vdots & \ddots & \cdots & \vdots \\ v(x_m^*, x_1) & v(x_m^*, x_2) & \cdots & v(x_m^*, x_n) & v(x_m^*, x_1^*) & v(x_m^*, x_2^*) & \cdots & w + v(x_m^*, x_m^*) \end{pmatrix}$$

$$\begin{aligned} \boldsymbol{\Sigma}_{11} &= \rho^2 \begin{pmatrix} w + v(x_1, x_1) & v(x_1, x_2) & \cdots & v(x_1, x_n) \\ v(x_2, x_1) & w + v(x_2, x_2) & \cdots & v(x_2, x_n) \\ \vdots & \ddots & \cdots & \vdots \\ v(x_n, x_1) & v(x_n, x_2) & \cdots & w + v(x_n, x_n) \end{pmatrix} \\ &= \rho^2 (wI + A), \end{aligned} \quad (\text{A.6})$$

$$\begin{aligned}
\Sigma_{12} &= \Sigma_{12}^t & (A.7) \\
&= \rho^2 \begin{pmatrix} v(x_1^*, x_1) & v(x_1^*, x_2) & \cdots & v(x_1^*, x_n) \\ \vdots & \ddots & \cdots & \vdots \\ v(x_m^*, x_1) & v(x_m^*, x_2) & \cdots & v(x_m^*, x_n) \end{pmatrix} \\
&= \rho^2 (t(x_1^*), t(x_2^*), \dots, t(x_m^*)) \\
&= \rho^2 L,
\end{aligned}$$

$$\begin{aligned}
\Sigma_{22} &= \rho^2 \begin{pmatrix} w + v(x_1^*, x_1^*) & v(x_1^*, x_2^*) & \cdots & v(x_1^*, x_m^*) \\ \vdots & \ddots & \cdots & \vdots \\ v(x_m^*, x_1^*) & v(x_m^*, x_2^*) & \cdots & w + v(x_m^*, x_m^*) \end{pmatrix} & (A.8) \\
&= \rho^2 (wI + A^*).
\end{aligned}$$

Suponiendo la variable  $y_1$  conocida e igual a  $y$  se puede usar el resultado del Teorema A.2 para encontrar un estimador de  $y^*$  dados los parámetros. Con base en la expresión (A.9) dicho estimador es  $H^*\beta - LA^{-1}(y - H\beta)$ ; sin embargo, no es el estimador que se busca pues éste aún depende de los parámetros,

$$[y^* | \beta, \rho^2, y] \sim N_m(H^*\beta + L(wI + A)^{-1}(y - H\beta), \rho^2 W), \quad (A.9)$$

donde

$$W = (wI + A)^* - L(wI + A)^{-1}L^t.$$

Si se observa detenidamente las expresiones (A.9) y (3.14) se nota una similitud entre ellas. De hecho la única diferencia es que la primera es para una muestra de tamaño  $m$  y la segunda para una sólo observación. Esta expresión aún depende de  $\beta$ , pero bajo los supuestos de la Sección 3.2.1 y aplicando de nuevo el resultado del teorema A.2 se obtiene (3.15) para una muestra de  $m$  observaciones. Lo anterior se muestra a continuación:

$$[y^* | \rho^2, y] \sim N_m(H^*b_1 + L(wI + A)^{-1}(y - Hb_1), \rho^2 W), \quad (A.10)$$

donde

$$W^* = W + (H^{*t} - H(wI + A)^{-1}L)^t B_1^{-1} (H^{*t} - H(wI + A)^{-1}L).$$

Estos resultado brindan una gran ayuda para estimar, pues computacionalmente siempre es más rápido trabajar con operaciones matriciales.



# Apéndice B

## Modelo General

En este apartado se pretende abordar un modelo general de regresión con el fin de hacer más claros algunos de los resultados del cuerpo principal de este trabajo.

### B.1. Modelo

Suponga un modelo escrito en forma matricial donde  $Y$  es una variable respuesta y  $H$  una variable explicativa:

$$Y = H\beta + \delta + \epsilon, \quad (\text{B.1})$$

donde  $H_{n \times q}$  y

$$\begin{aligned} \beta &\sim N_q(b, B^{-1}), \\ \delta &\sim N_n(0, D^{-1}), \\ \epsilon &\sim N_n(0, E^{-1}). \end{aligned} \quad (\text{B.2})$$

Por lo tanto  $Y_{n \times 1}$  y se puede reescribir como

$$Y = C\theta + \epsilon, \quad (\text{B.3})$$

con  $C = [H \quad I_n]$  y  $\theta = [\beta^t \quad \delta^t]^t$ .

