



CIMAT

Centro de Investigación en Matemáticas, A.C.

ASPECTOS TEÓRICOS Y DE IMPLEMENTACIÓN EN LA CUANTIFICACIÓN DE LA INCERTIDUMBRE BAYESIANA: DE SIERPINSKI A SCIPY

T E S I S

Que para obtener el grado de
Maestro en Ciencias
con Orientación en
Probabilidad y Estadística

Presenta

María Fernanda Méndez Lemus

Director de Tesis:

Dr. J. Andrés Christen Gracia

Autorización de la versión final

Guanajuato, Gto., 28 de Julio de 2018

ASPECTOS TEÓRICOS Y DE
IMPLEMENTACIÓN EN
CUANTIFICACIÓN DE LA
INCERTIDUMBRE BAYESIANA: DE
SIERPIŃSKI A SCIPY

2019

CIMAT

Agradecimientos

A mi mamá, por siempre creer en mi y apoyarme incondicionalmente.

A mi madrina, por recordarme siempre quien soy.

A Delia. Gracias a ella pude llegar a Guanajuato.

A Delio, fuente de infinita inspiración y respeto, por hacer de mis últimos meses en Guanajuato los mejores y crear el ambiente propicio para finalizar este trabajo.

Al Dr. Andrés Christen por su dirección, apoyo y paciencia durante mi último semestre.

Al Dr. Juan Carlos Pardo Millán y al Dr. Joaquin Ortega, ya que su ayuda fue crucial para la culminación y desarrollo de mis estudios de maestría.

A Ehyter, por invertir tiempo fuera del aula para continuar enseñándome.

A mis sinodales por haber accedido a invertir su valioso tiempo en la revisión de este texto.

A Roberto Pichardo, por todas sus enseñanzas sobre escritura de textos científicos.

A *los de siempre*, por no dejarme varada en una ciudad desconocida.

A Jorge Luis, por seguirme los pasos hasta Guanajuato.

A Santiago y Camilo, siempre dispuestos a charlar y a compartir una buena taza de café.

A Liz Lemus, en quien siempre encontré una compañera de trabajo.

A mis compañeros de cubo: Sonny y Sebastián; por compartirme de su vasta experiencia académica. Su presencia hizo más feliz mi estancia en Guanajuato.

A mi compañero de casa: Jorge Javier; por hacerme sentir en un hogar.

A CONACyT por su contribución económica durante mis 24 meses de estudio.

A CIMAT.

Índice general

Agradecimientos	2
1. Introducción	6
2. Preliminares	8
2.1. Funciones medibles en espacios producto	8
2.1.1. Espacios Polacos	10
2.2. Desintegración	14
2.3. Variación Total	15
2.3.1. Convergencia débil	18
2.4. Lemas auxiliares	19
3. Aspectos teóricos	25
3.1. Existencia de la distribución posterior	25
3.2. Convergencia y consistencia	26
3.3. Cotas de error	33
3.4. Truncamientos	38
4. Implementación	42
4.1. Deconvolución	42
4.2. Resultados	44
4.2.1. Caso $k = 10$	44
4.2.2. Caso $k = 7$	45
4.2.3. Caso $k = 8$	45

<i>Índice general</i>	5
-----------------------	---

5. Conclusiones	62
------------------------	-----------

Bibliografía	63
---------------------	-----------

Capítulo 1

Introducción

Sea $y \in \mathbb{R}^m$ una muestra observada, estamos interesados en determinar el parámetro θ del modelo \mathcal{F} , donde

$$y_i = \mathcal{F}(\theta) + \epsilon_i; \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2).$$

A \mathcal{F} se le conoce como regresor o Forward Map (FM) y en la expresión anterior ϵ_i representa el ruido con el que se colectan las observaciones y . En el enfoque Bayesiano la solución a este problema es la obtención de una medida de probabilidad, llamada distribución posterior, con la que se cuantifica la incertidumbre sobre θ . Es por ello que no sólo es de interés asegurar la existencia de dicha distribución, sino además poder implementar algoritmos que simulen de ella y nos permitan obtener cantidades de interés. Algunas de las dificultades que esto representa, son

- sólo se cuenta con una versión numérica de \mathcal{F} ,
- θ puede ser infinito-dimensional. En cuyo caso debemos encontrar una manera de trabajar con una distribución a priori alternativa π_k que sea computable.

Al trabajar bajo las condiciones anteriores obtenemos una versión numérica de la medida posterior, cuya calidad depende de las aproximaciones anteriores. En este trabajo se revisan y exponen resultados aparecidos en [2] encaminados a la selección de un nivel óptimo de discretización para \mathcal{F} y un truncamiento para la distribución apriori. Todo esto con la finalidad de controlar el error entre la distribución posterior

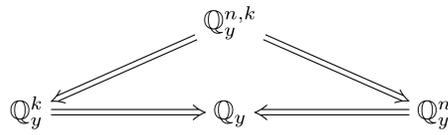


Figura 1.1: Convergencia débil entre las diferentes medidas posteriores.

teórica y numérica.

Este trabajo está dividido en dos partes; la primera de ellas, que es la parte teórica, se encuentra dedicada a la fundamentación y la segunda a algunos aspectos de implementación. Los capítulos que comprenden la teoría son el 2 y 3, dejando el capítulo 4 dedicado a un breve reporte sobre la implementación realizada. En los preliminares se encuentran todos los resultados que son herramienta fundamental para las pruebas de los capítulos subsiguientes.

En la primera sección del capítulo 3 se discuten condiciones suficientes para la existencia de la medida posterior y se establece un marco general donde dichas condiciones son satisfechas.

En la sección 3.2 se muestra la convergencia (en distribución o convergencia débil) de la media posterior numérica $\mathbb{Q}_y^{n,k}$ a la teórica \mathbb{Q}_y , de acuerdo al diagrama 1.1.

Además de probar la convergencia, es importante conocer la velocidad con la que se da. En esta sección también se obtienen las tasas de convergencia haciendo uso de la distancia Varición Total (VT).

Respecto a la sección 3.3, ésta contiene el resultado principal de [2], una cota para controlar el error entre $\mathbb{Q}_y^{n,k}$ y \mathbb{Q}_y .

Finalmente en el último capítulo, mediante una implementación se muestra un ejemplo de deconvolución, con el cual se discute la selección de la discretización vía la cota anterior.

Capítulo 2

Preliminares

El objetivo de este capítulo es coleccionar algunos resultados y definiciones que estaremos utilizando a lo largo del texto. También se indicará la notación que emplearemos.

En las secciones de Desintegración y Convergencia débil, encontraremos resultados que servirán como herramienta para pruebas del capítulo siguiente, por lo que sólo aparecerán las demostraciones que no rebasen el objetivo central de esta tesis; y para las pruebas que fueron omitidas, se incluyen referencias bibliográficas donde el lector pueda consultarlas.

2.1. Funciones medibles en espacios producto

Con el fin de reducir la longitud de las pruebas, introducimos la siguiente notación que será de utilidad los subsecuentes resultados.

Definición 2.1.1. Sean A, B, C conjuntos no vacíos y $f : A \times B \rightarrow C$ una función. Para cualesquiera $a \in A, b \in B$ fijos, definimos las funciones $f_a : B \rightarrow C, f^b : A \rightarrow C$ mediante

$$f_a(b) \mapsto f(a, b) \quad f^b(a) \mapsto f(a, b).$$

Sabemos que si f es una función medible de dos variables, entonces f_x y f^y re-

sultan medibles; de manera natural surge la pregunta sobre si el converso es cierto. La respuesta a esta interrogante resulta ser negativa.

En 1920 Sierpiński muestra en [7] la existencia de un conjunto no medible $E \subseteq \mathbb{R}^2$ que satisface las siguientes propiedades.

- (a) Todo subconjunto cerrado de \mathbb{R}^2 con medida de Lebesgue positiva tiene intersección no vacía con E .
- (b) E no contiene tercias colineales.

De la existencia de este conjunto se desprende el siguiente

Contraejemplo 2.1.2. *Existe una función $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ cuyas funciones f_x y f_y son medibles para cualesquiera $x, y \in \mathbb{R}$, pero f no es una función medible en \mathbb{R}^2 .*

Sean E el conjunto descrito anteriormente y ℓ cualquier recta paralela al eje de las ordenadas. La propiedad (a) implica que $\ell \cap E \neq \emptyset$; y a su vez, se sigue de (b) que $|\ell \cap E| \leq 2$, ya que de lo contrario habríamos encontrado tres puntos colineales, lo cual llevaría a una contradicción.

Llamemos L a la colección de las rectas verticales cuya intersección con E tiene exactamente dos puntos. Definimos al conjunto E_1 (posiblemente vacío) como aquel formado por los puntos $e = (e_1, e_2)$ tales que

1. para alguna $\ell \in L$, $e \in \ell \cap E$ y
2. si $(e'_1, e'_2) \in \ell \cap E$, entonces $e'_2 \leq e_2$.

De este modo podemos particionar al conjunto E en E_1 y E_2 , donde $E_2 := E \setminus E_1$. Si los conjuntos E_1 y E_2 son medibles, entonces E también lo sería y obtendríamos de nuevo una contradicción. En conclusión, alguno de ellos no es medible; llamémosle G .

Para cualquier $x_0 \in \mathbb{R}$ denotemos mediante ℓ_{x_0} a la recta definida por la ecuación $x = x_0$. Se tienen dos casos, $\ell_{x_0} \cap G = \emptyset$ o $\ell_{x_0} \cap G = \{(x_0, y_0)\}$, para algún $y_0 \in \mathbb{R}$. Entonces la regla

$$x \mapsto \begin{cases} y_0 & \text{si } \ell_{x_0} \cap G \neq \emptyset, \\ 0 & \text{en otro caso} \end{cases}$$

define una función en \mathbb{R} . Observemos que la gráfica de esta función es un conjunto no medible de \mathbb{R}^2 , pues contiene a G . Haciendo f idéntica a la función indicadora de dicha gráfica se sigue que f_x y f_y son medibles; sin embargo, f como función de \mathbb{R}^2 no lo es.

2.1.1. Espacios Polacos

Como fue discutido en la sección anterior, no basta con medibilidad *entrada a entrada* para obtener medibilidad conjunta. Resulta que bajo el contexto de espacios Polacos, añadiendo continuidad como hipótesis obtenemos medibilidad conjunta. Esto quedará específicamente enunciado en el teorema 2.1.5. Comenzaremos con una breve introducción a espacios métricos.

Una *métrica* o *distancia* en un conjunto S no vacío es una función $d : S \times S \rightarrow \mathbb{R}^+$ que satisface las siguientes propiedades para cualesquiera $p, q, r \in S$:

1. $d(p, q) = 0$ si, y sólo si, $p = q$ y
2. (desigualdad triangular) $d(p, q) \leq d(p, r) + d(r, q)$,

donde \mathbb{R}^+ es el conjunto de los números reales positivos, incluyendo al 0. El ejemplo más natural de distancia es la función valor absoluto $|\cdot|$ en \mathbb{R} , definida por $d(p, q) = |p - q|$ y cuya generalización a \mathbb{R}^m está dada mediante la relación

$$d((p_1, \dots, p_m), (q_1, \dots, q_m)) = \sqrt{\sum_{i=1}^m (p_i - q_i)^2}.$$

A esta función distancia se le conoce como *métrica usual* o *euclidiana*.

Sean S es un conjunto y d una función distancia, entonces a la pareja (S, d) le llamaremos *espacio métrico*. Ahora, si (S, d) es un espacio métrico, r un número real positivo y $s \in S$, definimos la *bola abierta con centro en s y radio r* como el conjunto

$$B_r(s) := \{q \in S : d(s, q) < r\}.$$

Si declaramos como abierto a cualquier A subconjunto de S que verifique para

cualquier $s \in A$ la existencia de un real positivo r tal que $B_r(s) \subseteq A$. Entonces fácilmente se prueba que $\tau_d := \{A \subseteq S : A \text{ es abierto}\}$ satisface las propiedades

1. los subconjuntos \emptyset y S pertenecen a τ_d ;
2. la intersección de cualesquiera dos elementos en τ_d es también un elemento de τ_d ; y
3. la unión de cualquier colección de subconjuntos abiertos pertenece a τ_d .

Es decir, τ_d genera una topología para S .

Si (S, τ) es un espacio topológico, diremos que es *metrizable* si podemos encontrar una métrica d en S tal que $\tau_d = \tau$, donde τ_d sigue la definición del párrafo anterior. Si además S equipado con la métrica d resulta ser completo, o sea que cualquier sucesión de Cauchy definida en él sea convergente, entonces diremos que el espacio (S, τ) es *completamente metrizable*.

Definición 2.1.3. Un espacio topológico será llamado *espacio polaco* si contiene un subconjunto denso numerable (es separable) y es completamente metrizable.

Claramente \mathbb{R} equipado con la topología usual es un espacio polaco.

Algunos resultados elementales sobre espacios Polacos se encuentran en los siguientes enunciados.

1. Todo subespacio cerrado de un espacio polaco es, él mismo, un espacio polaco.
2. El producto topológico de una familia numerable de espacios polacos es un espacio polaco.

Este último nos permiten concluir que \mathbb{R}^m con lo topología usual es también un espacio polaco para todo $m \in \mathbb{N}$.

Proposición 2.1.4. Sea (S, d) un espacio completamente metrizable y $f : S \rightarrow \mathbb{R}$ una función continua. Si $f(s_0) > a$, entonces existen $l, m \in \mathbb{N}$ tales que

$$f\left(B_{\frac{1}{l}}(s_0)\right) \subseteq \left(a + \frac{1}{m}, \infty\right).$$

Demostración. Sean S y f como en las hipótesis. Elegimos $\epsilon > 0$ de manera que $\epsilon < |f(s_0) - a|$. Dado que f es continua, existe $\delta > 0$ para el cual la contención

$$f(B_\delta(x_0)) \subseteq B_\epsilon(f(x_0))$$

es cierta.

Por hipótesis y por la elección de ϵ , si hacemos $r = f(s_0) - \epsilon - a$, entonces $0 < r$. La propiedad arquimediana implica que existe $m \in \mathbb{N}$ suficientemente grande tal que $\frac{1}{m} < r$. Observemos que de lo anterior se sigue que $a + \frac{1}{m} < f(s_0) - \epsilon$. Ahora, si $q \in B_\epsilon(f(x_0))$, entonces es cierto que $f(s_0) - \epsilon < q$ y por transitividad, $a + \frac{1}{m} < q$. Esto último prueba que

$$B_\epsilon(f(s_0)) \subseteq \left(a + \frac{1}{m}, \infty\right).$$

Elegiendo $l \in \mathbb{N}$ de tal manera que $\frac{2}{l} < \delta$ se obtiene el resultado deseado. □

Concluimos la sección con uno de los resultados centrales del capítulo.

Teorema 2.1.5. *Sean (S, \mathcal{A}) un espacio medible y (T, \mathcal{B}) un espacio Polaco equipado con la σ -álgebra de los borelianos. Si $f : S \times T \rightarrow \mathbb{R}$ es una función que satisface lo siguiente*

1. *para todo $s \in S$, f_s es continua; y*
2. *f^t es \mathcal{A} -medible para cualquier $t \in T$.*

Entonces f es $\mathcal{A} \times \mathcal{B}$ medible.

Demostración. Como T es un espacio separable, podemos considerar $D \subseteq T$ denso numerable. Sea $\{d_n : n \in \mathbb{N}\}$ una enumeración de sus elementos, para cada $n \in \mathbb{N}$ definimos la colección

$$\mathcal{U}_n := \{B_{1/k}(d_n) : k \in \mathbb{N}\}.$$

Veamos ahora que $\bigcup_{n \in \mathbb{N}} \mathcal{U}_n$ es cubierta de T . Si $t \in T$, entonces $B_1(t) \cap D$ es no vacía; sea $n \in \mathbb{N}$ tal que $d_n \in B_1(t) \cap D$. Luego, $t \in B_1(d_n)$, es decir $t \in \mathcal{U}_n$.

Para mostrar la medibilidad de f , probaremos que $f^{-1}(a, \infty)$ es un elemento de $\mathcal{A} \times \mathcal{B}$. Esto lo haremos demostrando la siguiente afirmación.

Afirmación 2.1.6.

$$f^{-1}(a, \infty) = \bigcup_{m \geq 1} \bigcup_{l \geq 1} \bigcap_{k \geq l} \bigcup_{n \geq 1} \left[(f^{d_n})^{-1} \left(a + \frac{1}{m}, \infty \right) \times B_{1/k}(d_n) \right]$$

Supongamos que la afirmación anterior es cierta. Entonces hemos terminado, pues la σ -álgebra $\mathcal{A} \times \mathcal{B}$ es generada por la colección $\{A \times B : A \in \mathcal{A}, B \in \mathcal{B}\}$. Como f^{d_n} es \mathcal{A} -medible para cualquier $n \in \mathbb{N}$, entonces $(f^{d_n})^{-1} \left(a + \frac{1}{m}, \infty \right) \in \mathcal{A}$ y es claro que cualquier bola abierta pertenece a los borelianos \mathcal{B} . Al considerar sólo uniones e intersecciones, a lo más, numerables, aseguramos que $f^{-1}(a, \infty) \in \sigma(\mathcal{A} \times \mathcal{B})$ y así f es $\mathcal{A} \times \mathcal{B}$ medible. Procederemos ahora a demostrar la afirmación.

Sea (s_0, t_0) tal que $f(s_0, t_0) > a$, entonces $f_{s_0}(t_0) > a$. Como f_{s_0} es continua, se sigue de la proposición 2.1.4 que existen naturales l_0 y m_0 de tal manera que

$$f_{s_0} \left(B_{2/l_0}(t_0) \right) \subseteq \left(a + \frac{1}{m_0}, \infty \right).$$

Sea $k \geq l_0$ y supongamos que $t_0 \in B_{1/k}(d_n)$, para algún $n \in \mathbb{N}$ dependiente de k . Si $t \in B_{1/k}(d_n)$, entonces $|t - t_0| \leq |t - d_n| + |d_n - t_0| < \frac{2}{k} \leq \frac{2}{l_0}$, lo cual muestra que $B_{1/k}(d_n) \subset B_{2/l_0}(t_0)$. Esto último implica que $f_{s_0} \left(B_{1/k}(d_n) \right) \subseteq \left(a + \frac{1}{m_0}, \infty \right)$. Como k ha sido arbitrario, lo anterior es válido para todo $k \geq l_0$, lo cual demuestra que el lado izquierdo de la ecuación está contenido en el derecho.

