



UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales
Departamento de Matemática

Tesis de Licenciatura

**Estimación no paramétrica de la densidad en variedades
Riemannianas**

Andrés Leandro Muñoz

Directores: Dr. Guillermo S. Henry y Dra. Daniela Rodriguez

14 de Abril de 2011

Índice general

1. Introducción	1
2. Estimación no paramétrica de la densidad	5
2.1. Estimación de la densidad por núcleos	6
2.2. Estimación de la densidad por vecinos más cercanos con núcleo	9
3. Preliminares Geométricos	11
3.1. Variedades Diferenciables	11
3.2. Conexiones y geodésicas	16
3.3. Variedades Riemannianas	20
3.4. Probabilidades y estadística en variedades Riemannianas	24
4. Estimación no paramétrica de la densidad en variedades Riemannianas	27
4.1. Estimación de la densidad por núcleos en variedades Riemannianas . .	27
4.2. Estimación de la densidad por vecinos más cercanos con núcleo en variedades Riemannianas	33

Capítulo 1

Introducción

La inferencia estadística comúnmente se focaliza sobre funciones de distribución que son puramente paramétricas o puramente no paramétricas. En los modelos paramétricos se comienza haciendo supuestos rígidos sobre la estructura de los datos para luego estimar de la manera más eficiente posible los parámetros que definen su estructura. Un modelo paramétrico razonable produce inferencias precisas mientras que un modelo erróneo posiblemente conducirá a conclusiones equivocadas.

Sin embargo, en la mayoría de las aplicaciones, los modelos paramétricos constituyen una aproximación al modelo subyacente, y la búsqueda de un modelo adecuado suele no ser sencilla. Es aquí donde, las técnicas de estimación no paramétricas surgen como una alternativa más flexible a los modelos paramétricos. Como punto en común, los métodos no paramétricos explotan la idea de suavizado local, que solamente utiliza las propiedades de continuidad o diferenciabilidad local de la función a estimar. El éxito del suavizado local depende de la presencia de una cantidad suficiente de observaciones alrededor de cada punto de interés, para que éstas puedan proveer la información adecuada para la estimación. Así mismo, los procedimientos de estimación no paramétricos pueden ayudar en el inicio de la investigación a descubrir la estructura probabilística que gobierna los datos de modo que los supuestos del análisis paramétrico estén bien fundamentados.

Entre los métodos no paramétricos de estimación para la función de densidad se encuentran los estimadores basados en núcleos que fueron introducidos por Rosenblatt [21] y Parzen [18]. Estos estimadores se construyen en cada punto del eje real de acuerdo con los valores muestrales más cercanos al mismo, es decir se considera un entorno alrededor de cada punto donde se desea estimar la densidad y basados en las observaciones que se encuentran en ese entorno se construye el estimador, dándole mayor peso a aquellas observaciones más cercanas y menor peso aquellas más alejadas, dentro del entorno. Para establecer los pesos se suele utilizar diversas funciones de ponderación llamadas núcleos. Los entornos están dados a partir de un parámetro

de suavizado o ventana, para hacernos una idea de los mismos podemos imaginarnos una bola centrada en el punto a estimar cuyo radio corresponde justamente al ancho de banda o ventana.

Las propiedades del estimador de la densidad dependen de la elección del núcleo y del ancho de la ventana. La combinación de la función de ponderación, el ancho de la ventana y el tamaño de muestra hacen a la bondad de la estimación resultante. Ventanas demasiado pequeñas derivarán en estimadores muy variables ya que en cada punto los entornos carecerán de suficientes observaciones en las cuales basar la estimación. Por otra parte, un ventana demasiado grande producirá estimadores muy suaves, que no lograrán captar la estructura local de la densidad dando lugar a estimadores sesgados.

Si aplicamos este estimador a datos procedentes de distribuciones con colas pesadas, con una ventana suficientemente pequeña para estimar bien la parte central de la distribución no lograremos estimar correctamente las colas de la distribución. Mientras que con un valor de ventana grande para la correcta estimación de las colas no podremos ver los detalles que ocurren en la parte principal de la distribución. Para superar estos defectos, Loftsgaaren y Quesenberry [16] y Wagner [22] propusieron un estimador conceptualmente similar al estudiado por Rosenblatt [21] y Parzen [18] pero cuyos entornos no son fijos sino que se adaptan al punto en el cual se está estimando. Estos estimadores se conocen con el nombre de estimadores por vecinos más cercanos con núcleos.

En muchas aplicaciones, las variables aleatorias toman valores en otros espacios en lugar de \mathbb{R} , un ejemplo sencillo aparece si estudiamos el viento y consideramos su velocidad conjuntamente con su dirección. En este contexto es natural pensar que estas variables descansan en un cilindro en lugar de \mathbb{R}^2 . Una gran cantidad de ejemplos pueden encontrarse en la astronomía, la geología, el análisis de imágenes, entre otros campos, donde las variables a estudiar se encuentran en la esfera, en grupos de Lie, y en diversas variedades diferenciables, ver por ejemplo Joshi *et.al.* [12] y Goh y Vidal [7]. Es por eso que resulta interesante estudiar una propuesta de estimación de la densidad cuando las variables involucradas toman valores en una variedad Riemanniana en lugar de \mathbb{R}^d , y por lo tanto esta estructura debe tenerse en cuenta a la hora de generar procedimientos de estimación e inferencia.

En este contexto, en los últimos años han surgido diferentes aportes, particularmente en el caso donde la variedad es la esfera (ver Mardia [17] y Fisher *et.al.* [6]). Estimadores no paramétrico con datos direccionales fueron estudiados en Hall *et.al.* [8], Bai *et.al.* [1] y más recientemente en Di Marzio *et.al.* [4], se ha propuesto un estimador basado en núcleos en el caso donde las variables toman valores en el toro. En Bhattacharya y Patrangenaru [2], Hendriks y Landsman [9] y Pennec [20] se concientran los primeros trabajos de inferencia estadística sobre variedades Riemannianas, ya sea desde la definición de las medidas de posición y escala hasta su estimación en

versiones clásicas y robustas.

Esta tesis se divide en tres capítulos. En el Capítulo 2 se dará un breve resumen de los estimadores no paramétricos de la función de densidad de variables aleatorias reales. El Capítulo 3 contiene las definiciones y preliminares geométricos necesarios para el estudio del estimador de densidad cuando consideramos variables aleatorias en variedades Riemannianas. Finalmente, en el Capítulo 4 se estudiarán dos familias de estimadores de la densidad de variables aleatorias que toman valores en una variedad Riemanniana. Por un lado, se estudiará el estimador propuesto por Pelletier [19] y algunas de sus propiedades. Por otro, se introducirán estimadores de vecinos más cercanos en este contexto de variedades y se probará que el estimador propuesto es uniforme y fuertemente consistente sobre un compacto de la variedad.

Capítulo 2

Estimación no paramétrica de la densidad

Una característica básica que describe el comportamiento de una variable aleatoria X es su función de densidad. El conocimiento de la función de densidad nos ayuda en muchos aspectos. Por ejemplo, si tenemos un conjunto de observaciones generadas a partir de la densidad f y queremos conocer cuantas observaciones caen en un conjunto podemos calcular a partir de la función de densidad f la probabilidad de que la variable aleatoria X pertenezca a ese determinado conjunto como una integral sobre dicho conjunto, es decir

$$P(X \in A) = \int_A f(x)dx.$$

Si este valor es alto para un cierto conjunto A comparado con la probabilidad sobre otro conjunto B , de manera informal se podría decir que dado un conjunto de observaciones, hay una alta probabilidad de encontrar una observación en la región A y baja en la region B , es decir, la función de densidad nos dirá donde las observaciones ocurren más frecuentemente.

En la mayoría de los estudios prácticos no se conoce la función de densidad de X directamente. Y en su lugar sólo contamos con un conjunto de observaciones X_1, \dots, X_n que suponemos independientes, idénticamente distribuidas y con función de densidad f desconocida. Nuestro objetivo es estudiar como estimar la función de densidad basándonos en la muestra aleatoria X_1, \dots, X_n .

Los métodos de estimación no paramétricos han surgido con el objetivo de dar una respuesta a este problema y han sido ampliamente estudiados. En este Capítulo estudiaremos dos propuestas para la estimación de la función de densidad: los estimadores de Rosenblatt–Parzen y los estimadores basados en los vecinos más cercanos.

2.1. Estimación de la densidad por núcleos

Sea X_1, \dots, X_n una muestra aleatoria con función de densidad $f(x)$. Como mencionamos anteriormente, el problema consiste en estimar $f(x)$ a partir de las observaciones. En primer lugar, intentaremos dar una idea intuitiva de la estimación de la función de densidad por núcleos.

Si X es una variable aleatoria con densidad f , luego dado $h > 0$

$$\frac{1}{2h}P(x - h < X < x + h) = \frac{1}{2h} \int_{x-h}^{x+h} f(x)dx$$

y claramente, tenemos que $\frac{1}{2h}P(x - h < X < x + h) \rightarrow f(x)$ si $h \rightarrow 0$. Por otro lado un estimador natural de $P(x - h < X < x + h)$ es simplemente considerar la proporción de la muestra que cae en el intervalo $(x - h, x + h)$. Entonces dado un h suficientemente pequeño podemos deducir el siguiente estimador de $f(x)$,

$$\tilde{f}(x) = \frac{1}{2h} \frac{\#\{X_i : X_i \in (x - h, x + h)\}}{n}.$$

Otra forma de expresar este estimador es de la siguiente manera, si definimos la función w como $w(x) = \frac{1}{2}I_{(|x|<1)}$, es fácil ver que $\tilde{f}(x)$ es equivalente a

$$\tilde{f}(x) = \sum_{i=1}^n \frac{1}{hn} w\left(\frac{x - X_i}{h}\right). \quad (2.1)$$

Notemos que $w \geq 0$, $\int w(s)ds = 1$, además, para cada $1 \leq i \leq n$ tenemos que $w\left(\frac{x-X_i}{h}\right) = \frac{1}{2}$ si y solo si $X_i \in (x - h, x + h)$, es decir la función w le otorga un peso uniforme a cada observación X_i en el entorno $(x - h, x + h)$. De esta forma, una manera de generalizar (2.1) sería reemplazar la función de peso o núcleo w por una función K no negativa que verifique la condición $\int K(x)dx = 1$. Así, si consideramos una función de pesos K con mayor suavidad obtendríamos un estimador más suave.

De esta manera obtenemos el estimador definido por Rosenblatt y Parzen ([21] y [18]) que constituye uno de los estimadores no paramétricos mas estudiados y que definimos de la siguiente manera,

$$\tilde{f}(x) = \frac{1}{hn} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (2.2)$$

donde K es una función núcleo, $h = h_n$ es llamado el *parametro de suavizado* o *ancho de ventana* y satisface $h_n \rightarrow 0$ si $n \rightarrow \infty$.

El paraméto de suavizado suele ser un punto crucial en el proceso de estimación, ya que como su nombre lo indica se encuentra altamente relacionado con el nivel de

suavización que se introduce en la estimación. En general los pesos utilizados decrecen de manera suave, dándole así menor pesos a las observaciones más alejadas del punto x . Algunas opciones posibles de núcleos, podrían ser

Núcleo Gaussiano:

$$K(t) = e^{-\frac{1}{2}t^2}.$$

Núcleo Epanechnicov:

$$K(t) = \begin{cases} \frac{3}{4}(1-t^2) & |t| \leq 1 \\ 0 & |t| > 1. \end{cases}$$

Núcleo Tricúbico:

$$K(t) = \begin{cases} (1-|t|^3)^3 & |t| \leq 1 \\ 0 & |t| > 1. \end{cases}$$

Las siguientes proposiciones establecen que el estimador de Rosenblatt–Parzen es efectivamente una función de densidad y asintóticamente insesgado.

Proposición 2.1.1. Si $\int_{+\infty}^{-\infty} K(x)dx = 1$ entonces $\int_{+\infty}^{-\infty} \tilde{f}(x)dx = 1$

Demostración.

$$\begin{aligned} \int_{+\infty}^{-\infty} \tilde{f}(x)dx &= \int_{+\infty}^{-\infty} \sum_{i=1}^n \frac{1}{hn} K\left(\frac{x-X_i}{h}\right) dx = \\ &= \frac{1}{n} \sum_{i=1}^n \int_{+\infty}^{-\infty} \frac{1}{h} K\left(\frac{x-X_i}{h}\right) dx = \frac{1}{n} \sum_{i=1}^n \int_{+\infty}^{-\infty} K(s)ds = 1 \end{aligned}$$

□

Proposición 2.1.2. Bajo los siguientes supuestos

- i) f es r -veces derivable, $f^{(r)}$ es una función acotada.
- ii) $\int K = 1$, $\int K(s)s^j ds = 0$ para todo $j = 1, \dots, r-1$ y $\int K(s)|s^r| ds < \infty$.

Tenemos que $E[\tilde{f}(x)] = f(x) + O(h^r)$ si $h \rightarrow 0$ para cada x .