Considerando  $\epsilon$  como el error aleatorio, lo que interesa es hacer inferencia sobre  $\theta$ ; por consiguiente se le asigna una distribución inicial expresada como

$$p(\theta) = N_m(\theta \mid t, T^{-1}), \quad (\text{B.4})$$

con  $m = n + q$ ,  $t = \begin{pmatrix} b \\ 0 \end{pmatrix}$  y  $T = \begin{pmatrix} B & 0 \\ 0 & D \end{pmatrix}$ .

Suponga un experimento de  $n$  observaciones de la variable  $Y$ , es decir,  $\{y_i\}_{i=1}^n$ . Se denota  $y = (y_1, \dots, y_n)^t$  como el vector de los datos observados usados para calcular la distribución final del parámetro  $\theta$ . Se observa que la expresión (B.3) tiene la forma del problema de regresión para el cual hay resultados ya conocidos.

En

$$p(\theta | y) = N_m(\theta | t_1, T_1^{-1}) \quad (\text{B.5})$$

se expresa la distribución final, la cual se obtiene al usar el teorema de Bayes  $p(\theta | y) \propto p(\theta)L(\theta; y)$  donde  $L$  expresa la función de verosimilitud, con

$$\begin{aligned} t_1 &= [C^t EC + T]^{-1}(C^t Ey + Tt) \\ T_1 &= C^t EC + T. \end{aligned} \quad (\text{B.6})$$

La expresión (B.5) se escribe en términos de  $\beta$  y  $\delta$  como se muestra en

$$p(\beta | y) = N_q(\beta | b_1, B_1^{-1}) \quad (\text{B.7})$$

y

$$p(\delta | y) = N_n(\delta | d_1, D_1^{-1}) \quad (\text{B.8})$$

respectivamente, donde

$$\begin{aligned} b_1 &= [H^t E(D + E)^{-1} DH + B]^{-1}(H^t E(D + E)^{-1} Dy + Bb), \\ B_1 &= H^t E(D + E)^{-1} DH + B, \end{aligned} \quad (\text{B.9})$$

$$\begin{aligned} D_1 &= D + E - EH(H^t EH + B)^{-1} H^t E, \\ d_1 &= [D + E - EH(H^t EH + B)^{-1} H^t E]^{-1} E[y - H(H^t EH + B)^{-1}(H^t Ey + Bb)]. \end{aligned} \quad (\text{B.10})$$

Las expresiones en (B.10) se pueden seguir desarrollando, al hacerlo se llega al resultado

$$\begin{aligned} d_1 &= y - [D + E - EH(H^t EH + B)^{-1} H^t E]^{-1} (Dy + EH(H^t EH + B)^{-1} Bb), \\ D_1^{-1} &= (D + E)^{-1} \\ &\quad + (D + E)^{-1} EH[(H^t DH + B) + (H^t E(D + E)^{-1} DH - H^t DH)]^{-1} H^t E(D + E)^{-1}. \end{aligned} \quad (\text{B.11})$$

La covarianza final entre estas dos variables es:

$$\begin{aligned} \text{Cov}(\beta, \delta | y) &= -[H^t E(D + E)^{-1} DH + B]^{-1} H^t E(D + E)^{-1} \\ &= -B_1^{-1} H^t E(D + E)^{-1}. \end{aligned} \quad (\text{B.12})$$

## B.2. Predicción

En esta sección se hace predicción manteniendo  $C$ , con esto se pretende encontrar predicciones sobre el conjunto de los datos que se han observado.