Conversamente, supongamos que (s_0, t_0) pertenece al lado derecho de la ecuación. Es decir, existen $m_0, l_0 \in \mathbb{N}$ tales que para cualquier $k \geq l_0$ existe un natural $n(k)$ de modo que $s_0 \in (f^{d_{n(k)}})^{-1} \left(a + \frac{1}{m_0}, \infty \right)$ y $t_0 \in B_{1/k}(d_{n(k)})$. Observemos que la colección $\{d_{n(k)}\}_k$ define una sucesión convergente a t_0 . De la continuidad de f_{s_0} se tiene que $f_{s_0}(t_0) \geq a + \frac{1}{m_0} > a$. Luego, $(s_0, t_0) \in^{-1} (a, \infty)$. Por tanto, se concluye la igualdad deseada. \square

2.2. Desintegración

Para lo que resta de esta sección estaremos trabajando con la notación siguiente. Sea S un conjunto no vacío y \mathcal{A} una σ -álgebra en S . Denotaremos mediante $\mathcal{M}(S)$ a la colección de todas las medidas definidas en el espacio (S, \mathcal{A}) ; y como $\mathcal{P}(S)$ al conjunto de medidas de probabilidad sobre el respectivo espacio. Llamaremos $M^+(S)$ a la colección de todas aquellas funciones $f : S \rightarrow \mathbb{R}$ que son positivas y \mathcal{A} -medibles; y naturalmente a $M(S)$ cuando simplemente sean funciones medibles.

Definición 2.2.1. Sean $(S, \mathcal{A}), (T, \mathcal{B})$ espacios medibles, y λ, μ medidas σ -finitas en \mathcal{A}, \mathcal{B} respectivamente. Sea $X : S \rightarrow T$ es una función medible. Diremos que la colección $\{\lambda_t\}_{t \in T} \subseteq \mathcal{M}(\mathcal{A})$ es una desintegración de λ respecto a X y μ , si para cualesquiera $t \in T$ y $f \in M^+(S)$

1. λ_t es σ -finita y $\lambda_t(\{X \neq t\}) = 0$ μ -casi donde sea
2. La asignación $t \mapsto \int f(s)\lambda_t(ds)$ es una función medible
3. $\int f(s)\lambda(ds) = \int \int f(s)\lambda_t(ds)\mu(dt)$

Cuando μ coincide con F_X , la distribución de X bajo λ , simplemente diremos que es una X -desintegración.

Un caso de particular interés surge al considerar a λ como una medida en un espacio producto. Esto se ilustrará en el siguiente ejemplo.

Ejemplo 2.2.2. Supongamos que $\lambda = \nu \otimes \mu$ en el espacio producto $S \times T$, y consideremos a X como la función cuya regla de correspondencia es $(s, t) \mapsto t$. Por ejemplo, λ puede corresponder a la medida de Lebesgue en \mathbb{R}^2 y μ a la medida de Lebesgue en \mathbb{R} . Si pensamos que $S \times t$ es una copia de S que vive encajada en el espacio producto, entonces λ_t representaría la copia respectiva de ν . Con ayuda del Teorema de Fubini, podemos mostrar que $\{\lambda_t\}$ es una (X, μ) desintegración de λ . Volviendo al caso donde λ representa la medida de Lebesgue en \mathbb{R}^2 , se tiene que $\lambda \circ X^{-1}$ no será σ -finita a menos que ν sea finita. Es por ello que resulta conveniente que en la definición de una (X, μ) desintegración, μ no necesariamente sea la distribución de X ($\lambda \circ X^{-1}$). Más aún, no hay manera de obtener una desintegración si ν no es finita.

Como podemos observar en el ejemplo anterior, la existencia de una desintegración no está siempre garantizada. Es por ello que necesitaremos del siguiente resultado.

Teorema 2.2.3. (*Existencia*) Sean S un espacio métrico equipado con una medida de Radon σ -finita λ , y (T, \mathcal{B}) un espacio medible. Sea $X : S \rightarrow T$ una función medible, y sea μ medida σ -finita en \mathcal{B} tal que $F_X \ll \mu$. Si \mathcal{B} puede ser generada por una colección numerable de conjuntos y $\{\{t\} : t \in T\} \subseteq \mathcal{B}$. Entonces λ tiene $(T-\mu)$ -desintegración. Además, si $\{\lambda_t^*\}_{t \in T}$ es otra $(T-\mu)$ -desintegración, debe tenerse que

$$\mu(\{t \in T : \lambda_t^* \neq \lambda_t\}) = 0.$$

Si estamos desintegrando una medida producto y nuestras medidas ya cuentan con densidades respecto a otras medidas, las densidades de la desintegración quedan de la siguiente manera.

Teorema 2.2.4. (*En el contexto anterior*) Sea P una medida de probabilidad con densidad f respecto a una medida σ -finita λ . Si λ tiene una (X, μ) desintegración $\{\lambda_t : t \in T\}$, entonces se cumple lo siguiente

1. La distribución de X bajo P ($P \circ X^{-1}$) tiene densidad $q(t) := \int f(d)\lambda_t(ds)$ con respecto a μ .
2. Si f_t es una función $\mathcal{A} \times \mathcal{B}$ medible tal que $q(t)f_t = f$, entonces P_t definida como $\frac{dP_t}{d\lambda_t} := f_t$ es una medida de probabilidad para F_X casi todo t y P tiene distribuciones condicionales $\{P_t : t \in T\}$.

2.3. Variación Total

Dado (S, \mathcal{A}) un espacio medible, en la presente sección definiremos una distancia sobre el espacio de funciones $\mathcal{P}(S)$ y discutiremos sobre su relación con el concepto de convergencia en distribución.

Definición 2.3.1. Para cualesquiera $\mu, \nu \in \mathcal{P}(S)$ definimos la distancia *variación total* entre ellas como sigue,

$$\|\mu - \nu\|_{VT} = \sup_{A \in \mathcal{A}} |\mu(A) - \nu(A)|$$

Recordemos que la expresión $\mu(A) - \nu(A)$ es equivalente a la resta de integrales $\int_A \mu_1(ds) - \int_A \mu_2(ds)$. Entonces, si hacemos $B := \{s \in S : \mu(s) > \nu(s)\}$ se tienen las siguientes desigualdades,

$$\begin{aligned} \int |\mu(s) - \nu(s)|(ds) &= \int_B (\mu(s) - \nu(s))(ds) + \int_{S \setminus B} (\nu(s) - \mu(s))(ds) \\ &= \int_B (\mu(s) - \nu(s))(ds) - \int_{S \setminus B} (\mu(s) - \nu(s))(ds) \quad (2.1) \\ &\leq 2 \sup_{A \in \mathcal{A}} \left| \int_A (\mu(s) - \nu(s))ds. \right| \end{aligned}$$

Por otro lado, para cualquier $A \in \mathcal{A}$

$$\begin{aligned} \left| \int_A (\mu(s) - \nu(s))ds \right| &= \max \left\{ \int_A (\mu(s) - \nu(s))(ds), \int_A (\nu(s) - \mu(s))(ds) \right\} \\ &= \leq \max \left\{ \int_{A \cap B} (\mu(s) - \nu(s))(ds), \int_{A \cap (S \setminus B)} (\nu(s) - \mu(s))(ds) \right\} \\ &\leq \max \left\{ \int_B (\mu(s) - \nu(s))(ds), \int_{S \setminus B} (\nu(s) - \mu(s))(ds) \right\} \\ &= \int_B (\mu(s) - \nu(s))ds \\ &= \frac{1}{2} \int_S |\mu(s) - \nu(s)|ds. \end{aligned} \quad (2.2)$$

La última desigualdad es consecuencia de que $\int_S (\mu(s) - \nu(s))ds = 0$, lo cual implica que debe tenerse $\int_B (\mu(s) - \nu(s))ds = \int_{S \setminus B} (\nu(s) - \mu(s))ds$. Como A fue arbitrario, entonces podemos concluir que

$$\sup_{A \in \mathcal{A}} \left| \int_A (\mu(s) - \nu(s))ds \right| \leq \frac{1}{2} \int_S |\mu(s) - \nu(s)|ds$$

y en conjunto con 2.1 obtenemos

$$\sup_{A \in \mathcal{A}} \left| \int_A (\mu(s) - \nu(s)) ds \right| = \frac{1}{2} \int_S |\mu(s) - \nu(s)| ds.$$

Ahora, si $h : S \rightarrow \mathbb{R}$ es una función \mathcal{A} medible tal que $|h| \leq 1$ tenemos que,

$$\begin{aligned} \frac{1}{2} \left| \int h(s) \mu(ds) - \int h(s) \nu(ds) \right| &\leq \frac{1}{2} \int |h(s)| |\mu - \nu|(ds) \\ &\leq \frac{1}{2} \int |\mu - \nu|(ds) \\ &= \sup_{A \in \mathcal{A}} \left| \int_A (\mu(s) - \nu(s)) ds \right|. \end{aligned} \tag{2.3}$$

Si tomamos el supremo, se conserva la desigualdad y entonces

$$\frac{1}{2} \sup_{|h| \leq 1} \left| \int h(s) \mu(ds) - \int h(s) \nu(ds) \right| \leq \sup_{A \in \mathcal{A}} \left| \int_A (\mu(s) - \nu(s)) ds \right|.$$

Usando 2.2 se obtiene la desigualdad conversas

$$\frac{1}{2} \sup_{|h| \leq 1} \left| \int h(s) \mu(ds) - \int h(s) \nu(ds) \right| \geq \sup_{A \in \mathcal{A}} \left| \int_A (\mu(s) - \nu(s)) ds \right|.$$

Con esto hemos demostrado expresiones equivalentes para encontrar la variación total entre dos medidas de probabilidad, lo cual se resume en la siguiente

Proposición 2.3.2. *Sean μ, ν medidas de probabilidad definidas en el mismo espacio (S, \mathcal{A}) . Los siguientes enunciados son ciertos.*

1. $\|\mu - \nu\|_{VT} = \frac{1}{2} \sup_{|h| \leq 1} \left| \int h(s) \mu(ds) - \int h(s) \nu(ds) \right|.$
2. $\|\mu - \nu\|_{VT} = \frac{1}{2} \int |\mu - \nu|(ds).$

A lo largo del capítulo 3 estaremos usando indistintamente cualquiera de estas expresiones para el cálculo de variación total, según sea conveniente.

2.3.1. Convergencia débil

Sean $\mu, \{\mu_n : n \in \mathbb{N}\}$ medidas en un espacio métrico (S, d) equipado con la σ -álgebra de borel $\mathcal{B}(S)$. Decimos que la sucesión $\{\mu_n : n \in \mathbb{N}\}$ *converge débilmente* a μ , lo cual denotaremos como $\mu_n \Rightarrow \mu$, si para cualquier $f \in C_b(S)$ se cumple que

$$\lim_{n \rightarrow \infty} \int f(x) \mu_n(dx) = \int f(x) \mu(dx),$$

donde $C_b(S) := \{f : S \rightarrow \mathbb{R} : f \text{ es función continua y acotada}\}$.

Si $S = \mathbb{R}$ y hacemos

$$F_n(x) = \mu_n((-\infty, x]) \quad (2.4)$$

$$F(x) = \mu((-\infty, x]) \quad (2.5)$$

para $x \in \mathbb{R}$. Entonces F_n y F resultan ser funciones de distribución. Esta correspondencia motiva la siguiente definición que después relacionaremos con convergencia débil.

Definición 2.3.3. Sean $\{F_n : n \in \mathbb{N}\}$ y F , funciones de distribución. Diremos que la sucesión $\{F_n\}_{n \in \mathbb{N}}$ *converge en distribución* a F si

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

para cualquier punto de continuidad x de F . Cuando esto suceda lo denotaremos mediante $F_n \rightarrow^d F$.

La relación entre estas dos nociones de convergencia se encuentra plasmada en el siguiente resultado.

Teorema 2.3.4. Sean $\{\mu_n : n \in \mathbb{N}\}$ y μ una familia de medidas de probabilidad definidas en \mathbb{R} equipado con la σ -álgebra de los borelianos. Si F_n y F están definidas como en 2.4, entonces las siguientes afirmaciones son equivalentes.

1. $F_n \rightarrow^d F$
2. $\mu_n \Rightarrow \mu$

La prueba del teorema anterior puede ser consultada en [5].

Como toda función de distribución induce naturalmente una medida de probabilidad y con base en la equivalencia anterior, probaremos el lema siguiente.

Lema 2.3.5. Sean $\mu, \{\mu_n : n \in \mathbb{N}\}$ medidas de probabilidad definidas sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ tales que $\lim_{n \rightarrow \infty} \|\mu_n - \mu\|_{VT} = 0$. Entonces $\mu_n \Rightarrow \mu$.

Demostración. Hagamos $A_1 = (\infty, x]$ y observemos que

$$|\mu_n(A_1) - \mu(A_1)| \leq \sup_{A \in \mathcal{A}} |\mu_n(A) - \mu(A)|.$$

Por hipótesis el lado derecho de la igualdad anterior es igual a cero cuando $n \rightarrow \infty$, lo cual implica que al tomar el límite del lado izquierdo, también será cero, concluyendo así que $\mu_n \Rightarrow \mu$.

□

Teorema 2.3.6. (Scheffe) $\{f_n\}_{n \in \mathbb{N}}$ una colección de funciones de densidad en un espacio medible (S, \mathcal{A}, μ) . Supongamos que $f_n \rightarrow f$ μ -casi donde quiera. Entonces,

$$\lim_{n \rightarrow \infty} \int |f_n - f| d\mu = 0.$$

La demostración de este hecho puede ser consultada en [5].

En el teorema 2.3.6 para cada $n \in \mathbb{N}$ la integral $\int f_n d\mu$ define una medida; si la llamamos μ_n el teorema anterior nos permite concluir que $\lim_{n \rightarrow \infty} \|\mu_n - \mu\| = 0$, lo que implica la convergencia débil de μ_n μ . Dicho de otro modo, bajo estas suposiciones, la convergencia de densidades implica convergencia en sus distribuciones.

2.4. Lemas auxiliares

Para facilitar la lectura de las pruebas del capítulo siguiente, hemos condensado un par de resultados en esta sección. Las pruebas aparecen en [2].

Proposición 2.4.1. Sea $(x_n)_{n \in \mathbb{N}}$ una sucesión de números reales, tal que $x_n \rightarrow x$ y $x > 0$. Existe $n \in \mathbb{N}$ suficientemente grande tal que

$$\frac{|x_n^{-1} - x^{-1}|}{x^{-1}} \approx \frac{|x_n - x|}{x}.$$

Demostración. La expansión en series de Taylor de primer orden para x^{-1} alrededor de x_0 es

$$\frac{1}{x} = \frac{1}{x_0} - \frac{1}{x_0^2}(x - x_0) + R,$$

donde R es el residuo (de Lagrange) y es igual a $\frac{1}{x_1^3}(x - x_0)^2$ para algún x_1 que se encuentre entre x y x_0 .

Se tiene la siguiente igualdad:

$$\frac{1}{x} - \frac{1}{x_0} = -\frac{1}{x_0^2}(x - x_0) + R.$$

Luego, multiplicando por $\frac{1}{x_0^{-1}}$

$$\frac{x^{-1} - x_0^{-1}}{x_0^{-1}} = -\frac{1}{x_0}(x - x_0) + \frac{R}{x_0^{-1}}$$

$$\frac{x^{-1} - x_0^{-1}}{x_0^{-1}} = -\frac{1}{x_0}(x - x_0) + \frac{x_0^2}{x_0^2 x_1^3}(x - x_0)^2.$$

Tomamos valores absolutos

$$\frac{|x^{-1} - x_0^{-1}|}{x_0^{-1}} = \frac{|x - x_0|}{x_0} + \left| \frac{x_0}{x_1} \right|^3 \left(\frac{x - x_0}{x_0} \right)^2.$$

Ahora sólo debemos verificar que el término $\left| \frac{x_0}{x_1} \right|^3 \left(\frac{x - x_0}{x_0} \right)^2$ puede hacerse tan pequeño como se quiera.

Sea $\epsilon > 0$ tal que $|x - x_0| < \epsilon$. Para x_1 se cumple que $|x_1 - x_0| < |x - x_0| < \epsilon$. Multiplicando por $\frac{1}{x_0}$, $\frac{|x_1 - x_0|}{x_0} < \frac{|x - x_0|}{x_0} < \frac{\epsilon}{x_0}$ y por transitividad del orden, $\frac{|x_1 - x_0|}{x_0} < \frac{\epsilon}{x_0}$, o equivalentemente $-\frac{\epsilon}{x_0} < \frac{x_1 - x_0}{x_0} < \frac{\epsilon}{x_0}$. Sumando 1, obtenemos $1 - \frac{\epsilon}{x_0} < \frac{x_1}{x_0} < 1 + \frac{\epsilon}{x_0}$. Entonces

$$\left(1 + \frac{\epsilon}{x_0} \right)^{-3} < \left(\frac{x_0}{x_1} \right)^3 < \left(1 - \frac{\epsilon}{x_0} \right)^{-3},$$

pues x^{-3} es decreciente.

Por tanto

$$\left| \frac{x_0}{x_1} \right|^3 \left(\frac{x - x_0}{x_0} \right)^2 < \left(\frac{\epsilon}{x_0} \right)^2 \left(1 - \frac{\epsilon}{x_0} \right)^{-3}.$$

Eligiendo ϵ de tal modo que la razón $\frac{\epsilon}{x_0}$ sea suficientemente pequeña, el término anterior se vuelve despreciable y obtenemos la aproximación deseada.

□

Para el lema 2.4.3 utilizaremos el teorema de convergencia dominada, el cual anunciamos a continuación y cuya demostración puede consultarse en [4].

Teorema 2.4.2. (*Teorema de convergencia dominada en medida*) Sean (S, \mathcal{A}, μ) un espacio de medida y $\{f_n : n \in \mathbb{N}\}$ una colección de funciones integrables respecto a μ tal que $f_n \rightarrow f$, donde $f \in M(S)$. Si existe una función positiva g integrable de modo que para cualquier $n \in \mathbb{N}$ se tenga $|f_n| \leq g$ μ -casi donde sea. Entonces f es una función integrable y además

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu.$$

Lema 2.4.3. Para cada $n \in \mathbb{N}$, sean $b_n : \Theta \rightarrow \mathbb{R}^+$ funciones acotadas y π -integrables. Sea $b : \Theta \rightarrow \mathbb{R}^+$ su límite, tal que para todo $\theta \in \Theta$ y $n > N$, $|b_n(\theta) - b(\theta)| < kn^{-p}$; con $k > 0$ y $p > 1$ fijos. Si definimos $z_n = \int b_n(\theta)\pi(d\theta)$ y $z = \int b(\theta)\pi(d\theta)$, entonces $z_n \rightarrow z$ y $\frac{b_n(\theta)}{z_n} \rightarrow \frac{b(\theta)}{z}$ con radios de convergencia respectivos

$$|z_n - z| < kn^{-p}$$

y

$$\left| \frac{b_n(\theta)}{z_n} - \frac{b(\theta)}{z} \right| < \frac{b(\theta)kn^{-p}}{z^2} + \frac{kn^{-p}}{z},$$

para cualquier $\theta \in \Theta$ y n suficientemente grande.