Demostración.

$$E[\tilde{f}(x)] = E\left(\frac{1}{hn} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)\right) = \frac{1}{n} \sum_{i=1}^n E\left(\frac{1}{h} K\left(\frac{x-X_i}{h}\right)\right)$$

$$= E\left(\frac{1}{h}K\left(\frac{x-X_i}{h}\right)\right) = \frac{1}{h} \int_{+\infty}^{-\infty} K\left(\frac{x-u}{h}\right)f(u)du = \int_{+\infty}^{-\infty} K_h(y)f(x-y)dy$$

Donde $K_h(y) = \frac{1}{h}K\left(\frac{y}{h}\right)$. Haciendo el desarrollo de Taylor de f de orden r centrado en x tenemos

$$f(x-y) = f(x) - f'(x)y + \dots + \frac{f^{(r-1)}(x)}{(r-1)!}(-y)^{r-1} + \frac{f^{(r)}(\zeta_y)}{r!}(-y)^r.$$

Entonces

$$\begin{aligned} & \int_{+\infty}^{-\infty} K_h(y)(f(x) - f'(x)y + \dots + \frac{f^{(r-1)}(x)}{(r-1)!}(-y)^{r-1} + \frac{f^{(r)}(\zeta_y)}{r!}(-y)^r)dy = \\ & f(x) \int_{+\infty}^{-\infty} K_h(y)dy + \sum_{j=1}^{r-1} \frac{f^{(j)}(x)}{j!} \int_{+\infty}^{-\infty} K_h(y)(-y)^j dy + \int_{+\infty}^{-\infty} K_h(y) \frac{f^{(r)}(\zeta_y)}{r!}(-y)^r dy = \\ & f(x) + \frac{(-1)^r h^r}{r!} \int_{+\infty}^{-\infty} K(s) f^{(r)}(\zeta_s) s^r ds \\ & \left| E[\tilde{f}(x)] - f(x) \right| \leq \frac{h^r}{r!} \int_{+\infty}^{-\infty} K(s) |s^r| |f^{(r)}(\zeta_s)| ds \leq CO(h^r) \end{aligned}$$

concluyendo la demostración. \square

Este resultado muestra que si la ventana es mayor, el sesgo aumentará y para obtener menor sesgo habría entonces que considerar ventanas más pequeñas. Notemos que por la demostración anterior si tomamos $r = 2$ tenemos que $Sesgo(\tilde{f}(x)) = O(h^2)$. Usando argumentos similares a los considerados anteriormente se puede ver que $Var(\tilde{f}(x)) = O(\frac{1}{nh})$. Luego, si $h \rightarrow 0$ y $nh \rightarrow \infty$ tenemos que $\tilde{f}(x)$ es débilmente consistente a $f(x)$ para cada x .

Se pueden obtener resultados más fuertes de consistencia e incluso estudiar la distribución asintótica de $\tilde{f}(x)$. En el Capítulo 4 estudiaremos la extensión de este estimador en un espacio más general y probaremos su consistencia fuerte y uniforme sobre compactos. El estimador $\tilde{f}(x)$, pasará a ser un caso particular del estimador propuesto en el Capítulo 4 y por lo tanto, también tendremos resultados de consistencia más fuerte para el estimador expuesto anteriormente.

Respecto a la distribución asintótica, si bien existen resultados tanto para $\tilde{f}(x)$ como para su generalización a variedades Riemannianas, sus demostraciones no fueron objeto de estudio de esta tesis.

2.2. Estimación de la densidad por vecinos más cercanos con núcleo

Como mencionamos anteriormente el problema de escoger el valor de ancho de banda es no trivial. Pues un h demasiado pequeño tiene como efecto que la varianza del estimador aumente demasiado ya que son pocas las observaciones considerados en cada punto. Mientras que un valor demasiado alto da resultados con un alto sesgo debido a que se promedian demasiadas observaciones que no logran captar la tendencia o forma de la curva a estimar. A este compromiso en la elección del valor de h se le denomiña compromiso sesgo-varianza.

Una manera de dar una solución a este problema es considerar entornos variables. Es decir, en lugar de fijar un ancho de ventana y a partir de los valores muestrales que caen en él, estimar la función de densidad, la idea sería construir en cada punto donde deseamos estimar entornos que contengan una cantidad fija de observaciones. Más precisamente, sea $d(x, y) = |x - y|$ la distancia entre dos puntos x, y . Consideremos para cada valor de x las distancias $d(x, X_i)$ para $1 \leq i \leq n$ y llamemos $d_i(x)$ a las distancias ordenadas, es decir $d_i(x) = (d(x, X_i))^{(i)}$, el estadístico de orden i de las distancias al punto x .

Definimos el estimador de densidad por el método del *k-ésimo vecinos más cercanos* como

$$\hat{f}(x) = \frac{k}{2nd_k(x)}. \quad (2.3)$$

Con el fin de comprender un poco mejor esta definición, recordemos que, por lo visto en (2.1), para una muestra de tamaño n , uno esperaría aproximadamente $2hnf(x)$ observaciones dentro del intervalo $[x - h, x + h]$ para cada $h > 0$. Por otro lado exactamente k observaciones caerán dentro del intervalo $[x - d_k(x), x + d_k(x)]$, entonces es razonable esperar que k sea aproximadamente como $2d_k(x)nf(x)$. Y de aquí obtenemos el estimador de k vecinos más cercanos propuesto en (2.3).

Mientras que (2.1) está basado en un número de observaciones que yacen en un intervalo de longitud fija centrado en el punto de interés, el estimador de *k-ésimo vecinos más cercanos* es inversamente proporcional al tamaño del intervalo que contiene un número k de observaciones dado. Es posible generalizar el estimador de *k-ésimo vecinos más cercanos* combinando (2.1) con (2.3) obteniendo así el siguiente estimador

$$\hat{f}(x) = \frac{1}{nd_k(x)} \sum_{i=1}^n K\left(\frac{x - X_i}{d_k(x)}\right) \quad (2.4)$$

donde K es una función núcleo con las mismas propiedades que las definidas anteriormente, $k = k_n$ es una sucesión tal que $k_n \rightarrow \infty$ si $n \rightarrow \infty$ y $d_k(x)$ es la distancia entre x y el k -ésimo vecino más cercano.

Al igual que lo comentado en la Sección anterior en el Capítulo 4, propondremos una extensión del estimador de vecinos más cercanos con núcleos en el contexto de variedades Riemannianas. Y daremos resultados de consistencia fuerte y uniforme sobre un conjunto compacto. Como se verá más adelante este resultado se aplicará también al estimador $\hat{f}(x)$.

Capítulo 3

Preliminares Geométricos

3.1. Variedades Diferenciables

Definición 3.1.1. Una variedad topológica de dimensión $d \in \mathbb{N}$ es un espacio topológico (M, τ) Hausdorff, de base numerable, que es localmente homeomorfo a \mathbb{R}^d . Es decir, para cada $p \in M$ existe un abierto $U \in \tau$ y un abierto A de \mathbb{R}^d , tal que $p \in U$ y existe un homomorfismo $\varphi : U \rightarrow A$.

En la situación de la definición, introducimos los siguientes conceptos:

- (U, φ) es un entorno coordenado de p . También llamaremos a (U, φ) una carta alrededor de p .
- Sea $\pi^i : \mathbb{R}^d \rightarrow \mathbb{R}$, dadas por $\pi^i(x_1, \dots, x_d) = x_i$. A las funciones $\varphi_i = \pi^i \circ \varphi$ las llamaremos *coordenada i-ésima*
- Una colección de entornos coordinados $\mathcal{A} = \{(U_\alpha, \varphi_\alpha) : \alpha \in \mathcal{I}\}$ es un atlas si $M = \cup_\alpha U_\alpha$.

Observación 3.1.2. La dimensión de una variedad topológica está bien definida. Pues supongamos que tenemos V entorno de p , B un abierto de $\mathbb{R}^{d'}$ y $\psi : V \rightarrow B$ un homeomorfismo, luego $\psi \circ \varphi^{-1} : \varphi(U \cap V) \rightarrow \psi(U \cap V)$ es homeomorfismo entre abiertos de \mathbb{R}^d y $\mathbb{R}^{d'}$, con lo cual necesariamente $d = d'$.

Sea (M, τ) una variedad topológica de dimensión d y sean (U, φ) y (V, ψ) dos cartas. Diremos que son compatibles si $U \cap V = \emptyset$ o bien si la función cambio de coordenadas es un difeomorfismo, es decir, que las siguientes aplicaciones sean diferenciables como funciones entre abiertos de \mathbb{R}^d :

$$\begin{aligned}\psi \circ \varphi^{-1} : \varphi(U \cap V) &\longrightarrow \psi(U \cap V), \\ \varphi \circ \psi^{-1} : \psi(U \cap V) &\longrightarrow \varphi(U \cap V).\end{aligned}$$

Definición 3.1.3. Diremos que un atlas $\mathcal{A} = \{(U_\alpha, \varphi_\alpha) : \alpha \in \mathcal{I}\}$ es diferenciable si sus cartas son compatibles entre sí. Si un atlas diferenciable \mathcal{D} es maximal lo llamaremos una estructura diferenciable de la variedad M . Con maximal queremos decir lo siguiente: Si (U, φ) es una carta de M que es compatible con todas las cartas de \mathcal{D} entonces $(U, \varphi) \in \mathcal{D}$.

Definición 3.1.4. Una variedad diferenciable de dimensión d es una terna (M, τ, \mathcal{D}) donde (M, τ) es una variedad topológica de dimensión d y \mathcal{D} una estructura diferenciable.

Cuando digamos que M es una variedad diferenciable realmente nos estaremos refiriendo a la terna antes mencionada. A partir de ahora supondremos que todas las variedades son conexas, lo cual no significa una restricción importante para la teoría. Así mismo trabajaremos con variedades sin bordes.

Ejemplo 3.1.5. El espacio euclídeo d dimensional con la estructura diferenciable inducida por la carta global (\mathbb{R}^d, id) es una variedad diferenciable. También \mathbb{R}^d es una variedad diferenciable, distinta a la anterior, si consideramos la estructura diferenciable inducida por la carta (\mathbb{R}^d, ψ) , con $\psi(x) = (x_1^3, \dots, x_d^3)$.

Ejemplo 3.1.6. Sea S^d la esfera de \mathbb{R}^{d+1} . El conjunto de cartas (U_i^l, ρ_i^l) donde $i = 1, \dots, d+1$ y $l = -1, 1$, $U_i^l = \{x \in S^d : l \cdot x_i > 0\}$ y $\rho_i^l(x) = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_{d+1})$ es un atlas diferenciable. Luego, S^d dotado de la estructura diferenciable inducida por este atlas es una variedad diferenciable de dimensión d .

Ejemplo 3.1.7. El caso anterior es un caso particular del siguiente hecho. Sea $f : \mathbb{R}^d \rightarrow \mathbb{R}$ diferenciable. Sea $b \in \mathbb{R}$ un valor regular de f (i.e. el rango de Df_q es 1 para todo $q \in f^{-1}(b)$) tal que $f^{-1}(b) \neq \emptyset$. Luego, $M = f^{-1}(b)$ es una variedad diferenciable de dimensión $d-1$. La demostración de este hecho, así como sus generalizaciones, puede verse en Boothby [3].

Ejemplo 3.1.8. En S^d consideramos la relación de equivalencia $x \sim y$ si y sólo si $x = \pm y$. Luego, el conjunto $\mathbb{RP}^d = S^d / \sim$ tiene estructura de variedad diferenciable de dimensión d llamada el espacio proyectivo real de dimensión d .

Definición 3.1.9. Sean M y \tilde{M} dos variedades diferenciables de dimensiones d y \tilde{d} respectivamente. Diremos que una aplicación $f : M \rightarrow \tilde{M}$ es diferenciable en un punto $p \in M$ si para un par de cartas (U, φ) y $(\tilde{U}, \tilde{\varphi})$ alrededor de p y de $f(p)$ respectivamente, tales que $f(U) \subseteq \tilde{U}$ se tiene que la aplicación $\tilde{\varphi} \circ f \circ \varphi^{-1} : \varphi(U) \rightarrow \tilde{\varphi}(\tilde{U})$ es diferenciable en $\varphi(p)$. $f : M \rightarrow \tilde{M}$ es diferenciable si lo es para todo $p \in M$.

La definición anterior no depende de los entornos coordenados elegidos. En efecto, sean (V, ψ) y $(\tilde{V}, \tilde{\psi})$ otras cartas alrededor de p y $f(p)$ respectivamente tales que $f(V) \subseteq \tilde{V}$, entonces

$$\tilde{\psi} \circ f \circ \psi^{-1} = (\tilde{\psi} \circ \tilde{\varphi}^{-1}) \circ (\tilde{\varphi} \circ f \circ \varphi^{-1}) \circ (\varphi \circ \psi^{-1})$$

en un entorno de $\psi(p)$. Como $\tilde{\psi} \circ \tilde{\varphi}^{-1}$ y $\varphi \circ \psi^{-1}$ son diferenciables en todo su dominio por ser cambios de coordenadas y $\tilde{\varphi} \circ f \circ \varphi^{-1}$ es diferenciable en $\varphi(p)$, entonces se tiene que $\tilde{\psi} \circ f \circ \psi^{-1}$ es diferenciable en $\psi(p)$.

Notamos con $\mathcal{F}(M)$ a las funciones diferenciables de M en \mathbb{R} .

Consideremos una variedad diferenciable M de dimensión d y fijemos un punto $p \in M$. Sea $C_p = \{\gamma : (-\epsilon_\gamma, \epsilon_\gamma) \rightarrow M : 0 < \epsilon_\gamma, \gamma(0) = p, \gamma \text{ diferenciable}\}$. Diremos que dos curvas γ_1, γ_2 en C_p están relacionadas si para alguna carta (U, φ) alrededor de p se verifica que $\frac{d}{dt}(\varphi \circ \gamma_1)(t) \Big|_{t=0} = \frac{d}{dt}(\varphi \circ \gamma_2)(t) \Big|_{t=0}$. Es fácil verificar que es una relación de equivalencia y que es independiente de la carta elegida.

Definición 3.1.10. Un vector tangente a M en p es una de las clases de equivalencia definidas de C_p / \sim . Al conjunto de clases C_p / \sim lo llamamos fibra tangente de M en p y lo notamos con $T_p M$.

Notaremos con $\gamma'(0)$ al vector tangente inducido por la curva $\gamma \in C_p$.

