

Análisis intrínseco de la estimación puntual

José Manuel Corcuera Valverde

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tesisenxarxa.net) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tesisenred.net) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tesisenxarxa.net) service has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized neither its spreading and availability from a site foreign to the TDX service. Introducing its content in a window or frame foreign to the TDX service is not authorized (framing). This rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

Análisis intrínseco de la estimación puntual

Memoria presentada para optar al título de Doctor en Matemáticas por la Universidad de Barcelona, por

José Manuel Corcuera Valverde

Visto bueno
El Director,

Dr. Josep María Oller Sala,
Catedrático de Universidad.
Departamento de Estadística.
Universidad de Barcelona.

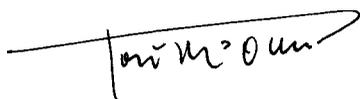
Barcelona, 11 de Mayo de 1994

Análisis intrínseco de la estimación puntual

Memoria presentada para optar al título de Doctor en Matemáticas por la Universidad de Barcelona, por

José Manuel Corcuera Valverde

Visto bueno
El Director,

A handwritten signature in black ink, reading "Josep Maria Oller Sala". The signature is written in a cursive style and is positioned above a horizontal line that extends to the left.

Dr. Josep Maria Oller Sala,
Catedrático de Universidad.
Departamento de Estadística.
Universidad de Barcelona.

Barcelona, 11 de Mayo de 1994

Índice

Prologo	1
1 Introduccion	2
1 1 Inconvenientes de las medidas clasicas de estimacion puntual	2
1 2 Enfoques invariantes	5
1 3 Las divergencias y la distancia de Rao	9
1 4 La geometria de los modelos parametricos	11
2 Objetos aleatorios en una variedad	13
2 1 Momentos y valores medios	13
3 El sesgo intrinseco y la distancia de Rao cuadratico media	24
3 1 Condiciones de regularidad	24
3 2 Sesgo intrinseco	25
3 3 La distancia de Rao cuadratico media	28
3 4 Ejemplos	28
4 Comportamiento local de un estimador	36
4 1 Cota inferior de la distancia de Rao cuadratico media	36
4 2 Desigualdad tensorial	44
4 3 La condicion de eficiencia	47
5 Eficiencia global de un estimador	51

5 1	Cotas del promedio Riemmaniano de la distancia de Rao cuadrático media	51
5 2	Metodos variacionales aplicados a la obtencion de cotas globales	58
6	Como mejorar el comportamiento de un estimador	64
6 1	Valores medios condicionados de funciones a valores en una variedad	64
6 2	Rao-Blackwell intrínseco	66
7	Propiedades asintóticas	69
7 1	Definiciones y resultados básicos	69
7 2	Normalidad asintótica en una variedad Estimación máximo-verosímil	73
7 3	Aproximación del tensor de sesgo en un sistema de coordenadas	76
8	Desarrollos de Taylor invariantes	79
8 1	Introducción	79
8 2	Principales resultados	80
8 3	Un teorema de Pitágoras para yokes	85
9	Eficiencias asintóticas	89
9 1	Introducción	89
9 2	Eficiencia intrínseca de primer orden	94
9 3	Comportamiento asintótico del sesgo intrínseco	97
10	Conclusiones	101
10 1	Resumen de resultados	101
10 2	Perspectivas del Análisis Intrínseco	102
11	Apendice	105
11 1	Derivada general covariante	106
11 2	La aplicación exponencial	107

11.3 Campos de Jacobi	110
11.4 Condiciones de geometría convexa	117
11.5 Coordenadas esféricas geodésicas	120
11.6 Volúmenes y teoremas de comparación	123
Bibliografía	126

Prologo

Una silla o una estrella *no* son en lo mas minimo aquello que parecen ser y como maximo conseguimos que sus borrosos contornos se vayan perfilando hasta convertirse en el haz de sensaciones que nos delimita todo objeto fisico. Por el contrario 2 o 317 no tienen nada que ver con ensacion alguna y sus propiedades se nos manifiestan con mayor claridad cuanto mas profundizamos y delimitamos nuestro examen.

G. H. Hardy

Los modelos estadísticos paramétricos con condiciones de regularidad convenientes tienen una estructura natural de variedad Riemanniana, dada por la métrica informativa. Ya que los parámetros son meras etiquetas de las medidas de probabilidad, una proposición inferencial debería formularse mediante objetos intrínsecos, invariantes bajo reparametrizaciones. En este contexto los estimadores serán objetos aleatorios a valores en la variedad correspondiente al modelo estadístico. A pesar de estas consideraciones, medidas clásicas e importantes que sirven para evaluar el comportamiento de un estimador, como el sesgo y el error cuadrático medio, son claramente dependientes de la parametrización del modelo estadístico.

En la presente memoria hemos abordado el problema de establecer medidas intrínsecas en estimación puntual. Para ello y en primer lugar se definen y estudian las propiedades relevantes de objetos aleatorios valorados en una variedad C^∞ n -dimensional. Así el concepto de momento de una variable aleatoria es generalizado al de campos aleatorios y la noción de valor medio es extendida para objetos aleatorios valorados en una variedad Hausdorff y conexa equipada con una conexión afin. En particular se considera el caso Riemanniano. Esta extensión se aplica posteriormente al sesgo y el error cuadrático medio en la teoría de estimación puntual.

Bajo esta óptica se han obtenido diversos resultados: cotas globales y locales para la distancia de Rao cuadrática media, las cuales dependen del sesgo intrínseco y de la curvatura del modelo estadístico. Asimismo se ha estudiado el comportamiento de un estimador cuando condicionamos respecto de un estadístico suficiente y completo, obteniéndose versiones intrínsecas de los teoremas de Rao-Blackwell y Lehmann-Scheffé. El estudio de propiedades asintóticas, desarrollos invariantes y eficiencias asintóticas completan el trabajo.

Introducción

1.1 Inconvenientes de las medidas clásicas de estimación puntual

La estimación se puede definir como la teoría que se ocupa de hacer inducciones a partir de los datos y de hacer inferencias sobre las inducciones. En la teoría de estimación paramétrica hacemos inducciones, a partir de la muestra, proponiendo medidas de probabilidad que pertenecen a una familia paramétrica, los parámetros son únicamente un “nombre” y por tanto no desempeñan para nosotros ningún papel en el proceso de inducción. Las inferencias se dan en la forma de estimaciones puntuales e intervalos de confianza y no importa eventualmente que inferencias específicas se necesiten. En este enfoque los estimadores suministran diferentes métodos de inducción.

Por otra parte, como es bien conocido, el sesgo y el error cuadrático son las medidas más ampliamente usadas para medir el comportamiento de un estimador. No obstante estas medidas son claramente dependientes del sistema de coordenadas o parametrización del modelo. No surgirían problemas si propiedades importantes, como la ausencia de sesgo o uniformemente mínima varianza, se preservasen bajo cambios en el sistema de coordenadas, esto es, cambios en la parametrización del modelo. Desgraciadamente esta no es la situación y esto es esencialmente debido, al carácter no tensorial del sesgo y del error cuadrático medio. Veamos un ejemplo que ilustre la situación que acabamos de comentar.

Ejemplo 1 1 1 Sea el modelo estadístico definido a través de la familia de densidades

$$p(x, \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)} \exp\{-\beta x\} \quad x, \beta \in \mathbb{R}^+,$$

donde $\alpha > 0$ es una constante conocida. Teniendo en cuenta que, para un tamaño muestral k ,

$$S = \sum_{i=1}^k X_i$$

es un estadístico suficiente para el modelo, el estimador, digamos estimador \mathcal{W} que es insesgado y UMV para β , con $k\alpha > 1$, viene dado por

$$\beta(\mathcal{W}) = \frac{k\alpha - 1}{S}$$

Pero parametrizando el mismo modelo estadístico como

$$p(x, \alpha, \lambda) = \frac{x^{\alpha-1}}{\lambda^\alpha \Gamma(\alpha)} \exp\left\{-\frac{x}{\lambda}\right\} \quad x, \lambda \in \mathbb{R}^+,$$

donde $\alpha > 0$ es de nuevo una constante conocida, el anterior estimador \mathcal{W} proporcionaría la estimación de λ

$$\lambda(\mathcal{W}) = \frac{S}{k\alpha - 1}$$

que es sesgada. El correspondiente estimador, digamos \mathcal{U} , que es ahora insesgado y UMV para λ viene dado por

$$\lambda(\mathcal{U}) = \frac{S}{k\alpha}$$

Sin embargo, si utilizáramos este estimador para dar una estimación de β obtendríamos

$$\beta(\mathcal{U}) = \frac{k\alpha}{S}$$

que es sesgada.

Por último, si calculamos el error cuadrático medio de estos estimadores, \mathcal{W} y \mathcal{U} , en ambas parametrizaciones, la siguiente tabla, suponiendo $k\alpha > 2$, resume la discusión anterior.

Parametros	Estimadores	
	\mathcal{W}	\mathcal{U}
β	$\text{Sesgo}(\beta(\mathcal{W})) = 0$ $\text{ECM}(\beta(\mathcal{W})) = \frac{\beta^2}{k\alpha - 2}$ <p>(No alcanzando la cota de Cramer Rao $\frac{\beta^2}{k\alpha}$)</p>	$\text{Sesgo}(\beta(\mathcal{U})) = \frac{\beta}{\alpha - 1}$ $\text{ECM}(\beta(\mathcal{U})) = \frac{(k\alpha + 2)\beta^2}{(k\alpha - 1)(k\alpha - 2)}$ <p>El estimador \mathcal{W} es preferible al estimador \mathcal{U}</p>
λ	$\text{Sesgo}(\lambda(\mathcal{W})) = \frac{\lambda}{k\alpha - 1}$ $\text{ECM}(\lambda(\mathcal{W})) = \frac{(k\alpha + 1)\lambda^2}{(k\alpha - 1)^2}$ <p>(Alcanzando la cota de Cramer Rao $\frac{\lambda^2}{k\alpha}$)</p>	$\text{Sesgo}(\lambda(\mathcal{U})) = 0$ $\text{ECM}(\lambda(\mathcal{U})) = \frac{\lambda^2}{k\alpha}$ <p>El estimador \mathcal{U} es preferible al estimador \mathcal{W}</p>

Este ejemplo muestra algunos de los problemas, paradojas o inconsistencias de la clasica estimacion insesgada uniformemente de minima varianza, esencialmente debidos a la dependencia respecto del sistema de coordenadas, o parametrizacion del modelo, de este criterio estadistico. Por tanto parece deseable modificar estas nociones clasicas en un sentido invariante o intrinseco.

En esta situacion surge una pregunta natural ¿se pueden formular nociones analogas al sesgo y al error cuadratico que dependan solo del procedimiento de estimacion empleado? Hay, en principio, varias formas de intentar alcanzar el proposito expuesto. Primero podemos privilegiar un sistema de coordenadas, aunque fuera dificil justificar la eleccion, y a continuacion definir una funcion de perdida extrinseca y proceder como en Lehmann, vease [42]. Este procedimiento puede resultar razonable desde un punto de vista teorico, pero desde el punto de vista de la inferencia estadistica puede resultar mas conveniente trabajar exclusivamente con conceptos intrinsecos al modelo estadistico.

El espíritu de lo que llamaremos *Analisis Intrínseco* de la estimación estadística es desarrollar una teoría de estimación análoga a la clásica, basada en las estructuras geométricas de los modelos estadísticos. Así uno de los objetivos del *Analisis Intrínseco* es suministrar herramientas invariantes que permitan analizar el comportamiento de un estimador, y otro es obtener resultados análogos a los clásicos y establecer conexiones entre las medidas clásicas no invariantes y las medidas intrínsecas obtenidas aquí.

En esta memoria, teniendo en cuenta la estructura Riemanniana de los modelos paramétricos regulares una medida de sesgo *intrínseco* es obtenida al considerar el valor medio de aplicaciones medibles a valores en una variedad. El promedio cuadrático de la distancia Riemanniana, o distancia de Rao, es el invariante análogo del error cuadrático medio.

La primera parte de esta memoria se dedica al estudio de los momentos de un campo aleatorio en una variedad real n -dimensional C^∞ , y también al concepto de valor medio de objetos aleatorios que toman valores en una (Hausdorff y conexa) variedad equipada con una conexión afin, a través de la aplicación exponencial. Se enfatizan las analogías y diferencias entre momentos y valores medios, y se considera, en particular, el caso Riemanniano.

La segunda parte es la aplicación de estos resultados al estudio de algunas medidas invariantes análogas al sesgo y al error cuadrático medio correspondientes a un estimador. La tercera y cuarta partes se dedican al desarrollo de versiones intrínsecas de cotas globales y locales de Cramer–Rao. En la quinta parte estudiamos el comportamiento del promedio cuadrático de la distancia de Rao cuando condicionamos respecto de un estadístico suficiente y completo, obteniéndose versiones intrínsecas de los teoremas de Rao Blackwell y Lehmann–Scheffé. A continuación se estudian algunas propiedades asintóticas especialmente las relacionadas con el estimador máximo-verosímil. Por último se propone un método para obtener desarrollos de Taylor invariantes y se aplican, entre otros casos, al estudio de las eficiencias asintóticas.

1.2 Enfoques invariantes

Aunque la Estadística diste mucho de estar axiomatizada de forma satisfactoria y los principios estadísticos sean como un conjunto de criterios recomendables pero

no exigibles y aunque un ilustre estadístico como L. Le Cam [40] llegue a decir al final de su artículo *Maximum Likelihood An Introduction*

If the hallowed maximum likelihood principle leads us to difficulties may
 be some other principle will save us
 There is indeed such a principle It is as follows
Basic Principle 0 Do not trust any principle

aunque proceda con tal excepticismo decimos, vamos a mencionar varios principios relacionados con la filosofía que subyace en la presente memoria. No pretende ser esta sección, como toda la introducción, excesivamente detallada desde el punto de vista matemático, mas bien pretende ser *metamatemática* y tener así la posibilidad de expresar mas claramente el contexto y los objetivos de la memoria.

Dado un modelo estadístico paramétrico $\{(\mathcal{X}, \mathbf{a}, P_\theta), \theta \in \Theta\}$, entenderemos que la teoría de estimación puntual, como ya hemos indicado en la sección anterior, se preocupa por determinar el mecanismo probabilístico que ha generado una muestra y del cual unicamente se sabe que pertenece al modelo anterior. Aunque resulte ambiguo, se tratara de estimar tal mecanismo de la forma mas "certera" posible. Si existe una biyección, cosa por otra parte natural, entre los valores de $\theta \in \Theta$ y las medidas probabilísticas, el problema anterior sera equivalente a determinar el valor verdadero de $\theta \in \Theta$, tales modelos se suelen denominar *identificables* y nosotros nos restringiremos a ellos. Sobre este punto deberiamos establecer el siguiente criterio de invarianza.

Principio de invarianza de la estimación bajo reparametrización Si θ_1 es la estimación de θ_1 y $g: \Theta_1 \rightarrow \Theta_2$ es una biyección entonces $\theta_2 = g(\theta_1)$ es la estimación de $\theta_2 = g(\theta_1)$.

Este principio resulta obvio si se entiende la estimación puntual en el sentido señalado al comienzo. No resultaria tan obvio si uno entendiera que el proposito es estimar el verdadero valor *del parametro*. Tal principio se verifica por ejemplo en el caso de la estimación máximo verosímil y se suele conocer esta propiedad como *invarianza funcional* aunque no emplearemos tal terminología pues resultaria superflua al considerar unicamente estimaciones, y por tanto estimadores, con esta propiedad.