Demostración. Sea $n \in \mathbb{N}$, definimos

$$M_n := \max\{b_n(\theta) : \theta \in \Theta\},$$

el cual está bien definido pues por hipótesis cada b_n es función acotada. Sea $M := \sup\{M_n : n \in \mathbb{N}\}$, entonces M existe pues la sucesión converge a una función acotada b .

Ahora, definimos la función $g : \Theta \rightarrow \mathbb{R}^+$ como la constante M . Definida de esta forma g es

1. positiva,
2. integrable, pues $\int g(\theta)\pi(d\theta) = M$; y
3. para cualquier $n \in \mathbb{N}$, $|b_n| \leq g$.

Luego, por el teorema de convergencia dominada se sigue que

$$\lim_{n \rightarrow \infty} \int b_n(\theta)\pi(d\theta) = \int b(\theta)\pi(d\theta).$$

Esta última igualdad es equivalente a $z_n \rightarrow z$.

Para verificar la velocidad de convergencia bastan las siguientes desigualdades

$$\left| \int b_n(\theta)\pi(d\theta) - \int b(\theta)\pi(d\theta) \right| \leq \int |b_n(\theta) - b(\theta)|\pi(d\theta) < kn^{-p},$$

de donde se sigue que $\frac{|z_n - z|}{z} < \frac{kn^{-p}}{z}$.

De la proposición 2.4.1 se sigue la existencia de algún $n \in \mathbb{N}$ suficientemente grande (y supongámoslo independiente de θ) tal que $\frac{|z_n^{-1} - z|}{z^{-1}} \approx \frac{|z_n - z|}{z}$. Esto nos es de utilidad, pues ahora podemos ocupar la tasa de convergencia que obtuvimos en la parte anterior de la prueba. Considerando que $\frac{|z_n^{-1} - z^{-1}|}{z^{-1}} < \frac{kn^{-p}}{z}$ o equivalentemente:

$$\frac{1}{z} - \frac{kn^{-p}}{z^2} < \frac{1}{z_n} < \frac{1}{z} + \frac{kn^{-p}}{z^2}. \quad (2.6)$$

Por hipótesis $|b_n(\theta) - b| < kn^{-p}$, pero resultará conveniente trabajar con la expresión $b(\theta) - kn^{-p} < b_n(\theta) < b(\theta) + kn^{-p}$.

Como resultado de multiplicar la última expresión por (2.6) obtenemos,

$$\frac{b(\theta)}{z} - \frac{b(\theta)kn^{-p}}{z^2} - \frac{kn^{-p}}{z} + \frac{k^2n^{-2p}}{z^2} < \frac{b_n(\theta)}{z_n} < \frac{b(\theta)}{z} + \frac{b(\theta)kn^{-p}}{z^2} + \frac{kn^{-p}}{z} + \frac{k^2n^{-2p}}{z^2}.$$

Ignorando los dos términos de orden $2p$ en la desigualdad anterior, hemos mostrado que

$$\left| \frac{b_n(\theta)}{z_n} - \frac{b(\theta)}{z} \right| < \frac{b(\theta)kn^{-p}}{z^2} + \frac{kn^{-p}}{z},$$

lo que concluye la prueba del teorema. \square

Corolario 2.4.4. Sea $h : \Theta \rightarrow \mathbb{R}$ un función medible, para cada $n \in \mathbb{N}$ definimos

$$\hat{h}_n = \int h(\theta) \frac{b_n(\theta)}{z_n} \pi(d\theta),$$

y

$$\hat{h} = \int h(\theta) \frac{b(\theta)}{z} \pi(d\theta).$$

Si las integrales anteriores existen, entonces

$$|\hat{h}_n - \hat{h}| < E_y[|h|] \frac{kn^{-p}}{z} + E_0[|h|] \frac{kn^{-p}}{z},$$

donde $E_y[|h|] = \int |h(\theta)| \frac{b(\theta)}{z} \pi(d\theta)$ y $E_0[|h|] = \int |h(\theta)| \pi(d\theta)$. Si además h es no negativa y acotada, entonces \hat{h}_n y \hat{h} definen medidas de probabilidad p_n y p respectivamente, para las cuales se cumple que

$$\|p_n - p\|_{VT} < \frac{kn^{-p}}{z}.$$

Demostración. Observemos que

$$|\hat{h}_n - \hat{h}| \leq \int |h(\theta)| \left| \frac{b_n(\theta)}{z_n} - \frac{b(\theta)}{z} \right| \pi(d\theta).$$

Por el lema anterior debe ser claro que

$$\int |h(\theta)| \left| \frac{b_n(\theta)}{z_n} - \frac{b(\theta)}{z} \right| \pi(d\theta) < \int |h(\theta)| \left(\frac{b(\theta)kn^{-p}}{z^2} + \frac{kn^{-p}}{z} \right) \pi(d\theta),$$

y utilizando la linealidad de la integral obtenemos el resultado.

Supongamos ahora que $|h(\theta)| \leq 1$. Entonces $E_y[|h|] \leq 1$ y $E_0[|h|] \leq 1$. Con esto obtenemos que

$$\left| \int h(\theta)p_n(d\theta) - h(\theta)p(d\theta) \right| < 2\frac{k}{z}n^{-p},$$

es decir

$$\frac{1}{2} \max_{|h| \leq 1} \left| \int h(\theta)p_n(d\theta) - h(\theta)p(d\theta) \right| < \frac{k}{z}n^{-p}.$$

Por lo tanto, $\|p_n - p\|_{VT} < \frac{k}{z}n^{-p}$.

□

Capítulo 3

Aspectos teóricos

3.1. Existencia de la distribución posterior

Sea $Y \in \mathcal{Y} \subseteq \mathbb{R}^m$ una colección de datos y $\{P_\theta : \theta \in \Theta\}$ una familia de modelos probabilísticos para Y , los cuales supondremos que cuentan con densidad $f_\theta(y)$ con respecto a una medida σ -finita λ . Si equipamos a \mathcal{Y} con la topología relativa a la usual en \mathbb{R}^m , es cierto que \mathcal{Y} resulta ser un espacio separable y completamente metrizable; por esta razón, en lo subsecuente supondremos que \mathcal{Y} es Polaco.

Primero, necesitaremos una medida en el espacio $\mathcal{Y} \times \Theta$, pues es donde vivirá nuestro modelo conjunto. Para ello, necesitaremos el siguiente lema.

Lema 3.1.1. *Sea $g : \mathbb{R}^m \times \Theta \rightarrow \mathbb{R}^+$ un función $\lambda \times \pi$ medible. Si $g(y, \theta)f_\theta(y)$ es una función medible. Entonces \mathbb{Q} define una medida de probabilidad conjunta en $\mathbb{R}^m \times \Theta$, donde*

$$\begin{aligned} \int g(y, \theta)\mathbb{Q}(dy, d\theta) &:= \int \int g(y, \theta)P_\theta(dy)\pi(d\theta) \\ &= \int \int g(y, \theta)f_\theta(y)\lambda(dy)\pi(d\theta) \end{aligned}$$

Demostración. Dado que π es una medida finita, pues es de probabilidad, es σ -finita. Como λ también lo es, se sigue del teorema de Tonelli que $\theta \rightarrow \int g(y, \theta)f(y|\theta)\lambda(dy)\pi(d\theta)$ es una función medible y que $\int g(y, \theta)\mathbb{Q}(dy, d\theta)$ coincide con las integrales iteradas. Por lo tanto, es una media en el espacio producto.

Además, haciendo g idéntica a la función constante 1 se obtiene que $\mathbb{Q}(\mathbb{R}^m \times \Theta) = 1$, lo que muestra que es una medida de probabilidad. \square

Lema 3.1.2. (Teorema de Bayes) *Sea Θ un espacio de Banach separable. Si para cualquier $y \in \mathcal{Y}$ se tiene que $\theta \mapsto f_\theta(y)$ es una función continua. Entonces se cumplen las siguientes afirmaciones,*

1. Existe \mathbb{Q} la medida conjunta en $\mathcal{Y} \times \Theta$.
2. \mathbb{Q} tiene θ -desintegración \mathbb{Q}_θ , con densidad f_θ .
3. La y -desintegración \mathbb{Q}_y de \mathbb{Q} existe y es la distribución posterior. Además, si $g \in M^+(\Theta)$, se tiene que

$$\int g(\theta) \mathbb{Q}_y(d\theta) = \frac{\int g(\theta) f_\theta(y) \pi(d\theta)}{\int f_\theta(y) \pi(d\theta)},$$

es decir, $\frac{\partial \mathbb{Q}_y}{\partial \pi} \propto f_\theta(y)$, para todo $y \in \mathcal{Y}$.

Demostración. Por hipótesis, $\theta \mapsto f_\theta(y)$ es una función continua. Aplicando el teorema 2.1.5, obtenemos que $f_\theta(y)$ es una función $\lambda \times \pi$ medible. Luego, si $g(y, \theta)$ es positiva y $\lambda \times \pi$ medible, se sigue que $g(y, \theta) f_\theta(y)$ también lo es. Lo anterior nos permite definir una medida de probabilidad \mathbb{Q} en el espacio producto $\mathbb{R}^m \times \Theta$, via el lema anterior. Si r y p son las funciones proyección en \mathbb{R}^m y Θ respectivamente, entonces tienen como distribución conjunta a \mathbb{Q} .

Por otro lado, $\mathcal{Y} \times \Theta$ es un espacio Polaco al ser producto de espacios Polacos, lo cual implica que \mathbb{Q} es una medida de Radon. Además las distribuciones de r y p respecto a \mathbb{Q} son absolutamente continuas respecto a π y λ . Con esto, se cumplen las hipótesis del Teorema de existencia de Desintegración (2.2.3). Por lo tanto, existen la θ -desintegración y la y -desintegración.

Respecto a las densidades de los puntos 2 y 3, esto se sigue del teorema 2.2.4. \square

3.2. Convergencia y consistencia

Para la presente sección trabajaremos bajo las suposiciones siguientes. Θ y V denotarán al espacio parametral y al espacio que contiene a nuestras variables de

estado, respectivamente. Supondremos que ambos tienen estructura de espacios de Banach. A su vez, supondremos que \mathcal{F} y \mathcal{H} son funcionales Borel medibles, donde $\mathcal{F} : \Theta \rightarrow V$ representará nuestro regresor y $\mathcal{H} : V \rightarrow \mathcal{Y}$ al funcional observacional, para algún \mathcal{Y} subconjunto de \mathbb{R}^{m+s} . Con dichas suposiciones obtenemos que $\mathcal{H} \circ \mathcal{F}$ es una función Borel medible entre el espacio de parámetros Θ y el espacio que contiene la muestra observada \mathbb{R}^m . Aquí m es el tamaño de muestra y s representa la presencia de posibles parámetros adicionales.

También necesitaremos que nuestra colección de datos, la cual estaremos denotando como y , siga una distribución que sea absolutamente continua respecto a la medida de Lebesgue λ . Es decir, que cuente con una densidad $f_o(y|\eta)$ para cualquier. La familia paramétrica de modelos para nuestra muestra y se encontrará definida mediante la familia de densidades, llamémosla $f(y|\theta)$. Así

$$f(y|\theta) = f_o(y|\mathcal{H}(\mathcal{F}(\theta))).$$

A manera de ejemplo, consideremos el caso usual de ruido Gaussiano. En este caso $y_j = \mathcal{H}(\mathcal{F}(\theta)) + \sigma\epsilon_j$, con $\epsilon_j \sim \mathcal{N}(0, 1)$. Entonces $y_j \sim \mathcal{N}(\mathcal{H}(\mathcal{F}(\theta)), \sigma^2)$ y por lo tanto

$$f_o(y|\eta) = \prod_{j \leq m} \sigma^{-1} \rho \left(\frac{y_j - \eta_j}{\sigma} \right),$$

$$\text{donde } \rho(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Típicamente solo contamos con una discretización $\mathcal{F}^{\alpha(n)}$ de nuestro regresor. Esta discretización puede depender de un refinamiento o tamaño de paso n y dicha dependencia quedará expresada en el término $\alpha(n)$. En consecuencia, llamaremos $f^n(y|\theta)$ a la versión numérica de la verosimilitud de nuestro datos que resulta al considerar esta aproximación, es decir

$$f^n(y|\theta) = f_o(y|\mathcal{H}(\mathcal{F}^{\alpha(n)}(\theta))).$$

Adicionalmente, consideraremos a π_k como una aproximación de nuestra distribución a priori π . La construcción mediante la cual obtenemos π_k se hará explícita en la última sección del capítulo. Por ahora sólo estaremos bajo la suposición de que π_k es una distribución sobre Θ . Cabe mencionar que la notación acordada en los párrafos

anteriores será la que utilicemos a lo largo de los subsecuentes capítulos.

Adicionalmente y en lo sucesivo, trabajaremos con los siguientes supuestos. Para cada $y \in \mathcal{Y}$

- (1) el modelo observacional $f_o(y|\eta)$ es uniformemente Lipschitz en su dominio, es decir que para cualesquiera η, η' se da la desigualdad

$$|f_o(y|\eta) - f_o(y|\eta')| < L|\eta - \eta'|,$$

con $L > 0$; y

- (2) $f_o(y|\eta)$ es acotado λ -casi seguramente.
- (3) $\mathcal{H} \circ \mathcal{F}$ y $\mathcal{H} \circ \mathcal{F}^{\alpha(n)}$ son asignaciones continuas.
- (4) Para algún funcional $|\cdot|$,

$$||\mathcal{H}(\mathcal{F}(\theta)) - \mathcal{H}(\mathcal{F}^{\alpha(n)}(\theta))|| < K_0|\alpha(n)|^p. \quad (3.1)$$

Esto significa que somos capaces de controlar el error en la discretización de nuestro regresor y más aún, que esta cota es independiente de θ . Observemos que como consecuencia de (1), si hacemos $K_1 = LK_0$ se sigue que

$$|f^n(y|\theta) - f(y|\theta)| = |f_o(y|\mathcal{H}(\mathcal{F}^{\alpha(n)}(\theta))) - f_o(y|\mathcal{H}(\mathcal{F}(\theta)))| < K_1|\alpha(n)|^p. \quad (3.2)$$

Siendo ésta última una cota para la verosimilitud, que de igual manera es independiente de θ pues K_0 lo es.

Notemos que el punto (4) es condición suficiente para asegurar la existencia de las medidas posteriores, ya que ésta implica que $\theta \mapsto f(y|\theta)$ y $\theta \mapsto f^n(y|\theta)$ son asignaciones continuas.

Al considerar la verosimilitud teórica y su versión numérica, obtenemos las respectivas medidas posteriores a las que denotaremos como \mathbb{Q}_y y \mathbb{Q}_y^n ; y $\mathbb{Q}_y^{n,k}$ si adicionalmente consideramos el truncamiento π_k ; y \mathbb{Q}_y^k como resultado de utilizar a π_k como distribución apriori y la verosimilitud teórica.

Si nuestra a priori es π_k y la verosimilitud es la teórica, entonces obtendremos la medida posterior \mathbb{Q}_y^k .

Definimos también $Z(y) = \int f(y|\theta)\pi(d\theta)$ la constante de normalización de la medida posterior y análogamente $Z^n(y)$, $Z^{n,k}(y)$ y $Z^k(y)$.

A continuación probaremos la consistencia al usar la discretización y la aproximación π_k , y también de qué manera se da la convergencia entre $\mathbb{Q}_y^{n,k}$, \mathbb{Q}_y^n , \mathbb{Q}_y^k y \mathbb{Q}_y . Se probará que esta convergencia es en distribución y para la velocidad de convergencia utilizaremos la distancia variación total que hemos definido en los preliminares.

Teorema 3.2.1.

1. $\mathbb{Q}_y^{n,k} \Rightarrow \mathbb{Q}_y^k$ y $\mathbb{Q}_y^n \Rightarrow \mathbb{Q}_y$, cuando $n \rightarrow \infty$
2. Supongamos que $\pi_k \Rightarrow \pi$. Entonces, $\mathbb{Q}_y^{n,k} \Rightarrow \mathbb{Q}_y^n$ y $\mathbb{Q}_y^k \Rightarrow \mathbb{Q}_y$

Demostración. Bajo el supuesto de que $f(y|\theta)$ es una función acotada, se sigue que también es π_k integrable, pues esta última es medida finita. Luego, como $f^n(y|\theta) \rightarrow f(y|\theta)$, por el teorema de convergencia dominada se tiene que

$$\lim_{n \rightarrow \infty} \int f^n(y|\theta)\pi_k(d\theta) = \int f(y|\theta)\pi_k(d\theta),$$

en otras palabras $Z^{n,k}(y) \rightarrow Z^k(y)$.

Luego, como $Z^k(y)$ y $Z^{n,k}(y)$ son ambos positivos, existe el límite (cuando $n \rightarrow \infty$) de $[Z^{n,k}(y)]^{-1}$ y es igual a $[Z^k(y)]^{-1}$. De aquí podemos concluir que

$$\frac{f^n(y|\theta)}{Z^{n,k}(y)} \rightarrow \frac{f(y|\theta)}{Z^k(y)}.$$

Observemos que $\frac{f^n(y|\theta)}{Z^{n,k}(y)}$ y $\frac{f(y|\theta)}{Z^k(y)}$ son densidades de $\mathbb{Q}_y^{n,k}$ y \mathbb{Q}_y^k respectivamente, lo que nos permite aplicar el lema de Scheffe para concluir que $\mathbb{Q}_y^{n,k} \Rightarrow \mathbb{Q}_y^k$; y con esto se concluye la prueba de 1.

Omitiremos la prueba de que $\mathbb{Q}_y^n \Rightarrow \mathbb{Q}_y$ ya que es completamente análoga a la anterior.