Se denota  $Y^*$  la variable de predicción, es decir  $Y^* = C\theta + \epsilon^*$ , donde  $\epsilon^* \sim N_n(0, E^{-1})$ . La distribución inicial de esta variable es la misma que para (B.3) y usando esto se obtiene la final, dicha expresión es

$$[Y^* | y] \sim N_n(m_y, V_y), \quad (\text{B.13})$$

donde

$$\begin{aligned} m_y &= E(Y^* | y) \\ &= E(C\theta | y) \\ &= CE(\theta | y) \\ &= Ct_1 \\ &= HE(\beta | y) + E(\delta | y) \\ &= Hb_1 + d_1, \end{aligned} \quad (\text{B.14})$$

$$\begin{aligned} V_y &= E^{-1} + C\text{Var}(\theta | y)C^t \\ &= E^{-1} + CT^{-1}C^t \\ &= E^{-1} + H\text{Var}(\beta | y)H^t + \text{Var}(\delta | y) + 2HC\text{ov}(\beta, \delta | y) \\ &= E^{-1} + HB_1^{-1}H^t + D_1^{-1} + 2HC\text{ov}(\beta, \delta | y) \\ &= E^{-1} + D_1^{-1} + HB_1^{-1}H^t - 2HB_1^{-1}H^tE(D + E). \end{aligned} \quad (\text{B.15})$$

Todo lo anterior es en un caso general, a continuación se hará el desarrollo bajo la suposición de dos casos particulares:

- a)  $E = \frac{1}{\sigma^2}I_n$ ,  $D = \frac{1}{\rho^2}A^{-1}$  y  $B = \frac{1}{\rho^2}B_0$ ,
- b)  $E = \frac{1}{\sigma^2}I_n$ ,  $D = \frac{1}{\rho^2}A^{-1}$  y  $B = B_0$ .

## B.3. Caso a)

En esta sección se expondrán expresiones particulares para (B.9), (B.10) y (B.11) suponiendo  $E = \frac{1}{\sigma^2}I_n$ ,  $D = \frac{1}{\rho^2}A^{-1}$  y  $B = \frac{1}{\rho^2}B_0$ .



Para (B.9):

$$\begin{aligned}
b_1 &= [H^t E(D + E)^{-1} DH + B]^{-1} (H^t E(D + E)^{-1} Dy + Bb) & (B.16) \\
&= \left\{ \frac{1}{\rho^2} [H^t (A + \frac{\sigma^2}{\rho^2} I_n)^{-1} H + B_0] \right\}^{-1} \left( \frac{1}{\sigma^2 \rho^2} H^t \left( \frac{1}{\rho^2} A^{-1} + \frac{1}{\sigma^2} I_n \right)^{-1} A^{-1} y + \frac{1}{\rho^2} B_0 b \right) \\
&= [H^t (A + \frac{\sigma^2}{\rho^2} I_n)^{-1} H + B_0]^{-1} (H^t (A + \frac{\rho^2}{\sigma^2} I_n)^{-1} y + B_0 b), \\
B_1 &= H^t E(D + E)^{-1} DH + B \\
&= \frac{1}{\sigma^2 \rho^2} H^t \left( \frac{1}{\rho^2} A^{-1} + \frac{1}{\rho^2} I_n \right)^{-1} A^{-1} H + \frac{1}{\rho^2} B_0 \\
&= \frac{1}{\rho^2} [H^t (A + \frac{\sigma^2}{\rho^2} I_n)^{-1} H + B_0].
\end{aligned}$$

Para (B.10):

$$\begin{aligned}
D_1 &= D + E - EH(H^t EH + B)^{-1} H^t E & (B.17) \\
&= \frac{1}{\rho^2} A^{-1} + \frac{1}{\sigma^2} I_n - \frac{1}{\sigma^4} H \left( \frac{1}{\sigma^2} H^t H + \frac{1}{\rho^2} B_0 \right)^{-1} H^t \\
&= \frac{1}{\rho^2} \left[ A^{-1} + \frac{\rho^2}{\sigma^2} (I_n - H(H^t H + \frac{\sigma^2}{\rho^2} B_0)^{-1} H^t) \right], \\
d_1 &= [D + E - EH(H^t EH + B)^{-1} H^t E]^{-1} E[y - H(H^t EH + B)^{-1} (H^t E y + Bb)] \\
&= [D_1]^{-1} \frac{1}{\sigma^2} \left[ y - H \left( \frac{1}{\sigma^2} H^t H + \frac{1}{\rho^2} B_0 \right)^{-1} \left( \frac{1}{\sigma^2} H^t y + \frac{1}{\rho^2} B_0 b \right) \right] \\
&= \frac{\rho^2}{\sigma^2} \left[ A^{-1} + \frac{\rho^2}{\sigma^2} (I_n - H(H^t H + \frac{\sigma^2}{\rho^2} B_0)^{-1} H^t) \right]^{-1} \left[ y - H(H^t H + \frac{\sigma^2}{\rho^2} B_0)^{-1} (H^t y + \frac{\sigma^2}{\rho^2} B_0 b) \right].
\end{aligned}$$