Sea $\gamma \in C_p$ y consideremos su vector tangente asociado $\gamma'(0)$. Si $f : M \rightarrow \mathbb{R}$ es una aplicación diferenciable entonces podemos considerar $f \circ \gamma : (-\epsilon_\gamma, \epsilon_\gamma) \rightarrow \mathbb{R}$ y calcular $\frac{d}{dt}(f \circ \gamma)(t) \Big|_{t=0}$. A este número real lo llamaremos la derivada direccional de f en la dirección del vector tangente $\gamma'(0)$. Se puede ver sin dificultad que si $\gamma_1 \sim \gamma_2$ entonces

$$\frac{d}{dt}(f \circ \gamma_1)(t) \Big|_{t=0} = \frac{d}{dt}(f \circ \gamma_2)(t) \Big|_{t=0}$$

Por lo tanto, cada clase de equivalencia proporciona una única derivada direccional en p y esta es la otra forma en la que veremos a los vectores tangentes. Es decir, como velocidades de curvas o bien como derivadas direccionales. Si fijamos un vector tangente en p , $v = \gamma'(0)$, definimos la aplicación: $D_v : \mathcal{F}(M) \rightarrow \mathbb{R}$ como $f \mapsto D_v(f) = v(f)$ con $v(f) = \frac{d}{dt}(f \circ \gamma)(t) \Big|_{t=0}$.

La aplicación D_v verifica las propiedades de una derivación:

- i) $D_v(f + g) = D_v(f) + D_v(g)$.
- ii) $D_v(f \cdot g) = D_v(f)g(p) + D_v(g)f(p)$.

Por lo tanto, tenemos una definición equivalente, y muy útil (ver Keilhauer [14]), de vector tangente:

Definición 3.1.11. Sea M una variedad diferenciable y $p \in M$, un vector tangente en p a M es una aplicación $D_v : \mathcal{F}(M) \rightarrow \mathbb{R}$ que satisface las propiedades *i)* y *ii)*.

Observación 3.1.12. *Las derivaciones tienen su dominio natural en el germen de funciones, este conjunto es el que resulta de cocientar el conjunto de funciones diferenciables en un entorno de p y por la relación de equivalencia $f \sim g$ si y sólo si $f|_W = g|_W$ en W entorno de p , ver Warner [23].*

La fibra tangente o espacio tangente $T_p M$ es un espacio vectorial real de igual dimensión que la variedad M . Podemos construir una base de este de la siguiente manera. Sea $\dim M = d$ y $(U, \varphi = \varphi_1, \dots, \varphi_d)$ una carta alrededor de p . Sea $\{e_1, \dots, e_d\}$ la base canónica de \mathbb{R}^d , es decir $e_i = (0, \dots, 1, \dots, 0)$ es el vector con un 1 en la posición i -ésima y cero en las demás. Consideremos la recta $r_i : I \subseteq \mathbb{R} \rightarrow \mathbb{R}^d$ dada por $r_i(t) = \varphi(p) + t.e_i$, luego, levantándolas a M mediante el homeomorfismo φ^{-1} , tenemos las siguientes curvas en M

$$\gamma_i = \varphi^{-1}(\varphi(p) + t.e_i)$$

Se puede ver fácilmente que $\{\gamma'_1(0), \dots, \gamma'_d(0)\}$ es una base de $T_p M$. Notaremos con $\frac{\partial}{\partial \varphi_i}|_p$ al vector tangente $\gamma'_i(0)$. Veamos como es la derivación asociada al vector tangente $\frac{\partial}{\partial \varphi_i}|_p$. Por definición tenemos que

$$\begin{aligned} D_{\frac{\partial}{\partial \varphi_i}|_p}(f) &= \frac{d}{dt}(f \circ \gamma_i)(t) \Big|_{t=0} \\ &= \frac{d}{dt}(f \circ \varphi^{-1}) \circ (\varphi \circ \gamma_i)(t) \Big|_{t=0} = \sum_{j=1}^d \frac{\partial(f \circ \varphi^{-1})}{\partial u_j} \Big|_{\varphi(p)} \frac{d}{dt}(\varphi_j \circ \gamma_i) \Big|_{t=0} \\ &= \sum_{j=1}^d \frac{\partial(f \circ \varphi^{-1})}{\partial u_j} \Big|_{\varphi(p)} \frac{d}{dt}(\pi^j \circ \varphi)(\varphi^{-1}(\varphi(p) + t.e_i)) \Big|_{t=0} \\ &= \sum_{j=1}^d \frac{\partial(f \circ \varphi^{-1})}{\partial u_j} \Big|_{\varphi(p)} \frac{d}{dt} \pi^j(\varphi(p) + t.e_i) \Big|_{t=0}, \end{aligned}$$

Por lo tanto, $D_{\frac{\partial}{\partial \varphi_i}|_p} : \mathcal{F}(M) \rightarrow \mathbb{R}$ está dada por

$$f \longrightarrow \frac{\partial f}{\partial \varphi_i}|_p := \frac{\partial(f \circ \varphi^{-1})}{\partial u_i} \Big|_{\varphi(p)}.$$

De ahora en adelante nos referiremos a $\{\frac{\partial}{\partial \varphi_1}|_p, \dots, \frac{\partial}{\partial \varphi_d}|_p\}$ como la base de vectores tangentes inducida por la carta (U, φ) .

Si $v \in T_p M$, entonces podemos escribir a v en la base de vectores tangentes inducidos por la carta, $v = \sum_{i=1}^d v_i \frac{\partial}{\partial \varphi_i}|_p$. La d -upla (v_1, \dots, v_d) son las coordenadas de v con respecto a la carta (U, φ) . Si pensamos al vector v como derivación y evaluamos en φ_i , tenemos que $v(\varphi_i) = D_v(\varphi_i) = v_i$, es decir $v = \sum_{i=1}^d v_i(\varphi_i) \frac{\partial}{\partial \varphi_i}|_p$.

El fibrado tangente de una variedad M es la unión disjunta de todas las fibras tangentes $TM = \bigcup_{p \in M} T_p M$. Este conjunto tiene estructura de variedad diferenciable de dimensión $2d$ si $\dim M = d$. Sea $\pi : TM \rightarrow M$ la proyección canónica del fibrado tangente a M , i.e. $\pi(v) = q$ si $v \in T_q M$. La topología de TM es la que hace a π una función continua. Las cartas de M inducen de manera natural cartas en TM . Si (U, φ) es una carta de M , sea $(TU, \bar{\varphi})$, dada por $TU : \{v \in TM : \pi(v) \in U\}$, el cual es un abierto de TM por ser la pre-imagen vía la proyección de π del abierto U , y $\bar{\varphi}(v) = (\varphi(\pi(v)), v(\varphi_1), \dots, v(\varphi_d)) \in I\!\!R^{2d}$.

Definición 3.1.13. Un campo vectorial tangente \mathbf{X} sobre M es una aplicación que asigna a cada punto p un vector tangente a M en ese punto, $\mathbf{X}(p) \in T_p(M)$. Un campo tangente es una función $\mathbf{X} : M \rightarrow TM$ entre la variedad y el fibrado tangente de modo que el siguiente diagrama commuta

$$\begin{array}{ccc} & TM & \\ \mathbf{X} \nearrow & \downarrow \pi_1 & \\ M & \xrightarrow{id_M} & M \end{array}$$

donde id_M es la identidad de M . Un campo tangente \mathbf{X} se dice diferenciable si lo es como aplicación entre variedades.

El conjunto de campos tangentes diferenciables lo denotaremos con $\mathcal{X}(M)$. A menos que se diga lo contrario, cuando nos referimos a un campo sobre una variedad nos estamos refiriendo a un campo tangente diferenciable.

Ejemplo 3.1.14. Sea $M = S^n \times I\!\!R$ un cilindro con n impar, luego si definimos $\mathbf{X}(p) = (-p_2, p_1, \dots, -p_{n+1}, p_n, 1)$, \mathbf{X} es un campo tangente diferenciable.

Si $\mathbf{X} \in \mathcal{X}(M)$, y (U, φ) es una carta de M entonces la expresión local del campo en U esta dada por

$$\mathbf{X}|_U(q) = \sum_{i=1}^d \mathbf{X}_i(q) \frac{\partial}{\partial \varphi_i}|_q$$

donde $\mathbf{X}_i : U \rightarrow I\!\!R$ son funciones diferenciables.

3.2. Conexiones y geodésicas

Definición 3.2.1. Sea M una variedad diferenciable. Una conexión lineal ∇ sobre M es una aplicación $\nabla : \mathcal{X}(M) \times \mathcal{X}(M) \rightarrow \mathcal{X}(M)$ que verifica

- 1) $\nabla_{\mathbf{X}+\mathbf{Y}}\mathbf{Z} = \nabla_{\mathbf{X}}\mathbf{Z} + \nabla_{\mathbf{Y}}\mathbf{Z}$
- 2) $\nabla_{\mathbf{X}}(\mathbf{Y} + \mathbf{Z}) = \nabla_{\mathbf{X}}\mathbf{Y} + \nabla_{\mathbf{X}}\mathbf{Z}$
- 3) $\nabla_{f\mathbf{X}}\mathbf{Y} = f\nabla_{\mathbf{X}}\mathbf{Y}$
- 4) $\nabla_{\mathbf{X}}f\mathbf{Y} = \mathbf{X}(f)\mathbf{Y} + f\nabla_{\mathbf{X}}\mathbf{Y}$

donde f y g están en $\mathcal{F}(M)$, $\nabla_{\mathbf{X}}\mathbf{Y}$ se lee la derivada covariante de \mathbf{Y} con respecto a \mathbf{X} .

Sea ∇ una conexión lineal sobre M y (U, φ) una carta de M . Restringiendo la conexión a U podemos calcular $\nabla_{\frac{\partial}{\partial\varphi_i}}(\frac{\partial}{\partial\varphi_j})$ que resulta ser un campo sobre U . Por lo tanto, lo podemos escribir en la base inducida por la carta (U, φ) como

$$\nabla_{\frac{\partial}{\partial\varphi_i}}\frac{\partial}{\partial\varphi_j} = \sum_{k=1}^d \Gamma_{ij}^k \frac{\partial}{\partial\varphi_k}.$$

Las d^3 funciones diferenciables $\Gamma_{ij}^k : U \rightarrow \text{IR}$ con $i, j, k \in \{1, \dots, d\}$ se llaman símbolos de Christoffel de la conexión ∇ en U .

Consideremos una variedad dotada de una conexión lineal ∇ y sean \mathbf{X} y \mathbf{Y} dos campos cuyas expresiones locales con respecto a la carta (U, φ) son $\mathbf{X}|_U(q) = \sum_{i=1}^d \mathbf{X}_i(q)\frac{\partial}{\partial\varphi_i}|_q$ y $\mathbf{Y}|_U(q) = \sum_{j=1}^d \mathbf{Y}_j(q)\frac{\partial}{\partial\varphi_j}|_q$. Luego, de las propiedades de conexión 1), 2), 3), 4) se tiene que

$$\nabla_{\mathbf{X}}\mathbf{Y}|_q = \sum_{j=1}^d \mathbf{X}(q)(\mathbf{Y}_j)\frac{\partial}{\partial\varphi_j}|_q + \sum_{ijk=1}^d \mathbf{X}_i(q)\mathbf{Y}_j(q)\Gamma_{ij}^k(q)\frac{\partial}{\partial\varphi_k}|_q. \quad (3.1)$$

Observación 3.2.2. La conexión en una carta queda completamente determinada por los símbolos de Christoffel.

Proposición 3.2.3. Sea M dotada de una conexión lineal ∇ . Si $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ son campos de M tales que $X(p) = Y(p)$ para algún $p \in M$, entonces $\nabla_{\mathbf{X}}\mathbf{Z}|_p = \nabla_{\mathbf{Y}}\mathbf{Z}|_p$. Por otro lado, sea $c : I \rightarrow M$ una curva en M tal que $c(0) = p$ y $c'(0) = \mathbf{X}(p)$, si $\mathbf{Y} \circ c = \mathbf{Z} \circ c$, entonces $\nabla_{\mathbf{X}}\mathbf{Y}|_p = \nabla_{\mathbf{X}}\mathbf{Z}|_p$.

Demostración. La proposición se deduce de la fórmula para la expresión local (3.1). \square

Observación 3.2.4. Usualmente notaremos $\nabla_X \mathbf{Y}|_p = \nabla_{X(p)} \mathbf{Y}$.

Ejemplo 3.2.5. Sea $M = \mathbb{R}^d$, consideremos la carta global $(\mathbb{R}^d, Id_{\mathbb{R}^d})$ y $\{\frac{\partial}{\partial u_1}, \dots, \frac{\partial}{\partial u_d}\}$ los vectores tangentes inducidos por la carta global. Definimos $\nabla : \chi(\mathbb{R}^d) \times \chi(\mathbb{R}^d) \rightarrow \chi(\mathbb{R}^d)$ como $\nabla_X \mathbf{Y} = \sum_i \mathbf{X}(Y_i) \frac{\partial}{\partial u_i}$, donde $\mathbf{X} = \sum_{i=1}^d X_i \frac{\partial}{\partial u_i}$ y $\mathbf{Y} = \sum_{j=1}^d Y_j \frac{\partial}{\partial u_j}$. ∇ resulta una conexión lineal y los símbolos de Christoffel son nulos en la carta que hemos considerado $(\mathbb{R}^d, Id_{\mathbb{R}^d})$.