Para probar la segunda afirmación primero debemos observar que $f^n(y|\theta) \in C_b(\Theta)$ lo cual, de acuerdo con la definición de convergencia débil, implica que

$$\lim_{k \rightarrow \infty} \int f^n(y|\theta) \pi_k(d\theta) \rightarrow \int f^n(y|\theta) \pi(d\theta).$$

Es decir que $Z^{n,k}(y) \rightarrow Z^n(y)$. De esto podemos concluir que también se da la convergencia de los inversos,

$$[Z^{n,k}(y)]^{-1} \rightarrow [Z^n(y)]^{-1}.$$

Ahora, si $g(\theta)$ es una función perteneciente a $C_b(\Theta)$, entonces el producto $g(\theta)f(y|\theta)$ es también un elemento de $C_b(\Theta)$. Nuevamente utilizando la definición de convergencia débil, obtenemos que

$$\lim_{k \rightarrow \infty} \int g(\theta) f^n(y|\theta) \pi_k(d\theta) \rightarrow \int g(\theta) f^n(y|\theta) \pi(d\theta),$$

de donde se deduce

$$[Z^{n,k}(y)]^{-1} \int g(\theta) f^n(y|\theta) \pi_k(d\theta) \rightarrow \int g(\theta) f^n(y|\theta) \pi(d\theta) [Z^n(y)]^{-1}.$$

Por lo tanto $\mathbb{Q}_y^{n,k} \Rightarrow \mathbb{Q}_y^n$.

Análogamente se obtiene la prueba de $\mathbb{Q}_y^k \Rightarrow \mathbb{Q}_y$.

□

Teorema 3.2.2. *Bajo los supuestos anteriores y para n suficientemente grande, se cumplen las siguientes desigualdades*

$$\|\mathbb{Q}_y^{n,k} - \mathbb{Q}_y^k\|_{VT} < \frac{K_1}{Z^k(y)} |\alpha(n)|^p$$

$$\|\mathbb{Q}_y^n - \mathbb{Q}_y\|_{VT} < \frac{K_1}{Z(y)} |\alpha(n)|^p$$

Demostración. Para la primera afirmación basta con hacer $b_n = f^n(y|\theta)$, $b = f(y|\theta)$,

$z_n = Z^{n,k}(y)$ y $z = Z^k(y)$ en el lema 2.4.3. Luego, por la segunda parte del corolario 2.4.4 se obtiene el resultado.

Análogamente para la segunda afirmación, haciendo $z_n = Z^n(y)$ y $z = Z(y)$. \square

Teorema 3.2.3. *Si $\|\pi_k - \pi\|_{VT} \rightarrow 0$, entonces para $k \in \mathbb{N}$ suficientemente grande, se cumplen las siguientes desigualdades,*

$$1. \|\mathbb{Q}_y^{n,k} - \mathbb{Q}_y^n\|_{VT} < \frac{f^n(y|\hat{\theta}_n)}{Z^n(y)} \|\pi_k - \pi\|_{VT}.$$

$$2. \|\mathbb{Q}_y^k - \mathbb{Q}_y^n\|_{VT} < \frac{f(y|\hat{\theta})}{Z(y)} \|\pi_k - \pi\|_{VT}.$$

Donde $\hat{\theta}_n$ y $\hat{\theta}$ son los valores que maximizan $f^n(y|\cdot)$ y $f(y|\cdot)$ respectivamente.

Demostración. Realizaremos la prueba sólo de la primera desigualdad ya que la prueba es análoga para el segundo resultado.

Sea $h : \Theta \rightarrow \mathbb{R}$ una función medible tal que $|h| \leq 1$. Entonces,

$$\begin{aligned} & \left| \int h(\theta) f^n(y|\theta) \pi_k(d\theta) - \int h(\theta) f^n(y|\theta) \pi(d\theta) \right| = \left| \int h(\theta) f^n(y|\theta) (\pi_k - \pi)(d\theta) \right| \\ & \leq \int |h(\theta)| f^n(y|\theta) |\pi_k - \pi|(d\theta) \leq \int f^n(y|\hat{\theta}_n) |\pi_k - \pi|(d\theta) = f^n(y|\hat{\theta}_n) \|\pi_k - \pi\|_{VT}. \end{aligned}$$

Lo anterior significa que $|b_k - b| \leq f^n(y|\hat{\theta}_n) \|\pi_k - \pi\|_{VT}$.

Notemos ahora que

$$\left| \int f^n(y|\theta) \pi_k(d\theta) - \int f^n(y|\theta) \pi(d\theta) \right| \leq \int f^n(y|\theta) |\pi_k - \pi|(d\theta) < f^n(y|\hat{\theta}_n) \|\pi_k - \pi\|_{VT}.$$

Esto quiere decir que $|Z^{n,k}(y) - Z^n(y)| < f^n(y|\hat{\theta}_n) \|\pi_k - \pi\|_{VT}$.

Al igual que en 2.4.1, elegimos n de modo que

$$\frac{|(Z^{n,k}(y))^{-1} - (Z^n(y))^{-1}|}{(Z^n(y))^{-1}} \approx \frac{|Z^{n,k}(y) - Z^n(y)|}{Z^n(y)},$$

o equivalentemente

$$\left| \frac{1}{Z^{n,k}(y)} - \frac{1}{Z^n(y)} \right| \approx \frac{|Z^{n,k}(y) - Z^n(y)|}{(Z^n(y))^2}.$$

En lo sucesivo de la prueba, denotemos mediante C a $f^n(y|\hat{\theta}_n)||\pi_k - \pi||_{VT}$.

De las igualdades anteriores, tenemos que

$$\frac{b}{Z^n(y)} - \frac{C}{Z^n(y)} - \frac{Cb}{(Z^n(y))^2} + \frac{C^2}{(Z^n(y))^2} < \frac{b_k}{Z^{n,k}(y)} < \frac{b}{Z^n(y)} + \frac{C}{Z^n(y)} + \frac{Cb}{(Z^n(y))^2} + \frac{C^2}{(Z^n(y))^2},$$

e ignorando los términos de orden cuadrático, se sigue que

$$\frac{b}{Z^n(y)} - \left[\left(\frac{1}{Z^n(y)} + \frac{1}{(Z^n(y))^2} \right) C \right] < \frac{b_k}{Z^{n,k}(y)} < \frac{b}{Z^n(y)} + \left[\left(\frac{1}{Z^n(y)} + \frac{1}{(Z^n(y))^2} \right) C \right].$$

Una expresión equivalente a la anterior es la siguiente,

$$\left| \frac{b_k}{Z^{n,k}(y)} - \frac{b}{Z^n(y)} \right| < \left(\frac{b}{Z^n(y)} \frac{1}{Z^n(y)} + \frac{1}{Z^n(y)} \right) C.$$

Luego,

$$\left(\frac{b}{Z^n(y)} \frac{1}{Z^n(y)} + \frac{1}{Z^n(y)} \right) C \leq \left| \left(\frac{b}{Z^n(y)} \frac{1}{Z^n(y)} + \frac{1}{Z^n(y)} \right) C \right|.$$

Como $\left| \frac{b}{Z^n(y)} \right| \leq 1$ y además los términos $Z^n(y)$ y C son positivos, se sigue que

$$\begin{aligned} \left| \left(\frac{b}{Z^n(y)} \frac{1}{Z^n(y)} + \frac{1}{Z^n(y)} \right) C \right| &\leq \left(\left| \frac{b}{Z^n(y)} \right| \frac{1}{Z^n(y)} + \frac{1}{Z^n(y)} \right) C \\ &\leq \frac{2}{Z^n(y)} C. \end{aligned}$$

En conclusión,

$$\left| \int h(\theta) \mathbb{Q}_y^{n,k}(d\theta) - \int h(\theta) \mathbb{Q}_y^n(d\theta) \right| \leq \frac{2}{Z^n(y)} C,$$

lo cual implica que

$$\frac{1}{2} \max_{|h| \leq 1} \left| \int h(\theta) \mathbb{Q}_y^{n,k}(d\theta) - \int h(\theta) \mathbb{Q}_y^n(d\theta) \right| \leq \frac{1}{2} \frac{2}{Z^n(y)} C,$$

con lo que se muestra el resultado. \square

Teorema 3.2.4. *Existen n y k suficientemente grandes, tales que se satisface la siguiente desigualdad $\|\mathbb{Q}_y^{n,k} - \mathbb{Q}_y\|_{VT} < \frac{K_1}{Z^k(y)} |\alpha(n)|^p + \frac{f(y, \hat{\theta})}{Z(y)} \|\pi_k - \pi\|_{VT}$*

Demostración. Trivialmente la siguiente igualdad es cierta.

$$\|\mathbb{Q}_y^{n,k} - \mathbb{Q}_y\|_{VT} = \|\mathbb{Q}_y^{n,k} - \mathbb{Q}_y^k + \mathbb{Q}_y^k - \mathbb{Q}_y\|_{VT}$$

Luego, por propiedades de norma, se cumple que

$$\|\mathbb{Q}_y^{n,k} + \mathbb{Q}_y^k - \mathbb{Q}_y^k - \mathbb{Q}_y\|_{VT} \leq \|\mathbb{Q}_y^{n,k} - \mathbb{Q}_y^k\|_{VT} + \|\mathbb{Q}_y^k - \mathbb{Q}_y\|_{VT}.$$

Finalmente, con las tasas de convergencia obtenidas en 3.2.2 y 3.2.3 se obtiene el resultado. \square

Corolario 3.2.5. $\mathbb{Q}_y^{n,k}$ converge débilmente a \mathbb{Q}_y .

3.3. Cotas de error

Uno de los objetivos principales de este texto, es conocer cómo seleccionar un nivel de discretización n y de aproximación k . La estrategia que seguiremos para lograr esto, será la de comparar los modelos teórico \mathbb{Q}_y y aproximado $\mathbb{Q}_y^{n,k}$ mediante el Factor de Bayes (BF). El BF es la razón posterior de un modelo contra otro; en este caso, asumiendo igual probabilidad. Esto es, el Factor de Bayes es $\frac{p}{1-p}$, donde p representa la probabilidad posterior del modelo numérico. Entonces $p = \frac{Z^{n,k}(y)}{Z^{n,k}(y) + Z(y)}$. Por lo tanto, simplificando la expresión anterior, obtenemos que el BF se reduce a $\frac{Z^{n,k}(y)}{Z(y)}$. Una expresión alternativa y conveniente es

$$\frac{1}{2} \left| \frac{Z^{n,k}(y)}{Z(y)} - 1 \right|,$$

que será con la que trabajemos en la sección.

Lo que haremos será acotar la expresión anterior. Como queremos que la cota no dependa de la muestra, lo que vamos a acotar será el valor esperado del Factor de Bayes Absoluto (EABF). O sea que, acotaremos el término

$$\int \frac{1}{2} \left| \frac{Z^{n,k}(y)}{Z(y)} - 1 \right| Z(y) \lambda(dy).$$

Con esto queremos lograr mantener el término anterior suficientemente pequeño para que el BF esté cercano a 1 y entonces nuestro modelo numérico sea prácticamente igual al teórico.

Teorema 3.3.1. *Definimos $\phi_y(\eta) = -\log f_o(y|\eta)$. Si $\|\pi_k - \pi\|_{VT} \rightarrow 0$ y $\phi_y \in C^1$ λ -casi seguramente, entonces se cumple la siguiente desigualdad*

$$\frac{1}{2} \int |Z^{n,k}(y) - Z^k(y)| \lambda(dy) < \frac{K_0 |\alpha(n)|^p}{2} \int \int f(y|\theta) \|\nabla \phi_y(\mathcal{H}(\mathcal{F}(\theta)))\|_1 \pi_k(d\theta) \lambda(dy).$$

Demostración. Notemos que

$$\left| \frac{Z^{n,k}(y)}{Z(y)} - 1 \right| Z(y) = \left| \frac{Z^{n,k}(y)}{Z(y)} - 1 \right| |Z(y)| = |Z^{n,k}(y) - Z(y)|.$$

Luego, por la desigualdad del triángulo, obtenemos que $|Z^{n,k}(y) - Z(y)| \leq |Z^{n,k}(y) - Z^k(y)| + |Z^k(y) - Z(y)|$. Finalmente, por la linealidad de la integral se concluye que

$$\begin{aligned} \int \frac{1}{2} \left| \frac{Z^{n,k}(y)}{Z(y)} - 1 \right| Z(y) \lambda(dy) &\leq \int \frac{1}{2} [|Z^{n,k}(y) - Z^k(y)| + |Z^k(y) - Z(y)|] \lambda(dy) \\ &= \int \frac{1}{2} |Z^{n,k}(y) - Z^k(y)| \lambda(dy) + \int \frac{1}{2} |Z^k(y) - Z(y)| \lambda(dy). \end{aligned} \tag{3.3}$$

De modo que si queremos acotar el lado izquierdo de 3.3, podemos acotar las dos integrales del lado derecho. Este es el camino que seguiremos para completar la prueba.

Primero, daremos una cota para el término $\int \frac{1}{2} |Z^k(y) - Z(y)| \lambda(dy)$. Para ello recordemos que $|Z^k(y) - Z(y)| = \left| \int f(y|\theta) (\pi_k - \pi)(d\theta) \right|$. Luego, por propiedades de

la integral

$$\left| \int f(y|\theta)(\pi_k - \pi)(d\theta) \right| \leq \int f(y|\theta)|\pi_k - \pi|(d\theta).$$

Por lo tanto

$$\int |Z^k(y) - Z(y)|\lambda(dy) \leq \int \int f(y|\theta)|\pi_k - \pi|(d\theta)\lambda(dy).$$

Intercambiando el orden de integración podemos calcular la integral anterior, es decir

$$\int \int f(y|\theta)|\pi_k - \pi|(d\theta)\lambda(dy) = \int \int f(y|\theta)\lambda(dy)|\pi_k - \pi|(d\theta).$$

Además $\int f(y|\theta)\lambda(dy) = 1$, pues $f(y|\theta)$ es densidad respecto a la medida λ , por tanto

$$\int |Z^k(y) - Z(y)|\lambda(dy) \leq \int |\pi_k - \pi|d\theta = 2\|\pi_k - \pi\|_{VT}.$$

O sea, que $\int \frac{1}{2}|Z^k(y) - Z(y)|\lambda(dy) \leq \|\pi_k - \pi\|_{VT}$.

Volviendo a la desigualdad 3.3 y con la cota que encontramos anteriormente, podemos concluir que

$$\int \frac{1}{2} \left| \frac{Z^{n,k}(y)}{Z(y)} - 1 \right| Z(y)\lambda(dy) \leq \int \frac{1}{2} |Z^{n,k}(y) - Z^k(y)|\lambda(dy) + \|\pi_k - \pi\|_{VT}.$$

Ahora buscaremos acotar el término $\int \frac{1}{2} |Z^{n,k}(y) - Z^k(y)|\lambda(dy)$. Llamemos $R_n(\theta)$ al cociente $\frac{f^n(y|\theta)}{f(y|\theta)}$. De esta manera podemos reescribir la diferencia $Z^{n,k}(y) - Z^k(y)$ como sigue:

$$|Z^{n,k}(y) - Z^k(y)| = \left| \int f(y|\theta)(R_n(\theta) - 1)\pi_k(d\theta) \right|.$$

Para la función $\log(x)$, al rededor de 1, la aproximación lineal está dada mediante

$$\log(x) \approx \log(1) + 1(x - 1),$$

pues la derivada de $\log(x)$ es la función $\frac{1}{x}$. Esto quiere decir que para valores cercanos

a 1, $\log(x)$ y la recta tangente $x - 1$ son muy parecidos.

Si η y η_1 son suficientemente cercanos, entonces el cociente $\frac{f_o(y|\eta)}{f_o(y|\eta_1)}$ se asemeja a 1. Utilizando la aproximación lineal de la función logaritmo tenemos que

$$\log\left(\frac{f_o(y|\eta)}{f_o(y|\eta_1)}\right) \approx \log(1) + \left(\frac{f_o(y|\eta)}{f_o(y|\eta_1)} - 1\right)$$

$$\log\left(\frac{f_o(y|\eta)}{f_o(y|\eta_1)}\right) \approx \frac{f_o(y|\eta)}{f_o(y|\eta_1)} - 1.$$

Como $\log\left(\frac{f_o(y|\eta)}{f_o(y|\eta_1)}\right) = \log(f_o(y|\eta)) - \log(f_o(y|\eta_1)) = \phi_y(\eta) - \phi_y(\eta_1)$, se tiene entonces que

$$|R_n(\theta) - 1| \approx |\phi_y(\eta_1) - \phi_y(\eta)| = |\phi_y(\eta) - \phi_y(\eta_1)|.$$

Utilizando la aproximación de Taylor de primer orden para $\phi_y(\eta)$ al rededor de η_1 , obtenemos que

$$|\phi_y(\eta) - \phi_y(\eta_1)| = |[\nabla\phi_y(\mathcal{H}(\mathcal{F}(\theta))) \cdot (\mathcal{H}(\mathcal{F}^{\alpha(n)}(\theta)) - \mathcal{H}(\mathcal{F}(\theta)))] + R|$$

Ignorando los términos de orden superior en el residuo,

$$\begin{aligned} |\nabla\phi_y(\mathcal{H}(\mathcal{F}(\theta))) \cdot (\mathcal{H}(\mathcal{F}^{\alpha(n)}(\theta)) - \mathcal{H}(\mathcal{F}(\theta)))| &= \left| \sum \frac{\partial}{\partial\eta_i} \phi_y(\mathcal{H}(\mathcal{F}(\theta))) (\mathcal{H}(\mathcal{F}^{\alpha(n)}(\theta)) - \mathcal{H}(\mathcal{F}(\theta))) \right| \\ &\leq \sum \left| \frac{\partial}{\partial\eta_i} \phi_y(\mathcal{H}(\mathcal{F}(\theta))) \right| |(\mathcal{H}(\mathcal{F}^{\alpha(n)}(\theta)) - \mathcal{H}(\mathcal{F}(\theta)))| \\ &< \sum \left| \frac{\partial}{\partial\eta_i} \phi_y(\mathcal{H}(\mathcal{F}(\theta))) \right| K_0 |\alpha(n)|^p \\ &= K_0 |\alpha(n)|^p \|\nabla\phi_y(\mathcal{H}(\mathcal{F}(\theta)))\|_1. \end{aligned}$$

Sustituyendo se obtiene

$$\int \frac{1}{2} |Z^{n,k}(y) - Z^k(y)| \lambda(dy) < \frac{K_0 |\alpha(n)|^p}{2} \int \int f(y|\theta) \|\nabla\phi_y(\mathcal{H}(\mathcal{F}(\theta)))\|_1 \pi_k(d\theta) \lambda(dy).$$

□

Teorema 3.3.2. *Supongamos que los datos se distribuyen de acuerdo a una familia de localización y escala, es decir*

$$f_o(y|\eta) = \prod_{i=1}^m \sigma^{-1} \rho \left(\frac{y_i - \eta_i}{\sigma} \right),$$

con ρ una función simétrica, acotada y clase C^1 que es densidad respecto a la medida de Lebesgue en \mathbb{R} ; y además $\int_{-\infty}^{\infty} x^2 \rho(x) dx = 1$.