Para (B.11):

$$\begin{aligned}
D_1^{-1} &= (D + E)^{-1} & (B.18) \\
&\quad + (D + E)^{-1} EH[(H^t DH + B) + (H^t E(D + E)^{-1} DH - H^t DH)]^{-1} H^t E(D + E)^{-1} \\
&= \left( \frac{1}{\rho^2} A^{-1} + \frac{1}{\sigma^2} I_n \right)^{-1} \\
&\quad + \frac{1}{\sigma^4} \left( \frac{1}{\rho^2} A^{-1} + \frac{1}{\sigma^2} I_n \right)^{-1} H [B_1]^{-1} H^t \left( \frac{1}{\rho^2} A^{-1} + \frac{1}{\sigma^2} I_n \right)^{-1} \\
&= \rho^2 \left\{ (A^{-1} + \frac{\rho^2}{\sigma^2} I_n)^{-1} \right. \\
&\quad \left. + \frac{\rho^2}{\sigma^4} (A^{-1} + \frac{\rho^2}{\sigma^2} I_n)^{-1} H [H^t (A + \frac{\sigma^2}{\rho^2} I_n)^{-1} H + B_0]^{-1} H^t (A^{-1} + \frac{\rho^2}{\sigma^2} I_n)^{-1} \right\}, \\
d_1 &= y - [D + E - EH(H^t EH + B)^{-1} H^t E]^{-1} (Dy + EH(H^t EH + B)^{-1} Bb) \\
&= y - [D_1]^{-1} \left( \frac{1}{\rho^2} A^{-1} y + \frac{1}{\rho^2 \sigma^2} H \left( \frac{1}{\sigma^2} H^t H + \frac{1}{\rho^2} B_0 \right)^{-1} B_0 b \right) \\
&= y - \left[ A^{-1} + \frac{\rho^2}{\sigma^2} (I_n - H(H^t H + \frac{\sigma^2}{\rho^2} B_0)^{-1} H^t) \right]^{-1} \left( A^{-1} y + H(H^t H + \frac{\sigma^2}{\rho^2} B_0)^{-1} B_0 b \right).
\end{aligned}$$

## B.4. Caso b)

En esta sección se expondrán expresiones particulares para (B.9), (B.10) y (B.11) suponiendo  $E = \frac{1}{\sigma^2}I_n$ ,  $D = \frac{1}{\rho^2}A^{-1}$  y  $B = B_0$ .

Para (B.9):

$$\begin{aligned}
b_1 &= [H^t E(D + E)^{-1} D H + B]^{-1} (H^t E(D + E)^{-1} D y + B b) & (B.19) \\
&= \left\{ \frac{1}{\rho^2} [H^t (A + \frac{\sigma^2}{\rho^2} I_n)^{-1} H] + B_0 \right\}^{-1} \left( \frac{1}{\sigma^2 \rho^2} H^t \left( \frac{1}{\rho^2} A^{-1} + \frac{1}{\sigma^2} I_n \right)^{-1} A^{-1} y + B_0 b \right) \\
&= \left[ \frac{1}{\rho^2} H^t (A + \frac{\sigma^2}{\rho^2} I_n)^{-1} H + B_0 \right]^{-1} \left( \frac{1}{\rho^2} H^t (A + \frac{\sigma^2}{\sigma^2} I_n)^{-1} y + B_0 b \right), \\
B_1 &= H^t E(D + E)^{-1} D H + B \\
&= \frac{1}{\sigma^2 \rho^2} H^t \left( \frac{1}{\rho^2} A^{-1} + \frac{1}{\rho^2} I_n \right)^{-1} A^{-1} H + B_0 \\
&= \frac{1}{\rho^2} [H^t (A + \frac{\sigma^2}{\rho^2} I_n)^{-1} H] + B_0.
\end{aligned}$$