Si la manipulacion de los parametros, o de la parametrizacion, no deben alterar la estimacion, lo mismo puede decirse respecto a la manipulacion de los propios datos

Principio de invarianza bajo transformaciones biyectivas (y medibles) de los datos Si $\theta(x_1, \dots, x_k)$ es el valor estimado de θ a partir de una muestra x_1, \dots, x_k este mismo valor se debe obtener con $\tilde{\theta}(y_1, \dots, y_k)$ donde y_1, \dots, y_k resultan de transformar de forma (medible y) biyectiva los datos

Es tambien facil de ver que el estimador maximo-verosimil tambien tiene esta propiedad En cambio este principio *no* es seguido por el *metodo de los momentos* Consideremos, por ejemplo, una distribucion de Poisson $P(\lambda)$ y una muestra de tamaño k , x_1, \dots, x_k Supongamos que queremos estimar $\eta = \lambda^2 + \lambda$, tal metodo sugiere utilizar como estimacion $\eta = \bar{x}^2 + \bar{x} = \left(\frac{1}{k} \sum_{i=1}^k x_i\right)^2 + \frac{1}{k} \sum_{i=1}^k x_i$, en cambio si ahora consideramos la transformacion $y = x^2$ resulta que la estimacion de η seria $\tilde{\eta} = \bar{y} = \frac{1}{k} \sum_{i=1}^k y_i$ que obviamente no coincide con la anterior

Aqui es necesario mencionar otra propiedad de invarianza que algunos autores, al unirla a la propiedad de invarianza bajo reparametrizacion y bajo transformaciones biyectivas, conocen como *equivarianza* del estimador, nosotros exigiremos tal propiedad y asi tendremos el siguiente principio

Principio de invarianza Si dos problemas de estimacion tienen la misma estructura formal (el mismo modelo estadistico) entonces el estimador (la regla de estimacion) debe ser el mismo

Conviene aclarar la posible falta de obviedad del anterior enunciado Consideremos una familia de $N(\mu, 1)$ n dimensional y una muestra de tamaño k , x_1, \dots, x_k Sea $\mathcal{U}(x_1, \dots, x_k)$ una estimacion de μ Si ahora añadimos la cantidad a a los datos anteriores, tendremos una muestra de tamaño k $y_1 = x_1 + a, \dots, y_k = x_k + a$ de la familia $N(\mu + a, 1)$ Al no haber variado la familia el estimador debera ser el mismo de modo que la estimacion de $\mu + a$ debera ser $\mathcal{U}(y_1, \dots, y_k)$ Si ahora aplicaramos el principio de invarianza bajo reparametrizacion la estimacion de $\mu + a$ debera ser $\mathcal{U}(x_1, \dots, x_k) + a$ Por ultimo como la transformacion de los datos es biyectiva ambas estimaciones deberan coincidir,

esto es

$$\mathcal{U}(x_1 + a, \dots, x_k + a) = \mathcal{U}(x_1, \dots, x_k) + a$$

Esta propiedad tan natural y que en nuestro contexto debe cumplir todo estimador no la cumple el estimador de James-Stein [29] de la media de una $N_n(\mu, I)$

$$\mu = \left(1 - \frac{n-2}{k(\bar{x}_1^2 + \dots + \bar{x}_n^2)}\right) \bar{X}$$

que sin embargo tiene menor riesgo si tomamos como funcion de perdida el cuadrado de la norma Euclidea de las desviaciones de las estimaciones respecto del verdadero valor, para dimension $n \geq 3$

Parece coherente exigir que la “certidumbre”, “plausibilidad” o riesgo del estimador no dependa de la parametrización que utilicemos para representar la familia de probabilidades ni de posibles transformaciones biyectivas de los datos y no cambie bajo transformaciones que no alteren el modelo. En este sentido podemos establecer el siguiente principio

Principio de invarianza de la funcion de perdida La funcion de perdida que utilicemos para medir el riesgo de un estimador debe ser invariante bajo reparametrizaciones, transformaciones biyectivas y medibles de los datos y bajo transformaciones que dejen invariante el modelo

La ausencia de estas condiciones conlleva, como hemos visto en el ejemplo de la seccion anterior, situaciones paradójicas. Por supuesto estas condiciones son muy generales y nuestro proposito es desarrollar una teoria de la inferencia en la que no necesitemos introducir o añadir nuevos objetos al modelo estadístico. Es por eso que utilizaremos una medida de certidumbre de la estimación que a la vez de intrínseca en el sentido anterior lo será en el sentido de no añadir nada externo al modelo. Esto es, la funcion de perdida se deducirá del propio modelo

Una funcion de perdida con las propiedades anteriores será introducida en la proxima seccion

1 3 Las divergencias y la distancia de Rao

Dado un modelo estadístico paramétrico $\{(\mathcal{X}, \mathbf{a}, P_\theta), \theta \in \Theta\}$ existen muchas formas de medir la discrepancia (o divergencia) entre dos medidas pertenecientes a $M = \{P_\theta, \theta \in \Theta\}$. No obstante podríamos acordar que tales medidas de discrepancia deben tener la forma de una aplicación

$$D: M \times M \rightarrow \mathbb{R}_+ \cup \{0\}$$

con la propiedad adicional de que $\phi(\theta, \theta') = 0$ si $\theta = \theta'$. Automáticamente si nos restringimos a funciones “suaves” podremos expresar esta condición de mínimo en la diagonal de $M \times M$ de la forma

$$(i) \quad \partial D(\theta, \theta) = 0$$

(ii) la matriz

$$\{\partial^2 D(\theta, \theta)\}$$

es definida positiva

y esto para todo $\theta \in \Theta$, donde la derivada se calcula con respecto al segundo argumento. Puesto que la parametrización es irrelevante, en el sentido de que los parámetros son meras etiquetas que nos permiten distinguir una ley de otra dentro de la familia M , lo anterior deberá formularse de forma independiente de la elección de la parametrización. Tal cosa se consigue si pensamos en M como una variedad donde las diferentes parametrizaciones juegan el papel de sistema de coordenadas en la variedad. Entonces, las propiedades anteriores de D deben entenderse enunciadas para un sistema, local o global, de coordenadas en M .

Las funciones con estas propiedades se denominan *divergencias* y diferentes tipos de divergencias, medidas de diversidad y *entropías*, conceptos estrechamente relacionados con el anterior, han sido propuestos en el ámbito de la Estadística y de la Teoría de la Información. Véase por ejemplo Bahattacharyya [11], Shannon [56], Kullback y Leibler [38], Renyi [55] y Rao [53] entre otros. Desde un punto de vista general y abstracto y con otro propósito que el de distanciar poblaciones, estas funciones fueron consideradas por Barndorff Nielsen, véase [5], y tales funciones cambiadas de signo las denomina *yokes*.

Las condiciones anteriores son muy generales y la posible forma de D resulta muy arbitraria. Si consideramos un modelo dominado por una medida de referencia μ , los elementos de M se pueden considerar como las funciones de densidad del modelo, en definitiva un subconjunto de $L^1(\mu)$ y tendremos que D define un funcional con las propiedades 1) y 11) No existe, hasta la fecha, una caracterización de los mismos y en cualquier caso deberíamos imponer que tal funcional fuera invariante bajo cambios en la medida de referencia μ . Si consideramos la clase de funcionales

$$D_\phi : L^1(\mu) \times L^1(\mu) \rightarrow \mathbf{R}^+$$

$$(f, g) \mapsto D_\phi(f, g) = \int_{\mathcal{X}} \phi(f, g) d\mu$$

con $\phi : \mathbf{R}^+ \times \mathbf{R}^+ \rightarrow \mathbf{R}^+$ una función "suave", tendremos que la invarianza bajo cambios en la medida de referencia hara necesario que

$$\phi(x, y) = \phi(1, y/x)x, \quad \text{para todo } x, y, z \in \mathbf{R}^+$$

esto conduce a las discrepancias de la forma

$$D_\phi = \int_{\mathcal{X}} \phi\left(\frac{f}{g}\right)g d\mu$$

que se conocen como ϕ -divergencias. Las condiciones $D_\phi(f, g) \geq D_\phi(f, f) = 0$ equivalen a

$$E_g \left(\phi \left(\frac{f}{g} \right) \right) \geq \phi(1) = \phi \left(E_g \left(\frac{f}{g} \right) \right) = 0$$

Si queremos una $\phi(x)$ que valga para cualquier familia, lo anterior se cumplira si $\phi(x)$ es convexa y $\phi(1) = 0$

Si suponemos las condiciones de suavidad y regularidad necesarias para poder intercambiar la derivacion con respecto al parametro con la integracion en \mathcal{X} resultara

$$\partial_{\alpha\beta} D(\theta, \theta) = \int_{\mathcal{X}} \phi''(1) (\partial_\alpha \log f_\theta) (\partial_\beta \log f_\theta) f_\theta d\mu = \phi''(1) g_{\alpha\beta}$$

donde $g_{\alpha\beta}$ es la *matriz de informacion de Fisher*. Por tanto

$$D(\theta, \theta + \Delta\theta) = \frac{1}{2} \phi''(1) g_{\alpha\beta} \Delta\theta^\alpha \Delta\theta^\beta + \dots,$$

Esto es, *localmente* todas las φ divergencias son equivalentes. Si consideramos en el conjunto de densidades todos los cambios suaves de coordenadas, por ejemplo C^∞ , tenemos una variedad C^∞ en la que, como es fácil de comprobar, $g_{\alpha\beta}$ se comporta como un tensor 2-covariante. Esto proporciona una métrica natural en el conjunto de densidades. Tal métrica se denomina *métrica informativa* y la distancia a la que da origen se denomina *distancia de Rao*.

1.4 La geometría de los modelos paramétricos

El elemento infinitesimal anterior $ds^2 = g_{\alpha\beta} d\theta^\alpha d\theta^\beta$, tiene las siguientes importantes propiedades

- Es invariante bajo transformaciones *suficientes*
- Es invariante bajo cambios en la medida de referencia
- Es invariante bajo reparametrizaciones
- Es, salvo un constante multiplicativa, localmente equivalente a todas las φ -divergencias

Todo esto hace pensar que ds^2 es el elemento infinitesimal de distancia natural entre las distribuciones de un modelo estadístico paramétrico.

La idea de considerar la matriz de información de Fisher como medida local de la distancia entre distribuciones llevó a Rao [54], en 1945, a introducir la métrica anterior en M , dotando a la variedad de una estructura Riemanniana. La distancia obtenida puede considerarse como una generalización de la distancia de Mahalanobis [44] pues coincide con esta en el caso de una familia $N(\mu, \Sigma)$ donde Σ es conocida. Cabe citar aquí el trabajo de Jeffreys [30] quien al darse cuenta del carácter tensorial de la matriz de Fisher, la utiliza para definir distribuciones a priori *no* informativas e invariantes.

Aproximadamente treinta años más tarde, Cencov [20] introdujo una familia de conexiones afines en M , que difieren de la conexión de Levi-Civita asociada a la métrica Riemanniana introducida por Rao, revelando ciertas propiedades geométricas de la familia exponencial. Paralelamente Efron [24] en su estudio de la eficiencia del estimador máximo-verosímil para una subfamilia exponencial

curvada uniparametrica, encontro una relacion entre la perdida de informacion y ciertas curvaturas de la subfamilia en cuestion, aqui tambien se utilizaron conexiones no Riemannianas aunque no explicitamente. La generalizacion del trabajo de Efron fue llevada a cabo por Madsen [43] y, de forma bastante exhaustiva, por Amari [2] aunque ciñéndose en parte a las familias exponenciales. Amari tambien analiza a fondo la familia uniparametrica de conexiones que surge en los modelos estadisticos parametricos. En relacion con la geometria informativa de los modelos, es decir su estudio descriptivo: calculo de geodesicas, curvaturas y distancias, se podrian destacar los trabajos de Burbea [13], Atkinson y Mitchell [3], Oller [49], Oller y Cuadras [51] y Calvo y Oller [17] entre otros.

A pesar de las buenas propiedades de la matriz de informacion de Fisher como tensor metrico natural para distanciar probabilidades de una familia parametrica, algunos autores han utilizado otras metricas, aunque no siempre con la finalidad de "distanciar" poblaciones, destacando en este sentido los trabajos de Burbea y Rao [16] y especialmente las llamadas *geometrias observadas* de Barndorff-Nielsen, vease por ejemplo [6], donde podria decirse que la informacion de Fisher es estimada a partir de la muestra proporcionando una metrica variable que asintoticamente coincide con la geometria informativa y que en este contexto se denomina *geometria esperada*.

Objetos aleatorios en una variedad

2.1 Momentos y valores medios

Sea (χ, \mathcal{a}, P) un espacio de probabilidad, donde χ es el espacio muestral, \mathcal{a} es una σ -álgebra de subconjuntos de χ y P es una medida de probabilidad en \mathcal{a} . Sea (M, \mathfrak{A}) una variedad real C^∞ n dimensional, donde \mathfrak{A} es el atlas de M .

Sea f una aplicación medible, $f: \chi \rightarrow M$, también llamada un *objeto aleatorio* en M , esto es, una aplicación medible tal que para todo conjunto abierto $W \subset M$, $f^{-1}(W) \in \mathcal{a}$. Introduciremos ahora las nociones de valor medio y momentos de f , con las suposiciones más débiles posibles y manteniendo la noción intuitiva de medida de *centralidad*, en íntima relación con la idea de *centro de masas* como veremos más tarde (en relación con el presente tema véase Karcher [33], Kobayashi y Nomizu [37], Kendall [36] [35] y Emery y Mokobodzki [25]).

Si existiese una carta global (M, ϕ) podríamos intentar definir el valor medio de f como

$$E(f) = \phi^{-1} \left(\int_{\chi} (\phi \circ f)(x) P(dx) \right),$$

pero esta ingenua solución no resulta satisfactoria ya que $E(f)$ depende, en general, del sistema de coordenadas. Solamente las transformaciones lineales preservarían $E(f)$.

Para poder resolver este problema, necesitamos introducir primero algunos conceptos. Sea A un subconjunto de M , y $\mathcal{F}_A^{(p,q)}$ el conjunto de todos los campos tensoriales C^∞ en un subconjunto abierto de A , de orden $p+q$, p veces contravariante y q veces covariante. Si fijamos $m \in A$, cualquier aplicación X de χ a $\mathcal{F}_A^{(p,q)}$

induce una aplicacion X_m , tal que $X_m: \chi \rightarrow T_q^p(M_m)$ con $X_m(x) = (X(x))_m$, donde $T_q^p(M_m)$ simboliza el espacio de (p, q) tensors en el espacio tangente en m , M_m , teniendo la estructura natural de un espacio vectorial topologico finito-dimensional. Considerando la σ -algebra de Borel en $\mathcal{F}_A^{(p, q)}$ inducida por las σ -algebras de Borel de los M_m , tenemos una definicion simple,

Definicion 2 1 1 *Un C^∞ (p, q) campo aleatorio tensorial en A , X , es una aplicacion medible de χ a $\mathcal{F}_A^{(p, q)}$*

De la definicion se sigue que $\forall m \in A$, la aplicacion X_m es una aplicacion medible en (χ, \mathcal{A})

Ademas, cualquier campo tensorial aleatorio se puede caracterizar por sus $n^{(p+q)}$ componentes con respecto a un sistema de coordenadas, $\theta^1, \dots, \theta^n$,

$$X_{\beta_1}^{\alpha_1} \dots \beta_q^{\alpha_p}(x, \theta^1, \dots, \theta^n) = \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q = 1, \dots, n,$$

las cuales son claramente, fijado x , funciones C^∞ de $\theta^1, \dots, \theta^n$, y, fijado θ real valued measurable funciones reales medibles en (χ, \mathcal{A})

Sea \otimes el producto de campos tensoriales. En el presente contexto es natural definir

Definicion 2 1 2 *El momento de orden k de un campo tensorial aleatorio X es un (kp, kq) -tensor ordinario en A definido por*

$$\mathcal{M}^k(X) = \int_{\chi} X(x) \overbrace{\otimes \dots \otimes}^k X(x) P(dx), \quad k \in \mathbf{N},$$

supuesta la existencia de la anterior integral

Notese que $\mathcal{M}^k(X)$ puede ser calculado explicitamente a traves de sus componentes en un sistema de coordenadas. Las componentes de $\mathcal{M}^k(X)$, con respecto a un sistema de coordenadas $\theta^1, \dots, \theta^n$, vendra dado por

$$M_{\beta_1}^{\alpha_1} \dots \beta_q^{\alpha_p}(\theta) = \int_{\chi} \left(X_{\beta_1}^{\alpha_1} \dots \beta_q^{\alpha_p}(x, \theta) \right) P(dx)$$

Esta es de hecho la forma mas simple y tambien la extension mas natural del momento de orden k a un campo tensorial aleatorio. En particular, el momento de primer orden es la esperanza de X

Podemos escribir tambien

$$\mathcal{M}^k(X) = E(X \overbrace{\otimes \otimes}^k X),$$

donde el producto tensorial de campos tensoriales aleatorios se define de forma natural a partir producto tensorial de campos tensoriales ordinarios

De forma similar podriamos definir los *momentos centrales*, los cuales poseeran las propiedades clasicas, por ejemplo

$$\mathcal{M}_c^2(X) = E((X - EX) \otimes (X - EX)) = E(X \otimes X) - E(X) \otimes E(X)$$

En el caso que X sea un campo vectorial, las componentes de este tensor, con respecto a un sistema de coordenadas, se pueden escribir en notacion matricial, obteniendose la matriz de covarianzas, Σ_X ,

$$\Sigma_X = E(XX') - E(X)E(X)',$$

identificando, en la anterior ecuacion, los vectores con sus componentes y siendo X un vector columna y X' el correspondiente vector fila

De cara a la definicion de valor medio de un objeto aleatorio, tenemos que introducir una estructura adicional sobre la variedad supondremos que esta dotada de una conexion afin. Tipicos ejemplos de variedades con una conexion afin son las variedades Riemannianas

Asociada con una conexion afin existe una aplicacion, llamada la aplicacion exponencial, $\exp_p: M_p \rightarrow M$. Ella esta definida para todo v en un entorno abierto estrellado de $0_p \in M_p$. Asimismo, es tambien bien sabido que esta aplicacion, en general, no posee inversa, aunque haya casos importantes donde existe. De cualquier forma siempre podemos restringir la aplicacion a un entorno abierto de $0_p \in M_p$, tal que la aplicacion inversa este bien definida, siendo entonces la aplicacion exponencial un difeomorfismo local. Informacion adicional puede encontrarse en (11.2) del apendice

Vamos a precisar el tipo de entornos que consideramos apropiados para definir el valor medio de un objeto aleatorio

Definicion 2.1.3 Diremos que un entorno $B(p)$ de $p \in M$ es normal si $B(p)$ es la imagen difeomorfica, por la aplicacion exponencial, de un entorno abierto estrellado de $0_p \in M_p$

Notese que un entorno normal $B(p)$ de p tiene la propiedad de que cualquier $q \in B(p)$ puede ser unido a p por una *única* geodesica contenida en $B(p)$

En el espacio vectorial M_p consideraremos entornos estrellados, $A(p)$, tales que $A(p) = -A(p)$, en el caso que tengamos solamente una conexión afin, y esferas en el caso Riemanniano. Este tipo de entornos seran denominados *bolas* con centro 0_p , incluso en el caso afin

Definición 2 1 4 *La imagen, $B(p)$, mediante la aplicacion exponencial, de una bola abierta $A(p)$ con centro 0_p , se denominara una bola normal con centro p si $B(p)$ es un entorno normal de p*

Observese que, en el caso Riemanniano, la geodesica *mas corta* que une p con cualquier $q \in B(p)$, siendo $B(p)$ una bola normal con centro p , es única en M y se encuentra en $B(p)$. No obstante, podemos considerar entornos mas generales con esta propiedad

Definición 2 1 5 *Diremos que un conjunto abierto $B(p)$ es un entorno regular normal de p si y solo si su interseccion con cualquier bola normal de centro p sigue siendo normal*

En el caso Riemanniano podemos asegurar la existencia de este tipo de entornos. Puesto que cada punto p tiene un entorno donde la aplicacion exponencial es un difeomorfismo, podemos obtener un entorno normal de $0_p \in M_p$ como sigue: sea $A(p)$ el correspondiente entorno estrellado en M_p , consideremos entonces una bola con centro $0_p \in M_p$ donde la aplicacion exponencial sea inyectiva, si restringimos ahora la aplicacion a la interseccion con $A(p)$ obtenemos un entorno regular normal de p . Es facil ver que en el caso Riemanniano un entorno $B(p)$ de p es regular normal si y solo si la geodesica mas corta que une p con cualquier otro punto de $B(p)$ es única y se encuentra en $B(p)$, entonces los entornos regulares normales son una generalizacion de entornos con esta propiedad al caso afin