Entonces,

$$\int \frac{1}{2} \left| \frac{Z^{n,k}(y)}{Z(y)} - 1 \right| Z(y) \lambda(y) < \rho(0) \frac{K_0 |\alpha(n)|^p}{\sigma} m + \|\pi_k - \pi\|_{VT}$$

Demostración. Para obtener este resultado, basta con cambiar el orden de integración en la integral doble del teorema 3.3.1 y calcular $\int f(y|\theta) \|\nabla \phi_y(\mathcal{H}(\mathcal{F}(\theta)))\|_1 \lambda(dy)$.

Calcular esta integral es equivalente a calcular $\sum_{i \leq m} \int \left| \frac{\partial}{\partial \eta_i} \phi_y(\eta) \right| f_o(y|\eta) \lambda(dy)$.

Haciendo $\rho(x) = e^{V(x)}$, se tiene que $\phi_y = -\log(f_o(y|\eta)) = V(x)$. Entonces

$$\left| \frac{\partial}{\partial \eta_i} \phi_y(\eta) \right| = \left| \sigma^{-1} V' \left(\frac{y_i - \eta_i}{\sigma} \right) \right|$$

Por lo tanto

$$\int \left| \frac{\partial}{\partial \eta_i} \phi_y(\eta) \right| f_o(y|\eta) \lambda(dy) = \int_{-\infty}^{\infty} \left| \sigma^{-1} V' \left(\frac{y_i - \eta_i}{\sigma} \right) \right| \sigma^{-1} \rho \left(\frac{y_i - \eta_i}{\sigma} \right) dy_i.$$

Por la simetría de ρ y haciendo el cambio de variable $x = \frac{y_i - \eta_i}{\sigma}$, la última integral puede calcularse como $2\sigma^{-1} \int_0^{\infty} V'(x) \rho(x) dx$. Pero $\rho'(x) = V'(x) \rho(x)$, entonces $\int_0^{\infty} V'(x) \rho(x) dx = \rho(0)$. Y con esto último obtenemos el resultado. □

Para controlar el error entre la distribución posterior numérica y teórica solo debemos ocuparnos de mantener el EABF por debajo de un umbral pequeño b . De este modo, si el error en la discretización del regresor $K_0 |\alpha(n)|^p$ lo denotamos con

K , lo anterior se traduce en asegurar que

$$K < \frac{\sigma b - \|\pi_k - \pi\|_{VT}}{m \rho(0)}. \quad (3.4)$$

3.4. Truncamientos

Supongamos, como lo hemos venido haciendo, que Θ es un espacio de Banach equipado con una norma $\|\cdot\|$. Típicamente, un espacio de Banach, es un espacio de funciones de dimensión infinita, como por ejemplo el conjunto de funciones continuas, reales y con dominio compacto. Supongamos ahora que contamos con una base $\{\phi_i\}_{i \in \mathbb{N}}$ para este espacio, y más aún, que todos sus elementos son de norma 1. Entonces cualquier elemento $\phi(t)$ cuenta con una representación con elementos de esta base. Dicha representación puede escribirse como

$$\theta(t) = \theta_0(t) + \sum_{i \in \mathbb{N}} \beta_i \phi_i(t),$$

con $\{\beta_i\}_{i \in \mathbb{N}} \subseteq \mathbb{R}$ y $\theta_0(t) \in \Theta$ fijo. Observemos que si $\{\beta_i\}_{i \in \mathbb{N}}$ es una colección de variables aleatorias y μ una medida de probabilidad en \mathbb{R}^∞ , entonces esta última induce una distribución F sobre nuestro conjunto de variables aleatorias. Si definimos la función $g : \mathbb{R}^\infty \rightarrow \Theta$ mediante la siguiente regla de correspondencia

$$\{\beta_i\}_{i \in \mathbb{N}} \mapsto \theta_0(t) + \sum_{i \in \mathbb{N}} \beta_i \phi_i(t),$$

esta resulta ser medible. Luego, equipando a Θ con la medida $\mu \circ g^{-1}$, naturalmente podemos inducir una distribución π . Una aproximación π_k para dicha distribución se obtiene al considerar la distribución marginal F_k inducida por la función proyección de los primeros k elementos. Esto es lo que consideraremos como un truncamiento y lo denotaremos mediante θ_k .

De acuerdo con Rosalsky y Rosenblatt (1997), si $\{X_i : i \in \mathbb{N}\}$ es una colección de variables aleatorias en un espacio de Banach real separable χ , que satisfacen

$\sum_{i \in \mathbb{N}} \|X_i\| < \infty$. Entonces existe $X \in \chi$ tal que

$$\sum_{i \leq n} X_i \rightarrow X;$$

siendo esta última convergencia casi segura, lo cual en espacios cuya medida total sea igual a 1 se traduce como convergencia en probabilidad.

Siguiendo el resultado anterior, en nuestro caso es suficiente garantizar que

$$\sum_{i \in \mathbb{N}} E[|\beta_i \phi_i|] < \infty,$$

como elegimos los elementos de nuestra base de modo que $\|\phi_i\| = 1$, lo anterior es equivalente a que $\sum_{i \in \mathbb{N}} E|\beta_i| < \infty$.

Sea θ el límite de la sucesión θ_k cuando $k \rightarrow \infty$. Recordando que convergencia en probabilidad implica convergencia débil, entonces se sigue que $\theta_k \Rightarrow \theta$. Además $F_k \Rightarrow F$, pues es la sucesión de distribuciones marginales. Dado que g es una función continua, por el teorema del mapeo continuo obtenemos que $\pi_k \Rightarrow \pi$. Con esto último concluimos que $\mathbb{Q}_y^{n,k} \Rightarrow \mathbb{Q}_y^n$ y $\mathbb{Q}_y^k \Rightarrow \mathbb{Q}_y^n$ como es deseado.

Un problema que puede presentarse al considerar este enfoque es al momento de controlar la tasa de las últimas convergencias. Pues recordando el resultado 3.3.2, necesitamos conocer $\|\pi_k - \pi\|_{VT}$. Una manera de acotar esta distancia es mediante la desigualdad siguiente

$$\|\pi_k - \pi\|_{VT} \leq P(\theta_k \neq \theta).$$

Observemos que $P(\theta_k \neq \theta) = P(\sum_{i \geq k+1} \beta_i \phi_i(t)) = P(\|\sum_{i \geq k+1} \beta_i \phi_i(t)\| > 0)$, por lo que

$$\|\pi_k - \pi\|_{VT} \leq 1 - F(\beta_i = 0; i \geq k + 1).$$

Sin embargo, al considerar $\{\beta_i\}_{i \in \mathbb{N}}$ como un conjunto de variables aleatorias continuas e independientes, la cota anterior es obsoleta ya que $F(\beta_i = 0; i \geq k + 1) = 0$.

En lo sucesivo de la sección discutiremos una manera de solucionar el problema anterior, enfocándonos en la construcción de una distribución F para la colección $\{\beta_i\}_{i \in \mathbb{N}}$ que cumpla con dos condiciones,

1. $\sum_{i \in \mathbb{N}} E|\beta_i| < \infty$ y
2. $F(\beta_i = 0; i \geq k + 1) > 0$.

Para cada $i \in \mathbb{N}$ consideramos a γ_i como una variable aleatoria cuya función de distribución es $\psi(\cdot)$ y es absolutamente continua respecto a la medida de Lebesgue en \mathbb{R} . Además, requerimos que sea de media cero y varianza 1. Supongamos también que $\{\xi_i\}_{i \in \mathbb{N}}$ es una sucesión de números reales positivos tal que $\sum_{i \in \mathbb{N}} \xi_i < \infty$. Sea ϑ una variable aleatoria que toma valores en los números naturales. Definiremos F de modo que, condicional a ϑ , la colección $\{\gamma_i\}_{i \in \mathbb{N}}$ tenga cola de ceros. Hacemos esto definiendo sus distribuciones finito dimensionales de la siguiente manera,

$$P(\gamma_i < x | \vartheta) = \begin{cases} \psi(x) & \text{si } i \leq \vartheta, \\ I_{(-\infty, 0]}(x) & \text{si } i > \vartheta. \end{cases}$$

Como consecuencia de la definición anterior,

$$P(\gamma_i < x_i; i \leq k) = \sum_{l \in \mathbb{N}} P(\vartheta = l) \prod_{i \leq k} P(\gamma_i < x_i | \vartheta).$$

Entonces $F(\beta_i = 0; i \geq k + 1) = P(\vartheta \leq k)$ y esta última se convierte en $\sum_{j \geq k+1} P(\vartheta = j)$ que es un término positivo, como se quería.

Por otro lado, la condición número 1 se sigue de las siguientes desigualdades

$$\begin{aligned} \sum_{i \in \mathbb{N}} E|\beta_i| &= \sum_{i \in \mathbb{N}} E|\gamma_i \xi_i| \\ &= \sum_{i \in \mathbb{N}} E|\gamma_i| \xi_i \\ &\leq \sum_{i \in \mathbb{N}} E[\gamma_i^2] \xi_i \\ &\leq \sum_{i \in \mathbb{N}} \xi_i < \infty. \end{aligned}$$

La segunda desigualdad se sigue del hecho de que $E[\gamma_i^2] = 1$, pues por hipótesis

su media es cero.

De aquí que, la cota encontrada en 3.4 para EABF , se convierte en

$$\rho(0) \frac{K}{\sigma} m + \sum_{j \geq k+1} P(\vartheta = j),$$

donde $\|\mathcal{H}(\mathcal{F}(\theta)) - \mathcal{H}(\mathcal{F}^{\alpha(n)}(\theta))\| < K$.

Finalmente la distribución posterior finito dimensional considerando el truncamiento de la construcción anterior, está dada por

$$\pi(\gamma_1, \dots, \gamma_k, \vartheta | y) \propto \sigma^{-m} \prod_{j \leq m} \rho \left(\frac{y_j - \mathcal{H}_j(\mathcal{F}^{\alpha(n)}(\theta_k))}{\sigma} \right) I_{(l \leq k)}(\vartheta) P(\vartheta = l) \prod_{i \leq k} P(\gamma_i < x_i | \vartheta),$$

sujeto a la cota anterior.

Incluir las variables aleatorias $\{\gamma_i\}_{i \leq k}$ y ϑ no debe representar ningún problema, pues bajo cualquier método de Monte Carlo (MC), como MCMC, es usual incluir variables aleatorias no observables.

Capítulo 4

Implementación

4.1. Deconvolución

La *convolución* de dos funciones f y g , denotada por $f * g$, se encuentra definida mediante la integral

$$(f * g)(t) = \int g(y - x)f(y)dy.$$

Supongamos ahora que $c(z) := \frac{1}{2\alpha}1_{[-\alpha,\alpha]}(z)$ y $\theta(t) := \beta_0 + \sum_{i=1}^4 \beta_i \phi_i(t)$, donde $\phi_i(t) = \cos(2\pi it)$ y β_i fijos para cada $i \in \{0, 1, 2, 3, 4\}$. Entonces la convolución entre estas dos funciones es el resultado de la integral siguiente:

$$\int_0^1 c(y - x)\theta(y)dy.$$

Y realizando el cambio de variable $u = y - x$, la expresión anterior se convierte en

$$\int_{-x}^{1-x} c(u)\theta(u + x)du.$$

Sustituyendo c en la última integral,

$$\int_x^{1-x} \frac{1}{2\alpha}1_{[-\alpha,\alpha]}(u)\theta(u + x).$$

Observemos que la función $\frac{1}{2\alpha}1_{[-\alpha,\alpha]}(u)$ solo toma valores distintos de cero cuando $u \in [a, b]$, donde $a = \max\{-x, -\alpha\}$ y $b = \min\{1-x, \alpha\}$. Haciendo un último cambio de variable, con $u+x=z$ obtenemos como expresión equivalente

$$\int_{\max\{x-\alpha,0\}}^{\min\{x+\alpha,1\}} \frac{1}{2\alpha}\theta(z)dz.$$

Fijando α , la integral anterior tiene solución analítica. Pues si $b = \min\{x+\alpha, 1\}$ y $a = \max\{x-\alpha, 0\}$, entonces

$$\frac{1}{2\alpha} \int_a^b \cos(2\pi i\nu) d\nu = \frac{1}{2\alpha 2\pi} [\text{sen}(2\pi ib) - \text{sen}(2\pi ia)].$$

En la figura 4.1 se pueden observar las gráficas de θ y su convolución con c .

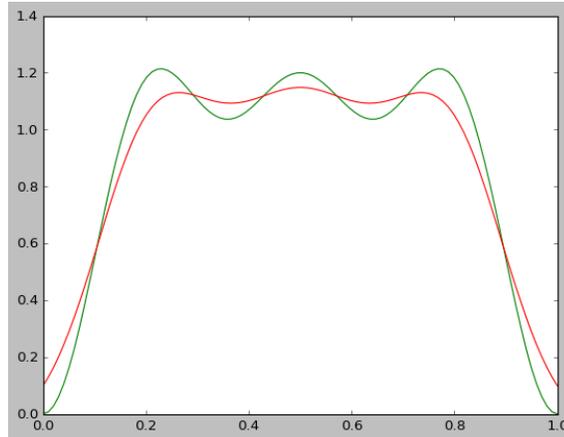


Figura 4.1: En verde la gráfica de la función θ . En rojo la convolución entre θ y c . Con $\alpha = 0,1$ y $\beta = [0,900, -0,400, -0,300, -0,200]$.

Supongamos ahora que contamos con m observaciones $\{y_i\}_{i \leq m}$ provenientes de la convolución anterior, pero que desconocemos θ . Mediante una aplicación de Cuantificación de la incertidumbre Bayesiana obtendremos una aproximación de la función faltante θ .

Como distribución apriori vamos a considerar a θ_k definida como

$$\theta_k(t) := \beta_0 + \sum_{i \leq k} \beta_i \phi_i(t),$$

donde, para cada $i \leq k$, $\phi_i(t) = \cos(2\pi it)$ y β_i es una variable aleatoria que se distribuye Normal de media cero y varianza $\frac{\sigma}{i}$, con σ conocida. Si además presuponemos que sólo contamos con una versión numérica del regresor o Forward Map $\mathcal{F}^n[\theta]$, entonces debemos elegir el nivel de aproximación de ésta para no perder calidad en la distribución posterior. Para ello nos basaremos en la cota encontrada en el capítulo anterior y cuya expresión se encuentra en la ecuación 3.4.

$$K < \frac{\sigma}{m} \frac{b}{\rho(0)}.$$

En este caso $\rho(0) = \frac{1}{\sqrt{2\pi}}$ y la discretización de $\mathcal{F}[\theta_k]$ se realizará mediante la regla de Simpson. Como contamos con la versión analítica del regresor, K puede calcularse como el valor absoluto de la diferencia entre $\mathcal{F}[\theta_k]$ y $\mathcal{F}^n[\theta_k]$.

4.2. Resultados

Mediante una implementación en Python-Scipy pudimos seleccionar el nivel de discretización en el regresor. En dicha implementación nos ocupamos de mantener K por debajo de la cota establecida para $b = \frac{1}{20}$. Uno de los parámetros del programa es el nivel de discretización n ; una vez establecido, se corre el programa y si la desigualdad no es satisfecha, se arroja un mensaje que nos informa que la cota está siendo rebasada. De esta manera sabemos que debemos refinar la discretización, es decir, aumentar n . Todas las simulaciones obtenidas de la distribución posterior se realizaron con *t-walk*.

Generamos una muestra $\{y_i\}_{i \leq m}$ para $m = 10$. Dichas observaciones se obtienen de la manera siguiente. Hagamos $\mathcal{F}[\theta](x) = \int_0^1 c(y-x)\theta(y)dy$. Entonces

$$y_i = \mathcal{F}[\theta](x) + \epsilon_i, \epsilon_i \sim N(0, 1).$$

4.2.1. Caso $k = 10$

Como primer caso, consideramos $k = 10$. Para $n \leq 17$ el error aún está por encima de lo que podemos tolerar, es en $n = 18$ el primer momento en el que se satisface la cota. A continuación se muestran gráficos para diferentes valores de n , a

saber 5, 10, 15, 17, 18 y 20. Para cada uno de estos casos se obtuvieron los estimadores MAP y la media posterior, con ayuda de los cuales se graficaron las funciones θ_{MAP} y θ_{PMN} respectivamente. De manera análoga, con estos estimadores se obtienen las gráficas de aproximaciones del regresor $\mathcal{F}^n[\theta_{MAP}]$, $\mathcal{F}^n[\theta_{PMN}]$. Desde la figura 4.2 hasta 4.9 se muestran los resultados.

4.2.2. Caso $k = 7$

Como segundo caso, consideramos $k = 7$. Ahora pudimos observar que para $n \leq 13$ el error aún está por encima de lo que podemos tolerar. Como primer momento en el que se satisface la cota está $n = 14$; o sea que ahora que hemos disminuido k también se ha reducido n . A continuación se muestran gráficos para diferentes valores de n , a saber 10, 14 y 15. Para cada uno de estos casos se obtuvieron los estimadores MAP y la media posterior, con ayuda de los cuales se graficaron las funciones θ_{MAP} y θ_{PMN} respectivamente. De manera análoga, con estos estimadores se obtienen las gráficas de aproximaciones del regresor $\mathcal{F}^n[\theta_{MAP}]$, $\mathcal{F}^n[\theta_{PMN}]$. También se muestran las gráficas de las marginales obtenidas para $\beta = (\beta_0, \dots, \beta_6)$. Desde la figura 4.10 hasta 4.13 se muestran los resultados.

4.2.3. Caso $k = 8$

Por último, consideramos $k = 8$, donde pudimos observar también como primer momento en el que se satisface la cota a $n = 14$. A continuación se muestran gráficos para diferentes valores de n , a saber 10, 14 y 15. Para cada uno de estos casos se obtuvieron los estimadores MAP y la media posterior, con ayuda de los cuales se graficaron las funciones θ_{MAP} y θ_{PMN} respectivamente. De manera análoga, con estos estimadores se obtienen las gráficas de aproximaciones del regresor $\mathcal{F}^n[\theta_{MAP}]$, $\mathcal{F}^n[\theta_{PMN}]$. También se muestran las gráficas de las marginales obtenidas para $\beta = (\beta_0, \dots, \beta_7)$. De la figura 4.14 hasta 4.17 se muestran los resultados.