Definición 3.2.6. Sea $c : I \rightarrow M$ una curva y $\mathbf{X} : I \rightarrow TM$, decimos que \mathbf{X} es un campo tangente diferenciable a lo largo de la curva c , si es una aplicación diferenciable y si $\mathbf{X}(t) \in T_{c(t)}M$ para todo $t \in I$. Es decir el siguiente diagrama comuta

$$\begin{array}{ccc} & TM & \\ & \nearrow \mathbf{x} & \downarrow \pi_1 \\ I & \xrightarrow{c} & M \end{array}$$

Al conjunto de campos tangentes diferenciables a largo de c los notamos con χ_c .

Ejemplo 3.2.7. Sea $c : I \rightarrow M$ una curva de M , luego la curva del fibrado tangente $c'(t)$ dada por las velocidades de c en t es un campo a lo largo de c .

Observación 3.2.8. Si $\mathbf{X} \in \chi(M)$, entonces $\mathbf{X} \circ c$ es un campo a lo largo de c . Por otro lado, si $\mathbf{X} \in \chi_c$ podemos extender localmente \mathbf{X} a un campo de $\mathbf{Y} \in \chi(M)$ de tal forma que $\mathbf{Y} \circ c = \mathbf{X}$. La extensión no es única, pero si \mathbf{Z} es otra extensión, como $\mathbf{Y} \circ c = \mathbf{Z} \circ c$, tenemos que $\nabla_{c'(t)} \mathbf{Y} = \nabla_{c'(t)} \mathbf{Z}$.

Ahora estamos en condiciones de definir la derivada covariante a lo largo de curvas. Esta noción nos permite generalizar el concepto de aceleración de una curva.

Definición 3.2.9. Sea M dotada de una conexión ∇ , $c : I \rightarrow M$ una curva y $\mathbf{X} \in \chi_c$. Llamaremos derivada covariante de \mathbf{X} a lo largo de c a la aplicación $\frac{D\mathbf{X}}{dt} : I \rightarrow T(M)$ definida por $\frac{D\mathbf{X}}{dt}|_{t_0} = \nabla_{c'(t_0)} \mathbf{Y}$ donde \mathbf{Y} es una extensión a $\chi(M)$ de \mathbf{X} .

Notar que $\frac{D\mathbf{X}}{dt}$ es un campo a lo largo de c .

Observación 3.2.10. Sea (U, φ) una carta de M y \mathbf{X} un campo a lo largo de la curva c . Luego, $\mathbf{X} = \sum_{i=1}^d \mathbf{X}_i(t) \frac{\partial}{\partial \varphi}|_{c(t)}$, donde $\mathbf{X}_i : I \rightarrow \mathbb{R}$ son d funciones diferenciables. De las propiedades de la derivada covariante y de la definición de la derivada covariante a lo largo de una curva se deduce la siguiente expresión local:

$$\frac{D\mathbf{X}}{dt}(t) = \sum_{k=1}^d \left(\frac{d\mathbf{X}_k}{dt}|_t + \sum_{i,j=1}^d \mathbf{X}_i(t) \frac{\partial \varphi_j \circ c}{\partial t}|_t \Gamma_{ij}^k(c(t)) \right) \frac{\partial}{\partial \varphi_k}|_{c(t)}.$$

En particular si $\mathbf{X}(t) = c'(t)$, como $c'(t) = \sum_{i=1}^d c'(t)(\varphi_i) \frac{\partial}{\partial \varphi_i}|_{c(t)}$, tenemos que $\mathbf{X}_i(t) = \frac{\partial \varphi_i \circ c}{\partial t}|_t$, entonces la expresión local en este caso es

$$\frac{Dc'}{dt}(t) = \sum_{k=1}^d \left(\frac{\partial^2 \varphi_k \circ c}{\partial^2 t}|_t + \sum_{i,j=1}^d \frac{\partial \varphi_i \circ c}{\partial t}|_t \frac{\partial \varphi_j \circ c}{\partial t}|_t \Gamma_{ij}^k(c(t)) \right) \frac{\partial}{\partial \varphi_k}|_{c(t)}. \quad (3.2)$$

Ejemplo 3.2.11. Sea \mathbb{R}^d dotada de la conexión ∇ definida en el Ejemplo 3.2.5. Para una curva $c(t) = (c_1(t), \dots, c_d(t))$ en \mathbb{R}^d se tiene que $\frac{Dc'}{dt}(t) = \sum_{i=1}^d \frac{\partial^2 c_i}{\partial^2 t}|_t \frac{\partial}{\partial u_i}|_{c(t)}$.

El campo $\frac{Dc'}{dt}(t)$ es el campo aceleración de c . Las curvas de \mathbb{R}^d de aceleración nula son las rectas. La generalización de este hecho es el concepto de geodésica.

Definición 3.2.12. Diremos que una curva c es una geodésica si $\frac{Dc'}{dt}(t) = 0$.

De la fórmula (3.2) tenemos que c es una geodésica si y sólo si para toda carta se tiene que

$$\frac{Dc'}{dt}(t) = \sum_{k=1}^d \left(\frac{\partial^2 \varphi_k \circ c}{\partial^2 t}|_t + \sum_{i,j=1}^d \frac{\partial \varphi_i \circ c}{\partial t}|_t \frac{\partial \varphi_j \circ c}{\partial t}|_t \Gamma_{ij}^k(c(t)) \right) \frac{\partial}{\partial \varphi_k}|_{c(t)} = 0$$

En consecuencia c será una geodésica si para todo $1 \leq k \leq d$:

$$\frac{\partial^2 \varphi_k \circ c}{\partial^2 t}|_t + \sum_{i,j=1}^d \frac{\partial \varphi_i \circ c}{\partial t}|_t \frac{\partial \varphi_j \circ c}{\partial t}|_t \Gamma_{ij}^k(c(t)) = 0 \quad (3.3)$$

Teorema 3.2.13. Si $v \in T_p M$, existe un intervalo I de \mathbb{R} tal que $0 \in I$ y una geodésica $\gamma_v : I \rightarrow M$ que satisface:

i) $\gamma_v(0) = p$ y $\gamma'_v(0) = v$.

ii) Si $\alpha : J \rightarrow M$ es otra geodésica que satisface i), entonces $\gamma_v|_{I \cap J} = \alpha|_{I \cap J}$.

Demostración. La demostración de este teorema puede verse en Lee [15]. La idea de la demostración es la siguiente. Sea (U, φ) una carta alrededor de p . Consideremos el siguiente sistema de ecuaciones diferenciales de segundo orden

$$\frac{d^2 q^k}{dt^2}(t) + \sum_{i,j=1}^d \dot{q}^i(t) \dot{q}^j(t) \Gamma_{ij}^k(c(t)) = 0$$

donde $1 \leq k \leq d$, con condiciones iniciales $q^i(0) = \varphi^i(0)$ y $\dot{q}^i(0) = v_i$, con v_i la i -ésima coordenada del vector v con respecto a la carta (U, φ) . El teorema de existencia y unicidad de ecuaciones diferenciales ordinarias nos permite encontrar una solución $q = (q^1, \dots, q^d)$. Luego definiendo $\gamma_v(t) = \varphi^{-1}(q(t))$, se comprueba fácilmente que γ_v satisface la ecuación (3.3) y la propiedad *i*). La propiedad *ii*) se desprende de la unicidad de ecuaciones diferenciales. \square

Definición 3.2.14. Sea $c : I \rightarrow M$ una geodésica de M con respecto a una conexión ∇ , tal que $c(0) = p$. Decimos que c es maximal si para toda geodésica $\gamma : J \rightarrow M$ tal que con $I \subseteq J$ y $\gamma|_I = c$ se tiene que $I = J$.

Definición 3.2.15. Una variedad se dice completa si el intervalo donde está definida toda geodésica maximal es la recta real.

Sea $v \in T_p M$, vamos a notar a la geodésica maximal que en el instante cero pasa por p con velocidad v como $\gamma_v : I_v \rightarrow M$. Con $\widetilde{T_p M}$ notamos al abierto de $T_p M$ que contiene a todos los vectores tangentes v tales que $1 \in I_v$. Observamos que 0_p , el vector nulo de $T_p M$ siempre pertenece $\widetilde{T_p M}$ y que este conjunto es estrellado con respecto a 0_p . Cuando la variedad es completa, $\widetilde{T_p M} = T_p M$.

Ejemplo 3.2.16. *\mathbb{R}^d dotado con la conexión usual es una variedad completa. Si al espacio euclídeo le sacamos un punto $\mathbb{R}^d - \{p\}$ deja de ser completo. Una variedad compacta dotada de una conexión resulta completa.*

Definición 3.2.17. Sea M dotada de una conexión ∇ y sea $p \in M$. La aplicación exponencial en p es la aplicación $\exp_p : \widetilde{T_p M} \rightarrow M$ dada por

$$\exp_p(v) = \gamma_v(1)$$

Observación 3.2.18. *Por definición $\exp_p(tv) = \gamma_{tv}(1)$. Si consideramos la curva $c(s) = \gamma_v(s.t) : \frac{1}{t}I_v \rightarrow M$ podemos ver sin dificultad que resulta una geodésica, que $c(0) = p$ y que $c'(0) = tv$. Luego por unicidad de la geodésica, $\frac{1}{t}I_v \subseteq I_{tv}$ y $c = \gamma_{tv}$. En particular, $c(1) = \gamma_{tv}(1)$, con lo cual*

$$\exp_p(tv) = \gamma_v(t).$$

Es decir la función exponencial transforma segmentos de rectas del tangente en geodésicas de la variedad.

El siguiente Teorema es consecuencia de que el diferencial de la aplicación identidad en 0_p es la identidad y del Teorema de la función inversa para variedades. Su demostración puede verse en Do Carmo [5].

Teorema 3.2.19. *Sea M una variedad diferencial dotada de una conexión ∇ . Sea $p \in M$. Existe un entorno U_0 de $0_p \in T_p(M)$ y un entorno U_p de p tal que la aplicación exponencial restringida a estos entornos es un difeomorfismo.*

3.3. Variedades Riemannianas

Definición 3.3.1. Un métrica Riemanniana g sobre una variedad diferenciable M es un campo tensorial diferenciable de tipo $(0, 2)$ sobre M , simétrico y definido positivo. Es decir para cada punto $p \in M$, la restricción de la métrica a la fibra $T_p(M)$, que notamos con g_p , es un producto escalar definido positivo sobre el espacio tangente.

Definición 3.3.2. Una variedad Riemanniana es una variedad diferenciable M dotada de un métrica Riemanniana g , que denotaremos con (M, g) .

Ejemplo 3.3.3. Sea M una subvariedad de \mathbb{R}^d (cuando $d = 3$ hablamos de superficies). Luego podemos restringir la métrica canónica de \mathbb{R}^d , es decir el producto interno usual, a cada fibra tangente (plano tangente cuando $d = 3$), es decir $v, w \in T_p\mathbb{R}^d$ $g_p(v, w) = \langle v, w \rangle$.

Ejemplo 3.3.4. Sea $M = B_1(0) = \{x \in \mathbb{R}^2 : \|x\| < 1\}$ la bola euclídea de radio 1. Para $v, w \in T_p M$ sea

$$g_p(v, w) = \frac{1}{(1 - \|p\|^2)^2} \langle v, w \rangle.$$

donde $\|p\|$ es la norma euclídea de \mathbb{R}^2 . g es una métrica Riemanniana conocida con el nombre de hiperbólica.

Sea (M, g) una variedad Riemanniana y sea $\gamma : [a, b] \rightarrow M$ una curva diferenciable. La longitud de γ está dada por

$$L(\gamma) = \int_a^b \left\| \gamma'(t) \right\| dt = \int_a^b \left(g_{\gamma'(t)}(\gamma'(t), \gamma'(t)) \right)^{1/2} dt.$$

Definición 3.3.5. Sea (M, g) una variedad Riemanniana y dos puntos p y $q \in M$, definimos la distancia entre p y q como:

$$d_g(p, q) = \inf_{\gamma} \{L(\gamma) : \gamma : [0, 1] \rightarrow M, \gamma(0) = p, \gamma(1) = q\}.$$

El siguiente es el teorema fundamental de la geometría Riemanniana, su demostración puede verse en Do Carmo [5].

Teorema 3.3.6. Toda variedad diferenciable admite una métrica Riemanniana.

Teorema 3.3.7. Sea (M, g) una variedad Riemanniana, luego existe una única conexión lineal ∇ , tal que:

- i) $\nabla_X Y - \nabla_Y X = [X, Y]$ (simétrica o libre de torsión).
- ii) $X(g(Y, Z)) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)$ (compatible con la métrica g).

Para todo $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \in \mathcal{X}(M)$.

La demostración puede verse en Boothby [3] y Do Carmo [5], entre otros.

Definición 3.3.8. A esta conexión se la conoce con el nombre de conexión de Levi-Civita.

Dada una carta (U, φ) los símbolos de Christoffel están dados por (ver Do Carmo [5]):

$$\Gamma_{ij}^k(p) = \frac{1}{2} \sum_{l=1}^d \left(\frac{\partial g_{jl}}{\partial \varphi_i}|_p + \frac{\partial g_{il}}{\partial \varphi_j}|_p - \frac{\partial g_{ij}}{\partial \varphi_l}|_p \right) g^{lk}(p),$$

donde $g_{ij} : U \rightarrow I\!\!R$ es la función diferenciable dada por $g_{ij}(p) = g_p(\frac{\partial}{\partial \varphi_i}, \frac{\partial}{\partial \varphi_j})$ y g^{lk} es la entrada (l, k) de la matriz inversa de la matriz simétrica definida positiva $g = [g_{ij}]$. Llamamos a la matriz $g = [g_{ij}]$ la representación local de la métrica en la carta.

Del Teorema 3.3.7 podemos concluir que toda variedad Riemanniana admite una conexión. El siguiente resultado es el conocido teorema de Hopf-Rinow, ver Lee [15] y Do Carmo [5].