De esta forma, dado un objeto aleatorio f que toma valores en una (Hausdorff y conexa) variedad, equipada con una conexión afin (que puede ser la conexión de Levi Civita correspondiente a una variedad Riemanniana), habra una forma natural de definir un vector aleatorio, fijado $p \in M$, dado por $\exp_p^{-1}(f(x))$. Este

vector no estara necesariamente definido para todo $x \in \mathcal{X}$, pero si lo esta salvo en un conjunto de medida nula, podemos introducir la siguiente definicion de *valor medio*,

Definicion 2 1 6 *Un punto en la variedad $p \in M$ es un valor medio del objeto aleatorio f y escribiremos $p = \mathfrak{M}(f)$, si y solo si existe un entorno regular normal de p donde f toma valores casi seguramente $[P]$, y*

$$\int_{\mathcal{X}} \exp_p^{-1}(f(x)) P(dx) = 0_p$$

Este valor medio es lo que Emery y Mokobodzki [25] llaman *baricentro exponencial*

Observaciones Remarquemos que esta es una definicion de un *valor medio intrinseco*, independiente del sistema de coordenadas

Ademas, en el caso en que M sea una variedad Riemanniana completa, si P_f es la medida de probabilidad inducida por la aplicacion medible en M , y P_f esta dominada por la medida Riemanniana, para cualquier $p \in M$ tendremos un entorno regular normal de p , con probabilidad $[P]$ igual a uno, donde $\exp_p^{-1}(f(x))$ estara definida. Esto es una consecuencia inmediata del teorema de la medida imagen y de que el *cut locus* de p en M tiene medida Riemanniana cero, vease la seccion (11 2) del apendice

En lo que sigue utilizaremos la notacio $\exp_p^{-1}(\cdot)$ para indicar la inversa de la aplicacion exponencial en cierto entorno regular normal de p

Consideremos ahora varios ejemplos

Ejemplo 2 1 7 Sea M precisamente \mathbf{R}^n . Identificando los puntos con sus coordenadas correspondientes a la carta trivial, y considerando la usual conexion afin Euclideana, tenemos, para $z, m \in \mathbf{R}^n$, que $\exp_m^{-1}(z) = (z - m)_m$. Para encontrar el valor medio de una variable aleatoria f tendremos que resolver la ecuacion

$$\int_{\mathcal{X}} (f(x) - m)_m P(dx) = 0_m,$$

pero esta ecuacion tiene como solucion unica la solucion trivial

$$m = \int_{\mathcal{X}} f(x) P(dx),$$

supuesta la existencia de esta integral. Por tanto recuperamos la definicion clasica

$$\mathfrak{M}(f) = E(f) = \int_x f(x) P(dx)$$

Ademas, el momento central de segundo orden de $\exp_m^{-1}(f(x))$ se puede escribir, en notacion matricial y omitiendo el subindice m , como

$$\begin{aligned} \Sigma_f &= \mathcal{M}_c^2(\exp_m^{-1}(f(x))) = E((f(x) - m)(f(x) - m)') \\ &= E(ff') - E(f)E(f)', \end{aligned}$$

que es la matriz de covarianza usual

Ejemplo 2 1 8 Otro ejemplo interesante viene dado al considerar el valor medio de la distribucion de Von Mises. En este caso la variedad es la esfera Euclidea n -dimensional, con la conexi3n inducida por la sumersi3n natural en el espacio Euclideo \mathbf{R}^n . La medida de probabilidad inducida en la variedad es absolutamente continua con respecto a la medida de superficie en la esfera y la correspondiente funci3n de densidad viene dada por

$$p(x, \xi, \lambda) = \alpha_n(\lambda) \exp\{\lambda \xi' x\} \quad x, \xi \in S_n = \{z \in \mathbf{R}^n \mid z'z = 1\}, \lambda \in \mathbf{R}^+,$$

donde $\alpha_n(\lambda) = \lambda^{k/2-1}/(2\pi)^{k/2} I_{k/2-1}(\lambda)$ es una constante de normalizaci3n, y $I_{k/2-1}$ la funci3n de Bessel modificada de primer tipo y orden $k/2 - 1$. En este caso es clara la existencia de dos valores medios, dados por ξ y $-\xi$. Comparese este resultado con las *direcciones medias* definidas en Mardia *et al* [39, 424-451]. Vease tambien la exhaustiva y didactica exposici3n de Jupp y Mardia [32],

Ejemplo 2 1 9 Consideremos una variable aleatoria uniformemente distribuida en un circulo, con la conexi3n inducida por la sumersi3n natural en la variedad Euclidea \mathbf{R}^2 . Entonces, *todos* los puntos de la circunferencia son valores medios

Podemos suministrar, en el caso Riemanniano, una medida escalar de *dispersi3n* con respecto al valor medio m : el ordinario valor esperado de la distancia Riemanniana al cuadrado entre $f(x)$ y m , que puede ser vista como una versi3n invariante, independiente del sistema de coordenadas, de la varianza de una variable aleatoria real. Es tambien posible definir una medida de dispersi3n con respecto a un punto de referencia arbitrario de una variedad Riemanniana, como

el valor medio de la distancia Riemanniana al cuadrado entre $f(x)$ y el punto de referencia seleccionado

Podemos señalar también que, con esta extensión del concepto de valor medio, mantenemos el intuitivo y atractivo significado de *medida de centralidad*, a pesar de que ya no tenemos las propiedades de *linealidad* de la esperanza. Sin embargo, esto es absolutamente normal, ya que no podemos identificar, en general, M con sus espacios tangente. De forma similar tendremos una disociación entre el valor medio y el concepto de momento de primer orden. Los momentos de un objeto aleatorio que toma valores en M , se definirán entonces como

Definición 2.1.10 *El momento de orden k del objeto aleatorio f es un $(k, 0)$ campo tensorial en A definido por*

$$\mathcal{M}^k(f)_m = \int_x \exp_m^{-1}(f(x)) \overbrace{\otimes \otimes}^k \exp_m^{-1}(f(x)) P(dx), \quad \forall m \in A, \quad k \in \mathbb{N}$$

supuesta la existencia de la anterior integral

veamos que hay una relación entre el *valor medio* definido y el clásico *centro de masas*, \mathfrak{C} ,

$$\mathfrak{C} = \arg \min_{m \in M} \mathcal{H}_f(m),$$

donde $\mathcal{H}_f(m) = \int_x \rho^2(m, f(x)) P(dx)$. Antes de nada tenemos los siguientes resultados

Proposición 2.1.11 *Si existe algún $m_0 \in M$ para el cual $\mathcal{H}_f(m_0)$ este definido, entonces la función $\mathcal{H}_f(m)$ estará definida para todo $m \in M$*

Demostración

Por la desigualdad triangular

$$\mathcal{H}_f(m) \leq 2\mathcal{H}_f(m_0) + 2\rho^2(m, m_0),$$

de lo que se sigue la proposición ■

Supongamos ahora que existe un punto m_0 tal que $\mathcal{H}_f(m_0) < \infty$, entonces

Proposición 2.1.12 Siempre que $\exp_m^{-1}(\cdot)$ este bien definida para todo $p \in M$ P -casi seguramente la función $\mathcal{H}_f(m)$ es diferenciable y

$$X_m \mathcal{H}_f = -2 \langle X_m, \int_x \exp_m^{-1}(f(x)) P(dx) \rangle$$

Demostración

Para todo $X_m \in M_m$, ya que, fijando $q \in M$, $\rho^2(\cdot, q)$ es una función C^∞ , podemos escribir

$$\begin{aligned} X_m \rho^2(\cdot, q) &= X_m \|\exp_m^{-1}(q)\|^2 = 2 \langle \nabla_{X_m} \exp_m^{-1}(q), \exp_m^{-1}(q) \rangle \\ &= -2 \langle X_m, \exp_m^{-1}(q) \rangle, \end{aligned}$$

donde la última igualdad puede comprobarse fácilmente considerando un sistema de coordenadas geodesicas esféricas con origen q . Entonces, tenemos

$$|X_m \rho^2(\cdot, q)| \leq 2 \|X_m\| \rho(m, q),$$

por tanto, si U_m es un entorno de m con diámetro D , por la desigualdad triangular

$$|X_m \rho^2(\cdot, q)| \leq 2 \|X_m\| (\rho(m, q) + D) \quad \forall m' \in U_m \quad X_m \in M_m \quad (2.1)$$

Sea X un campo vectorial C^∞ tal que $X(m) = X_m$ y consideremos, en un entorno de m incluido en U_m , la curva integral de X , $\gamma(t)$, tal que $\gamma(0) = m$ y el vector tangente en m sea X_m , por teorema del valor medio

$$X_m \mathcal{H}_f = -2 \lim_{m \rightarrow m'} \int_x X_m \rho^2(\cdot, q) P(dx),$$

donde $X_m = X(m')$ y m' es on γ , entonces por el teorema de convergencia dominada

$$X_m \mathcal{H}_f = -2 \int_x \langle X_m, \exp_m^{-1}(f(x)) \rangle P(dx)$$

Finalmente, la continuidad de $X_m \mathcal{H}_f$ se sigue de la desigualdad (2.1) y, de nuevo, por el teorema de convergencia dominada ■

Estamos ahora en condiciones de establecer la mencionada relación entre valores medios y centros de masas

Proposición 2 1 13 Sea $(\mathcal{X}, \mathcal{A}, P)$ un espacio de probabilidad, (M, \mathfrak{A}) una variedad Riemannian completa $f: \mathcal{X} \rightarrow M$ una aplicación medible, tal que P_f este dominada por la medida Riemanniana V_R , $P_f \ll V_R$. Sea la función \mathcal{H}_f definida como

$$\mathcal{H}_f(m) = \int_{\mathcal{X}} \rho^2(m, f(x)) P(dx)$$

y que suponemos existe para todo $m \in M$. Entonces \mathcal{H}_f posee un punto crítico en $m \in M$ si y solo si $m = \mathfrak{M}(f)$

Demostración

Por la proposición anterior y la observación que sigue a la definición (2 1 6) \mathcal{H}_f tendrá un punto crítico en m si y solo si $X_m \mathcal{H}_f = 0 \quad \forall X_m \in M_m$, es decir

$$0 = X_m \mathcal{H}_f = -2 \langle X_m, \int_{\mathcal{X}} \exp_m^{-1}(f(x)) P(dx) \rangle \quad \forall X_m \in M_m,$$

lo cual se satisface si y solo si

$$\int_{\mathcal{X}} \exp_m^{-1}(f(x)) P(dx) = 0,$$

con lo que se sigue la proposición ■

Esta última proposición muestra que el *concepto de valor medio* es más débil que el *concepto de centro de masas*. Notese también que para definir el primero solo necesitamos una conexión afín, mientras que el segundo requiere una estructura métrica.

En este momento es natural preguntarse en qué condiciones existirá un valor medio. Vamos a dar condiciones suficientes para asegurar la existencia de valores medios. Previamente necesitamos algunas definiciones y resultados.

Definición 2 1 14 Sea M una variedad completa, diremos que un conjunto A es regular convexo si y solo si para cualesquiera $p, q \in A$ la geodésica más corta de p a q es única en M y se encuentra en A .

Notemos que un abierto regular convexo es un entorno regular normal de todos sus puntos.

Proposición 2 1 15 *Sea A un conjunto regular convexo en una variedad completa M . Entonces, cualquier geodesica minimal que une un punto $p \in \partial A$ y $q \in \overset{\circ}{A}$ no puede ser tangente a ∂A . Donde ∂A denota la frontera de A .*

Demostración

Sea $p \in \partial A$ el punto de tangencia de una geodesica tangente a ∂A , supongamos que exista q un punto en A , interior de A , proximo a p , unido por esta geodesica. Podemos suponer siempre q tan proximo a p como necesitemos, ya que podemos tomar p como el punto donde la linea geodesica deja la frontera de A . Sea $B_\epsilon(q)$ una bola abierta, con centro q y radio ϵ , como \exp_p^{-1} es un difeomorfismo en algun entorno de p que contiene $B_\epsilon(q)$, para $s = 1$ existira un $\delta > 0$ tal que para todo $v \in B_\delta(0_p)$

$$\tilde{q} = \exp_p(s(\exp_p^{-1}(q) + v))$$

estara en $B_\epsilon(q) \subset A$. Ahora bien, para s suficientemente pequeño, habra puntos de la geodesica mas corta que une p y \tilde{q} fuera de A , contradiciendo que A es un conjunto regular convexo. Esto es debido al hecho de que si la geodesica es tangencial a ∂A , podriamos encontrar lineas geodesicas con origen p y puntos fuera de A con vector tangente tan proximo a $\exp_p^{-1}(q)$ como quisieramos, de forma que la diferencia entre los vectores tangentes estuviese en $B_\delta(0_p)$ ■

Proposición 2 1 16 *Sea (χ, \mathfrak{a}, P) un espacio de probabilidad, (M, \mathfrak{A}) una variedad completa y $f: \chi \rightarrow M$ una aplicacion medible, sea A un conjunto regular convexo tal que $P\{f \in \overset{\circ}{A}\} = 1$. Supongamos*

$$\mathcal{H}_f(m) = \int_\chi \rho^2(m, f(x)) P(dx) < \infty$$

Entonces f tiene un valor medio $\mathfrak{M}(f) \in \overset{\circ}{A}$.

Demostración

Notemos primero que existira un conjunto compacto $C \subset \bar{A}$, donde $\bar{A} = A \cup \partial A$, tal que

$$\inf_{m \in A} \mathcal{H}_f(m) = \min_{m \in C} \mathcal{H}_f(m)$$

En caso contrario, sea $K \subset \bar{A}$ un conjunto compacto tal que $P(K) > 0$. Existirá entonces una sucesión $\{p_n\}_{n \in \mathbf{N}}$, $p_n \in \bar{A}$ tal que $\lim_{n \rightarrow \infty} \rho(p_n, K) = \infty$ y

$$\inf_{m \in A} \mathcal{H}_f(m) = \lim_{n \rightarrow \infty} \mathcal{H}_f(p_n) \geq \lim_{n \rightarrow \infty} \rho^2(p_n, K) P(K) = \infty,$$

contradiciendo que $\mathcal{H}_f(m)$ existe. Ahora, por la anterior proposición, si p pertenece a la frontera de A

$$\text{grad } \mathcal{H}_f(p) = - \int_A \exp_p^{-1}(f(x)) P(dx)$$

es un promedio de vectores señalando hacia afuera, de manera que p no será un extremo. Así el mínimo estará en el interior tal como queríamos probar. ■

El sesgo intrínseco y la distancia de Rao cuadrático media

Vamos ahora a aplicar los conceptos mencionados previamente para desarrollar medidas intrínsecas análogas al sesgo y el error cuadrático medio de un estimador

3.1 Condiciones de regularidad

Sea $\{\mathcal{X}, \mathcal{A}, P_\theta, \theta \in \Theta\}$ un modelo estadístico paramétrico, donde Θ , el espacio de parámetros, es n -variedad real C^∞ . Usualmente Θ es un conjunto abierto de \mathbb{R}^n y en este caso es costumbre usar el mismo símbolo, θ , para indicar puntos y coordenadas

Supondremos una aplicación uno a uno $\theta \mapsto p(\cdot, \theta)$ y consideraremos el conjunto de todas las medidas de probabilidad del modelo estadístico, M , con la estructura de n -variedad real C^∞ inducida por esta aplicación. Vamos a denotar esta variedad por (M, \mathfrak{A}) , siendo \mathfrak{A} el atlas inducido por las parametrizaciones, esto es las coordenadas en el espacio de parámetros

En el caso dominado, que es el que consideraremos, las medidas de probabilidad se pueden representar por funciones de densidad. Entonces supondremos, dada una medida de referencia σ finita μ , que $P_\theta \ll \mu, \forall \theta \in \Theta$ y denotaremos $p(\cdot, \theta)$ la función de densidad con respecto a μ , es decir, una cierta versión de la derivada de Radon-Nikodym $dP_\theta/d\mu$. Ahora, a través de la identificación $P_\theta \mapsto p(\cdot, \theta)$, los puntos en M pueden ser considerados bien densidades o bien medidas de probabilidad. Además, supondremos ciertas condiciones de regularidad

dad

- 1 (M, \mathfrak{A}) , es una variedad Hausdorff y conexa
- 2 Para x es fijo, la funcion real en M $\xi \mapsto p(x, \xi)$ es una funcion C^∞
- 3 Para cualquier carta local (W, θ) , las funciones de x , $\partial \log p(x, \theta) / \partial \theta^i$ $i = 1, \dots, n$, son linealmente independientes, y pertenecen a $L^\alpha(p(\cdot, \theta) d\mu)$ para un $\alpha > 0$ conveniente
- 4 Las derivadas parciales de orden suficiente

$$\partial / \partial \theta^i, \quad \partial^2 / \partial \theta^i \partial \theta^j, \quad \partial^3 / \partial \theta^i \partial \theta^j \partial \theta^k, \quad i, j, k = 1, \dots, n,$$

y la integracion con respecto a $d\mu$ de $p(x, \theta)$ se pueden intercambiar cuando sea preciso

Cuando se satisfagan todas estas condiciones, para alguna version de la funcion de densidad, diremos que el modelo estadístico paramétrico es *regular*, y en este caso la variedad (M, \mathfrak{A}) tiene una estructura Riemanniana natural, dada por su *metrica informativa*. En tal caso, tendremos tambien una conexión afin definida sobre la variedad, la conexión de Levi Civita, asociada de forma natural con el modelo estadístico. Para mas detalles, vease Amari [2], Atkinson y Mitchell [3], Barndorff Nielsen [6], Barndorff Nielsen y Blaesild [9], Burbea [13], Burbea y Rao [16], Castillo [19] y Oller [50], entre muchos otros.