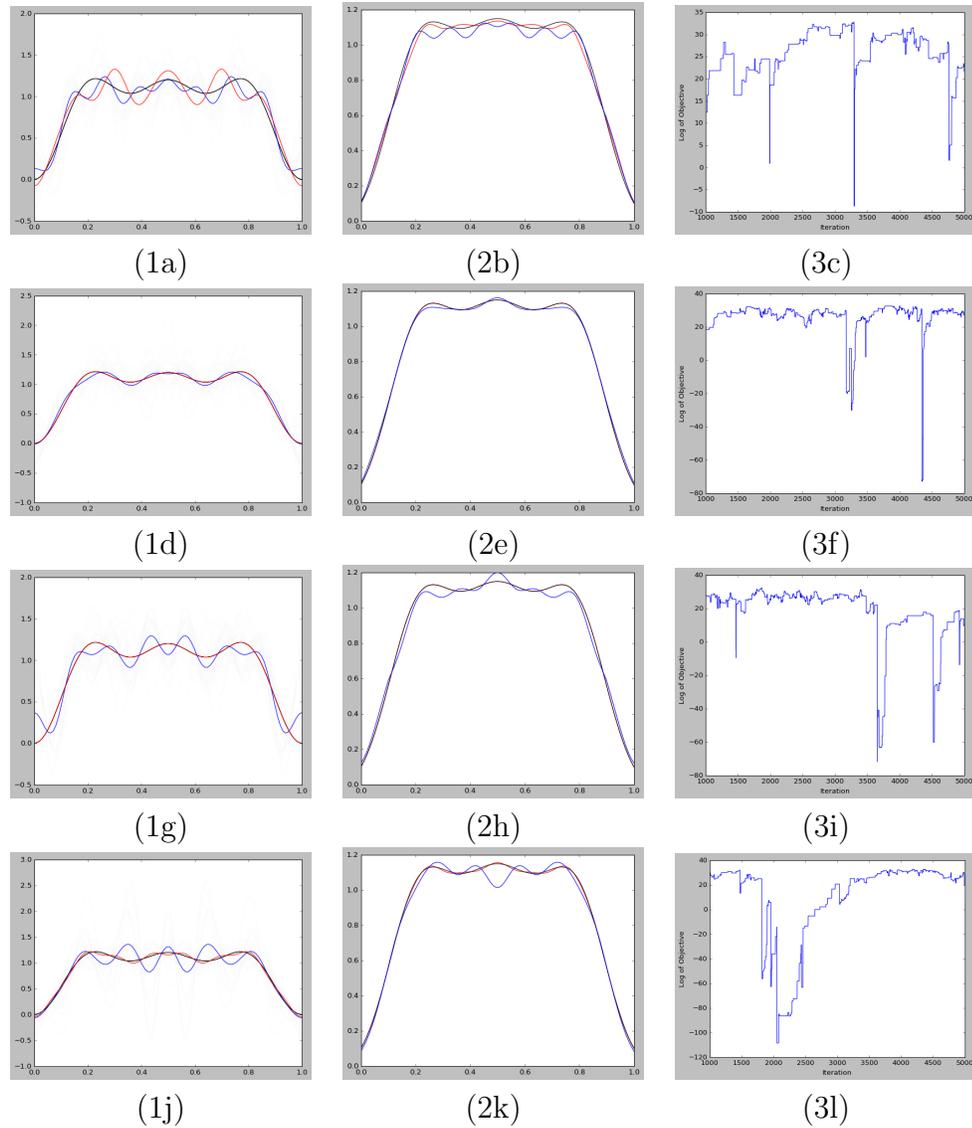


Figura 4.2: (1) En negro la función real θ , en azul θ_{PMN} y en rojo θ_{MAP} . Las transparencias en gris son funciones estimadas a través de las simulaciones.

(2) Negro en $\mathcal{F}[\theta]$, en azul $\mathcal{F}^n[\theta_{PMN}]$ y en rojo $\mathcal{F}^n[\theta_{MAP}]$.

(3) Convergencia de la simulación.

Caso $n = 5$ (a-c)

Caso $n = 10$ (d-f)

Caso $n = 15$ (g-i)

Caso $n = 17$ (j-l)

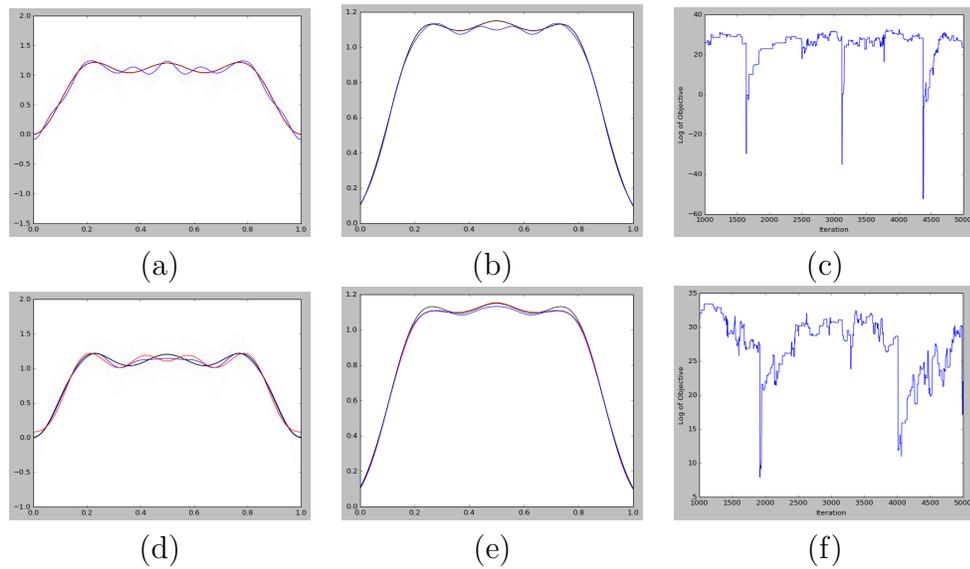


Figura 4.3: (1) En negro la función real θ , en azul θ_{PMN} y en rojo θ_{MAP} . Las transparencias en gris son funciones estimadas a través de las simulaciones.

(2) Negro en $\mathcal{F}[\theta]$, en azul $\mathcal{F}^n[\theta_{PMN}]$ y en rojo $\mathcal{F}^n[\theta_{MAP}]$.

(3) Convergencia de la simulación.

Caso $n = 18$ (a-c)

Caso $n = 20$ (d-f)

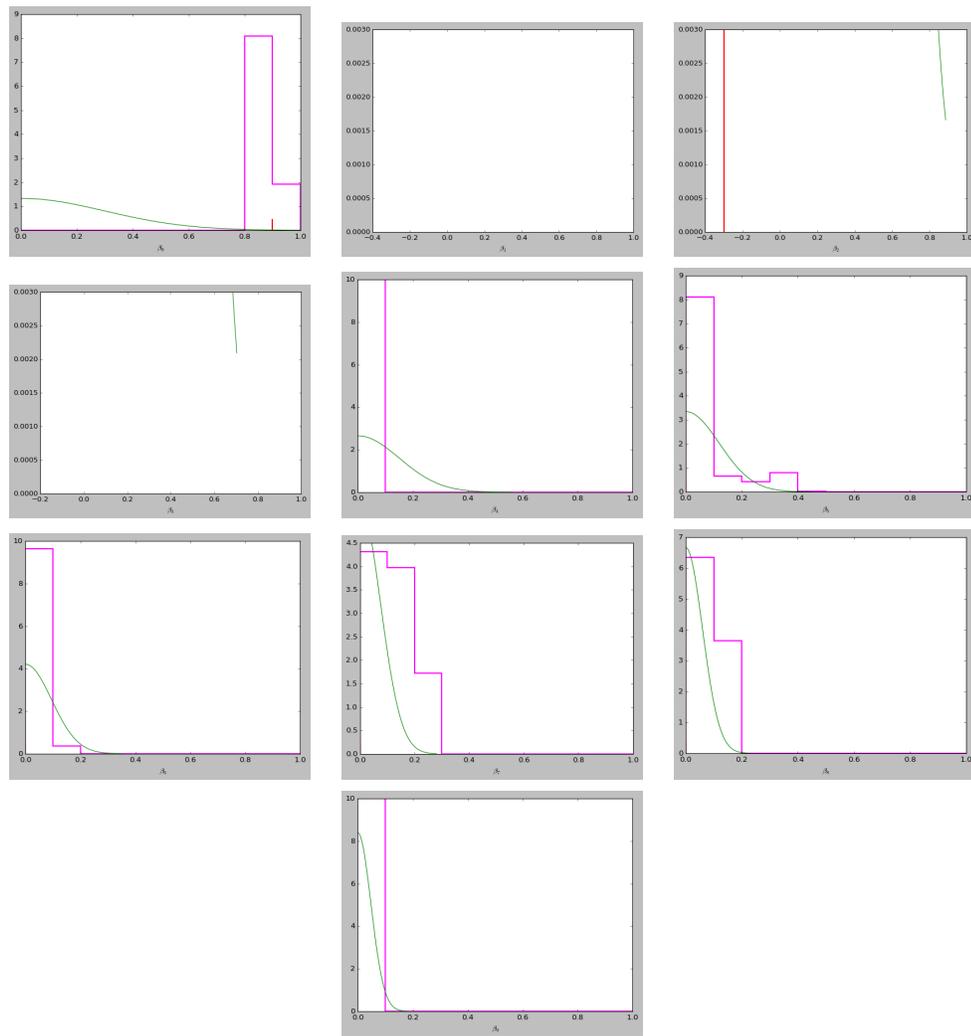


Figura 4.4: Caso $n = 5$.
 Gráficas de las marginales de β . En verde la distribución a priori de β_i y en magenta los histogramas de los valores simulados.

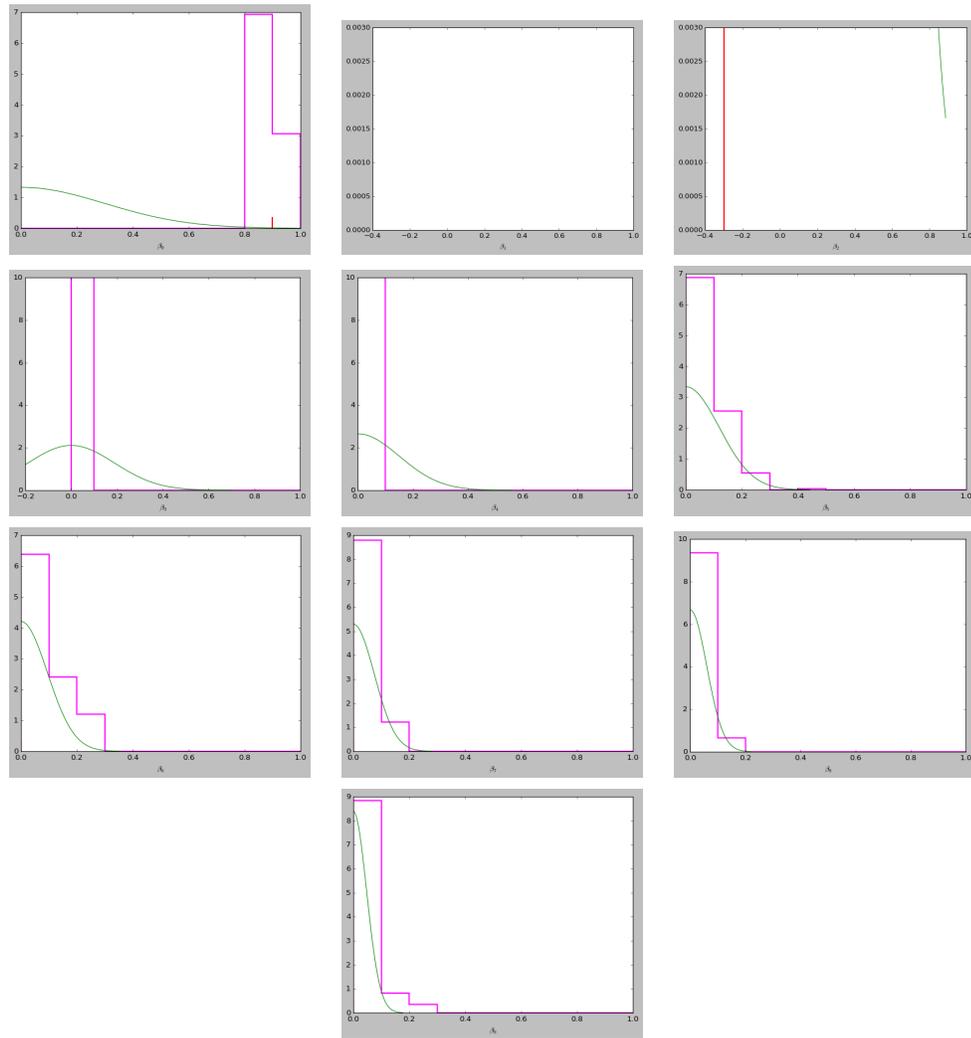


Figura 4.5: Caso $n = 10$.
 Gráficas de las marginales de β . En verde la distribución a priori de β_i y en magenta los histogramas de los valores simulados.

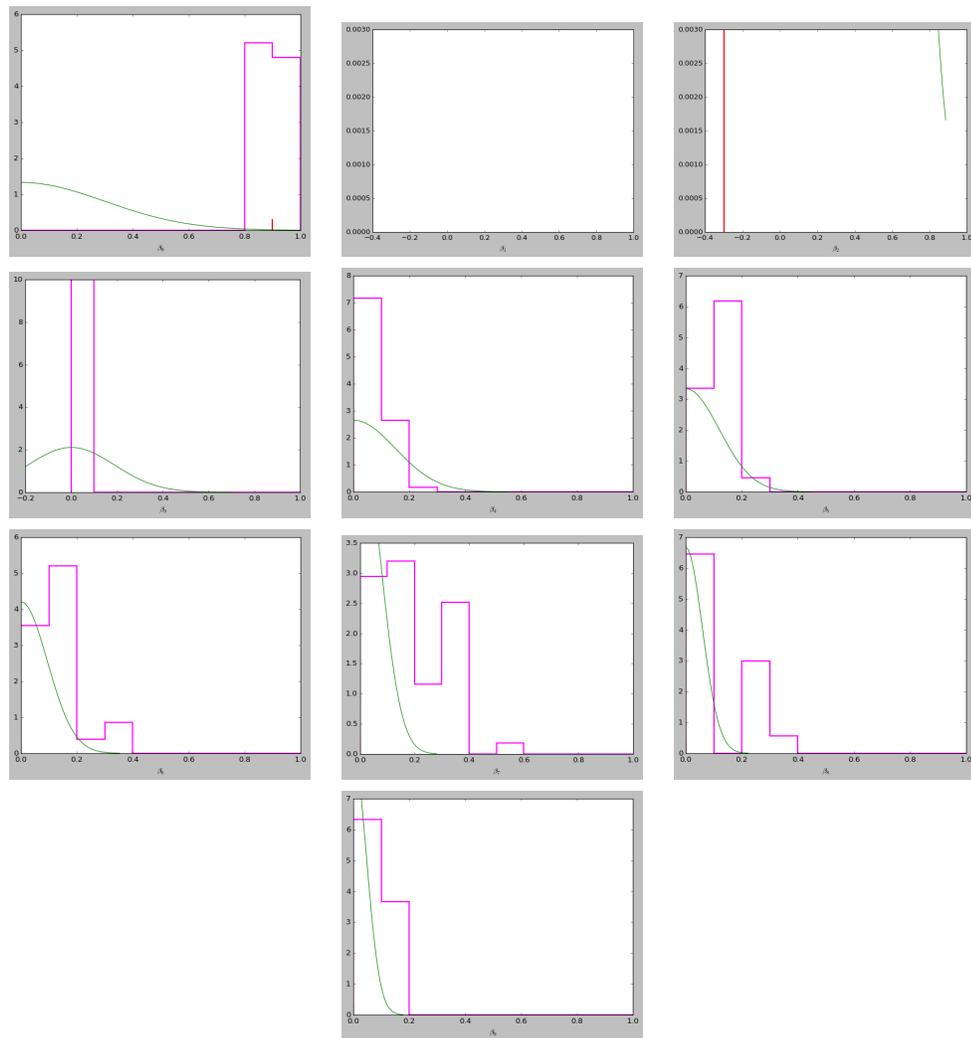


Figura 4.6: Caso $n = 15$.
 Gráficas de las marginales de β . En verde la distribución a priori de β_i y en magenta los histogramas de los valores simulados.

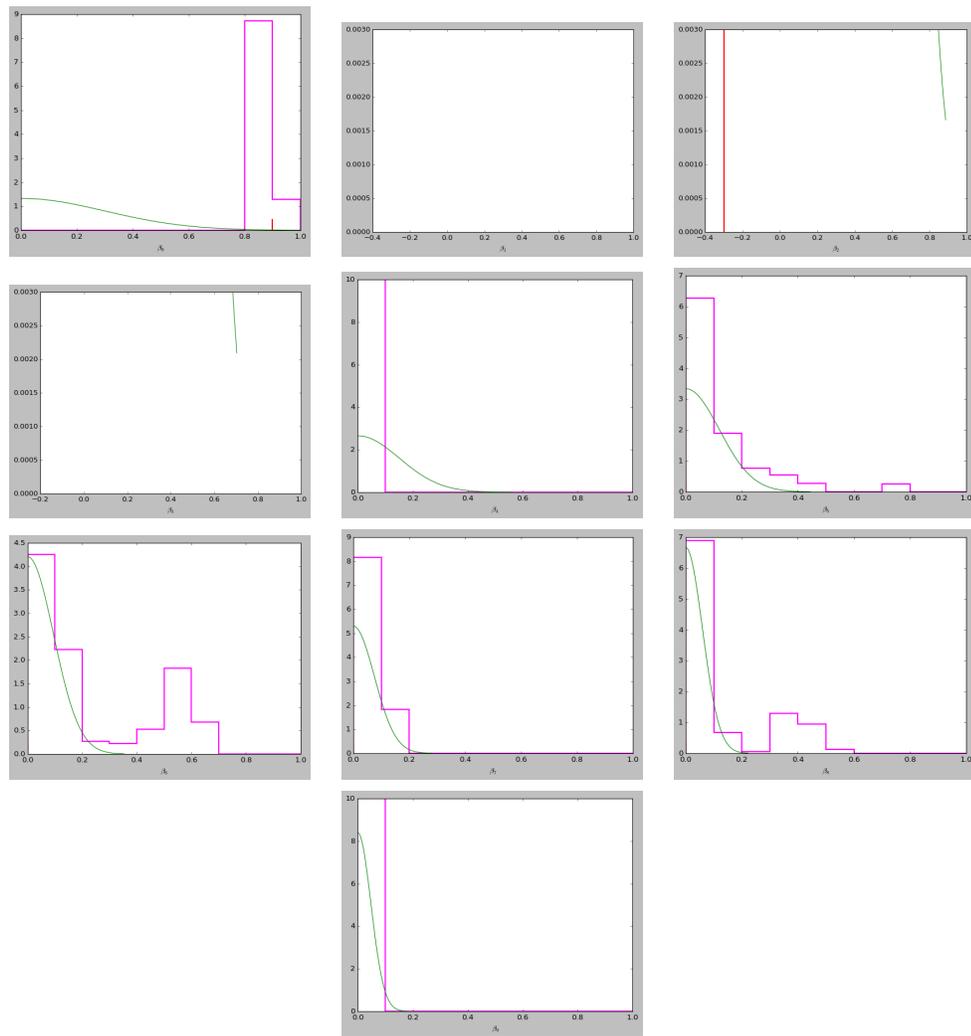


Figura 4.7: Caso $n = 17$.
 Gráficas de las marginales de β . En verde la distribución a priori de β_i y en magenta los histogramas de los valores simulados.