Teorema 3.3.9. Sea (M, g) una variedad Riemanniana dotada de su conexión de Levi-Civita. Son equivalentes:

- i) (M, g) es completa.
- ii) (M, d_g) es completo como espacio métrico.

Si (M, g) es completa entonces dado p y q existe una geodésica tal que $\gamma(0) = p$ y $\gamma(1) = q$ y $L(\gamma) = d_g(p, q)$.

Ejemplo 3.3.10. Sea M una superficie regular de $I\!\!R^3$ dotada con la métrica inducida por la métrica canónica de $I\!\!R^3$. Luego si \mathbf{X} y \mathbf{Y} son campos de M los podemos pensar como funciones diferenciables $\mathbf{X} : M \rightarrow I\!\!R^3$ y $\mathbf{Y} : M \rightarrow I\!\!R^3$. Sea $p \in M$, consideremos

$$\mathbf{Y}_{*_p}(\mathbf{X}(p)) \in I\!\!R^3$$

donde $\mathbf{Y}_{*_p} : T_p M \rightarrow I\!\!R^3$ es el diferencial de la función \mathbf{Y} aplicado al vector $\mathbf{X}(p)$. Ahora $\mathbf{Y}_{*_p}(\mathbf{X}(p))$ no tiene porque pertenecer al espacio tangente $T_p M$. Si denotamos con T la proyección ortogonal de $I\!\!R^3$ a $T_p M$ y con \perp la proyección al espacio normal, tenemos que $\mathbf{Y}_{*_p}(\mathbf{X}(p)) = (\mathbf{Y}_{*_p}(\mathbf{X}(p))^T + (\mathbf{Y}_{*_p}(\mathbf{X}(p)))^\perp)$. Si definimos

$$\nabla_{\mathbf{X}} \mathbf{Y}|_p = (\mathbf{Y}_{*_p}(\mathbf{X}(p))^T,$$

∇ es la conexión de Levi-Civita de la métrica inducida por la métrica canónica de $I\!\!R^3$. Si $c : I \rightarrow M$ es una curva y $\mathbf{X} \in \chi_c$, luego

$$\frac{D\mathbf{X}}{dt}|_t = (\mathbf{X}'(t))^T$$

es la derivada covariante de \mathbf{X} a lo largo de c , donde \mathbf{X}' es la derivada usual, coordenada a coordenada del campo \mathbf{X} , i.e. $\mathbf{X}'(t) = (\mathbf{X}_1'(t), \mathbf{X}_2'(t), \mathbf{X}_3'(t))$. En particular la derivada covariante del campo velocidad de una curva, es decir su aceleración, es la proyección ortogonal al tangente de su aceleración euclídea

$$\frac{Dc'}{dt}|_t = (c''(t))^T.$$

Una geodésica en una superficie es una curva que tiene toda su aceleración euclídea normal a la superficie.

Ejemplo 3.3.11. Del ejemplo anterior no es difícil ver que las geodesicas de S^n con respecto a la conexión de Levi-Civita de métrica inducida están contenidas en los círculos mayores, i.e. intersecciones de la esfera con planos que pasan por su centro.

En lo que sigue, a menos que se diga lo contrario, cuando consideramos una conexión en una variedad Riemanniana nos estaremos refiriendo a la conexión de Levi-Civita.

Proposición 3.3.12. Sea (M, g) una variedad Riemanniana y γ una geodésica, entonces $\|\gamma'(t)\|$ es constante.

Demostración.

$$\frac{d}{dt}(\|\gamma'(t)\|^2) = \frac{d}{dt}g(\gamma'(t), \gamma'(t)) = g\left(\frac{D\gamma'}{dt}, \gamma'\right) + g\left(\gamma', \frac{D\gamma'}{dt}\right) = 0.$$

□

Sea $\{v_1, \dots, v_d\}$ una base ortonormal de $T_p M$, es decir $g_p(v_i, v_j) = 1$ si $i = j$ y $g_p(v_i, v_j) = 0$ si $i \neq j$. Sea $v \in T_p M$ entonces $v = \sum_{i=1}^d x_i v_i$. Definimos la aplicación $\Psi : T_p M \rightarrow \mathbb{R}^d$ dada por

$$\Psi(v) = \Psi\left(\sum_{i=1}^d x_i v_i\right) := (x_1, \dots, x_d),$$

$(T_p M, \Psi)$ es una carta global de $T_p M$. Sea $B_s(0_p)$ una bola centrada en 0_p de radio s tal que la $\exp_p : B_s(0_p) \rightarrow \exp_p(B_s(0_p)) = B_s(p)$ resulte un difeomorfismo. Por el Teorema 3.2.19 sabemos que estos entornos existen. Consideramos la carta de M dada $\varphi : B_s(p) \rightarrow \mathbb{R}^d$ donde

$$\varphi = \Psi \circ \exp_p^{-1}$$

Este tipo de cartas las llamaremos cartas normales centradas en p inducidas por la base ortonormal $B = \{v_1, \dots, v_d\}$. Nos referiremos al abierto de la variedad $B_s(p)$ como la bola normal centrada en p de radio s .

El siguiente teorema nos muestra la propiedad de minimizar distancias que localmente poseen las geodésicas.

Teorema 3.3.13. *Sea $p \in M$ y U_p una bola normal centrada en p . Dado $q \in U_p$ existe una única geodésica $\alpha : [0, 1] \longrightarrow U_p$ tal que $\alpha(0) = p$, $\alpha(1) = q$ que satisface $d_g(p, q) = L(\alpha)$.*

Demostración. Ver Do Carmo [5]. □

Observación 3.3.14. *La geodésica α mencionada en el teorema anterior es de la forma $= \alpha(t) = \exp_p(tv)$, donde v es el vector tangente tal que $\exp_p(v) = q$.*

El siguiente hecho será de utilidad en el Capítulo 4.

Proposición 3.3.15. *Sea (U_p, φ) una carta normal centrada en p inducida por la base ortonormal $B = \{v_1, \dots, v_d\}$ de $T_p(M)$. Si $q = \exp_p(v) \in U_p$ entonces $d_g(p, q) = \|u\|$ donde $v = \sum_{i=1}^d u_i v_i$ y $\|u\|$ es la norma euclídea de \mathbb{R}^d .*

Demostración. Sea U_0 entorno normal de 0_p tal que $\exp_p(U_0) = U_p$ y sea $\alpha(t) = \exp_p(tv)$ la geodésica radial que minimiza la distancia entre p y q (ver teorema anterior)

$$d_g(p, q) = L(\alpha) = \int_0^1 \|\alpha'\| dt.$$

Como α es una geodésica, $\|\alpha'(t)\|$ es constante. Por lo tanto, $\|\alpha'(t)\| = \|\alpha'(0)\| = \|v\|$ para todo t . Con lo cual $d_g(p, q) = \|v\|$. Por otro lado, como $B = \{v_1, \dots, v_d\}$ es una base ortonormal de $T_p M$ y $v = \sum_{i=1}^d u_i v_i$ tenemos que

$$\|v\|^2 = g_p\left(\sum_{i=1}^d u_i v_i, \sum_{j=1}^d u_j v_j\right) = \|u\|^2,$$

con lo que la proposición queda demostrada. □

Definición 3.3.16. Sea (M, g) una variedad Riemanniana de dimensión d . Llamamos radio de inyectividad a

$$\text{inj}_g M = \inf_{p \in M} \sup\{s \in \mathbb{R} > 0 : B_s(p) \text{ es una bola normal}\}.$$

Ejemplo 3.3.17. *Si $M = \mathbb{R}^d$ con la métrica canónica entonces $\text{inj}_g M = \infty$. Si $M = \mathbb{R}^d - \{p\}$ con la métrica usual, entonces el radio de inyectividad es cero. Si $M = S^1 \times \mathbb{R}$ con la métrica inducida de \mathbb{R}^3 el radio de inyectividad es π .*

Proposición 3.3.18. *Si (M, g) es una variedad compacta entonces el radio de inyectividad es positivo .*

Demostración. Sea $f : M \rightarrow I\mathbb{R}$ con $f(p) = \sup\{s \in I\mathbb{R} \geq 0 : B_s(p) \text{ es una bola normal}\}$. Recordemos que B es una bola normal centrada en p si existe una bola V en $T_p(M)$ centrada en 0_p tal que $\exp_p(V) = B$ y $\exp_p|_V : V \rightarrow B$ es un difeomorfismo. Sabemos que $f(p) > 0$ para todo $p \in M$. Ahora, como f es continua y M es compacto existe $q \in M$ tal que $f(p) \geq f(q)$ para todo $p \in M$. Por lo tanto, $\text{inj}_g M = \inf_{p \in M} f(p) = f(q) > 0$. \square

Definición 3.3.19. Consideremos una carta normal centrada en p (U, φ) de (M, g) inducida por la base ortonormal $\{v_1, \dots, v_d\}$ de $T_p M$, con $U = B_s(p)$ una bola normal centrada en p . Definimos la función de densidad de volumen sobre (M, g) como:

$$\theta_p(q) = \left| \det g_q \left(\frac{\partial}{\partial \varphi_i}|_q, \frac{\partial}{\partial \varphi_j}|_q \right) \right|^{\frac{1}{2}},$$

Observación 3.3.20. Esta definición es independiente de la elección de la carta, ver Henry y Rodriguez [10] y [11].

Observación 3.3.21. Notar que la función $\theta_p(q)$ no está definida para todo p y q en M , pero si lo está para aquellos puntos que verifican $d_g(p, q) < \text{inj}_g M$. Claramente $\theta_p(q) : B_s(p) \rightarrow I\mathbb{R} > 0$ es una función diferenciable. Además satisface que para todo p y q para los cuales esta definida, $d_g(p, q) < \text{inj}_g M$, se tiene que $\theta_p(q) = \theta_q(p)$.

Terminemos esta sección mencionando rápidamente como se integra una función soportada en el dominio de una carta. Para una exposición más detallada del tema de integración en variedades ver Boothby [3] y Keilhauer [14].

Sea (M, g) una variedad Riemanniana orientada (ver Boothby [3]) de dimensión d . Notamos con $d\nu_g$ al elemento de volumen inducido por la métrica g . Recordamos que la expresión del elemento de volumen en una carta (U, φ) esta dada por

$$d\nu_g|_q = \sqrt{\det[g_{ij}(q)]} d\varphi_1|_q \wedge \cdots \wedge d\varphi_d|_q.$$

Si f es una función continua $f : M \rightarrow I\mathbb{R}$ soportada en el dominio de una carta (U, φ) . La integral de f sobre U está dada por

$$\int_M f d\nu_g = \int_U f d\nu_g = \int_{\varphi(U)} f \circ \varphi^{-1}(\mathbf{u}) \left(\sqrt{\det[g_{ij}]} \circ \varphi^{-1}(\mathbf{u}) \right) du_1 \dots du_d.$$

3.4. Probabilidades y estadística en variedades Riemannianas

En esta Sección introduciremos algunas nociones básicas de probabilidad y estadística sobre una variedad Riemanniana. El primer concepto a definir, es el de variable aleatoria.

Definición 3.4.1. Sea (Ω, \mathcal{A}, P) un espacio de probabilidad y sea M una variedad Riemanniana con su σ -álgebra de Borel, $\mathcal{B}(M)$, es decir la generada por la clase de conjuntos abiertos de M . Sea $X : \Omega \rightarrow M$, diremos que X es un variable aleatoria si para todo conjunto $\chi \in \mathcal{B}(M)$ se tiene que $X^{-1}(\chi) \in \mathcal{A}$.

Como en el caso real, podemos trabajar directamente con la medida de probabilidad inducida en M . Otra definición que necesitaremos es el concepto de densidad.

Definición 3.4.2. La variable aleatoria X tiene función de densidad f si para todo $\chi \in \mathcal{B}(M)$, $P(X^{-1}(\chi)) = \int_{\chi} f d\nu_g$.

En lo que resta de esta Sección nos centraremos en definir alguna medida de posición o valor central y dispersión de la distribución como así tambien alguna propuesta de estimación de las mismas.

Un problema que surge para poder definir la noción de esperanza es que no tenemos una manera de generalizar la integral de funciones cuya imagen está en M . Por lo tanto la esperanza se definirá de una manera *indirecta* pero sumamente natural.

En primer lugar notemos que si $\varphi(p)$ una función boreleana definida sobre M a valores reales y X una variable aleatoria con valores en M con función de densidad f . Luego $\varphi \circ X$ es una variable aleatoria real sobre M y su esperanza está dada por

$$E [\varphi \circ X] = \int_M \varphi(q) f(q) d\nu_g.$$

Definición 3.4.3. Dada X una variable aleatoria con valores en M con función de densidad f . La varianza $\sigma_X^2(y)$ es la esperanza de la distancia al cuadrado entre la variable aleatoria y el valor fijo $y \in M$, es decir

$$\sigma_X^2(y) = E [d_g^2(X, y)] = \int_M [d_g(y, q)]^2 f(q) d\nu_g.$$

Se puede demostrar que si la función de densidad f tiene soporte compacto entonces $\sigma_X^2(y)$ existe y es finita para todo valor de $y \in M$.

Definición 3.4.4. (*Media de Fréchet*) Dada X una variable aleatoria con valores en M con función de densidad f . Si la varianza $\sigma_X^2(y)$ es finita para todo valor de $y \in M$, cada \bar{X} que minimice a $\sigma_X^2(y)$ es llamado la esperanza de X . Por lo tanto

$$\bar{X} = E [X] = \arg \min_{y \in M} (E [d_g^2(X, y)]).$$

Si este valor existe llamamos varianza al mínimo valor $\sigma_X^2 = \sigma_X^2(\bar{X})$.