3.2 Sesgo intrinseco

De esta forma, un modelo estadístico paramétrico regular puede ser visto como una variedad Riemanniana. En este contexto, un *estimador \mathcal{U} de la funcion de densidad (o medida de probabilidad) verdadera* $p_0 = p(\cdot, \theta_0) \in M$ del modelo estadístico es una familia de aplicaciones medibles

$$\mathcal{U} = \{\mathcal{U}_k \quad \chi^k \mapsto M, \quad k \in \mathbb{N}\}$$

donde la verdadera medida de probabilidad en χ^k es $(P_0)_k(dx) = p_{(\kappa)}(x, \theta_0) \mu_k(dx) = \prod_{i=1}^k p(x_i, \theta_0) \mu(dx_i)$

Observese que, dado un estimador, tendremos una sucesion de objetos aleatorios que toman valores en una variedad representativa del modelo estadístico y viceversa

Ademas, si fijamos k , a todo estimador \mathcal{U} de la funcion de densidad verdadera $p_0 = p(\cdot, \theta_0)$, le podemos asociar un C^∞ campo vectorial (campo tensorial contravariante de primer orden) inducido en la variedad a traves de la inversa, supuesta su existencia, de la aplicacion exponencial correspondiente a la conexi3n Riemanniana $A_p(x) = \exp_p^{-1}(\mathcal{U}_k(x))$

Suponiendo que $p_0 = p(\cdot, \theta_0)$, es la verdadera funcion de densidad, podemos ahora introducir la siguiente definicion

Definicion 3 2 1 *Un estimador \mathcal{U} es intrinsecamente insesgado, si y solo si, p_0 es un valor medio de \mathcal{U}_k , $\forall k \in \mathbf{N}$ para cualquier $p_0 \in M$ que sea la funcion de densidad verdadera, es decir, $\mathfrak{M}_{p_0}(\mathcal{U}_k) = p_0$, donde \mathfrak{M}_{p_0} simboliza el valor medio de \mathcal{U}_k calculado con respecto a la medida de probabilidad verdadera $(P_0)_k$*

Notemos que la definicion de estimador insesgado, al contrario que la clasica, es invariante con respecto a cualquier cambio de coordenadas o reparametrizacion

Podriamos intentar calcular los campos tensoriales aleatorios correspondientes a un campo vectorial asociado a un estimador, supuesta su existencia, y obtener, para el momento de primer orden, el campo tensorial *esperanza del estimador* Sea $p_0 = p(x, \theta_0)$ la verdadera, aunque desconocida, funcion de densidad correspondiente a la verdadera medida de probabilidad P_{θ_0} , entonces tenemos $\mathcal{E}_p = E_{p_0}(A_p) = E_{p_0}(\exp_p^{-1}(\mathcal{U}_k))$ En coordenadas, con respecto a la parametrizacion dada por $\theta^1, \dots, \theta^n$, si escribimos $p_{(k)}(x, \theta_0)$ la correspondiente μ_k funcion de densidad para una muestra aleatoria simple de tamaño k , tendremos

$$\mathcal{E}^\alpha(\theta) = \int_{\mathcal{X}^k} A^\alpha(x, \theta) p_{(k)}(x, \theta_0) \mu_k(dx) \quad \alpha = 1, \dots, n,$$

donde $A^1(x, \theta), \dots, A^n(x, \theta)$ son las componentes de $A_p(x) = \exp_p^{-1}(\mathcal{U}_k(x))$, y la dependencia en k se ha omitido en la notacion. Notese que, para todo θ , la integral se calcula siempre con respecto a la misma medida de referencia μ_k

Resulta conveniente, a fin de medir el sesgo de un estimador, introducir la siguiente definicion

Definición 3 2 2 *El campo tensorial de sesgo se define como*

$$B_p = E_p \left(\exp_p^{-1}(\mathcal{U}_k) \right),$$

o en coordenadas,

$$B^\alpha(\theta) = \int_{\mathcal{X}^k} A^\alpha(x, \theta) p_{(k)}(x, \theta) d\mu_k(x) \quad \alpha = 1, \dots, n,$$

supuesta su existencia

Es interesante reseñar que $\|B\|^2$ proporciona una medida escalar e intrínseca del sesgo. Asimismo observemos que

$$B^\alpha(\theta_0) = \mathcal{E}^\alpha(\theta_0) \quad \alpha = 1, \dots, n,$$

Obviamente tenemos el siguiente resultado

Proposición 3 2 3 *Un estimador \mathcal{U} es intrínsecamente insesgado si y solo si su campo tensorial de sesgo es nulo, esto es*

$$B^\alpha(\theta) = 0 \quad \alpha = 1, \dots, n \quad \forall \theta \in \Theta$$

Observaciones Es interesante señalar que un estimador es intrínsecamente insesgado si y solo si es *estacionario* en el sentido de Hendriks [26] si tomamos el cuadrado de la distancia de Rao como función de pérdida.

En el caso Euclideo podemos establecer una clara relación entre la ausencia de sesgo intrínseco y de sesgo clásico en cierto sistema de coordenadas,

Teorema 3 2 4 *Consideremos un modelo estadístico regular tal que la variedad de funciones de densidad es simplemente conexa y completa, y supongamos además que todas las curvaturas seccionales son cero. Entonces existe un sistema de coordenadas global $\theta^1, \dots, \theta^n$ tal que las componentes del correspondiente tensor métrico son constantes y, bajo este sistema de coordenadas, un estimador \mathcal{U} es insesgado si y solo si es intrínsecamente insesgado.*

Demostración

La existencia de un sistema de coordenadas global $\theta^1, \dots, \theta^n$ tal que las componentes del correspondiente tensor métrico son constantes es un resultado bien conocido, véase por ejemplo Kobayashi y Nomizu [37, pag 105, vol II] Entonces la conclusión se sigue de que las geodésicas son líneas rectas, y la variedad es esencialmente como \mathbb{R}^n

■

3 3 La distancia de Rao cuadrático media

La distancia Riemanniana que se obtiene a partir de la métrica informativa es conocida como la distancia de Rao. La media del cuadrado de la distancia de Rao de la estimación al verdadero valor θ , que denominamos *distancia de Rao cuadrático media*, es la versión intrínseca natural del error cuadrático medio,

Definición 3 3 1 *Dado un estimador \mathcal{U} y una muestra de tamaño k denominamos distancia de Rao cuadrático media de \mathcal{U} a la función escalar en la variedad $E_\theta(\rho^2(\mathcal{U}_k, \theta))$*

Observaciones Si consideramos funciones que dependen del modelo estadístico y no de consideraciones externas, la distancia de Rao aparece de forma natural y con las propiedades deseables, como puede apreciarse en Oller [50] y como ya se señaló en la introducción, para ser la generalización intrínseca del error cuadrático

Es interesante señalar que un estimador es intrínsecamente insesgado si y solo si es *estacionario* en el sentido de Hendriks [26] si tomamos el cuadrado de la distancia de Rao como función de pérdida

3 4 Ejemplos

Presentamos aquí algunos ejemplos, en los que calculamos el sesgo y distancia de Rao cuadrático media de diferentes estimadores

Ejemplo 3 4 1 *La distribucion exponencial unidimensional*

Consideremos la funcion de densidad exponencial parametrizada de manera que

$$p(x, \lambda) = \frac{1}{\lambda} \exp\left\{-\frac{x}{\lambda}\right\} \quad x, \lambda \in \mathbf{R}^+$$

La componente del tensor metrico viene dada por $g_{11}(\lambda) = 1/\lambda^2$ Claramente, si escribimos $\theta = \log \lambda$, la nueva componentes sera $\bar{g}_{11}(\theta) = 1$ Consideremos ahora el estimador maximo-verosimil para el parametro λ calculado para una muestra de tamaño k dado por \bar{X}_k , la media muestral ordinaria El estimador maximo-verosimil para θ vendra dado por $\log \bar{X}_k$ Puesto que el tensor metrico es constante para el sistema de coordenadas dado por θ , el tensor de sesgo, escribiendo $S = k\bar{X}_k$, sera

$$B^1(\theta) = E\left(\log\left(\frac{S}{k}\right) - \theta\right) = \int_{\mathbf{R}^+} \left(\log\left(\frac{s}{k}\right) - \theta\right) \frac{s^{k-1}}{\Gamma(k) e^{k\theta}} \exp\left\{-\frac{s}{e^\theta}\right\} ds,$$

y haciendo el cambio $u = s/e^\theta$, obtenemos

$$B^1(\theta) = \frac{1}{\Gamma(k)} \int_{\mathbf{R}^+} \log(u) u^{k-1} e^{-u} du - \log k = \frac{\Gamma'(k)}{\Gamma(k)} - \log k = \Psi(k) - \log k,$$

donde $\Psi(k) = \Gamma'(k)/\Gamma(k)$, siendo Γ la usual funcion gamma Se trata por tanto de un estimador sesgado Aunque, podemos corregir facilmente el sesgo, obteniendo en este caso un estimador intrinsecamente insesgado Con respecto a la parametrizacion dada por θ , el estimador corregido sera

$$\theta = \log \bar{X}_k - \Psi(k) + \log k,$$

y con respecto a la parametrizacion original

$$\lambda = \frac{k\bar{X}_k}{e^{\Psi(k)}}$$

Veamos ahora cual es la distancia de Rao cuadratico media de este estimador,

$$\begin{aligned} E\left(\rho^2(\theta, \theta)\right) &= \int_{\mathbf{R}^+} \left(\log\left(\frac{s}{k}\right) - \Psi(k) + \log k - \theta\right)^2 \frac{s^{k-1}}{\Gamma(k) e^{k\theta}} \exp\left\{-\frac{s}{e^\theta}\right\} ds, \\ &= \frac{1}{\Gamma(k)} \int_{\mathbf{R}^+} (\log u - \Psi(k))^2 u^{k-1} \exp\{-u\} du \\ &= \frac{1}{\Gamma(k)} \left(\Gamma''(k) - 2\Gamma'(k)\Psi(k) + \Gamma(k)\Psi(k)^2\right) \\ &= \left(\frac{\Gamma'(k)}{\Gamma(k)}\right)' = \Psi'(k) \end{aligned}$$

Es interesante observar que la familia de exponenciales resulta invariante si consideramos el grupo de transformaciones en el espacio muestral $\{\mu \in X, \mu \in \mathbf{R}^+\}$ y como el grupo inducido en el espacio de parametros es conmutativo y actua transitivamente el estimador invariante que minimiza uniformemente la distancia de Rao cuadratico media es el intrinsecamente insesgado, vease Lehmann [42] para mas detalles. No obstante en este caso es facil demostrar este resultado directamente. Consideremos un estimador invariante \mathcal{U} funcion del estadistico suficiente S , debera cumplir $\mathcal{U}(\alpha S) = \alpha \mathcal{U}(S)$ para todo $\alpha \in \mathbf{R}^+$, entonces $\mathcal{U}(S) = \beta S$ para una cierta constante β . Se trata entonces de buscar β para la cual $E(\rho^2(\beta S, \theta))$ sea minimo. Derivando e igualando a cero,

$$\frac{2}{\beta} \int_{\mathbf{R}^+} (\log(\beta s) - \theta) \frac{s^{k-1}}{\Gamma(k) e^{k\theta}} \exp\{-\frac{s}{e^\theta}\} ds = 0$$

y esto conduce obviamente a $\beta = e^{-\Psi(k)}$, es evidente tambien que se trata de un minimo.

Ejemplo 3.4.2 *La distribucion de Poisson unidimensional*

Consideremos la funcion de densidad de una Poisson unidimensional parametrizada como

$$p(x, \lambda) = e^{-\lambda} \frac{\lambda^x}{x!} \quad \lambda \in \mathbf{R}^+, x \in \mathbf{N}$$

La componente del tensor metrico viene dada por $g_{11}(\lambda) = 1/\lambda$. Es claro que si hacemos $\theta = 2\sqrt{\lambda}$, la nueva componente del tensor metrico sera $\bar{g}_{11}(\theta) = 1$. Consideremos ahora el estimador maximo-verosimil para el parametro λ obtenido para una muestra de tamaño k dado por \bar{X}_k , de nuevo la media muestral ordinaria. El correspondiente estimador maximo verosimil para θ vendra dado por $2\sqrt{\bar{X}_k}$. Como el tensor metrico es constante bajo el sistema de coordenadas dado por θ , el tensor de sesgo, escribiendo $S = k\bar{X}_k$, sera

$$B^1(\theta) = E\left(2\sqrt{\frac{S}{k}} - \theta\right) = 2\left(\sum_{j=1}^{\infty} \sqrt{\frac{j}{k}} e^{-k\frac{\theta^2}{4}} \frac{(k\frac{\theta^2}{4})^j}{j!}\right) - \theta,$$

que es claramente no nulo. Ademas, ya que la ecuacion

$$E(f(S)) = \sum_{j=0}^{\infty} f(j) e^{-k\lambda} \frac{(k\lambda)^j}{j!} = 2\sqrt{\lambda}$$

equivale a

$$\sum_{j=0}^{\infty} f(j) \frac{(k\lambda)^j}{j!} = \frac{2}{\sqrt{k}} \sqrt{k\lambda} e^{k\lambda},$$

donde f es una funcion arbitraria, no tiene solucion puesto que $\sqrt{z}e^z$ no es una funcion analitica, concluimos que para una familia de distribuciones de Poisson unidimensionales no existe un estimador intrinsecamente insesgado basado en el estadistico suficiente S

Ejemplo 3 4 3 Consideremos las distribuciones de probabilidad elipticas multi variantes, con matriz de dispersion fija $\Sigma = \Sigma_0$, que es la familia parametrica con funciones de densidad, en \mathbf{R}^n con respecto a la medida de Lebesgue, dadas por

$$p(x, \mu) = \frac{\Gamma(n/2)}{\pi^{n/2}} |\Sigma_0|^{-1/2} F\left((x - \mu)' \Sigma_0^{-1} (x - \mu)\right),$$

donde Σ_0 es una matriz fija $n \times n$ estrictamente definida positiva, $\mu = (\mu_1, \dots, \mu_n)'$ es un parametro vectorial, $\Gamma(n/2)$ es la funcion gamma, y F es una funcion no negativa en $\mathbf{R}_+ = [0, \infty)$ satisfaciendo

$$\int_0^\infty r^{n/2-1} F(r) dr = 1$$

El vector μ y la matriz Σ_0 puede expresarse en terminos de $E(X)$ y $\text{Cov}(X)$, supuesta la existencia de esta ultima. De hecho, sea $t = (t_1, \dots, t_n)'$ y $\phi_F(t) = E(\exp\{t'X\})$ la funcion caracteristica de la anterior familia de distribuciones de probabilidad, que puede expresarse como

$$\phi_F(t) = \exp\{t'\mu\} \Lambda_F(t'\Sigma_0 t),$$

donde

$$\Lambda_F(s) = \Gamma(n/2) \int_0^\infty r^{n/2-1} F(r) K_{n/2-1}(rs) dr \quad s \in \mathbf{R},$$

con

$$K_\nu(s) = 2^\nu \frac{J_\nu(\sqrt{s})}{(\sqrt{s})^\nu} = \sum_{m=0}^\infty \frac{(-s)^m}{4^m m! \Gamma(m + \nu + 1)},$$

y donde J_ν es la funcion de Bessel ordinaria de orden ν

Por tanto, formalmente

$$E(X) = -i \left. \frac{\partial \phi_F(t)}{\partial t} \right|_{t=0} \quad \text{y} \quad E(XX') = - \left. \frac{\partial^2 \phi_F(t)}{\partial t \partial t'} \right|_{t=0}$$

Esto da $E(X) = \mu$ y $E(XX') = \mu\mu' + c_F \Sigma_0$, donde

$$c_F = -2\Lambda'_F(0) = \frac{1}{n} \int_0^\infty r^{n/2} F(r) dr,$$

y de aqui

$$\text{Cov}(X) = c_F \Sigma_0$$

En particular, $E(X)$ existe si y solo si $\int_0^\infty r^{n/2-1/2} F(r) dr < \infty$, adicionalmente $\text{Cov}(X)$ existe si y solo si tenemos $\int_0^\infty r^{n/2} F(r) dr < \infty$ en cuyo caso $0 < c_F < \infty$

Una distribucion normal multivariante no degenerada $N_n(\mu, \Sigma_0)$ es un ejemplo de distribucion eliptica multivariante con

$$F(s) = \frac{1}{2^{n/2} \Gamma(n/2)} \exp\{-s/2\}, \quad \Lambda_F(s) = \exp\{-s/2\}, \quad c_F = 1$$

Otras propiedades basicas de las distribuciones de probabilidad elipticas pueden ser encontradas en Kelker [34] y son sumarizadas en Muirhead [47, pp 32 40]

Vamos a suponer ademas que

$$a = \frac{4}{n} \int_0^\infty t^{n/2} (\mathcal{L}F)^2(t) F(t) dt < \infty,$$

donde $\mathcal{L}F = F'/F$, a fin de poder asegurar la existencia de la matriz de informacion de Fisher, que vendria dada por

$$E \left(\frac{\partial \log p}{\partial \mu} \frac{\partial \log p}{\partial \mu'} \right) = a \Sigma_0^{-1},$$

vease Mitchell y Krzanowski [46] y Burbea y Oller [14] para mas detalles

Por tanto, la metrica informativa para esta familia parametrica de distribuciones de probabilidad vendra dada por

$$ds^2 = a d\mu' \Sigma_0^{-1} d\mu$$

Como el tensor metrico es constante, la variedad es Euclidea, y las geodesicas son lineas rectas. Identificando los puntos de la variedad con sus coordenadas, la geodesica que pasa por μ_0 cuando $t = 0$ y alcanza μ_1 para $t = 1$ viene dada por

$$\mu(t) = (\mu_1 - \mu_0)t + \mu_0 \quad t \in \mathbf{R},$$

y si escribimos $p_0 = p(\cdot, \mu_0)$, y $p_1 = p(\cdot, \mu_1)$, tenemos

$$\exp_{p_0}^{-1}(p_1) = (\mu_1 - \mu_0)_{p_0},$$

donde, en la ultima ecuacion, hemos identificado los vectores tangentes en p_0 con sus componentes correspondientes a la base canonica inducida por el sistema de coordenadas. Considerando el estimador para μ dado por

$$\bar{X}_k = \frac{1}{k} \sum_{i=1}^k x_i,$$

y omitiendo el subindice p_0 para los vectores tangentes en p_0 , podemos escribir

$$E_{p_0} (\bar{X}_k - \mu_0) = \frac{1}{k} \sum_{i=1}^k E_{p_0}(x_i) - \mu_0 = 0$$