Para (B.10):

$$\begin{aligned}
D_1 &= D + E - E H (H^t E H + B)^{-1} H^t E & (B.20) \\
&= \frac{1}{\rho^2} A^{-1} + \frac{1}{\sigma^2} I_n - \frac{1}{\sigma^4} H \left( \frac{1}{\sigma^2} H^t H + B_0 \right)^{-1} H^t \\
&= \frac{1}{\rho^2} \left[ A^{-1} + \frac{\rho^2}{\sigma^2} (I_n - H (H^t H + \sigma^2 B_0)^{-1} H^t) \right], \\
d_1 &= [D + E - E H (H^t E H + B)^{-1} H^t E]^{-1} E [y - H (H^t E H + B)^{-1} (H^t E y + B b)] \\
&= [D_1]^{-1} \frac{1}{\sigma^2} [y - H \left( \frac{1}{\sigma^2} H^t H + B_0 \right)^{-1} \left( \frac{1}{\sigma^2} H^t y + B_0 b \right)] \\
&= \frac{\rho^2}{\sigma^2} \left[ A^{-1} + \frac{\rho^2}{\sigma^2} (I_n - H (H^t H + \sigma^2 B_0)^{-1} H^t) \right]^{-1} [y - H (H^t H + \sigma^2 B_0)^{-1} (H^t y + \sigma^2 B_0 b)].
\end{aligned}$$

Para (B.11):

$$\begin{aligned}
D_1^{-1} &= (D + E)^{-1} && \text{(B.21)} \\
&+ (D + E)^{-1}EH[(H^tDH + B) + (H^tE(D + E)^{-1}DH - H^tDH)]^{-1}H^tE(D + E)^{-1} \\
&= \left(\frac{1}{\rho^2}A^{-1} + \frac{1}{\sigma^2}I_n\right)^{-1} \\
&+ \frac{1}{\sigma^4}\left(\frac{1}{\rho^2}A^{-1} + \frac{1}{\sigma^2}I_n\right)^{-1}H[B_1]^{-1}H^t\left(\frac{1}{\rho^2}A^{-1} + \frac{1}{\sigma^2}I_n\right)^{-1} \\
&= \rho^2\left\{(A^{-1} + \frac{\rho^2}{\sigma^2}I_n)^{-1} \right. \\
&\quad \left. + \frac{1}{\sigma^4}(A^{-1} + \frac{\rho^2}{\sigma^2}I_n)^{-1}H\left[\frac{1}{\rho^2}[H^t(A + \frac{\sigma^2}{\rho^2}I_n)^{-1}H] + B_0\right]^{-1}H^t(A^{-1} + \frac{\rho^2}{\sigma^2}I_n)^{-1}\right\}, \\
d_1 &= y - [D + E - EH(H^tEH + B)^{-1}H^tE]^{-1}(Dy + EH(H^tEH + B)^{-1}Bb) \\
&= y - [D_1]^{-1}\left(\frac{1}{\rho^2}A^{-1}y + \frac{1}{\sigma^2}H\left(\frac{1}{\sigma^2}H^tH + \frac{1}{\rho^2}B_0\right)^{-1}B_0b\right) \\
&= y - \left[\frac{1}{\rho^2}(A^{-1} + \frac{\rho^2}{\sigma^2}(I_n - H(H^tH + \sigma^2B_0)^{-1}H^t))\right]^{-1}\left(\frac{1}{\rho^2}A^{-1}y + H(H^tH + \sigma^2B_0)^{-1}B_0b\right).
\end{aligned}$$