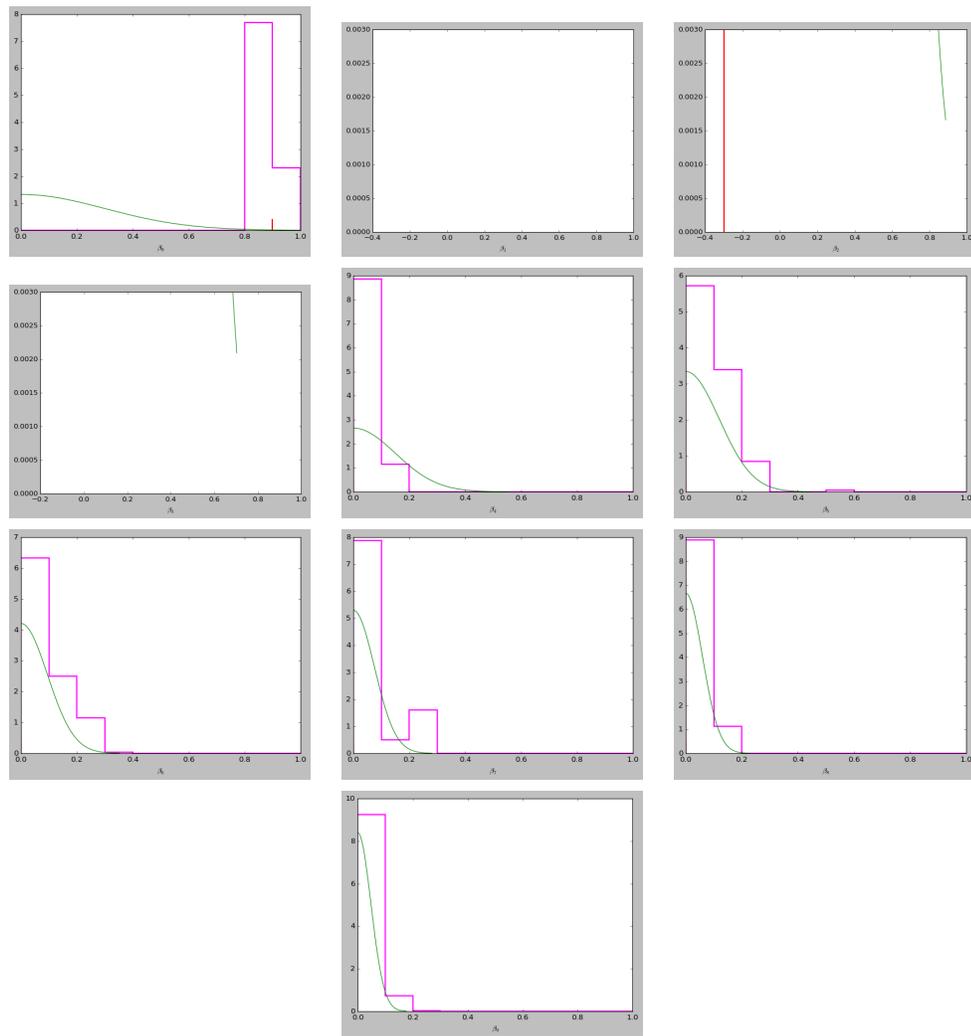


Figura 4.8: Caso $n = 18$.
 Gráficas de las marginales de β . En verde la distribución a priori de β_i y en magenta los histogramas de los valores simulados.

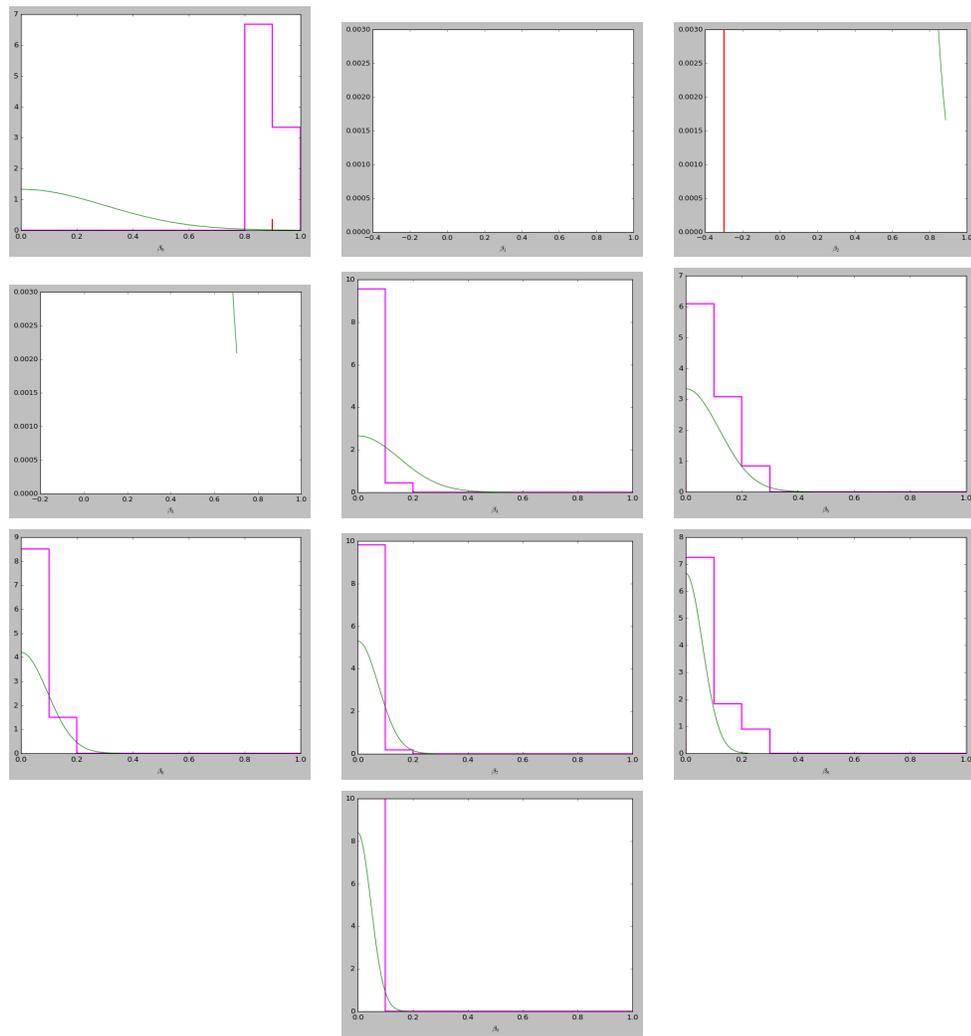


Figura 4.9: Caso $n = 20$.
 Gráficas de las marginales de β . En verde la distribución a priori de β_i y en magenta los histogramas de los valores simulados.

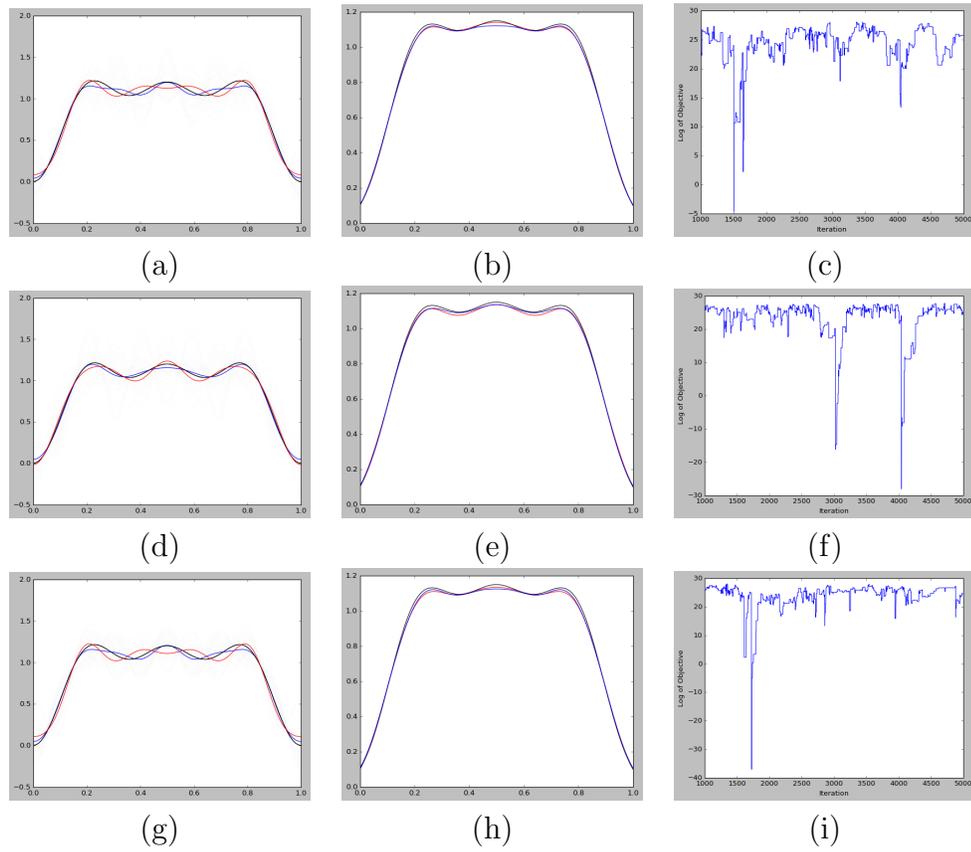


Figura 4.10: (1) En negro la función real θ , en azul θ_{PMN} y en rojo θ_{MAP} . Las transparencias en gris son funciones estimadas a través de las simulaciones.

(2) Negro en $\mathcal{F}[\theta]$, en azul $\mathcal{F}^n[\theta_{PMN}]$ y en rojo $\mathcal{F}^n[\theta_{MAP}]$.

(3) Convergencia de la simulación.

Caso $n = 10$ (a-c)

Caso $n = 14$ (d-f)

Caso $n = 15$ (g-i)

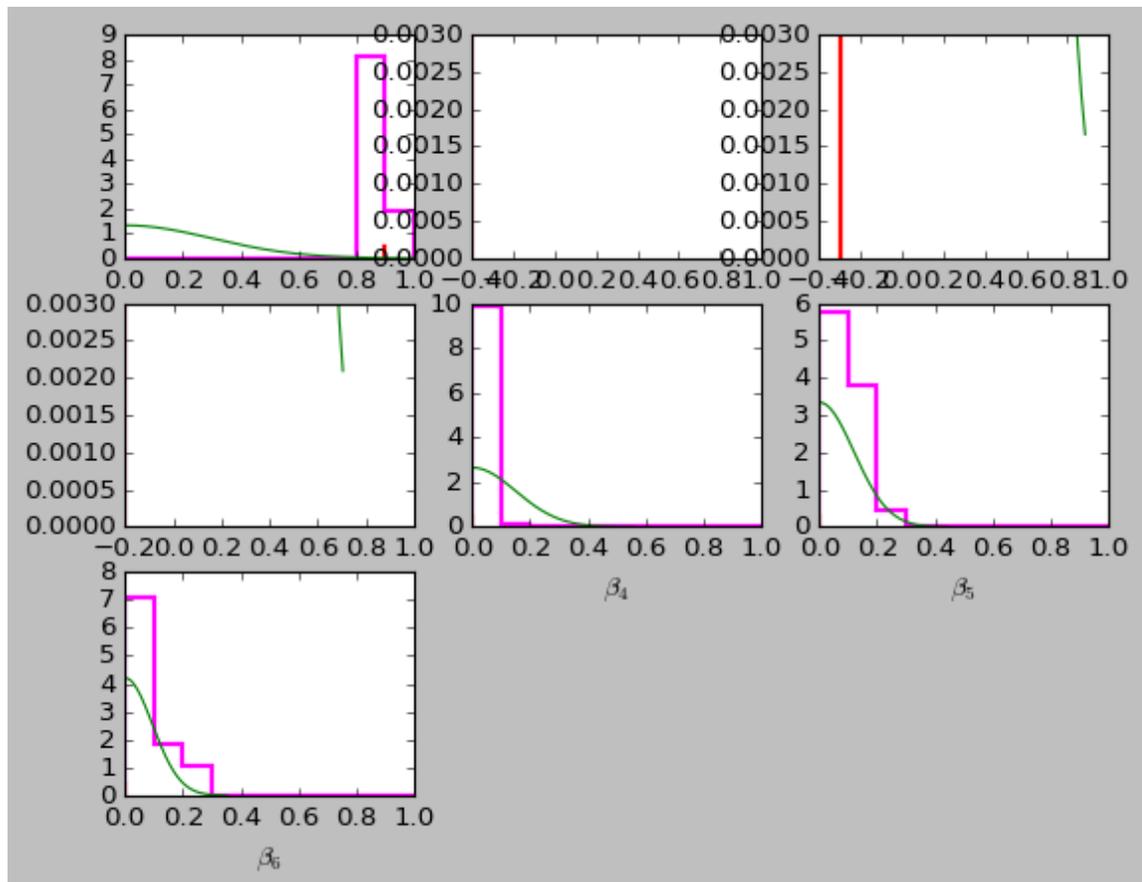


Figura 4.11: Caso $n = 10$.

Gráficas de las marginales de β . En verde la distribución a priori de β_i y en magenta los histogramas de los valores simulados.

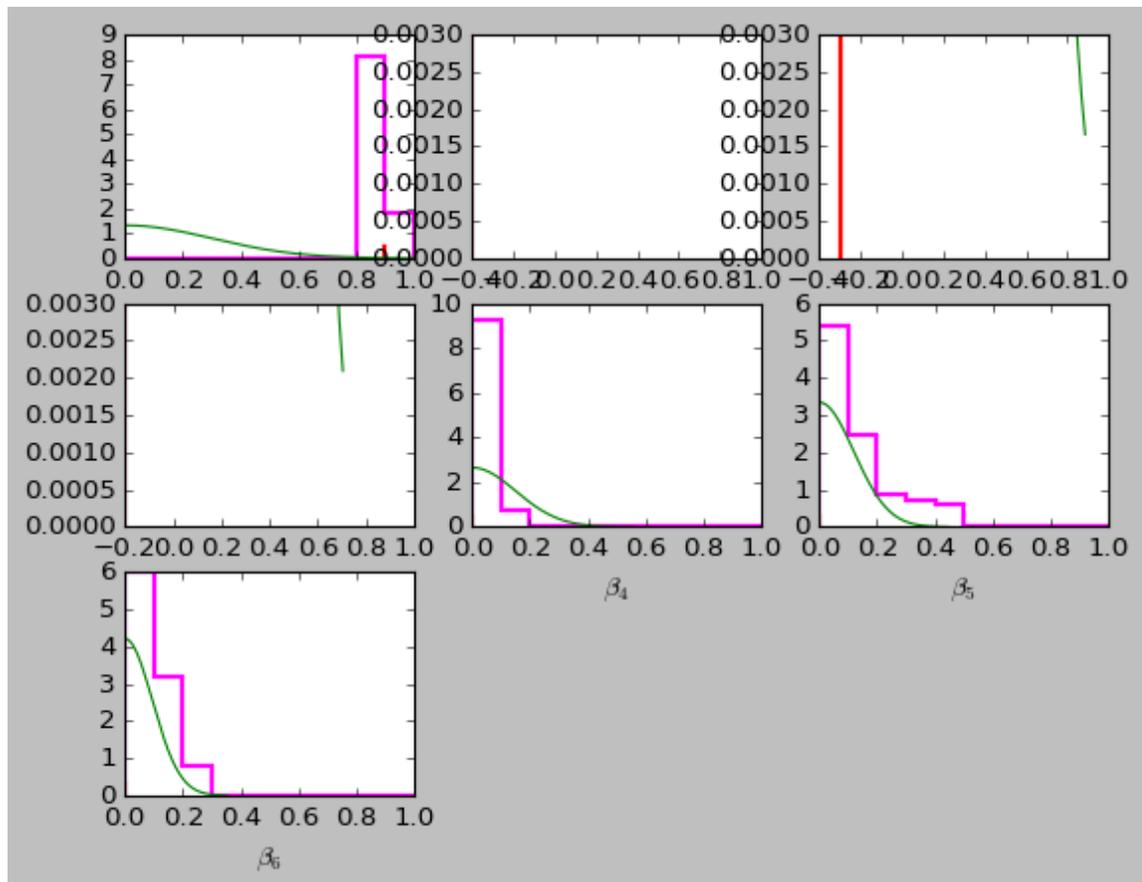


Figura 4.12: Caso $n = 14$.

Gráficas de las marginales de β . En verde la distribución a priori de β_i y en magenta los histogramas de los valores simulados.