Asi pues \bar{X}_k es intrinsecamente insesgado

Vamos a calcular su distancia de Rao cuadratico media

$$\begin{aligned} E(\rho^2(\bar{X}_k, \mu_0)) &= a E((\bar{X}_k - \mu_0)' \Sigma_0^{-1} (\bar{X}_k - \mu_0)) \\ &= a E(\text{tr}(\Sigma_0^{-1} (\bar{X}_k - \mu_0)(\bar{X}_k - \mu_0)')) \\ &= a \text{tr}(\Sigma_0^{-1} E((\bar{X}_k - \mu_0)(\bar{X}_k - \mu_0)')) = a c_F \text{tr}\left(\frac{1}{k} I_n\right) = a c_F \frac{n}{k} \end{aligned}$$

mas adelante veremos que $a c_F$ debera ser mayor que la unidad, e igual a la unidad si y solo si la distribucion eliptica es normal

El sencillo ejemplo que viene a continuacion muestra cuan diferente pueden llegar a funcionar, fijando una parametrizacion, la funcion de perdida proporcionada por el error cuadratico y la que se obtiene utilizando el cuadrado de la distancia de Rao

Ejemplo 3 4 4 Sea el modelo estadistico definido por la familia de densidades de Pascal

$$p(x, \theta) = (1 - \theta)^x \theta, \quad x \in \mathbf{N} \cup \{0\}, \quad \theta \in (0, 1)$$

Sea $\tilde{\theta}$ estimador insesgado en esta parametrizacion, para una muestra de tamaño $k = 1$. Entonces

$$E(\tilde{\theta}) = \sum_{x=0}^{\infty} \tilde{\theta}(x) (1 - \theta)^x \theta = \theta$$

implica que

$$\tilde{\theta}(0) = 1, \quad \tilde{\theta}(x) = 0, \quad \forall x \geq 1$$

Este es el estimador UMV insesgado para θ , pero, salvo esta importante propiedad, no parece un estimador razonable. Por otra parte, el estimador máximo-verosímil

$$\hat{\theta} = \frac{1}{x+1}, \quad x \in \mathbf{N} \cup \{0\},$$

parece más apropiado que el anterior. Sin embargo, si calculamos los errores cuadráticos medios, (ECM), a fin de comparar la precisión de estos estimadores obtenemos

$$\text{ECM}(\tilde{\theta}) = \theta(1 - \theta),$$

y

$$\begin{aligned} \text{ECM}(\theta) &= \sum_{x=0}^{\infty} \left(\frac{1}{x+1} - \theta \right)^2 (1-\theta)^x \theta = \frac{\theta}{1-\theta} \sum_{x=0}^{\infty} \frac{(1-\theta)^{x+1}}{(x+1)^2} \\ &\quad + \frac{2\theta^2}{1-\theta} \sum_{x=0}^{\infty} \frac{(1-\theta)^{x+1}}{x+1} + \theta^2 \\ &= \frac{\theta}{1-\theta} \int_{1-\theta}^0 \frac{\log(1-t)}{t} dt + \frac{2\theta^2}{1-\theta} \log \theta + \theta^2 \end{aligned}$$

Utilizando el programa *Mathematica*, versión 1.2, se obtiene que $\text{ECM}(\tilde{\theta}) - \text{MSE}(\theta)$ es una función positiva en $(\theta_0, 1)$ y negativa en $(0, \theta_0)$, donde aproximadamente $\theta_0 = 0.1606$.

En este sentido la función de pérdida “error cuadrático” no distingue claramente entre estos dos estimadores. De hecho, se puede ver que $\tilde{\theta}$ es un estimador admisible con respecto a la pérdida “error cuadrático”. En cambio si usamos el cuadrado de la distancia de Rao como función de pérdida obtendremos,

$$E \left(\rho^2(\tilde{\theta}, \theta) \right) = \infty$$

donde

$$\rho(\theta, \eta) = 2 \log \frac{1 + |\sqrt{1-\theta} - \sqrt{1-\eta}| - \sqrt{(1-\theta)(1-\eta)}}{\sqrt{\theta\eta}},$$

Por tanto $\tilde{\theta}$ es, con respecto a la nueva pérdida, un estimador inadmisibles. Además, como el lector puede comprobar fácilmente,

$$E(\rho^2(\theta, \theta)) = \sum_{x=0}^{\infty} \left\{ 2 \log \frac{1 + \left| \sqrt{1 - \frac{1}{x+1}} - \sqrt{1 - \theta} \right| - \sqrt{\left(1 - \frac{1}{x+1}\right)(1 - \theta)}}{\sqrt{\frac{\theta}{x+1}}} \right\}^2 (1 - \theta)^x \theta < \infty \quad \forall \theta > 0$$

que muestra la superioridad del estimador máximo verosímil

Comportamiento local de un estimador

4.1 Cota inferior de la distancia de Rao cuadrático media

En esta sección se estudia la relación entre el sesgo y la distancia de Rao cuadrático media entre la densidad, o la medida de probabilidad, estimada y la verdadera, obteniendo una versión intrínseca de la cota Cramer Rao, y cotas inferiores basadas en los teoremas de comparación de geometría Riemanniana, véase (11.3) del apéndice. Algunos resultados análogos aunque en un enfoque diferente pueden encontrarse en Hendriks [26].

Antes de nada es conveniente tener aquí presentes ciertas definiciones y resultados de geometría diferencial.

Denotemos por $\mathfrak{S}_p = \{\xi \in M_p \mid \|\xi\| = 1\}$, y para cada $\xi \in \mathfrak{S}_p$ definamos

$$C_p(\xi) = \sup\{s > 0 \mid \rho(p, \gamma_\xi(s)) = s\},$$

donde ρ es la distancia Riemanniana y γ_ξ es una geodésica definida en un intervalo abierto conteniendo el cero, tal que $\gamma_\xi(0) = p$ y con vector tangente en el origen ξ . Entonces si hacemos

$$\mathfrak{D}_p = \{s\xi \in M_p \mid 0 \leq s < C_p(\xi), \xi \in \mathfrak{S}_p\}$$

y

$$D_p = \exp_p(\mathfrak{D}_p),$$

es sabido que \exp_p aplica \mathfrak{D}_p difeomórficamente sobre D_p , véase (11.2) del apéndice. De hecho D_m es el entorno regular maximal de m en el sentido que cualquier otro entorno regular m está incluido en el

Asimismo conviene recordar la definicion de curvatura seccional

Definicion 4 1 1 *Dados dos campos X, Y en M tales que para todo $p \in M$ generan un subespacio de dimension 2, la curvatura Riemanniana seccional, K , se define como la funcion escalar*

$$K(X, Y) = \langle R(X, Y)X, Y \rangle / (\langle X, X \rangle \langle Y, Y \rangle - \langle X, Y \rangle^2),$$

donde $R(X, Y)$, es el operador (la curvatura)

$$R(X, Y) = \nabla_Y \nabla_X - \nabla_X \nabla_Y - \nabla_{[Y, X]}, \quad [Y, X] = YX - XY$$

y ∇ la conexion Riemanniana de M

En el caso bidimensional coincide con la *curvatura Gaussiana* de una superficie. En el caso general la curvatura seccional en el punto $p \in M$ es la curvatura Gaussiana de la superficie generada por las geodesicas que parten de p y son tangentes al subespacio bidimensional generado por $X(p)$ e $Y(p)$ en M_p . Si $n = 1$ adoptamos el convenio $K = 0$

Teorema 4 1 2 (Cota inferior intrinseca de Cramer-Rao) *Sea \mathcal{U} un estimador correspondiente a una familia parametrica n dimensional regular de funciones de densidad. Supongamos que $\nu_k(M \setminus D_p) = 0 \quad \forall p \in M$, donde ν_k es la medida de probabilidad inducida por \mathcal{U}_k y $p = p(x, \theta)$, en la variedad. Sea A el campo tensorial asociado al estimador y sea B el correspondiente tensor de sesgo, $B = E(A)$. Supongamos que la distancia de Rao cuadratica media entre la verdadera densidad p y una estimacion correspondiente a una muestra de tamaño k , $E(\rho^2(\mathcal{U}_k, p))$, existe para todo $k \in \mathbb{N}$, y que la derivada covariante de $E_p(A_p)$ existe y se puede obtener derivando bajo el signo integral. Entonces,*

1 *En general tenemos*

$$E(\|A\|^2) = E(\rho^2(\mathcal{U}_k, p)) \geq \frac{\{\text{div}(B) - E(\text{div}(A))\}^2}{kn} + \|B\|^2,$$

donde $\text{div}(\)$ consiste en el operador divergencia

2 *Si todas las curvaturas seccionales son cero, $K = 0$, entonces*

$$E(\rho^2(\mathcal{U}_k, p)) \geq \frac{(\text{div}(B) + n)^2}{kn} + \|B\|^2$$

3 Si todas las curvaturas seccionales son no positivas, $K \leq \mathcal{K} < 0$ y $-n < \operatorname{div}(B)$, entonces

$$E\left(\rho^2(\mathcal{U}_k, p)\right) \geq \frac{(\operatorname{div}(B) + n)^2}{kn} \mathcal{S}_n^2 + \|B\|^2,$$

donde

$$\mathcal{S}_n = 1 + \frac{(n-1) \left(\sqrt{-\mathcal{K}} \|B\| \coth \left(\sqrt{-\mathcal{K}} \|B\| \right) - 1 \right)}{\operatorname{div}(B) + n}$$

4 Si todas las curvaturas seccionales son menores o iguales que una constante positiva \mathcal{K} , y $d(M) < \pi/2\sqrt{\mathcal{K}}$, siendo $d(M)$ el diametro de la variedad, y $-1 \leq \operatorname{div}(B)$, entonces

$$E\left(\rho^2(\mathcal{U}_k, p)\right) \geq \frac{(\operatorname{div}(B) + n)^2}{kn} \mathcal{T}_n^2 + \|B\|^2,$$

donde

$$\mathcal{T}_n = \frac{2\sqrt{1 - \frac{4(n-1)\|B\|^2\mathcal{K}}{(n + \operatorname{div}(B))\pi^2}}}{1 + \sqrt{1 + 16(n-1)\mathcal{K} \frac{(n + \operatorname{div}(B) - 4(n-1)\frac{\|B\|^2\mathcal{K}}{\pi^2})}{kn\pi^2}}}$$

En particular, para estimadores intrinsecamente insesgados, tenemos

5 Si todas las curvaturas seccionales son no positivas, entonces

$$E\left(\rho^2(\mathcal{U}_k, p)\right) \geq \frac{n}{k}$$

6 Si todas curvaturas seccionales son menores o iguales que una constante positiva \mathcal{K} y $d(M) < \pi/2\sqrt{\mathcal{K}}$, entonces

$$E\left(\rho^2(\mathcal{U}_k, p)\right) \geq \frac{4}{\left(1 + \sqrt{1 + 16(n-1)\mathcal{K}/(k\pi^2)}\right)^2} \frac{n}{k},$$

donde las esperanzas, en cada punto p , estan calculadas con respecto a la medida de probabilidad correspondiente $(P)_k$

Demostracion

Sean $A^\alpha(x, \theta)$, $\alpha = 1, \dots, n$, las componentes de $\exp_p^{-1}(\mathcal{U}_k)$, $B^\alpha = E(A^\alpha)$ y $C^\alpha(x, \theta)$, $\alpha = 1, \dots, n$, las componentes de un tensor contravariante de primer orden. Entonces, por la desigualdad de Cauchy-Schwartz,

$$|\langle A - B, C \rangle| \leq \|A - B\| \|C\|,$$

donde $\langle \cdot, \cdot \rangle$ y $\| \cdot \|$ simbolizan el producto y la norma definida en cada espacio tangente. Además,

$$E(|\langle A - B, C \rangle|) \leq E(\|A - B\| \|C\|) \leq \sqrt{E(\|A - B\|^2)} \sqrt{E(\|C\|^2)},$$

de nuevo por la desigualdad de Cauchy-Schwartz, y donde las esperanzas, en cada punto p , son calculadas con respecto a la medida de probabilidad correspondiente $p_{(k)} d\mu_k$.

Sea $C(x, \theta) = \text{grad}(\log p_{(k)}(x, \theta))$, donde $\text{grad}(\cdot)$ es el operador gradiente. En coordenadas, y utilizando el convenio de sumacion de indices repetidos, podemos escribir

$$C^\alpha(x, \theta) = g^{\alpha\beta}(\theta) \frac{\partial \log p_{(k)}(x, \theta)}{\partial \theta^\beta},$$

donde $g^{\alpha\beta}(\theta)$ son las componentes del campo tensorial contravariante fundamental y donde p es la funcion de densidad conjunta. Por tanto tenemos

$$\|C\|^2 = g^{\alpha\beta} \frac{\partial \log p_{(k)}}{\partial \theta^\beta} \frac{\partial \log p_{(k)}}{\partial \theta^\alpha},$$

tomando esperanzas, y usando notacion matricial,

$$\begin{aligned} E(\|C\|^2) &= E(C'G^{-1}C) = E(\text{tr}(C'G^{-1}C)) = E(\text{tr}(G^{-1}CC')) = \\ &= \text{tr}(G^{-1}E(CC')) = k \text{tr}(G^{-1}G) = k \text{tr}I_n = kn \end{aligned}$$

Por otra parte tenemos tambien

$$|E(\langle A, C \rangle)| = |E(\langle A - B, C \rangle)| \leq E(|\langle A - B, C \rangle|)$$

y

$$E(\|A - B\|^2) = E(\|A\|^2) - \|B\|^2$$

De manera que

$$|E(\langle A, C \rangle)| \leq \sqrt{E(\|A\|^2) - \|B\|^2} \sqrt{k n},$$

pero $\|A\|^2 = \rho^2(p, \mathcal{U}_k)$, donde ρ es la distancia Riemanniana y en este caso la distancia de Rao. Entonces

$$|E(\langle A, C \rangle)| \leq \sqrt{E(\rho^2(p, \mathcal{U}_k)) - \|B\|^2} \sqrt{k n} \quad (4.1)$$

Por otro lado

$$\langle A, C \rangle = g_{\alpha\beta} A^\alpha C^\beta = g_{\alpha\beta} A^\alpha g^{\beta\gamma} \frac{\partial \log p_{(k)}}{\partial \theta^\gamma} = A^\alpha \frac{\partial \log p_{(k)}}{\partial \theta^\alpha},$$

por tanto,

$$E(\langle A, C \rangle) = \int_{\mathcal{X}^k} A^\alpha \frac{\partial \log p_{(k)}}{\partial \theta^\alpha} p_{(k)} d\mu_k = \int_{\mathcal{X}^k} A^\alpha \frac{\partial p_{(k)}}{\partial \theta^\alpha} d\mu_k$$

Notemos que $A^\alpha \partial p_{(k)} / \partial \theta^\alpha$ es una función de x que es independiente del sistema de coordenadas. Para x fija es una función escalar en la variedad.

Adicionalmente, puesto que

$$\int_{\mathcal{X}^k} A^\alpha p_{(k)} d\mu_k = B^\alpha \quad \alpha = 1, \dots, n,$$

tomando la derivada covariante obtenemos,

$$\int_{\mathcal{X}^k} \left\{ \frac{\partial A^\alpha}{\partial \theta^i} + \Gamma_{ij}^\alpha A^j \right\} p_{(k)} d\mu_k + \int_{\mathcal{X}^k} A^\alpha \frac{\partial p_{(k)}}{\partial \theta^i} d\mu_k = \frac{\partial B^\alpha}{\partial \theta^i} + \Gamma_{ij}^\alpha B^j$$

donde Γ_{ij}^α son los símbolos de Christoffel de segunda especie.