Observemos que, como la definición de la esperanza es el resultado de una minimización, su existencia y unicidad no está garantizada pues el mínimo global podría no ser alcanzado y de cualquier manera este no tendría porque ser único. Para librarse de este problema Kacher [13], propuso considerar el conjunto de mínimos locales de σ_X^2 , este conjunto se lo conoce como *medias de Kacher*. Claramente el conjunto de medias de Fréchet esta incluído en el conjunto de medias de Kacher. Usando esta extensión, Kacher [13] estableció condiciones sobre la variedad y la distribución que garantizan la existencia y unicidad de la media. Más precisamente, obtuvó el siguiente resultado.

Teorema 3.4.5. *Sea X una variable aleatoria en M con función de densidad f . Si el soporte de f esta incluído en una bola geodésica regular $B_r(y)$, luego existe un único \bar{X} sobre dicha bola.*

Con las definiciones dadas anteriormente para las medidas de posición y dispersión es natural definir estimadores de ellas, de la siguiente manera. Dado un conjunto de datos x_1, \dots, x_n se define el promedio empírico como

$$\bar{X}_n = \arg \min_{y \in M} y \in M \left(\frac{1}{n} \sum_{i=1}^n d_g^2(y, X_i) \right).$$

Luego, si este valor existe , llamamos varianza empírica al mínimo valor

$$s_n^2 = \frac{1}{n} \sum_{i=1}^n d_g^2(y, \bar{X}_n).$$

Capítulo 4

Estimación no paramétrica de la densidad en variedades Riemannianas

4.1. Estimación de la densidad por núcleos en variedades Riemannianas

Sean X_1, \dots, X_n un conjunto de variables aleatorias independientes e idénticamente distribuidas sobre una variedad Riemanniana (M, g) de dimensión d orientada, sin borde y con función de densidad f . Con el objetivo de poder estimar la función f basandose en las observaciones X_1, \dots, X_n Pelletier [19] propuso el siguiente estimador basado en núcleos,

$$f_n(p) = \frac{1}{nh^d} \sum_{i=1}^n \frac{1}{\theta_p(X_i)} K\left(\frac{d_g(p, X_i)}{h}\right) \quad (4.1)$$

donde $K : I\mathbb{R} \rightarrow I\mathbb{R}$ es una función no negativa, $\theta_p(q)$ denota la función de volumen sobre (M, g) , d_g la distancia inducida por la métrica g , $h = h_n$, el ancho de banda o ventana, es una sucesión de números no negativos que verifica $\lim_{n \rightarrow \infty} h = 0$ y $h < \text{inj}_g M$, para todo n . Este último requerimiento sobre h garantiza que el estimador esté definido para todo $p \in M$. Notemos que el estimador definido por Pelletier es consistente con el estimador definido por Rosenblatt [21] y Parzen [18] en el caso Euclídeo. Ya que en el caso Euclídeo, $\theta_p(q) = 1$ para todo $p, q \in I\mathbb{R}$ y la distancia geodésica es simplemente $d_g(p, q) = |p - q|$.

Sea U un conjunto abierto de M , denotamos por $C^k(U)$ al conjunto de las funciones k veces diferenciables definidas en U a valores en $I\mathbb{R}$.

En el trabajo de Pelletier [19], se estudiaron algunas propiedades esenciales del estimador definido en (4.1). Por ejemplo, una propiedad deseable es que f_n , el estimador de f , sea efectivamente una función de densidad, así como obtener una expresión para su esperanza a los efectos de decidir si el estimador propuesto es insesgado o asintóticamente insesgado. Por otro parte, es razonable que el estimador sea consistente, es decir al aumentar el tamaño muestral, f_n esté cerca en algún sentido a la densidad f a estimar. En un trabajo reciente , Henry y Rodriguez [11] estudiaron la consistencia fuerte uniforme sobre compactos para el estimador propuesto por Pelletier. Las siguientes proposiciones resumen estas propiedades y dan una idea de su demostración, para ello asumiremos las siguientes hipótesis.

H1. Sea M_0 un conjunto compacto de M tal que:

- i) f es una función acotada tal que $\inf_{p \in M_0} f(p) = A > 0$.
- ii) $\inf_{p,q \in M_0} \theta_p(q) = B > 0$.

H2. Dado cualquier abierto U_0 de M_0 tal que $M_0 \subset U_0$, f es de clase C^2 sobre U_0 .

H3. La sucesión h verifica que $h \rightarrow 0$ y $\frac{nh_n^d}{\log n} \rightarrow \infty$ cuando $n \rightarrow \infty$.

H4. $K : I\!\!R \rightarrow I\!\!R$ es una función de Lipschitz de orden uno, acotada no negativa con soporte compacto en el $[0, 1]$ satisfaciendo: $\int_{\mathbb{R}^d} K(\|\mathbf{u}\|)d\mathbf{u} = 1$, $\int_{\mathbb{R}^d} \mathbf{u}K(\|\mathbf{u}\|)d\mathbf{u} = \mathbf{0}$ y $0 < \int_{\mathbb{R}^d} \|\mathbf{u}\|^2 K(\|\mathbf{u}\|)d\mathbf{u} < \infty$.

Proposición 4.1.1. *Bajo H4, tenemos que el estimador definido en (4.1) es una función de densidad.*

Demostración. Por la elección del núcleo solo bastará demostrar que $\int_M f_n(p)d\nu_g = 1$.

$$\int_M f_n(p)d\nu_g = \frac{1}{nh^d} \sum_{i=1}^n \int_M \frac{1}{\theta_p(X_i)} K\left(\frac{d_g(p, X_i)}{h}\right) d\nu_g$$

Como $sop K$ esta incluído en el intervalo $[0, 1]$, para $0 < h \leq inj_g M$ suficiente pequeño se tiene que

$$\int_M f_n(p)d\nu_g = \frac{1}{nh^d} \int_{U_i} \sum_{i=1}^n \frac{1}{\theta_p(X_i)} K\left(\frac{d_g(p, X_i)}{h}\right) d\nu_g = \frac{1}{nh^d} \int_{U_i} \sum_{i=1}^n \frac{1}{\theta_{X_i}(p)} K\left(\frac{d_g(p, X_i)}{h}\right) d\nu_g$$

donde U_i es una bola normal de radio h centrada en X_i . Ahora, tomando una carta

normal (U_i, φ^i) , para cada $1 \leq i \leq n$, tenemos que

$$\begin{aligned}
& \int_{U_i} \frac{1}{h^d \theta_{X_i}(p)} K\left(\frac{d_g(p, X_i)}{h}\right) d\nu_g \\
&= \int_{U_i} \frac{1}{h^d \theta_{\varphi^{-1}(0)}(\varphi^{-1}(u))} K\left(\frac{d_g(\varphi^{-1}(u), \varphi^{-1}(0))}{h}\right) \left| \det g_{\varphi^{-1}(u)} \left(\frac{\partial}{\partial \varphi_j} \Big|_{\varphi^{-1}(u)}, \frac{\partial}{\partial \varphi_k} \Big|_{\varphi^{-1}(u)} \right) \right|^{\frac{1}{2}} du_1 \dots du_d \\
&= \int_{U_i} \frac{1}{h^d \theta_{\varphi^{-1}(0)}(\varphi^{-1}(u))} K\left(\frac{d_g(\varphi^{-1}(u), \varphi^{-1}(0))}{h}\right) \theta_{\varphi^{-1}(0)}(\varphi^{-1}(u)) du_1 \dots du_d \\
&= \int_{\varphi^i(U_i)} \frac{1}{h^d} K\left(\frac{d_g(\varphi_i^{-1}(0), \varphi_i^{-1}(u))}{h}\right) du_1 \dots du_d \\
&= \int_{\varphi_i(U_i)} \frac{1}{h^d} K\left(\frac{\|\mathbf{u}\|}{h}\right) d\mathbf{u} \\
&= \int_{\frac{1}{h}\varphi_i(U_i)} K(\|\mathbf{y}\|) d\mathbf{y} = 1
\end{aligned}$$

pues $B_1(0) \subseteq \frac{1}{h}\varphi_i(U_i)$. □

Proposición 4.1.2. *Bajo H2 a H4, el estimador definido en (4.1) es asintóticamente insesgado, más precisamente*

$$E(f_n(p)) = f(p) + O(h^2).$$

Demostración. Notemos que si $d_g(X_i, p) > h$ tenemos que $f_n(p) = 0$. Sea $B_h(p)$ una bola normal de radio h centrada en p . Entonces,

$$\begin{aligned}
E(f_n(p)) &= \frac{1}{nh^d} \sum_{i=1}^n E\left(\frac{1}{\theta_p(X_i)} K\left(\frac{d_g(p, X_i)}{h}\right)\right) \\
&= \frac{1}{nh^d} \sum_{i=1}^n \int_{B_h(p)} \frac{1}{\theta_p(q)} K\left(\frac{d_g(p, q)}{h}\right) f(q) d\nu_g \\
&= \int_{B_h(p)} \frac{1}{h^d} \frac{1}{\theta_p(q)} K\left(\frac{d_g(p, q)}{h}\right) f(q) d\nu_g \\
&= \frac{1}{h^d} \int_{\varphi(B_h(p))} K\left(\frac{d_g(\varphi^{-1}(u), \varphi^{-1}(0))}{h}\right) f \circ \varphi^{-1}(u) du_1 \dots du_d \\
&= \frac{1}{h^d} \int_{\varphi(B_h(p))} K\left(\frac{\|u\|}{h}\right) f \circ \varphi^{-1}(u) du_1 \dots du_d \\
&= \frac{1}{h^d} \int_{\varphi(B_h(p))} K\left(\frac{\|u\|}{h}\right) m(u) du_1 \dots du_d
\end{aligned}$$

donde $m(u) = f \circ \varphi^{-1}(u)$, haciendo un desarrollo de Taylor de orden 2 tenemos,

$$m(u) = m(0) + \sum_{i=1}^d \frac{\partial m}{\partial u_i} \Big|_{u=0} u_i + \sum_{i=1}^d \sum_{j=1}^d \frac{\partial m}{\partial u_i \partial u_j} \Big|_{u=\epsilon} u_i u_j.$$

Observemos que $m(0) = f(\varphi^{-1}(0)) = f(p)$.

Luego, como $\int_{\mathbf{R}^d} K(\|\mathbf{u}\|) d\mathbf{u} = 1$, $B_1(0) \subseteq \frac{1}{h} \varphi(B_h(p))$ y $sopK \subseteq [0, 1]$ tenemos

$$\frac{1}{h^d} \int_{\varphi(B_h(p))} K\left(\frac{\|\mathbf{u}\|}{h}\right) m(0) du_1 \dots du_d = f(p) \int_{\frac{1}{h} \varphi(B_h(p))} K(\|\mathbf{u}\|) du_1 \dots du_d = f(p).$$

Usando que $\int_{\mathbf{R}^d} \mathbf{u} K(\|\mathbf{u}\|) d\mathbf{u} = \mathbf{0}$ tenemos que

$$\begin{aligned} \frac{1}{h^d} \int_{\varphi(B_h(p))} K\left(\frac{\|\mathbf{u}\|}{h}\right) \sum_{i=1}^n \frac{\partial m}{\partial u_i} \Big|_{\mathbf{u}=0} u_i du_1 \dots du_d = \\ \sum_{i=1}^n \frac{\partial m}{\partial u_i} \Big|_{\mathbf{u}=0} h \int_{\frac{1}{h} \varphi(B_h(p))} K(\|\mathbf{u}\|) u_i du_1 \dots du_d = 0. \end{aligned}$$

Y, por otro lado por $H4$ obtenemos que

$$\begin{aligned} \sum_{i=1}^d \sum_{j=1}^d \frac{\partial m}{\partial u_i \partial u_j} \Big|_{\mathbf{u}=\epsilon} \frac{1}{h^d} \int_{\varphi(B_h(p))} K\left(\frac{\|\mathbf{u}\|}{h}\right) u_i u_j du_1 \dots du_d = \\ \sum_{i=1}^d \frac{\partial^2 m}{\partial u_i^2} \Big|_{\mathbf{u}=\epsilon} \frac{1}{h^d} \int_{\varphi(B_h(p))} K\left(\frac{\|\mathbf{u}\|}{h}\right) u_i u_i du_1 \dots du_d \leq \\ C h^2 \int_{\frac{1}{h} \varphi(B_h(p))} K(\|\mathbf{u}\|) \|\mathbf{u}\|^2 du_1 \dots du_d < \bar{C} h^2 \end{aligned}$$

donde $\bar{C} < \infty$. Por lo tanto concluimos que $E(f_n(p)) = f(p) + O(h^2)$ \square

Teorema 4.1.3. *Asumiendo que valen H1 a H4, tenemos*

$$\sup_{p \in M_0} |f_n(p) - f(p)| \xrightarrow{c.t.p} 0.$$

Demostración. Por la proposición anterior sabemos que

$$\sup_{p \in M_0} |E(f_n(p)) - f(p)| \leq c h^2 \int \|\mathbf{u}\|^2 K(\|\mathbf{u}\|) d\mathbf{u}.$$

Como $|f_n(p) - f(p)| \leq |f_n(p) - E(f_n(p))| + |E(f_n(p)) - f(p)|$, entonces para obtener la consistencia fuerte es suficiente mostrar que $\sup_{p \in M_0} |f_n(p) - E(f_n(p))| \xrightarrow{c.t.p.} 0$. Dado $p \in M$, denotaremos

$$V_i(p) = \frac{1}{\theta_p(X_i)} K \left(\frac{d_g(p, X_i)}{h} \right) - E \left(\frac{1}{\theta_p(X_i)} K \left(\frac{d_g(p, X_i)}{h} \right) \right),$$

Sea $S_n(p) = \sum_{i=1}^n V_i(p) = nh^d[f_n(p) - E(f_n(p))]$. Notemos que las variables $V_i = V_i(p)$ satisfacen,