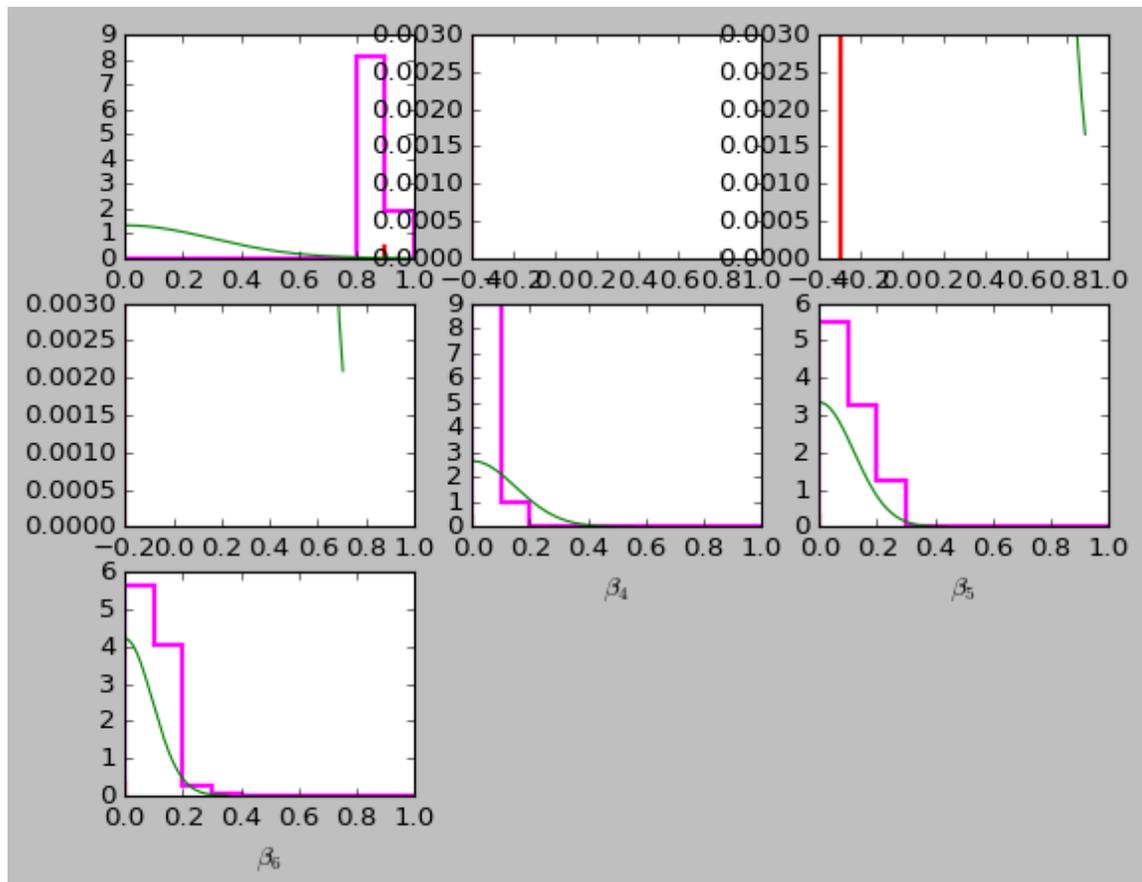


Figura 4.13: Caso $n = 15$.

Gráficas de las marginales de β . En verde la distribución a priori de β_i y en magenta los histogramas de los valores simulados.

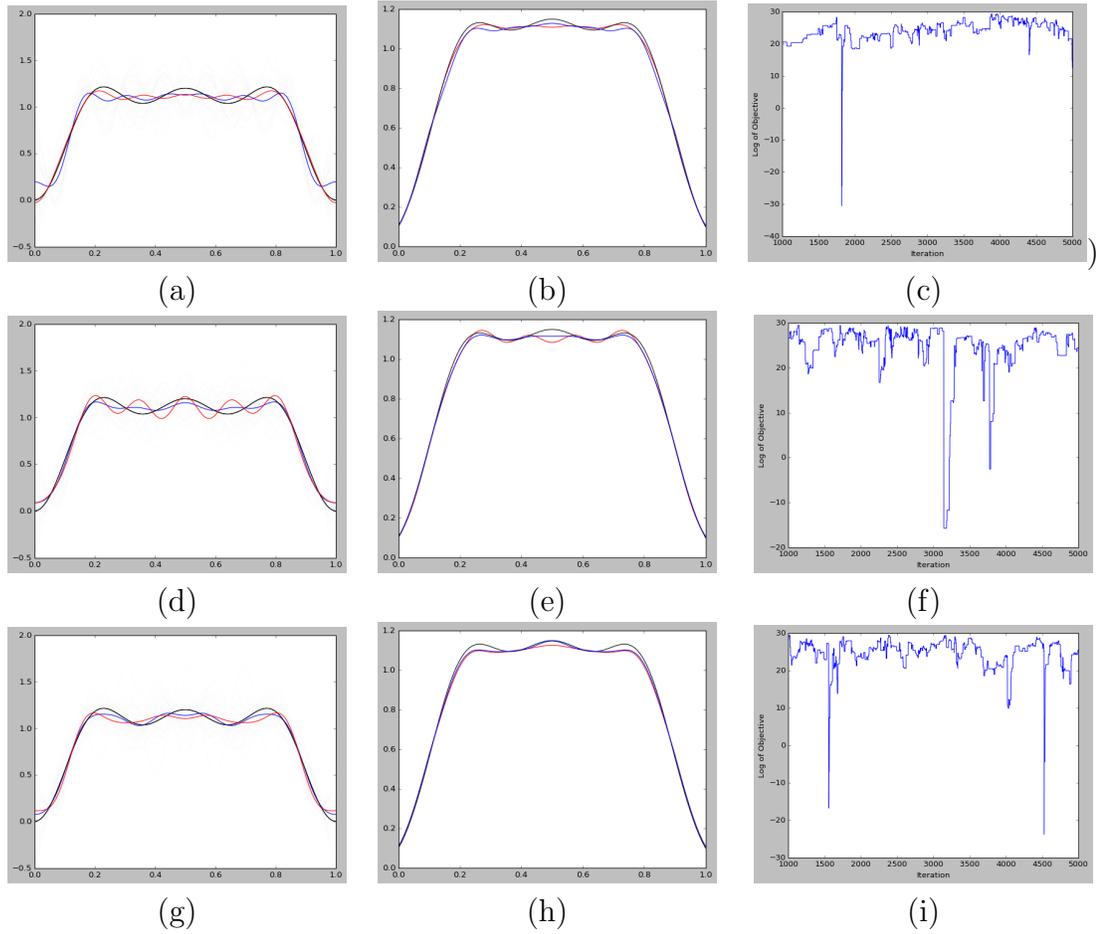


Figura 4.14: (1) En negro la función real θ , en azul θ_{PMN} y en rojo θ_{MAP} . Las transparencias en gris son funciones estimadas a través de las simulaciones.

(2) Negro en $\mathcal{F}[\theta]$, en azul $\mathcal{F}^n[\theta_{PMN}]$ y en rojo $\mathcal{F}^n[\theta_{MAP}]$.

(3) Convergencia de la simulación.

Caso $n = 10$ (a-c)

Caso $n = 14$ (d-f)

Caso $n = 15$ (g-i)

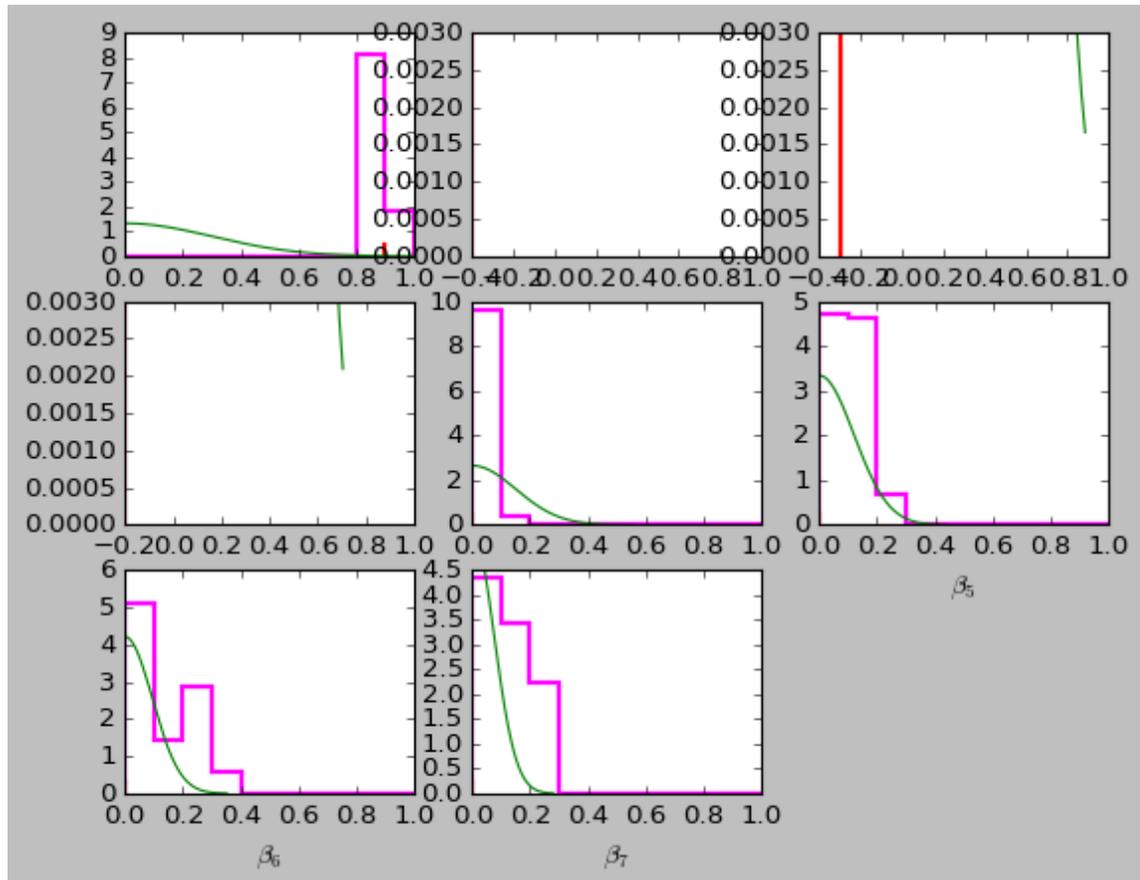


Figura 4.15: Caso $n = 10$.

Gráficas de las marginales de β . En verde la distribución a priori de β_i y en magenta los histogramas de los valores simulados.

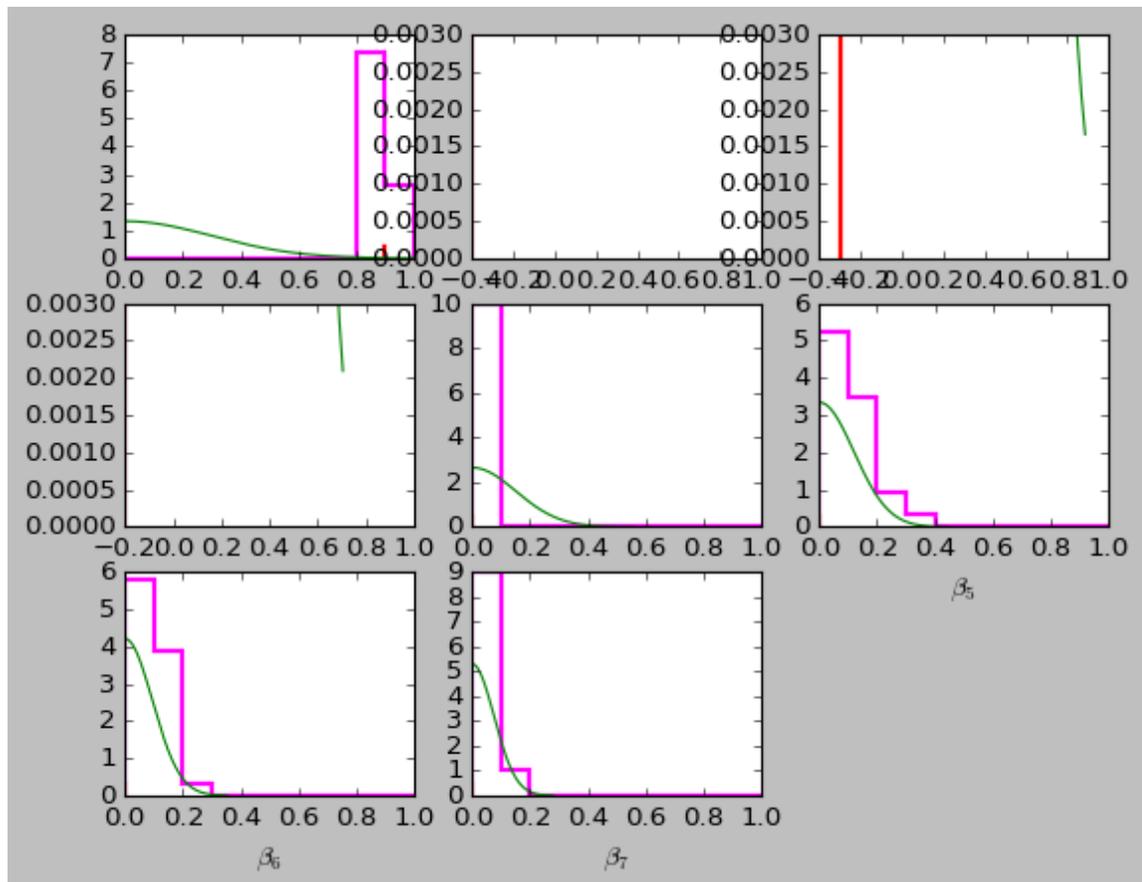


Figura 4.16: Caso $n = 14$.

Gráficas de las marginales de β . En verde la distribución a priori de β_i y en magenta los histogramas de los valores simulados.

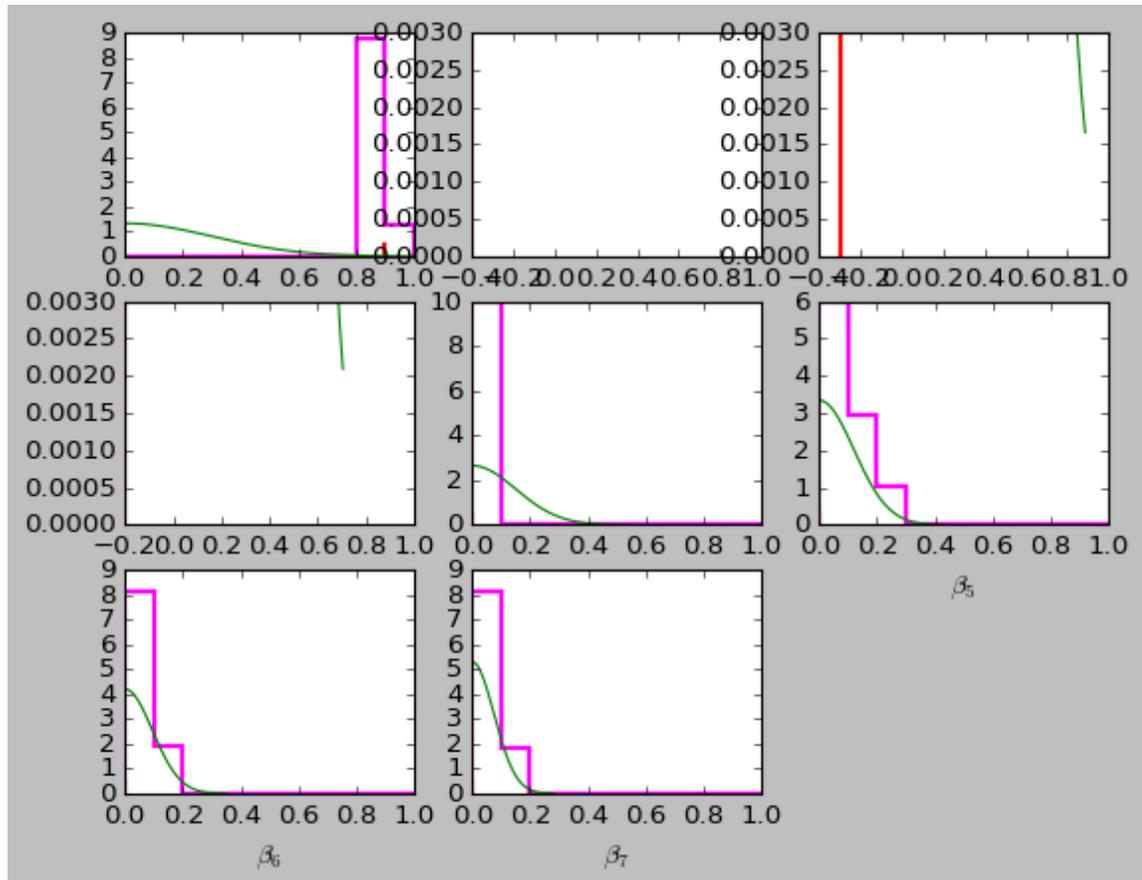


Figura 4.17: Caso $n = 15$.

Gráficas de las marginales de β . En verde la distribución a priori de β_i y en magenta los histogramas de los valores simulados.

Capítulo 5

Conclusiones

Es importante contar con fundamentación sólida si se desea hacer uso del enfoque bayesiano para atacar un problema inverso. Sin embargo, este no debe ser el único punto a considerar, pues no debe dejarse de lado el aspecto computacional ya que será el que resulte de utilidad. De aquí la importancia de un texto que cuente con ambos aspectos.

El marco en el cual se define la existencia de la distribución posterior resulta muy general, pues no es necesario verificar ninguna condición de regularidad sobre la distribución de los datos. Basta con verificar la continuidad de la verosimilitud, condición que no es difícilmente satisfecha. Por otro lado, trabajar en espacios Polacos bajo este contexto resulta, en la mayoría de los casos, muy natural y tampoco debe ser problemático. Fuera de estos dos aspectos, es importante recalcar que para todos los resultados obtenidos se usó solamente herramienta clásica de teoría de la Probabilidad, siendo Desintegración la más sofisticada de ellas debido en parte a la complejidad de su formalización.

Finalmente, aún cuando la construcción de un truncamiento π_k está bien fundamentada y cumple con las condiciones deseadas, puede resultar difícil la elección de una base para el espacio paramétrico Θ ; más aún cuando optamos por la representación de β_i que incluye a ψ y γ_i , pues además de esta dificultad el aspecto computacional puede complicarse, trayendo como consecuencia que sea un método difícil de abordar e implementar.

Bibliografía

- [1] J. T. Chang y D. Pollard. Conditioning as disintegration. *Statistica Neerlandica*, 51:287–317, 1997.
- [2] J. Andrés Christen. *Posterior distribution existence and error control in Banach spaces*. ArXiv e-prints, 2017.
- [3] K. Gowrisankaran. Measurability of functions in product spaces. *Proceedings of the American Mathematical Society*, 31:485–488, 1972.
- [4] Guillermo Grabinsky. *Teoría de la medida*. Las prensas de ciencias, 2013.
- [5] Soumendra N. Lahiri Krishna B. Athreya. *Measure Theory and Probability Theory*. Springer, 2006.
- [6] D. Pollard. *A User's Guide to Measure Theoretic Probability*. Cambridge Series in Statistical and Probabilistic Mathematics, 2002.
- [7] W. Sierpiński. Sur un probleme concernant les ensembles mesurables superficiellment. *Fundamenta Mathematicae*, 1:112–115, 1920.