Si contraemos índices obtenemos la ecuación escalar

$$\int_{\mathcal{X}^k} \left\{ \frac{\partial A^\alpha}{\partial \theta^\alpha} + \Gamma_{\alpha j}^\alpha A^j \right\} p_{(k)} d\mu_k + \int_{\mathcal{X}^k} A^\alpha \frac{\partial p_{(k)}}{\partial \theta^\alpha} d\mu_k = B^{\alpha, \alpha}$$

ahora bien la traza de la derivada covariante de un campo es por definición su divergencia (vease (11.1) del apéndice y (4.2) para más detalle), de modo que

$$E(\operatorname{div}(A)) + \int_{\mathcal{X}^k} A^\alpha \frac{\partial p_{(k)}}{\partial \theta^\alpha} d\mu_k = \operatorname{div}(B),$$

esto es,

$$E(\langle A, C \rangle) = \operatorname{div}(B) - E(\operatorname{div}(A))$$

y sustituyendo en (4.1) obtenemos 1

Veamos ahora como obtener diferentes cotas dependiendo de las curvaturas seccionales de M

Fijado x , vamos a escoger un sistema de coordenadas conveniente. Dado p y $\mathcal{U}_k(x)$, tomamos un sistema de coordenadas geodesicas esfericas con origen $\mathcal{U}_k(x)$, es decir un sistema (ρ, u) como el discutido en 11.5 del apendice, y definido casi seguramente, ya que $\nu_k(M \setminus D_{\mathcal{U}_k(x)}) = 0$

Es claro que las componentes del tensor A son $(-\rho, 0, 0, \dots, 0)$ cuando ρ , la distancia Riemanniana entre p y $\mathcal{U}_k(x)$, es la primera coordenada. Asimismo,

$$\frac{\partial A^\alpha}{\partial \theta^\alpha} = -1 \quad \text{and} \quad \Gamma_{\alpha_j}^\alpha A^j = -\rho \Gamma_{\alpha 1}^\alpha = -\frac{\partial \log \sqrt{g}}{\partial \rho} \rho,$$

donde g representa el determinante del tensor metrico. Entonces

$$\int_{\mathcal{X}^k} A^\alpha \frac{\partial p^{(k)}}{\partial \theta^\alpha} d\mu_k = \text{div}(B) + \int_{\mathcal{X}^k} \left\{ 1 + \rho \frac{\partial \log \sqrt{g}}{\partial \rho} \right\} p^{(k)} d\mu_k,$$

Consideremos diferentes casos

Caso 4.1.2.1 *Curvaturas seccionales nulas*

Como corolario del teorema de comparacion de Bishop, vease teorema (11.5.2) del apendice, o por calculos directos, tenemos

$$\frac{\partial \log \sqrt{g}}{\partial \rho} = \frac{n-1}{\rho},$$

dando lugar a

$$\int_{\mathcal{X}^k} A^\alpha \frac{\partial p^{(k)}}{\partial \theta^\alpha} d\mu_k = \text{div}(B) + n$$

Tenemos entonces

$$|\text{div}(B) + n| \leq \sqrt{E(\rho^2(\mathcal{U}_k, p)) - \|B\|^2} \sqrt{kn},$$

de lo que se deduce que

$$E(\rho^2(\mathcal{U}_k, p)) \geq \frac{(\text{div}(B) + n)^2}{kn} + \|B\|^2$$

Caso 4 1 2 2 *Todas las curvaturas seccionales son menores que cero, $K < \mathcal{K} < 0$ y $-n \leq \operatorname{div}(B)$*

Por (11.8) en la subseccion (11.5) del apéndice, tenemos

$$\frac{\partial \log \sqrt{g}}{\partial \rho} \geq (n-1)\sqrt{-\mathcal{K}} \coth(\sqrt{-\mathcal{K}}\rho),$$

originando que

$$\begin{aligned} \int_{X^k} A^\alpha \frac{\partial p_{(k)}}{\partial \theta^\alpha} d\mu_k &\geq \\ \operatorname{div}(B) + \int_{X^k} \left\{ 1 + (n-1)\sqrt{-\mathcal{K}}\|A\| \coth(\sqrt{-\mathcal{K}}\|A\|) \right\} p_{(k)} d\mu_k & \\ \geq \operatorname{div}(B) + 1 + (n-1)\sqrt{-\mathcal{K}}\|B\| \coth(\sqrt{-\mathcal{K}}\|B\|), & \end{aligned}$$

donde la segunda desigualdad es debida al hecho que la funcion $u \coth u$ es funcion convexa, y podemos aplicar la desigualdad de Jensen, y que $\|B\| \leq E\|A\|$, por la desigualdad de Cauchy-Schwartz. Por tanto si $-n \leq \operatorname{div}(B)$, como $1 < u \coth u$, $\forall u > 0$, tendremos que el segundo miembro de la desigualdad anterior es positivo y por tanto

$$E(\rho^2(\mathcal{U}_k, p)) \geq \frac{\left\{ \operatorname{div}(B) + 1 + (n-1)\sqrt{-\mathcal{K}}\|B\| \coth(\sqrt{-\mathcal{K}}\|B\|) \right\}^2}{kn} + \|B\|^2$$

Caso 4 1 2 3 *Todas las curvaturas seccionales son positivas y menores que una constante positiva $K \leq \mathcal{K}$, $d(M) < \pi/2\sqrt{\mathcal{K}}$, y $-1 \leq \operatorname{div}(B)$*

Por 11.9 en la subseccion 11.5 del apéndice tenemos

$$\frac{\partial \log \sqrt{g}}{\partial \rho} \geq (n-1)\sqrt{\mathcal{K}} \cot(\rho\sqrt{\mathcal{K}}),$$

de lo cual

$$\int_{X^k} A^\alpha \frac{\partial p_{(k)}}{\partial \theta^\alpha} d\mu_k \geq \operatorname{div}(B) + \int_{X^k} \left(1 + (n-1)\sqrt{\mathcal{K}}\rho \cot(\rho\sqrt{\mathcal{K}}) \right) p_{(k)} d\mu_k,$$

ahora bien $|u \cot u| \geq 1 - 4u^2/\pi^2$, $0 < u \leq \pi/2$, y por tanto, teniendo en cuenta que $0 \leq \rho \leq \pi/2\sqrt{\mathcal{K}}$, tenemos

$$\int_{X^k} A^\alpha \frac{\partial p_{(k)}}{\partial \theta^\alpha} d\mu_k \geq \operatorname{div}(B) + n - (n-1)\frac{4\mathcal{K}E(\rho^2)}{\pi^2},$$

ya que $(n-1)\frac{4\mathcal{K}E(\rho^2)}{\pi^2} - n \leq -1 \leq \operatorname{div}(B)$, el segundo miembro de la desigualdad anterior es positivo y por tanto

$$\left(\operatorname{div}(B) + n - (n-1)\frac{4\mathcal{K}E(\rho^2)}{\pi^2} \right)^2 \leq \left(E(\rho^2(\mathcal{U}_k, p)) - \|B\|^2 \right) kn,$$

finalmente, resolviendo la inecuacion,

$$E(\rho^2(\mathcal{U}_k, p)) \geq \frac{4(n + \operatorname{div}(B) - 4(n-1)\|B\|^2\mathcal{K}/\pi^2)^2}{\left(\sqrt{kn} + \sqrt{kn + 16(n-1)\mathcal{K}\frac{(n + \operatorname{div}(B) - 4(n-1)\|B\|^2\mathcal{K}/\pi^2)}{\pi^2}} \right)^2} + \|B\|^2$$

Los casos 5 y 6 se siguen trivialmente de los casos 2, 3 y 4, con $\operatorname{div}(B) = 0$ y $\|B\| = 0$ ■

Observaciones Notemos que todas las variedades unidimensionales correspondientes a familias uniparametricas de distribuciones de probabilidad son siempre Euclideas. Ademas, existen algunas familias bien conocidas de distribuciones de probabilidad que satisfacen las hipotesis del teorema anterior, como la multinomial, vease Atkinson y Mitchell [3], distribucion multinomial negativa, vease Oller y Cuadras [51], o distribuciones de valores extremos, vease Oller [49], entre muchas otras.

Asimismo, es facil comprobar que en caso normal multivariante, con matriz de covariancia conocida, la media muestral es un estimador que alcanza la cota inferior intrinseca de Cramer-Rao,

$$\begin{aligned} E(\rho^2(\bar{X}_k, \mu_0)) &= E\left((\bar{X}_k - \mu_0)' \Sigma^{-1} (\bar{X}_k - \mu_0)\right) \\ &= E\left(\operatorname{tr}(\Sigma^{-1} (\bar{X}_k - \mu_0)(\bar{X}_k - \mu_0)')\right) \\ &= \operatorname{tr}\left(\Sigma^{-1} E\left((\bar{X}_k - \mu_0)(\bar{X}_k - \mu_0)'\right)\right) = \operatorname{tr}\left(\frac{1}{k} I_n\right) = \frac{n}{k} \end{aligned}$$

Observemos tambien el efecto de las curvaturas seccionales en la precision de las estimaciones estadisticas. Por ultimo, como la distancia de Rao cuadratico media esta acotada superiormente por $d(M)^2$, donde $d(M)$ es el diametro de la variedad, resulta que

Proposicion 4.1.3 *Una condicion necesaria para poder tener un estimador in trinsecamente insesgado es $d(M) \geq \frac{1}{\sqrt{kn}}$*

Demostracion

Resulta inmediato del hecho de que $|\text{div}(A)| \geq 1$, en efecto para un estimador insesgado

$$\sqrt{kn} d(M) \geq \sqrt{kn E(\rho^2)} \|E(\text{div}(A))\| \geq 1$$

■

4.2 Desigualdad tensorial

A fin de poder expresar de forma tensorial las desigualdades que vienen a continuacion creemos conveniente recordar ciertas operaciones sobre tensores o campos tensoriales, cuyas definiciones y propiedades pueden encontrarse en Hicks [27]

En primer lugar dada una metrica Riemanniana tenemos una aplicacion lineal no singular G de M_m en M_m , dual de M_m , para cada $m \in M$, de manera que si $v_1, v_2 \in T_m$ $G(v_1)(v_2) \equiv \langle v_1, v_2 \rangle$. Designaremos por G la aplicacion inversa de G para todo $w \in M_m^*$ $\langle G(w), v_2 \rangle \equiv w(v_2)$. Podemos definir entonces las operaciones de “subir y bajar” indices

Definicion 4.2.1 Si $1 \leq i \leq p$, $1 \leq j \leq q + 1$ y T_q^p definimos la operacion de bajar el indice i a la posicion j como la que origina el tensor $G^{i,j}T$ de T_{q+1}^{p-1} tal que

$$(G^{i,j}T)(w_1, \dots, w_{p-1}, v_1, \dots, v_{q+1}) = T(w_1, \dots, w_{i-1}, G(v_j), w_i, \dots, w_{p-1}, v_1, \dots, v_j, \dots, v_{q+1})$$

donde $w_i \in M_m$ y $v_j \in M_m$ para todo i y todo j y donde v_j significa que falta v_j . Analogamente se define la operacion de subir el indice i a la posicion j , $G_*^{i,j}$

Definamos tambien la operacion de contraccion de un tensor

Definicion 4 2 2 Dado $T \in T_q^p$ definimos la operacion de contraccion del indice contravariante i con el contravariante j como la que origina el tensor $Tr^{i j}(T) \in T_{q-1}^{p-1}$ definido por

$$\left(Tr^{i j}(T)\right)\left(w_1, \dots, w_{p-1}, v_1, \dots, v_{p-1}\right) = \sum_{k=1}^n T\left(w_1, \dots, w_{i-1}, e_k, w_{i+1}, \dots, w_{p-1}, v_{j-1}, e_k, v_{j+1}, \dots, v_{p-1}\right)$$

donde $e_k, k = 1, \dots, n$ constituye una base de M_m y $e_k, k = 1, \dots, n$ su dual

Se comprueba que esta definicion no depende de la base elegida. Esta operacion se puede generalizar de forma obvia a $Tr^{i_1 i_2 \dots i_r j_1 j_2 \dots j_r}$ como la operacion que permite obtener un tensor al contraer los indices covariantes (contravariantes) i_1, i_2, \dots, i_r , con los contravariantes (covariantes) j_1, j_2, \dots, j_r . Señalar tambien que si $T \in T_1^1$ entonces $Tr^{1 1}(T)$ acostumbra a llamarse *traza* de T .

Como ya se indico en la seccion 2, dado un estimador \mathcal{U} podemos definir sus momentos de cualquier orden. A partir de ellos podemos construir el siguiente tensor

Definicion 4 2 3 El tensor de varianzas-covarianzas de un estimador \mathcal{U} , que escribiremos $Cov(\mathcal{U})$ se define como el momento centrado de segundo orden de \mathcal{U} , esto es

$$Cov(\mathcal{U})(p) = E_p((A - B) \otimes (A - B)) \quad \forall p \in M$$

donde $A = \exp_p^{-1}(\mathcal{U})$ y $B = E_p(A)$

Notese que

$$Tr^{1 2}\left(G^{2 2}Cov(\mathcal{U})\right) = E(\|A\|^2) - \|B\|^2$$

Con cierto abuso del lenguaje podemos decir que la traza de la covarianza de \mathcal{U} es la varianza de \mathcal{U} .

Tenemos el siguiente teorema para la covarianza de un estimador

Teorema 4 2 4 Con las condiciones del teorema (4 1 2) tenemos que la covarianza de un estimador \mathcal{U} esta acotada inferiormente, en el sentido de que la

diferencia con la cota es un tensor definido positivo La desigualdad viene dada por

$$\text{Cov}(\mathcal{U}) \geq \text{Tr}^{2^4} \left[G^{2^2} [D B - E(D A)] \otimes [D B - E(D A)] \right]$$

donde D es el operador derivada general covariante (vease apendice (11 1))

Demostracion

Consideremos, para un tamaño muestral k , el vector

$$Y = A - B - \frac{1}{k} \text{Tr}^{2^3} [E(A \otimes G(C)) \otimes C],$$

donde $C = \overline{\text{grad}(\log p_{(k)})}$ y $G(C)$ es la version covariante de C (vease las definiciones anteriores) Entonces como $E(Y \otimes Y) \geq 0$ tendremos

$$\begin{aligned} & E[(A - B) \otimes (A - B)] - \frac{1}{k} E \left[A \otimes \text{Tr}^{2^3} [E(A \otimes G(C)) \otimes C] \right] \\ & - \frac{1}{k} E \left[\text{Tr}^{2^3} [E(A \otimes G(C)) \otimes C] \otimes A \right] \\ & + \frac{1}{k^2} E \left[\text{Tr}^{2^3} [E(A \otimes G(C)) \otimes C] \otimes \text{Tr}^{2^3} [E(A \otimes G(C)) \otimes C] \right] \geq 0 \end{aligned}$$

de modo que, por las propiedades del operador traza

$$\begin{aligned} & E[(A - B) \otimes (A - B)] - \frac{2}{k} \text{Tr}^{3^4} E \left[A \otimes [E(A \otimes G(C)) \otimes C] \right] \\ & + \frac{1}{k^2} \text{Tr}^{2^3 4^5} [E(A \otimes G(C)) \otimes E(C \otimes C) \otimes E(G(C) \otimes A)] \geq 0 \end{aligned}$$

Ahora bien, como se comprueba facilmente tomando coordenadas

$$\text{Tr}^{1^2 3^4} [G(C) \otimes E(C \otimes C) \otimes E(G(C) \otimes A)] = k \text{Tr}^{1^3} C \otimes E(G(C) \otimes A)$$

de manera que

$$\begin{aligned} \text{Cov}(\mathcal{U}) & \geq \frac{1}{k} \text{Tr}^{2^4} [E(A \otimes C) \otimes E(A \otimes G(C))] \\ & = \frac{1}{k} \text{Tr}^{2^4} \left[G^{2^2} E(A \otimes G(C)) \otimes E(A \otimes G(C)) \right] \end{aligned}$$

Por ultimo como $G(C) = D \log p_{(k)}$ y $D B = E(D A) + E(A \otimes D \log p_{(k)})$, sustituyendo en la expresion anterior obtenemos el resultado deseado ■

Notemos que esta desigualdad coincide con la obtenida por Hendriks [26], teorema 3 2, para el caso $B = 0$, si tomamos el cuadrado de la distancia de Rao como funcion de perdida y la aplicacion de Hendriks ϕ como la identidad en M . Nuestro marco nos permite en cambio tratar el caso sesgado

4 3 La condicion de eficiencia

Analogamente a como se define la eficiencia clasicamente, tendremos la siguiente definicion de eficiencia intrinseca

Definicion 4 3 1 Diremos que un estimador \mathcal{U} intrinsecamente insesgado es intrinsecamente eficiente si su distancia de Rao cuadratico media alcanza la cota de Cramer–Rao, esto es

$$E(\rho^2(\mathcal{U}, \theta)) = \frac{n}{k}$$

Aunque el teorema 4 1 2 permite deducir en que condiciones se alcanzara la cota de Cramer–Rao, vamos a deducir la condicion de eficiencia de forma que se vean mas claramente todas las constantes en juego. El resultado toma la forma de la siguiente proposicion

Proposicion 4 3 2 Un estimador intrinsecamente insesgado \mathcal{U} sera eficiente si y solo si

$$A = -\frac{1}{kn}E(\operatorname{div}(A))C$$

donde $A(p) = \exp_p^{-1}(\mathcal{U})$ y $C = \operatorname{grad}(\log p_{(k)})$

Demostracion

Consideremos un estimador intrinsecamente insesgado \mathcal{U} y su campo vectorial asociado A . Entonces, teniendo en cuenta que $E(\langle A, C \rangle) = -\operatorname{div}(A)$,

$$\begin{aligned} E(\|A + \frac{1}{kn}E(\operatorname{div}(A))C\|^2) &= E(\|A\|^2) + \frac{2}{kn}E(\operatorname{div}(A))E(\langle A, C \rangle) \\ &\quad + \frac{1}{(kn)^2}E(\operatorname{div}(A))^2E(\|C\|^2) \\ &= E_p(\rho^2(\mathcal{U}, p)) - \frac{1}{kn}(E(\operatorname{div}(A)))^2 \geq 0 \end{aligned}$$

de manera que habra igualdad en el teorema 4.1.2 si $A = -\frac{1}{kn} E(\text{div}(A))C$ ■

Vamos a ver que ademas se debera cumplir que $E(\text{div}(A))$ sea independiente del punto $p \in M$ para que la igualdad pueda ser cierta. El resultado lo podemos expresar de la siguiente forma

Proposicion 4.3.3 *Una familia admite a \mathcal{U} como estimador eficiente si su densidad respecto de alguna medida de referencia se puede expresar*

$$p(x, \theta)_{(k)} = e^{-\frac{kn}{2\lambda} \rho^2(\mathcal{U}(x), \theta)}$$

donde λ es una constante positiva

Demostracion

La condicion de eficiencia la podemos escribir

$$\text{grad}(\rho^2(\mathcal{U}(x), \theta)) = 2\frac{1}{kn} E(\text{div}(A)) \text{grad}(\log p_{(k)}(x, \theta))$$

Vamos a considerar primero el caso unidimensional. En este caso $E(\text{div}(A)) = -1$ y resulta

$$\text{grad} \left\{ \rho^2(\mathcal{U}(x), \theta) + \frac{2}{k} \log p_{(k)}(x, \theta) \right\} = 0$$

de donde

$$p_{(k)}(x, \theta) = e^{-\frac{k}{2} \rho^2(\mathcal{U}(x), \theta)} h(x)$$

En el caso multimensional la condicion de eficiencia implicaria, derivandola, que

$$\frac{\partial E(\text{div}(A))}{\partial \theta^i} = \mu(\theta) \frac{\partial \log p_{(k)}(x, \theta)}{\partial \theta^i} \quad i = 1, \dots, n$$

y puesto que $E(\text{div}(A))$ no puede depender de x , $\mu(\theta) = 0$ y tampoco dependera de θ . De esta forma, si escribimos $E(\text{div}(A)) = -\lambda$ (de los teoremas de comparacion se deduce que $\lambda > 0$, vease apendice (11.3)) tendremos

$$\text{grad}(\rho^2(\mathcal{U}(x), \theta)) = -2\frac{\lambda}{kn} \text{grad}(\log p_{(k)}(x, \theta))$$

Asi que

$$p_{(k)}(x, \theta) = e^{-\frac{kn}{2\lambda} \rho^2(\mathcal{U}(x), \theta)} h(x)$$

■

Observacion En el caso unidimensional podemos encontrar un sistema de coordenadas o parametrización tal que

$$p_{(k)}(x, \theta) = e^{-\frac{k}{2}(\phi(\mathcal{U}(x)) - \phi(\theta))^2} h(x),$$

de manera que el estadístico suficiente $\phi(\mathcal{U}(x)) \sim N(\phi(\theta), 1/k)$