- 1) $V_i(p)$ son independientes.
- 2) $E(V_i) = 0$ para todo i .
- 3) de $H1$ y $H4$ tenemos que $|V_i(p)| < A$ para todo i .
- 4) $E(V_i^2) \leq \int_M \frac{1}{\theta_p(q)^2} K^2 \left(\frac{d_g(p, q)}{h} \right) f_X(q) d\nu_g \leq B^{-1} \|K\|_\infty h^d E(f_n(p)) \leq Ch^d$

la última desigualdad de 4) se desprende del hecho que $E(f_n(p)) \rightarrow f(p)$. Luego, la desigualdad de Bernstein implica que, dado $\epsilon > 0$ existe n_0 tal que $\forall n > n_0$ tenemos

$$P \left(\frac{1}{nh^d} |S_n(p)| > \epsilon \right) \leq 2 \exp \left(\frac{-nh^{2d}\epsilon^2}{2Ch^d + 2A\varepsilon h^d} \right) = 2 \exp(-nh^d\alpha),$$

con $\alpha = \frac{\epsilon^2}{2C+2A\varepsilon} > 0$, luego como la cota no depende de p tenemos que

$$\sup_{p \in M_0} P \left(\frac{1}{nh^d} |S_n(p)| > \epsilon \right) \leq 2 \exp(-nh^d\alpha). \quad (4.2)$$

Por otro lado, como M_0 es un conjunto compacto, podemos considerar un conjunto finito de bolas ($B_i = B_{h^\gamma}(p_i)$) con centros en $p_i \in M_0$ de radio h^γ con $\gamma > 2 + d$, que verifiquen $M_0 \subset \bigcup_{i=1}^l B_i$ con $l = O(h^{-\gamma d})$. Luego

$$|S_n(p)| \leq |S_n(p) - S_n(p_j)| + |S_n(p_j)| \quad \forall p \in M_0 \text{ y } \forall 1 \leq j \leq l \text{ entonces}$$

$$\sup_{p \in M_0} |S_n(p)| \leq \max_{1 \leq j \leq l} \sup_{p \in B_j} |S_n(p) - S_n(p_j)| + \max_{1 \leq j \leq l} |S_n(p_j)|. \quad (4.3)$$

Acotemos en cada B_j

$$\begin{aligned}
|S_n(p) - S_n(p_j)| &\leq \sum_{i=1}^n \frac{1}{\theta_p(X_i)} \left| K\left(\frac{d_g(p, X_i)}{h}\right) - K\left(\frac{d_g(p_j, X_i)}{h}\right) \right| \\
&\quad + K\left(\frac{d_g(p, X_i)}{h}\right) \left| \frac{1}{\theta_p(X_i)} - \frac{1}{\theta_{p_j}(X_i)} \right| \\
&\quad + E\left[\frac{1}{\theta_p(X_i)} \left| K\left(\frac{d_g(p, X_i)}{h}\right) - K\left(\frac{d_g(p_j, X_i)}{h}\right) \right| \right] \\
&\quad + E\left[K\left(\frac{d_g(p, X_i)}{h}\right) \left| \frac{1}{\theta_p(X_i)} - \frac{1}{\theta_{p_j}(X_i)} \right| \right]
\end{aligned}$$

Usando que K es una función Lipschitz con constante Lipschitz $\|K\|_L$, la cota para $\theta_p(q)$ y $p \in B_j$ tenemos que:

$$\left| \frac{1}{\theta_p(X_i)} \left| K\left(\frac{d_g(p, X_i)}{h}\right) - K\left(\frac{d_g(p_j, X_i)}{h}\right) \right| \right| \leq B^{-1} \|K\|_L \frac{1}{h} d_g(p, p_j) \leq B^{-1} \|K\|_L h^{\gamma-1}$$

Usando que K es una función acotada y la diferenciabilidad de $\theta_p(q)$ tenemos que:

$$\begin{aligned}
\left| K\left(\frac{d_g(p_j, X_i)}{h}\right) \right| \left| \frac{1}{\theta_p(X_i)} - \frac{1}{\theta_{p_j}(X_i)} \right| &\leq \|K\|_\infty \left| \frac{1}{\theta_{X_i}(p)} - \frac{1}{\theta_{X_i}(p_j)} \right| \\
&\leq \frac{\|K\|_\infty}{B^2} \mathbf{C} d_g(p, p_j) \leq \frac{\|K\|_\infty}{B^2} \mathbf{C} h^\gamma
\end{aligned}$$

Entonces,

$$\frac{1}{nh^d} |S_n(p) - S_n(p_j)| < 2h^{\gamma-(d+1)} (B^{-1} \|K\|_L + \|K\|_\infty \mathbf{C} B^{-2} h) \leq Ch^{\gamma-(d+1)}$$

para todo $p \in B_j$.

Como $\gamma - (d + 1) > 1$ y $h \rightarrow 0$ podemos afirmar que existe n_1 tal que para todo $n > n_1$ se verifica

$$\max_{1 \leq j \leq l} \sup_{p \in B_j} \frac{1}{nh^d} |S_n(p) - S_n(p_j)| < \varepsilon \quad (4.4)$$

Luego si $n > n_1$ tenemos que

$$\{\omega : \sup_{p \in M_0} \frac{1}{nh^d} |S_n(p)| > 2\varepsilon\} \subseteq \{\omega : \max_{1 \leq j \leq l} \frac{1}{nh^d} |S_n(p_j)| > \varepsilon\}$$

Finalmente, (4.2), (4.3) y (4.4) implica que, para $n > \max\{n_0, n_1\}$

$$\begin{aligned}
 P\left(\sup_{p \in M_0} \frac{1}{nh^d} |S_n(p)| > 2\varepsilon\right) &\leq P\left(\max_{1 \leq j \leq l} \frac{1}{nh^d} |S_n(p_j)| > \varepsilon\right) \\
 &\leq P\left(\bigcup_{1 \leq j \leq l} \frac{1}{nh^d} |S_n(p_j)| > \varepsilon\right) \\
 &\leq \sum_{i=1}^l P\left(\frac{1}{nh^d} |S_n(p_j)| > \varepsilon\right) \\
 &\leq l \sup_{p \in M_0} P\left(\frac{1}{nh^d} |S_n(p)| > \varepsilon\right) \\
 &\leq 2l \exp(-nh^d \alpha) = 2l n^{-\delta_n \alpha} \\
 &\leq C(nh^d)^{-\gamma} n^{-\delta_n \alpha + \gamma}
 \end{aligned}$$

donde $\delta_n = \frac{nh^d}{\log n}$. Como $\delta_n \rightarrow \infty$ y $nh^d \rightarrow \infty$, se tiene que $(nh^d)^{-\gamma} n^{-\delta_n \alpha + \gamma} \leq n^{-2}$ para todo $n > n_2$ y por lo tanto $P\left(\sup_{p \in M_0} \frac{1}{nh^d} |S_n(p)| > 2\varepsilon\right) \leq cn^{-2}$ si $n > \max\{n_0, n_1, n_2\}$. Lo cual muestra que $\sum_{n=1}^{\infty} P\left(\sup_{p \in M_0} \frac{1}{nh^d} |S_n(p)| > 2\varepsilon\right) < \infty$ y por el Lema de Borel-Cantelli se concluye la prueba. \square

Observación 4.1.4. Notar que para que el Teorema anterior sea cierto, se necesitaría que $\text{inj}_g M > h + h^\gamma$, lo cual se satisface cuando h^γ es pequeño.

4.2. Estimación de la densidad por vecinos más cercanos con núcleo en variedades Riemannianas

Al igual que en la Sección anterior, sean X_1, \dots, X_n v.a. i.i.d. sobre M con función de densidad f sobre una variedad Riemanniana (M, g) de dimensión d . Como discutimos en la introducción al igual que en el caso de variables reales, el parámetro de suavizado en el estimador definido por Pelletier tiene un fuerte efecto en la estimación. El hecho de que la ventana sea fija y no se adapte al punto donde estamos estimando puede conducir a entornos vacíos. Por ejemplo, en aquellas regiones de la variedad donde las variables tiene poca probabilidad de ocurrir posiblemente necesitaremos un tamaño de ventana más grande dando lugar a entornos que recoja una proporción de la muestra considerable para obtener estimaciones mas razonables. En cambio, en

aquellas regiones donde la probabilidad de ocurrencia es mayor posiblemente con un entorno más pequeño bastará para dar estimaciones confiables.

Al igual que cuando consideramos variables reales una solución para reducir este problema es considerar ventanas que se adapten al punto de la variedad en donde estamos estimando. De esta forma, con el objetivo de estimar f basandose en $X_1 \dots X_n$ combinando las ideas propuestas en el caso Euclídeo junto con las introducidas por Pelletier, se propone el siguiente estimador de vecinos más cercanos con núcleos,

$$\hat{f}_n(p) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\theta_p(X_i) H_n^d(p)} K \left(\frac{d_g(p, X_i)}{H_n(p)} \right) \quad (4.5)$$

donde $K : I\!\!R \rightarrow I\!\!R$ es una función no negativa, $\theta_p(q)$ denota la función de densidad de volumen sobre (M, g) y $H_n(p)$ es la distancia d_g entre p y el k -ésimo vecino más cercano a p con $k = k_n$ una sucesión que verifica $k_n \rightarrow \infty$.

Es decir, para estimar la función de densidad en el punto p consideraremos aquella ventana $H_n(p)$ que nos garantice que en la bola geodésica, $\{q : d_g(p, q) \leq H_n(p)\}$ haya al menos k observaciones entre las X_1, \dots, X_n . De esta manera, no corremos el riesgo de tener entornos vacíos al estimar la densidad en cada punto de la variedad.

Es importante notar que el estimador definido en (4.5) al igual que el definido en (4.1) depende de la distancia geodésica y la función de densidad de volumen, entonces para poder implementar el estimador en una determinada variedad es necesario tener expresiones explícitas de ellas. En la Sección 4.1, se observó que en el caso de variables reales $\theta_p(q) = 1$ para todo $p, q \in I\!\!R^d$ y $d_g(p, q) = \|p - q\|$. En Pelletier [19] y Henry y Rodriguez [10] se calculan en forma explícita los casos del cilindro y la esfera.

Sea (\mathcal{C}_1, g) el cilindro de radio 1 con la métrica inducida por la métrica canónica de $I\!\!R^3$ en este caso, en [10], se muestra que $\theta_p(q) = 1$ para todo $p, q \in \mathcal{C}_1$, $d_g(p, q) = d_g(\cos(r_1), \operatorname{sen}(r_1), s_1), (\cos(r_2), \operatorname{sen}(r_2), s_2)) = \|(r_1, s_1) - (r_2, s_2)\|$ si $\|(r_1, s_1) - (r_2, s_2)\| < \pi$.

El caso de la esfera de radio r fue estudiado en [19] y [10], en este caso $\theta_p(q) = rp\angle q$, donde $p\angle q$ denota el ángulo entre p y q . Y la función de densidad de volumen entre p y q esta dada por

$$\theta_p(q) = \frac{r|\operatorname{sen}(d_g(p, q)/r)|}{d_g(p, q)} \quad \text{para } q \neq p, -p.$$

Para finalizar, mostraremos el resultado principal de esta tesis. Este resultado es un análogo al Teorema 4.1.3 pero para el estimador propuesto con ventana aleatoria. Para ello será necesario el siguiente conjunto de hipótesis.

H1. Sea M_0 un conjunto compacto de M tal que:

- i) f es una función acotada tal que $\inf_{p \in M_0} f(p) = A > 0$.
- ii) $\inf_{p,q \in M_0} \theta_p(q) = B > 0$.

H2. Dado cualquier abierto U_0 de M_0 tal que $M_0 \subset U_0$, f es de clase C^2 sobre U_0 .

H3. La sucesión k_n es tal que $\lim_{n \rightarrow \infty} k_n = \infty$, $\lim_{n \rightarrow \infty} \frac{k_n}{n} = 0$ y $\lim_{n \rightarrow \infty} \frac{k_n}{\log n} = \infty$

H4. $K : I\!\!R \rightarrow I\!\!R$ es acotada no negativa, Lipschitz de orden uno, con soporte compacto en el $[0, 1]$ satisfaciendo, $\int_{I\!\!R^d} K(\|\mathbf{u}\|) d\mathbf{u} = 1$, $\int_{I\!\!R^d} \mathbf{u} K(\|\mathbf{u}\|) d\mathbf{u} = \mathbf{0}$ y $0 < \int_{I\!\!R^d} \|\mathbf{u}\|^2 K(\|\mathbf{u}\|) d\mathbf{u} < \infty$.