# Apéndice C

## Código en R

En este apéndice se presentan los códigos implementados para obtener los resultados del modelo de este trabajo.

```
#Función para interpolar usando procesos Gaussianos usando la función 'dist'
```

```
library(proxy)
library(maxLik)
library(optimx)
```

```
###Declaracion de Funciones
```

```
#Funciones para crear variables tipo listas que serán parámetros en la
#función principal
```

```
lista1<-function(bi,Bi,si,ai,w00)
{ #Lista para los parámetros de las distribuciones
  ii1=list(b0=bi,B0=Bi,s0=si,a0=ai,w=w00)#iniciales
  return(ii1)}
```

```
lista2<-function(lbdi,Vi)
{ #Lista para los parámetros de la función de
  ii2=list(lbd=lbdi,V=Vi) # correlación
  return(ii2)
}
```

```
#Funcion H
```

```
#Parámetros: Vector de observaciones x-s, función propuesta
Hx<-function(x,h) t(apply(x,1,h))
```

```
#Función principal, la salida de esta función es una lista con:
```

```

#Función esperanza final
#Función varianza final
#Valores de la esperanza final de los datos a interpolar
#Valores de la varianza final de los datos interpolar

#Los parámetros de la función son:

#Observaciones 'x'
#Observaciones de la función evaluada en 'x' ('y')
#Lista de parámetros iniciales de la distribución
#Lista de parámetros para la función de correlación inicial
#Función propuesta 'h'
#Valores a interpolar

#Para hacer uso de la distancia Mahalanobis
#se debe bajar la paquetería 'proxy'

proceso<-function(x,y,ini1,ini2,h,x)
{
  n=dim(x)[1] #Número de observaciones
  ne=dim(xe)[1] #Número de valores a interpolar
  d=dim(x)[2]#Dimensión del proceso
  xT=rbind(x,x) #Combinación entre los datos observados
  #y los que se interpolaran
  H=Hx(x,h)
  if (length(ini1$b==1)) {H=t(H)}
  He=Hx(xe,h)
  if (length(ini1$b==1)) {He=t(He)}
  AT1=-ini2$lbd*
  (as.matrix(dist(xT,method = "mahalanobis",upper=T,diag=T,cov=V)))^2
  #Usando la distancia de Mahalanobis
  AT=exp(as.matrix(AT1)) #Matriz de desitancias totales
  AT=AT+ini1$w*diag(n+ne)
  IA=solve(AT[1:n,1:n]) #Inversa de la matriz A
  print(dim(IA))
  L=AT[(n+1):(n+ne),1:n] #Matriz lamda
  K=L%*%IA
  Ae=AT[(n+1):(n+ne),(n+1):(n+ne)] #Matriz A estrella
  Be=ini1$B0+t(H)%*%IA%*%H
  #Matriz B, covarianzas finales para el parámetro beta
  IB=solve(Be)
  b=IB%*(ini1$B0%*%ini1$b0+ t(H)%*%IA%*%y )
  #Vector de esperanza final para el parámetro beta

```

```

Me=He**b+K**%(y-H**b)
#Vector de esperanza final para los valores a interpolar
W=(Ae-L**IA**t(L))
We=ini1$w*diag(ne)+W+t(t(He)-t(H)**t(K))**IB**%(t(He)-t(H)**t(K))
#Matriz W estrella
q=dim(matrix(ini1$b0))[1]
a=ini1$a0+n-q
s=ini1$s0+t(y)**(IA-IA**H**solve(t(H)**IA**H)**t(H)**IA)**y
vari=(s/(a-2))*diag(We)
#Vector de variaza final para los valores a interpolar

mod=list(m=Me,W=We,var=vari,B=Be,s=s,a=a,b=b)
return(mod)
}

```

```

#Función para poder graficar en el caso bivariado
#Esta función tiene como salida los vectores de 'x0', 'y0' y la matriz 'z0'
#parámetros necesarios para la función 'persp'

```

```

pgraf<-function(a,b)#Matriz de nx2 y vector columna de nx1
{
  bus<-function(z,m,v,n)
  {
    k=0
    for(j in 1:n)
    {
      if(z[1] == m[j,1] && z[2] == m[j,2] ) k<-j
    }
    if(k!=0) {return (v[k])}
    else {return(NA)}
  }
  n=dim(matrix(b))[1]
  x=sort(matrix(unique(a[,1])))
  y=sort(matrix(unique(a[,2])))
  nx=dim(matrix(x))[1]
  ny=dim(matrix(y))[1]
  d=max(nx,ny)
  M=matrix(NA, ncol=ny,nrow=nx)
  for (i in 1:nx)
  {
    for(j in 1:ny)