En el caso n -dimensional podemos afirmar lo siguiente

Proposición 4.3.4 *Supongamos que el espacio muestral $\Omega = \mathbf{R}^k$ que para $k > n$ existe $\mathcal{U}(x_1, \dots, x_k)$ estimador intrínsecamente eficiente de $\theta = (\theta_1, \dots, \theta_n)$ y que las densidades de la familia (regular) son C^1 con respecto a las observaciones, entonces existe una transformación suficiente de los datos tal que la familia resultante es normal con matriz de varianzas-covarianzas conocida*

Demostración

La condición de eficiencia

$$\text{grad}(\log p_{(k)}(x_1, \dots, x_k, \theta)) = \frac{nk}{\lambda} \exp^{-1}(\mathcal{U}(x_1, \dots, x_k))$$

implica que $\mathcal{U}(x_1, \dots, x_k)$ es un estadístico suficiente de dimensión $n < k$, de aquí se deduce que la familia es exponencial de orden $\leq n$, véase Barndorff-Nielsen and Pedersen [10], y por construcción el orden debe ser exactamente n y la familia completa (no curvada). De esta manera existirán n funciones ϕ_i , una función η y un sistema de coordenadas global tal que

$$\log p_{(k)}(x_1, \dots, x_k, \theta) = \sum_{i=1}^n \theta^i \phi_i(\mathcal{U}(x_1, \dots, x_k)) + \eta(\theta)$$

de aquí

$$[\exp^{-1}(\mathcal{U})]^i = \frac{\lambda}{nk} g^{ii}(\theta) ((\phi_i(\mathcal{U}) - \phi_i(\theta)))$$

con lo que

$$\rho^2(\mathcal{U}, \theta) = \frac{\lambda^2}{(nk)^2} ((\phi_i(\mathcal{U}) - \phi_i(\theta))) g^{ii}(\theta) ((\phi_i(\mathcal{U}) - \phi_i(\theta)))$$

de modo que por simetría $g^{il}(\theta)$ no depende en realidad de θ . De este modo la familia es euclídea y $\lambda = n$. Entonces resultará que el estadístico suficiente

$$(\phi_1(\mathcal{U}), \dots, \phi_n(\mathcal{U})) \sim N\left(\phi(\theta), \left(\frac{g^{il}}{k}\right)\right)$$

con $\phi(\theta) = (\phi_1(\theta), \dots, \phi_n(\theta))'$ ■

Observación Como vemos la condición de eficiencia resulta demasiado restrictiva para estudiar la bondad de diferentes estimadores. Es por ello que será necesario, como lo haremos en capítulos posteriores, estudiar el comportamiento asintótico de los mismos y en particular su eficiencia asintótica.

Eficiencia global de un estimador

Cualquiera que sea la función de pérdida que se considere, es bien sabido que, en general, no existe un estimador cuya función de riesgo sea uniformemente menor que la de cualquier otro. Por tanto, dado un estimador, parece razonable, a fin de medir su comportamiento en una cierta región del modelo estadístico, calcular la integral de la distancia de Rao cuadrático media, y entonces dividir esta cantidad por el volumen Riemanniano de la región considerada. Más precisamente, sea $\mathcal{B} \subset M$ un subconjunto medible, con $V(\mathcal{B}) \neq 0$, donde V es la medida Riemanniana, entonces denotaremos el *promedio Riemanniano de la distancia de Rao cuadrático media* por

$$\mathcal{R}_{\mathcal{U}_k}^2(\mathcal{B}) = \frac{\int_{\mathcal{B}} E(\rho^2(\mathcal{U}_k, p)) V(dp)}{\int_{\mathcal{B}} V(dp)}$$

el índice obtenido es un promedio ponderado de la distancia cuadrático media. Este enfoque es compatible con un punto de vista Bayesiano, si se supone una distribución a priori uniforme con respecto a volumen Riemanniano, véase Jeffreys [31], un enfoque similar se puede encontrar también en Prakasa Rao [52] y Cencov [20].

5.1 Cotas del promedio Riemanniano de la distancia de Rao cuadrático media

En esta sección vamos a obtener diferentes cotas inferiores para la anterior medida de bondad de un estimador. Empezaremos con algunos resultados de carácter

general

Proposición 5.1.1 *Sea X un campo vectorial C^∞ y f una función real C^∞ positiva casi seguramente, con respecto a la medida Riemanniana, V , definida al menos en una bola Riemanniana de centro p y radio $R > 0$, S_R . Supongamos que $\operatorname{div}(X) \leq -a$, donde a es un número real positivo. Entonces*

$$0 < \frac{a}{\operatorname{vol}(S_R)} \int_0^R \int_S f \, dV \, dr \leq \frac{1}{\operatorname{vol}(S_R)} \int_{S_R} f \|X\| \, dV + \\ + \frac{1}{\operatorname{vol}(S_R)} \int_0^R \int_S \|X\| \|\operatorname{grad}(f)\| \, dV \, dr$$

Demostración

Como

$$\operatorname{div}(fX) = f \operatorname{div}(X) + \langle X, \operatorname{grad}(f) \rangle, \quad (5.1)$$

y para $0 < r < R$,

$$\int_S f \operatorname{div}(X) \, dV \leq -a \int_S f \, dV,$$

tenemos

$$\int_S \operatorname{div}(fX) \, dV - \int_S \langle X, \operatorname{grad}(f) \rangle \, dV \leq -a \int_S f \, dV$$

Además, como consecuencia del *teorema de divergencia de Gauss*,

$$\int_S \operatorname{div}(fX) \, dV = \int_{\partial S} \langle fX, \nu \rangle \, dA,$$

donde ν indica un campo de vectores unitarios normales y hacia afuera, y dA la medida Riemanniana inducida en ∂S_r , y teniendo en cuenta, por la desigualdad de Cauchy–Schwartz, que

$$|\langle X, \operatorname{grad}(f) \rangle| \leq \|X\| \|\operatorname{grad}(f)\|,$$

$$|\langle fX, \nu \rangle| \leq f \|X\|,$$

resulta que

$$0 < a \int_S f dV \leq \int_{\partial S} f \|X\| dA + \int_S \|X\| \|\text{grad}(f)\| dV$$

Ahora, siguiendo reglas estandar en coordenadas polares, tenemos

$$\int_0^R \left(\int_{\partial S} f \|X\| dA \right) dr = \int_{S_R} f \|X\| dV,$$

y por tanto, variando r desde 0 a R , integrando y dividiendo por $\text{vol}(S_R) = \int_{S_R} dV$,

$$\begin{aligned} 0 < \frac{a}{\text{vol}(S_R)} \int_0^R \int_S f dV dr &\leq \frac{1}{\text{vol}(S_R)} \int_{S_R} f \|X\| dV + \\ &+ \frac{1}{\text{vol}(S_R)} \int_0^R \int_S \|X\| \|\text{grad}(f)\| dV dr \end{aligned}$$

■

Teorema 5 1 2 *Con las mismas hipotesis que en el teorema (4 1 2), si escribimos*

$$\mathcal{R}_{\mathcal{U}_k}^2(S_R) = \frac{1}{\text{vol}(S_R)} \int_{S_R} E(\rho^2(\mathcal{U}_k, p)) dV,$$

obtenemos la siguiente cota inferior para el promedio Riemanniano de la distancia de Rao cuadratica media

$$0 < \left\{ \frac{a \int_0^R \text{vol}(S_r) dr}{\text{vol}(S_R) + \sqrt{kn} \sqrt{\text{vol}(S_R)} \int_0^R \sqrt{\text{vol}(S_r)} dr} \right\}^2 \leq \mathcal{R}_{\mathcal{U}_k}^2(S_R), \quad (5 2)$$

donde $a = n$ si las curvaturas seccionales son no-positivas y $a = 1$ si son positivas, acotadas superiormente por \mathcal{K} , y el diametro de la variedad satisface $d(M) < \pi/2\sqrt{\mathcal{K}}$

Demostracion

Antes de nada, observemos que $A_p(x) = \exp_p^{-1}(\mathcal{U}_k(x))$ es un campo aleatorio vectorial C^∞ y $p = p(x, \theta)$, la verosimilitud que define el modelo estadístico, es una función aleatoria C^∞ . Entonces, siguiendo los mismos pasos que en el teorema (4.1.2), podemos escoger un sistema de coordenadas esféricas con origen $\mathcal{U}_k(x)$, bajo este sistema de coordenadas, y usando el convenio de sumación de índices repetidos, tenemos

$$\frac{\partial A^\alpha}{\partial \theta^\alpha} = -1 \quad \text{and} \quad \Gamma_{\alpha j}^\alpha A^j = -\rho \Gamma_{\alpha 1}^\alpha = -\frac{\partial \log \sqrt{g}}{\partial \rho} \rho,$$

donde g es el determinante del tensor métrico. Entonces

$$\operatorname{div}(A) = -1 - \rho \frac{\partial \log \sqrt{g}}{\partial \rho}$$

En el caso Euclídeoano

$$\frac{\partial \log \sqrt{g}}{\partial \rho} = \frac{n-1}{\rho}, \tag{5.3}$$

y por tanto $\operatorname{div}(A) = -n$

Cuando las curvaturas seccionales son no positivas, obtenemos

$$\frac{\partial \log \sqrt{g}}{\partial \rho} \geq \frac{n-1}{\rho},$$

y de esta forma $\operatorname{div}(A) \leq -n$

Finalmente, cuando las curvaturas seccionales son positivas, acotadas superiormente por \mathcal{K} , y el diámetro de la variedad satisface $d(M) < \pi/2\sqrt{\mathcal{K}}$, tenemos

$$\frac{\partial \log \sqrt{g}}{\partial \rho} \geq 0,$$

y en esta situación $\operatorname{div}(A) \leq -1$

En cualquier caso, $\operatorname{div}(A) < -a$ con $a = n$ or $a = 1$, dependiendo del signo de las curvaturas seccionales. Por tanto, podemos aplicar las fórmulas obtenidas previamente, y integrando con respecto a $d\mu$, la medida de referencia, se obtiene

$$0 < \frac{a}{\operatorname{vol}(S_R)} \int_{\mathcal{X}} \left(\int_0^R \int_S p \, dV \, dr \right) d\mu \leq \frac{1}{\operatorname{vol}(S_R)} \int_{\mathcal{X}} \left(\int_{S_R} p \|A\| \, dV \right) d\mu +$$

$$+ \frac{1}{\text{vol}(S_R)} \int_{\mathcal{X}} \left(\int_0^R \int_S \|A\| \|\text{grad}(p)\| dV dr \right) d\mu$$

Observando que

$$\int_{\mathcal{X}} p d\mu = 1, \quad \text{y} \quad \|\text{grad}(p)\| = \|\text{grad}(\log p)\| p,$$

e intercambiando integrales, se sigue que

$$\begin{aligned} 0 < \frac{a}{\text{vol}(S_R)} \int_0^R \text{vol}(S_r) dr &\leq \frac{1}{\text{vol}(S_R)} \int_{S_R} E(\|A\|) dV + \\ &+ \frac{1}{\text{vol}(S_R)} \int_0^R \int_S E(\|A\| \|\text{grad}(\log p)\|) dV dr, \end{aligned}$$

donde E es la esperanza con respecto a la medida de probabilidad $p d\mu$. Por la desigualdad de Cauchy–Schwartz,

$$E(\|A\| \|\text{grad}(\log p)\|) \leq \sqrt{E(\|A\|^2)} \sqrt{E(\|\text{grad}(p)\|^2)},$$

y puesto que

$$E(\|\text{grad}(\log p)\|^2) = kn,$$

obtenemos

$$\begin{aligned} 0 < \frac{a}{\text{vol}(S_R)} \int_0^R \text{vol}(S_r) dr &\leq \frac{1}{\text{vol}(S_R)} \int_{S_R} E(\|A\|) dV + \\ &+ \frac{\sqrt{kn}}{\text{vol}(S_R)} \int_0^R \left(\int_S \sqrt{E(\|A\|^2)} dV \right) dr \end{aligned}$$

Ademas, por la desigualdad de Jensen

$$E(\|A\|) \leq \sqrt{E(\|A\|^2)},$$

y

$$\frac{1}{\text{vol}(S_r)} \int_S \sqrt{E(\|A\|^2)} dV \leq \sqrt{\frac{1}{\text{vol}(S_r)} \int_S E(\|A\|^2) dV},$$

para $0 \leq r \leq R$, y entonces

$$0 < \frac{a}{\text{vol}(S_R)} \int_0^R \text{vol}(S_r) dr \leq \sqrt{\frac{1}{\text{vol}(S_R)} \int_{S_R} E(\|A\|^2) dV} + \frac{\sqrt{kn}}{\text{vol}(S_R)} \int_0^R \sqrt{\text{vol}(S)} \sqrt{\int_S E(\|A\|^2) dV} dr$$

Teniendo en cuenta que

$$\sqrt{\int_S E(\|A\|^2) dV},$$

es una función de r positiva y monótona creciente, ya que

$$E(\|A\|^2) = E(\rho^2(\mathcal{U}_k, p)),$$

si escribimos

$$\mathcal{R}_{\mathcal{U}_k}^2(S_R) = \frac{1}{\text{vol}(S_R)} \int_{S_R} E(\rho^2(\mathcal{U}_k, p)) dV,$$

entonces

$$0 < \frac{a}{\text{vol}(S_R)} \int_0^R \text{vol}(S_r) dr \leq \left(1 + \frac{\sqrt{kn}}{\sqrt{\text{vol}(S_R)}} \int_0^R \sqrt{\text{vol}(S_r)} dr \right) \sqrt{\mathcal{R}_{\mathcal{U}_k}^2(S_R)}$$

■

Observaciones Es interesante notar que mientras las cotas locales obtenidas en el teorema (4.1.2) podían anularse, la cota global, para bolas de radio mayor que

cero, es siempre positiva. Por otra parte los efectos de la curvatura se manifiestan aquí a través del volumen de una bola Riemanniana. La proposición (11.6.3) del apéndice implica, en variedades con curvatura seccional constante, que para bolas pequeñas, ya que el orden del numerador en la desigualdad (5.2) será el orden de la cota, la cota disminuirá al aumentar la curvatura. Por último pensamos que ulteriores investigaciones usando las expresiones obtenidas en la sección (11.6.2) para el volumen de una bola Riemanniana podrían revelar información más precisa de los efectos de la curvatura.

Corolario 5.1.3 *Cuando el modelo estadístico paramétrico es una variedad Euclídea tenemos la siguiente acotación inferior del promedio Riemanniano de la distancia de Rao cuadrático media*

$$0 < \left\{ \frac{n(n+2)R}{(n+1)(n+2+2\sqrt{kn}R)} \right\}^2 \leq \mathcal{R}_{U_k}^2(S_R)$$

Si la variedad Euclídea, M , es completa y simplemente conexa, obtenemos la siguiente acotación inferior sobre toda la variedad

$$\frac{n(n+2)^2}{4k(n+1)^2} \leq \lim_{R \rightarrow \infty} \mathcal{R}_{U_k}^2(S_R) \equiv \mathcal{R}_{U_k}^2(M)$$

Demostración

Como

$$\text{vol}(S) = \frac{2\pi^{n/2}r^n}{n\Gamma(n/2)},$$

tenemos

$$\int_0^R \sqrt{\text{vol}(S_r)} dr = \left(\frac{8\pi^{n/2}R^{n+2}}{n(n+2)^2\Gamma(n/2)} \right)^{1/2},$$

y

$$\int_0^R \text{vol}(S_r) dr = \frac{2\pi^{n/2}R^{n+1}}{n(n+1)\Gamma(n/2)},$$

entonces

$$0 < \left\{ \frac{n(n+2)R}{(n+1)(n+2+2\sqrt{knR})} \right\}^2 \leq \mathcal{R}_{\mathcal{U}_k}^2(S_R)$$

El segundo resultado se deriva tomando el limite cuando $R \rightarrow \infty$ ■

Ejemplo 5 1 4 Como ejemplo, consideremos la distribucion normal n -dimensional con matriz de covarianzas conocida, Σ . Dada una muestra de tamaño k , la densidad Riemanniana de la distancia de Rao cuadratica media correspondiente a la media muestral \bar{X}_k es

$$\mathcal{R}_{\mathcal{U}_k}^2(S_R) = \frac{n}{k},$$

que es claramente superior a

$$\frac{n(n+2)^2}{4k(n+1)^2}$$

5 2 Metodos variacionales aplicados a la obtencion de cotas globales

Otra forma de enfocar el problema de la obtencion de cotas globales es utilizando metodos variacionales. Un precedente de este procedimiento lo podemos encontrar en Cencov [20]. La idea es la siguiente, consideremos la integral de la cota que proporciona (4 1 2), esto es

$$\mathcal{Y}(B) = \int_{S_R} \left\{ \|B\|^2 + \frac{1}{kn}(\operatorname{div}(B) + a)^2 \right\} dV,$$

donde tomamos $a = n$ si las curvaturas seccionales son no-positivas y $a = 1$ en caso contrario. Entonces el anterior funcional depende unicamente de B y podemos entonces tratar de buscar el campo vectorial B, C^∞ , que lo hace minimo. El minimo que obtengamos corresponde a una clase de campos vectoriales mas amplia que la de tensores de sesgo C^∞ de modo que proporcionara una cota inferior del promedio del error intrinseco que puede que no se alcance para ningun estimador. De esta manera este procedimiento es un procedimiento complementario del visto anteriormente para la obtencion de cotas globales.