H5. El núcleo $K(u)$ verifica $K(uz) \geq K(u) \forall z \in (0, 1)$

Teorema 4.2.1. *Bajo H1 hasta H5 , tenemos el siguiente resultado*

$$\sup_{p \in M_0} |\widehat{f}_n(p) - f(p)| \xrightarrow{c.t.p.} 0.$$

Demostración. Notemos por $f_n(p, h)$ el estimador definido en (4.1) con ventana h . Para cada $\beta \in (0, 1)$ definimos

$$f_n^-(p, \beta) = \frac{1}{nD_n^+(\beta)^d} \sum_{i=1}^n \frac{1}{\theta_p(X_i)} K\left(\frac{d_g(p, X_i)}{D_n^-(\beta)}\right) = \frac{D_n^-(\beta)^d}{D_n^+(\beta)^d} f_n(p, D_n^-(\beta)) = \beta f_n(p, D_n^-),$$

$$f_n^+(p, \beta) = \frac{1}{nD_n^-(\beta)^d} \sum_{i=1}^n \frac{1}{\theta_p(X_i)} K\left(\frac{d_g(p, X_i)}{D_n^+(\beta)}\right) = \frac{D_n^+(\beta)^d}{D_n^-(\beta)^d} f_n(p, D_n^+(\beta)) = \beta^{-1} f_n(p, D_n^+).$$

Donde $D_n^-(\beta) = \beta^{\frac{1}{2d}} h_n$, $D_n^+(\beta) = \beta^{-\frac{1}{2d}} h_n$ y $h_n^d = \frac{k_n}{nf(p)\lambda(V_1)}$ con $\lambda(V_1)$ la medida de Lebesgue de la bola en $I\!\!R^d$ de radio 1 centrado en el origen. Notemos que $D_n^-(\beta) < D_n^+(\beta)$, y que $D_n^-(\beta)$ y $D_n^+(\beta)$ satisfacen la hipótesis *H3* de la Sección anterior luego por el Teorema 4.1.3 tenemos

$$\sup_{p \in M_0} |f_n^-(p, \beta) - \beta f(p)| \xrightarrow{c.t.p.} 0, \quad (4.6)$$

$$\sup_{p \in M_0} |f_n^+(p, \beta) - \beta^{-1} f(p)| \xrightarrow{c.t.p.} 0. \quad (4.7)$$

Para cada $\beta \in (0, 1)$ y para cada ϵ mayor a cero definamos los siguientes conjuntos

$$\begin{aligned}\mathbf{S}_n^-(\beta, \epsilon) &= \{\omega : \sup_{p \in M_0} |f_n^-(p, \beta) - f(p)| \leq \epsilon\}, \\ \mathbf{S}_n^+(\beta, \epsilon) &= \{\omega : \sup_{p \in M_0} |f_n^+(p, \beta) - f(p)| \leq \epsilon\}, \\ \mathbf{S}_n(\epsilon) &= \{\omega : \sup_{p \in M_0} |\hat{f}_n(p) - f(p)| \leq \epsilon\}, \\ \mathbf{A}_n(\beta) &= \{f_n^-(p, \beta) \leq \hat{f}_n(p) \leq f_n^+(p, \beta)\}.\end{aligned}$$

Veamos que

$$\mathbf{A}_n(\beta) \cap \mathbf{S}_n^-(\beta, \epsilon) \cap \mathbf{S}_n^+(\beta, \epsilon) \subseteq \mathbf{S}_n(\epsilon). \quad (4.8)$$

Para ello notemos que,

$$\text{Si } \omega \in \mathbf{S}_n^-(\beta, \epsilon) \text{ entonces } -\epsilon < f_n^-(p, \beta) - f(p) < \epsilon \quad \forall p \in M_0, \quad (4.9)$$

$$\text{Si } \omega \in \mathbf{S}_n^+(\beta, \epsilon) \text{ entonces } -\epsilon < f_n^+(p, \beta) - f(p) < \epsilon \quad \forall p \in M_0, \quad (4.10)$$

luego si $\omega \in \mathbf{A}_n(\beta)$, usando (4.9) y (4.10) tenemos que

$$-\epsilon < f_n^-(p, \beta) - f(p) < \hat{f}_n(p) - f(p) < f_n^+(p, \beta) - f(p) < \epsilon \quad \forall p \in M_0.$$

Entonces $|\hat{f}_n(p) - f(p)| < \epsilon \quad \forall p \in M_0$ y por lo tanto $\sup_{M_0} |\hat{f}_n(p) - f(p)| < \epsilon$ demostrando así (4.8).

Sean $\mathbf{B} = \sup_{p \in M_0} f(p)$, tomemos $\beta_\epsilon = 1 - \frac{\epsilon}{3\mathbf{B}}$ con $0 < \epsilon \leq \frac{3\mathbf{B}}{2}$ y consideremos los siguientes conjuntos

$$\begin{aligned}\mathbf{G}_n(\epsilon) &= \{\omega : D_n^-(\beta_\epsilon) \leq H_n(p) \leq D_n^+(\beta_\epsilon) \forall p \in M_0\}, \\ \mathbf{G}_n^-(\epsilon) &= \{\omega : \sup_{p \in M_0} |f_n^-(p, \beta_\epsilon) - \beta_\epsilon f(p)| \leq \frac{\epsilon}{3}\}, \\ \mathbf{G}_n^+(\epsilon) &= \{\omega : \sup_{p \in M_0} |f_n^+(p, \beta_\epsilon) - \beta_\epsilon^{-1} f(p)| \leq \frac{\epsilon}{3}\}.\end{aligned}$$

Veamos que

$$\mathbf{G}_n(\epsilon) \subset \mathbf{A}_n(\beta_\epsilon), \quad \mathbf{G}_n^-(\epsilon) \subseteq \mathbf{S}_n^-(\beta_\epsilon, \epsilon) \quad \text{y} \quad \mathbf{G}_n^+(\epsilon) \subseteq \mathbf{S}_n^+(\beta_\epsilon, \epsilon). \quad (4.11)$$

Para la primera inclusión,

$$\begin{aligned}f_n^+(p, \beta_\epsilon) &= \frac{1}{nD_n^-(\beta_\epsilon)^d} \sum_{i=1}^n \frac{1}{\theta_p(X_i)} K\left(\frac{d_g(p, X_i)}{D_n^+(\beta_\epsilon)}\right) \\ &= \left(\frac{H_n(p)}{D_n^-(\beta_\epsilon)}\right)^d \frac{1}{nH_n(p)^d} \sum_{i=1}^n \frac{1}{\theta_p(X_i)} K\left(\frac{H_n(p)d_g(p, X_i)}{D_n^+(\beta_\epsilon)H_n(p)}\right)\end{aligned}$$

luego, si $\omega \in \mathbf{G}_n(\epsilon)$, $\frac{H_n(p)}{D_n^+(\beta_\epsilon)} \leq 1$ y $\frac{H_n(p)}{D_n^-(\beta_\epsilon)} > 1$ entonces por $H5$ tenemos que,

$$K \left(\frac{H_n(p)d_g(p, X_i)}{D_n^+(\beta_\epsilon)H_n(p)} \right) \geq K \left(\frac{d_g(p, X_i)}{H_n(p)} \right)$$

y por lo tanto $f_n^+(p, \beta_\epsilon) \geq \hat{f}_n(p)$. Análogamente se puede ver que si $\omega \in \mathbf{G}_n(\epsilon)$, $f_n^-(p, \beta_\epsilon) \leq \hat{f}_n(p)$. Así, concluimos que $\mathbf{G}_n(\epsilon) \subset \mathbf{A}_n(\beta_\epsilon)$.

La segunda inclusión se sigue del hecho que

$$\sup_{p \in M_0} |f_n^-(p, \beta_\epsilon) - f(p)| \leq \sup_{p \in M_0} |f_n^-(p, \beta_\epsilon) - \beta_\epsilon f(p)| + \sup_{p \in M_0} |f(p)||\beta_\epsilon - 1| \leq \frac{\epsilon}{3} + \mathbf{B} \frac{\epsilon}{3\mathbf{B}} \leq \epsilon.$$

Y análogamente para la tercera inclusión,

$$\begin{aligned} \sup_{p \in M_0} |f_n^+(p, \beta_\epsilon) - f(p)| &\leq \sup_{p \in M_0} |f_n^+(p, \beta_\epsilon) - \beta_\epsilon^{-1} f(p)| + \sup_{p \in M_0} |f(p)||\beta_\epsilon^{-1} - 1| \\ &\leq \frac{\epsilon}{3} + \mathbf{B} \frac{\epsilon}{3\mathbf{B} - \varepsilon} = \frac{\epsilon}{3} + \frac{\epsilon}{3} \left(1 + \frac{\epsilon}{3\mathbf{B} - \epsilon} \right) \leq \frac{\epsilon}{3} + \frac{\epsilon}{3} 2 = \epsilon \end{aligned}$$

Por lo tanto, de (4.8) y (4.11) podemos concluir que $\mathbf{G}_n(\epsilon) \cap \mathbf{G}_n^-(\epsilon) \cap \mathbf{G}_n^+(\epsilon) \subseteq \mathbf{S}_n(\epsilon)$ y tomando complemento tenemos que $\mathbf{S}_n^c(\epsilon) \subseteq \mathbf{G}_n^c(\epsilon) \cup \mathbf{G}_n^-(\epsilon)^c \cup \mathbf{G}_n^+(\epsilon)^c$.

Luego dado $k \in N$ se verifica que $\bigcup_{k \leq n} \mathbf{S}_n^c(\epsilon) \subseteq \bigcup_{k \leq n} \mathbf{G}_n^c(\epsilon) \cup \bigcup_{k \leq n} \mathbf{G}_n^-(\epsilon)^c \cup \bigcup_{k \leq n} \mathbf{G}_n^+(\epsilon)^c$,

con lo cual,

$$\mathbf{P}\left(\bigcup_{k \leq n} \mathbf{S}_n^c(\epsilon)\right) \leq \mathbf{P}\left(\bigcup_{k \leq n} \mathbf{G}_n^c(\epsilon)\right) + \mathbf{P}\left(\bigcup_{k \leq n} \mathbf{G}_n^-(\epsilon)^c\right) + \mathbf{P}\left(\bigcup_{k \leq n} \mathbf{G}_n^+(\epsilon)^c\right).$$

Por (4.6) y (4.7) se tiene que $\lim_{k \rightarrow \infty} \mathbf{P}\left(\bigcup_{k \leq n} \mathbf{G}_n^-(\epsilon)^c\right) = 0$ y $\lim_{k \rightarrow \infty} \mathbf{P}\left(\bigcup_{k \geq n} \mathbf{G}_n^+(\epsilon)^c\right) = 0$. Por

otro lado, como se demostró en [10], usando que $\lim_{r \rightarrow 0} V(B_r(p))/r^d \lambda(\mathcal{V}_1) = 1$, donde $V(B_r(p))$ es el volumen de la bola geodésica centrada en p y radio r y argumentos análogos a los estudiados en Devroye y Wagner (1997) tenemos que

$$\sup_{p \in M_0} \left| \frac{k_n}{nf(p)H_n^d(p)\lambda(\mathcal{V}_1)} - 1 \right| \xrightarrow{c.t.p.} 0$$

y por lo tanto $\lim_{k \rightarrow \infty} \mathbf{P}\left(\bigcup_{k \leq n} \mathbf{G}_n^c(\epsilon)\right) = 0$ entonces $\lim_{k \rightarrow \infty} \mathbf{P}\left(\bigcup_{k \leq n} \mathbf{S}_n^c(\epsilon)\right) = 0$. Lo que concluye la demostración del Teorema. \square

Bibliografía

- [1] Bai, Z.D.; Rao, C. and Zhao, L. (1988). Kernel Estimators of Density Function of Directional Data. *J. Multivariate Anal.* **27**, 24-39.
- [2] Bhattacharya, R. and Patrangenaru, V. (2002). Nonparametric estimation of location and dispersion on Riemannian manifolds. *Journal of Statistical Planning and Inference*. **108**, 23-35.
- [3] Boothby, W. M. (1975). An introduction to differentiable manifolds and Riemannian geometry. *Academic Press, New York*.
- [4] Di Marzio, M., Panzeraa, A. and Taylor, C. (2011). Kernel density estimation on the torus. *Journal of Statistical Planning and Inference*. **141**, 6, 2156-2173.
- [5] Do Carmo, M. (1988). Geometria Riemaniana. *Proyecto Euclides, IMPA. 2da edición*.
- [6] Fisher, N.I., T. Lewis, and B. J. J. Embleton (1987). Statistical Analysis of Spherical Data. *New York: Cambridge University Press*.
- [7] Goh, A. and Vidal, R. (2008). Unsupervised Riemannian Clustering of Probability Density Functions. *Lecture Notes In Artificial Intelligence*. **5211**.
- [8] Hall, P. , Watson, G.S. and Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika* **74**, 751–762.
- [9] Hendriks, H. and Landsman, Z. (2007). Asymptotic data analysis on manifolds. *Annals of Statistics*, **35**, 1, 109-131.
- [10] Henry, G. and Rodriguez, D. (2009). Robust Nonparametric Regression on Riemannian Manifolds. *Journal of Nonparametric Statistics*. **21**, 5, 611-628.
- [11] Henry, G. and Rodriguez, D. (2009). Kernel Density Estimation on Riemannian Manifolds: Asymptotic Results. *Journal Math. Imaging Vis.* **43**, 235-639.

- [12] Joshi, J., Srivastava, A. and Jermyn, I. H. (2007). Riemannian Analysis of Probability Density Functions with Applications in Vision. *Proc. IEEE Computer Vision and Pattern Recognition*.
- [13] Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.*, **30**, 509–541.
- [14] Keilhauer, G. (1995). Geometría Diferencial I. Facículo 38, Cursos y Seminarios del Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, UBA.
- [15] Lee, J. (1997). Riemannian Manifolds. An introduction to Curvature. *Springer, 1st edition* .
- [16] Loftsgaarden, D. and Quesenberry, C. (1965). A nonparametric estimate of a multivariate density function. *Ann. Math. Statist.* **36**, 1049–1051.
- [17] Mardia, K. (1972). Statistics of Directional Data. *Academic Press, London*.
- [18] Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33**, 1065–1076.
- [19] Pelletier, B. (2005). Kernel Density Estimation on Riemannian Manifolds. *Statistics and Probability Letters*, **73**, 3, 297-304.
- [20] Pennec, X. (2006). Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements. *Journal Math. Imaging Vis.*, **25**, 127-154.
- [21] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27**, 832–837
- [22] Wagner, T. (1975). Nonparametric estimates of probability densities. *IEEE Trans. Information Theory IT*. **21**, 438–440.
- [23] Warner, F. (1983). Foundations of Differentiable Manifolds and Lie Groups. *Springer*.