```

```

    {
      M[i,j]=bus(c(x[i],y[j]),a,b,n)
    }

}

return(list(x=x,y=y,z=M))

}

#Función para graficar el proceso en el caso bivariado o univariado
plotproceso<-function(pro,x,y,xe){
  d=dim(x)[2]
  if (d==1){
    plot(xe,pro$m,main="",type='l',xlab='x',ylab='Estimación')
    title(sprintf("Estimación de y(.) \n No. de Observaciones=%d",dim(x)[1]))
    points(x,y)
  } else{
    dat=pgraf(xe,pro$m)
    op <- par(bg = "white")
    persp(dat$x,dat$y, dat$z, theta = 125, phi = 5, expand = 0.5,ticktype = "detailed",
    title(sprintf("Estimación de y(.) \n No. de Observaciones=%d",dim(x)[1]))
    dev.new()
    contour(dat$x,dat$y,dat$z)
  title(sprintf("Curvas de nivel para y(.) \n No. de Observaciones=%d",dim(x)[1]))
  }
}

plotvar<-function(pro,x,y,xe){
  dat=pro$m
  var=pro$var
  va1<-dat+qt(0.975,pro$a)*var
  va2<-dat-qt(0.975,pro$a)*var
  plot(xe,dat,ylab="Estimación",type='l',ylim=c(min(va2),max(va1)), xlab='x')
  title(sprintf("Estimación de y(.) \n No. de Observaciones=%d ",dim(x)[1]))
  lines(xe,va1, type = "l", col='blue')
  lines(xe,va2, type = "l", col='blue')
  points(x,y)
}

```

# Bibliografía

- [1] Antoine C. (1888) “Tensions des vapeurs; nouvelle relation entre les tensions et les temperatures”. *Comptes Rendus des Séances de l'Académie des Sciences*, **107**, 681–684, 778–780, 836–837.
- [2] Barber, D. (2012). *Bayesian Reasoning and Machine Learning*. Cambridge University Press.
- [3] Bernardo, J.M. (1979). “Reference Posterior Distributions for Bayesian Inference”. *Journal of the Royal Statistical Society B*, **41**, 113-147 (with discussion).
- [4] Bernardo, J. M., Smith, A. F. (2000) *Bayesian Theory*. Wiley.
- [5] Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- [6] Blight, B.J.N. y Ott, L. (1975) “A Bayesian approach to model inadequacy for polynomial regression”. *Biometrika* **62**, 79-88.
- [7] Bromeling Lyle D. (1985) *Bayesian Analysis of Linear Models*. Marcel Dekker
- [8] Ewing M., Sánchez J. (2004) “Vapor pressures of acetonitrile determined by comparative ebulliometry”. *J. Chem. Eng. Data*, **62**, 486-491.
- [9] Geoffrey, G. y Stirzaker D. (2001). *Probability and Random Processes*. Oxford University Press.
- [10] Gutiérrez-Peña, E. (1998) “Análisis Bayesiano de Modelos Jerárquicos Lineales”. *Serie Monografías*, **7**, No. 16. IIMAS, UNAM.
- [11] Mackay, David J.C. (1998) Introduction to Gaussian processes. En *Neural Networks and Machine Learning* (C.M. Bishop, ed.), **168** of NATO ASI Series, 133-165. Springer, Berlin.
- [12] Mardia, K. V., Kent, J. T., y Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press.
- [13] Mc Cullagh, P. y Nelder, J. (1989). *Generalized Linear Models* (2a ed.). Chapman and Hall/CRC.



- [14] O'Hagan, A. (1978). "Curve fitting and optimal design for prediction". *Journal of the Royal Statistical Society. Series B* **40** , 1-42.
- [15] O'Hagan, A. (1992). "Some Bayesian Numerical Analysis". *Bayesian Statistics* **4** , 345-363.
- [16] R Core Team (2013) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing; Vienna, Austria, URL = <http://www.R-project.org/>
- [17] Rasmussen C., Williams, C. (2006) *Gaussian Processes for Machine Learning*. Massachusetts Institute of Technology.
- [18] Stone, M.H. "The Generalized Weierstrass Approximation Theorem ". *Mathematics Magazine*. **21**, 167-184.
- [19] Tarantola A. (2005) *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM
- [20] Williams, C. K. I.(1998) "Prediction with Gaussian processes: From linear regression to linear prediction and beyond". In *Jordan, M.I., ed.: Learning in Graphical Models*. Kluwer Academic
- [21] Williams, C. K. I.(2002). "Gaussian processes". In *M. A. Arbib, editor, Handbook of Brain Theory and Neural Networks*, 466-470. The MIT Press, second edition.
- [22] Williams,C. K. I. and C. E. Rasmussen (1996) "Gaussian processes for regression". In *D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, editors, Advances in Neural Information Processing Systems 8*, 514-520. The MIT Press, Cambridge, MA.
- [23] Zellner, Arnold (1996) *An Introduction to Bayesian Inference in Econometrics*. Wiley