Proposicion 5 2 1 *El campo B , C^∞ , hace minimo el funcional*

$$y(B) = \int_{S_R} \left\{ \|B\|^2 + \frac{1}{kn} (\operatorname{div}(B) + a)^2 \right\} dV$$

si y solo si se verifica

$$B - \frac{1}{kn} \operatorname{grad}(\operatorname{div}(B)) = 0, \quad \forall p \in S_R \tag{5 4}$$

$$\operatorname{div}(B) + a = 0, \quad \forall \theta \in \partial S_R$$

y este valor minimo viene dado por

$$\mathcal{Y}^* = \frac{a^2}{kn} \operatorname{vol}(S_R) + \frac{a}{kn} \int_{S_R} \operatorname{div}(B) dV \tag{5 5}$$

con B verificando (5 4)

Demostracion

Consideremos la variacion primera $\delta\mathcal{Y}(B, \eta)$, donde η es un campo arbitrario. Entonces es facil ver que

$$\lim_{\epsilon \rightarrow 0} \frac{\mathcal{Y}(B + \epsilon\eta) - \mathcal{Y}(B)}{\epsilon} \equiv \delta y(B, \eta) = \int_{S_R} \left(2\langle B, \eta \rangle + \frac{2}{kn} \operatorname{div}(\eta) (\operatorname{div}(B) + a) \right) dV$$

Asimismo

$$\mathcal{Y}(B + \eta) - \mathcal{Y}(B) = \delta\mathcal{Y}(B, \eta) + \int_{S_R} \left\{ \|\eta\|^2 + \frac{1}{kn} (\operatorname{div}(\eta))^2 \right\} dV$$

con lo que el funcional es convexo, de modo que un punto estacionario correspondera a un minimo absoluto. Ahora la condicion $\delta\mathcal{Y}(B, \eta) = 0$ equivale a

$$\int_{S_R} \left\{ \langle B, \eta \rangle + \frac{1}{kn} \operatorname{div}(\eta) (\operatorname{div}(B) + a) \right\} dV = 0$$

Pero si tenemos en cuenta (5 1)

$$\begin{aligned} \frac{1}{kn} \operatorname{div}(\eta) (\operatorname{div}(B) + a) &= \operatorname{div} \left(\frac{1}{kn} \eta (\operatorname{div}(B) + a) \right) - \langle \operatorname{grad} \left(\frac{1}{kn} (\operatorname{div}(B) + a) \right), \eta \rangle \\ &= \frac{1}{kn} \operatorname{div}(\eta (\operatorname{div}(B) + a)) - \frac{1}{kn} \langle \operatorname{grad}(\operatorname{div}(B)), \eta \rangle \end{aligned}$$

podremos escribir la condicion de estacionariedad como

$$\int_{S_R} \left\{ \langle B, \eta \rangle + \frac{1}{kn} \operatorname{div}(\eta (\operatorname{div}(B) + a)) - \frac{1}{kn} \langle \operatorname{grad}(\operatorname{div}(B)), \eta \rangle \right\} dV = 0$$

y aplicando el teorema de la divergencia,

$$\int_{S_R} \langle B - \frac{1}{kn} \operatorname{grad}(\operatorname{div}(B)), \eta \rangle dV + \frac{1}{kn} \int_{\partial S_R} \langle (\operatorname{div}(B) + a)\eta, \nu \rangle dA = 0,$$

(5 4) se sigue entonces de que la igualdad anterior debe cumplirse para cualquier η

Veamos ahora la segunda parte de la proposicion Aplicando la primera con dicion de estacionariedad en (5 4) y teniendo en cuenta (5 1) obtenemos

$$\|B\|^2 = \frac{1}{kn} (\operatorname{div}(B \operatorname{div}(B)) - (\operatorname{div}(B))^2)$$

sutituyendo en $\mathcal{Y}(B)$ obtenemos

$$\begin{aligned} \mathcal{Y}^* &= \int_{S_R} \left\{ \frac{1}{kn} (\operatorname{div}(B \operatorname{div}(B)) + a^2 + 2a \operatorname{div}(B^*)) \right\} dV \\ &= \frac{a^2}{kn} \operatorname{vol}(S_R) + \frac{1}{kn} \int_{\partial S_R} \langle B \operatorname{div}(B) + 2aB, \nu \rangle dA \end{aligned}$$

aplicando ahora la segunda condicion de estacionariedad en (5 4) resultara

$$\mathcal{Y}^* = \frac{a^2}{kn} \operatorname{vol}(S_R) + \frac{a}{kn} \int_{\partial S_R} \langle B^*, \nu \rangle dA$$

y volviendo a aplicar el teorema de la divergencia obtenemos (5 5) ■

Observaciones Notemos que el valor mínimo de $y(B)$ solo depende de $\operatorname{div}(B^*)$, ahora bien $f^* \equiv \operatorname{div}(B^*)$ satisface la ecuacion

$$\Delta f = knf, \quad \text{con } f(p) = -a \quad \forall p \in \partial S_R \tag{5 6}$$

como se deduce facilmente de (5 4)

Un caso particular donde se puede resolver facilmente (5 6) es el caso Euclideo En tal caso notemos que si elegimos un sistema de coordenadas polares con centro p la simetria de (5 6) hace que la solucion solo dependa de la coordenada radial r Esto se puede deducir facilmente de la unicidad de la solucion en (5 6) y de

que si f es una solución, la función \tilde{f} que se obtiene a partir de f después de hacer un giro con centro p también lo es. Notemos que en dimensión uno esto significaría que la función solución es par. Asimismo tengamos en cuenta que en el caso Euclideo $a = n$.

Con estas consideraciones previas, tenemos la siguiente proposición

Proposición 5.2.2 *Si la variedad asociada al modelo paramétrico es Euclidea tenemos la siguiente acotación inferior del promedio Riemanniano de la distancia de Rao-cuadrático media*

$$0 < \frac{n}{k} \left(1 - \frac{{}_0F_1\left(\frac{n}{2} + 1, \frac{knR^2}{4}\right)}{{}_0F_1\left(\frac{n}{2}, \frac{knR^2}{4}\right)} \right) \leq \mathcal{R}_{\mathcal{U}_k}^2(S_R) \quad (5.8)$$

donde ${}_0F_1(a, z)$ es una función hipergeométrica generalizada, véase (11.11) del apéndice

Si la variedad Euclidea, M , es completa y simplemente conexa, obtenemos la siguiente acotación inferior sobre toda la variedad

$$\frac{n}{k} \leq \lim_{R \rightarrow \infty} \mathcal{R}_{\mathcal{U}_k}^2(S_R) \equiv \mathcal{R}_{\mathcal{U}_k}^2(M)$$

Demostración

Consideremos la ecuación diferencial

$$\Delta f = knf, \quad \text{con} \quad f(p) = -n, \forall p \in \partial S_R \quad (5.8)$$

en coordenadas polares con centro p . Sabemos, por la observación anterior, que f solo depende de r entonces, para $r \neq 0$, (véase Chavel [21, pp. 3-5])

$$\Delta f = \frac{1}{r^{n-1}} (r^{n-1} f'(r))' = (n-1) \frac{f'(r)}{r} + f''(r) \quad (5.9)$$

con lo que la ecuación a resolver será

$$r f''(r) + (n-1) f'(r) - knr f(r) = 0$$

Si buscamos una solución de la forma $f(r) = \sum_0^\infty a_j r^j$, obtenemos, sustituyendo en la ecuación,

$$(n-1)a_1 + \sum_{j=1}^\infty \{(n-1+j)(j+1)a_{j+1} - kna_{j-1}\} r^j = 0$$

de aquí si $n \neq 1$,

$$a_1 = 0, \quad a_{j+1} = \frac{kn}{(n-1+j)(j+1)} a_{j-1}, \quad j = 1, 2,$$

A la misma conclusión se llegaría para $n = 1$ aunque en este caso, como se indica en la observación anterior, partiríamos de una serie de potencias par. En definitiva

$$f(r) = a_0 \sum_{j=0}^{\infty} \frac{1}{(n/2)_j} \frac{(knr^2/4)^j}{j!}$$

donde $(n/2)_j \equiv (n/2)(n/2+1)(n/2+2) \dots (n/2+j-1)$. Y teniendo en cuenta la definición en (11.11) del apéndice podemos escribir más abreviadamente,

$$f(r) = a_0 {}_0F_1 \left(\frac{n}{2}, \frac{knr^2}{4} \right)$$

donde a_0 se obtendrá de la condición $f(R) = -n$

Necesitamos evaluar $\int_{S_R} f dV$. Efectuando el cálculo en coordenadas polares

$$\int_{S_R} f dV = A(S) \int_0^R r^{n-1} f(r) dr$$

donde $A(S)$ es el área de la superficie esférica n dimensional de radio unidad, esto es

$$A(S) = \frac{2\pi^{n/2}}{\Gamma(n/2)}$$

Ahora bien de (5.9) y (5.8)

$$\left(r^{n-1} f'(r) \right)' = knr^{n-1} f(r)$$

de modo que

$$\int_{S_R} f dV = \frac{2\pi^{n/2}}{kn\Gamma(n/2)} R^{n-1} f'(R)$$

Por último, derivando la expresión de $f(r)$ en serie de potencias, se obtiene,

$$f'(r) = a_0 kr {}_0F_1 \left(\frac{n}{2} + 1, \frac{knr^2}{4} \right)$$

con lo que, véase (5.5)

$$\mathcal{Y}^* = \frac{2\pi^{n/2}}{k\Gamma(n/2)} R^n \left(1 - \frac{{}_0F_1 \left(\frac{n}{2} + 1, \frac{knR^2}{4} \right)}{{}_0F_1 \left(\frac{n}{2}, \frac{knR^2}{4} \right)} \right) \leq \mathcal{R}_{\mathcal{U}_k}^2(S_R)$$

dividiendo ahora por el volumen de S_R obtenemos el resultado deseado

La segunda parte de la demostracion se sigue tomando el limite cuando $R \rightarrow \infty$ en (5.7) ■

Observaciones Es evidente, a la vista del ejemplo (5.1.4) que la cota anterior se llega a alcanzar y que ademas los resultados obtenidos por metodos variacionales son mas finos que los obtenidos al principio del capitulo, al menos en el caso Euclideo. Pero sin embargo no hay que olvidar que dado un campo C^∞ en M , este no correspondera en general al tensor de sesgo de un estimador de modo que no se puede asegurar que la cota obtenida aqui sea siempre alcanzable y tan grande como la obtenida por los otros metodos. Por ultimo pensamos que un resultado similar al anterior quizas se pueda obtener para el caso de curvaturas seccionales constantes utilizando en este caso las expresiones obtenidas en la seccion (11.6.2) para el volumen de una bola Riemanniana.

En el caso $n = 1$ la variedad es Euclidea de forma que podemos aplicar el resultado anterior y obtenemos como cota

$$\frac{\mathcal{Y}}{2R} = \frac{1}{k} - \frac{1}{k\sqrt{k}R} \tanh(\sqrt{k}R)$$

Como mejorar el comportamiento de un estimador

Hemos obtenido, en los capítulos precedentes, una cota inferior para la distancia de Rao cuadrático media, vamos ahora a estudiar como podemos hacer decrecer la distancia de Rao cuadrático media de un estimador dado

Clásicamente, esto se consigue por *Blackwellización* del estimador, esto es, tomando la esperanza condicionada respecto a un estadístico suficiente. Seguiremos un procedimiento similar aquí, pero ahora nuestros objetos aleatorios tomarán valores en una variedad y por tanto habrá que explicar el significado de valor medio condicionado en este contexto y entonces obtener versiones intrínsecas de los teoremas de Rao–Blackwell y Lehmann–Scheffé.

6.1 Valores medios condicionados de funciones a valores en una variedad

Sea $(\mathcal{X}, \mathbf{a}, P)$ un espacio de probabilidad. Sea (M, \mathfrak{A}) una n -variedad Riemanniana completa, (Hausdorff y conexa) C^∞ . Entonces M será un espacio métrico separable y completo (un “Polish space”) y tendremos versiones regulares de la probabilidad condicionada para cualquier cualquier objeto aleatorio, f , valorado en M con respecto a una σ algebra \mathcal{D} en el espacio muestral, \mathcal{X} .

Además si la distancia de Rao cuadrático media de f existe, tendremos

$$E(\rho^2(m, f)|\mathcal{D})(x) = \int_M \rho^2(m, t) P_{f|\mathcal{D}}(x, dt),$$

donde $P_{f|\mathcal{D}}(x, B)$ es una version regular de la probabilidad condicionada de f dada \mathcal{D} , $x \in \mathcal{X}$, donde B es un Borehano de M

Si dado $x \in \mathcal{X}$ hubiese uno y solo un extremo $p \in M$ de $E(\rho^2(m, f)|\mathcal{D})(x)$, o equivalentemente un punto $p \in M$ tal que

$$\int_M \exp_p^{-1}(t) P_{f|\mathcal{D}}(x, dt) = 0_p,$$

tendriamos una aplicacion de \mathcal{X} en M que asignaria un valor medio para cada x . Es claro que si la imagen de esta aplicacion fuese numerable, la aplicacion seria medible, pero ya que tenemos un conjunto denso numerable en M resulta que esta aplicacion es siempre medible. Esto justifica la siguiente definicion.

Definicion 6 1 1 Sea f un objeto aleatorio en M y \mathcal{D} una σ algebra en \mathcal{X} , definiremos el valor medio condicionado de f con respecto a \mathcal{D} como una aplicacion \mathcal{D} medible, Z , tal que

$$E(\exp_Z^{-1}(f(\cdot))|\mathcal{D}) = 0_Z$$

Escribiremos $\mathfrak{M}(f|\mathcal{D}) = Z$

Observaciones A partir de (2 1 16) una condicion suficiente para asegurar que el valor medio existe es tener un subconjunto regular convexo $N \subset M$ tal que $P\{f \in N\} = 1$. Asimismo podemos extender los resultados previos al caso en que M no sea completa, ya que N es difeomorfica a un conjunto abierto en \mathbb{R}^n y entonces existiran versiones regulares de la probabilidad condicionada de f dada \mathcal{D} .

Las proposiciones que vienen a continuacion son inmediatas.

Proposicion 6 1 2 Si f es una aplicacion \mathcal{D} medible entonces $\mathfrak{M}(f|\mathcal{D}) = f$
c s P

Proposicion 6 1 3 Si f es independiente de \mathcal{D} entonces $\mathfrak{M}(f|\mathcal{D}) = \mathfrak{M}(f)$
c s P

Observacion Es necesario señalar que, en general, $\mathfrak{M}(\mathfrak{M}(f|\mathcal{D})) \neq \mathfrak{M}(f)$, como observa Kendall [36] y como es facil de ver con sencillos contraejemplos.

6 2 Rao–Blackwell intrinseco

Vamos a aplicar estas nociones a la estimacion puntual. Dado un modelo estadístico paramétrico $\{\chi, \mathcal{A}, P_\theta, \theta \in \Theta\}$, sea $M = \{p \mid p = p(\cdot, \theta), \theta \in \Theta\}$ la variedad asociada con la métrica Riemanniana dada por la matriz de información de Fisher. Supondremos que el modelo es regular y que existe un abierto regular convexo $N \subset M$ tal que $\mu(M \setminus N) = 0$ (siendo μ la medida de referencia que domina el modelo)

Sea \mathcal{D} una σ algebra suficiente para el modelo estadístico. Dada una muestra de tamaño k y un estimador \mathcal{U}_k podemos ahora considerar el estimador $\mathfrak{M}(\mathcal{U}_k | \mathcal{D})$. Escribamos

$$\begin{aligned}\Delta_{\mathcal{U}_k}^2(p) &= E_p(\rho^2(\mathcal{U}_k, p)), \\ \Delta_{\mathfrak{M}(\mathcal{U}_k | \mathcal{D})}^2(p) &= E_p(\rho^2(\mathfrak{M}(\mathcal{U}_k | \mathcal{D}), p))\end{aligned}$$

Teniendo en cuenta que una función $h(q)$, $q \in M$ en la variedad se denomina convexa si $h(\gamma(t))$, $t \in \mathbb{R}$ es una función convexa en el sentido ordinario para cualquier geodésica $\gamma(t)$, tenemos los siguientes teoremas

Teorema 6 2 1 (Rao–Blackwell intrinseco) Si fijado $p \in N$ el cuadrado de la distancia de Rao $\rho^2(p, m)$ es una función convexa entonces $\Delta_{\mathfrak{M}(\mathcal{U}_k | \mathcal{D})}^2(p) \leq \Delta_{\mathcal{U}_k}^2(p)$

Demostración

Esta demostración es una adaptación de otra presente en Kendall [36]. Por convexidad, para todo t positivo

$$\begin{aligned}\rho^2(\gamma(t), p) &\geq \rho^2(\gamma(0), p) + \left. \frac{d\rho^2(\gamma(s), p)}{ds} \right|_{s=0} t \\ &= \rho^2(\gamma(0), p) + \left\langle \text{grad}(\rho^2)(0), \frac{d\gamma}{ds}(0) \right\rangle t,\end{aligned}$$

entonces escribiendo $m = \gamma(0)$ y $q = \gamma(t)$, como

$$\frac{d\gamma}{ds}(0) t = \exp_m^{-1}(q),$$

y

$$\text{grad}(\rho^2)(0) = -2 \exp_m^{-1}(p),$$

podemos escribir la anterior desigualdad

$$\rho^2(q, p) \geq \rho^2(m, p) - 2 \langle \exp_m^{-1}(p), \exp_m^{-1}(q) \rangle,$$

entonces tomando $m = \mathfrak{M}(\mathcal{U}_k | \mathcal{D})$ e integrando con respecto a $P_{\mathcal{U}_k | \mathcal{D}}(x, dq)$ obtenemos

$$\int_M \rho^2(q, p) P_{\mathcal{U}_k | \mathcal{D}}(x, dq) \geq \rho^2(\mathfrak{M}(\mathcal{U}_k | \mathcal{D}), p),$$

ya que

$$\int_M \exp_m^{-1}(q) P_{\mathcal{U}_k | \mathcal{D}}(x, dq) = 0_m$$

Finalmente tomando esperanzas obtenemos

$$\begin{aligned} \Delta_{\mathcal{U}_k}^2(p) &= E_p(\rho^2(\mathcal{U}_k, p)) = E_p(E_p(\rho^2(\mathcal{U}_k, p) | \mathcal{D})) \\ &\geq E_p(\rho^2(\mathfrak{M}(\mathcal{U}_k | \mathcal{D}), p)) = \Delta_{\mathfrak{M}(\mathcal{U}_k | \mathcal{D})}^2(p) \end{aligned}$$

■

Teorema 6.2.2 *Si las curvaturas seccionales en N son a lo sumo θ , o $\mathcal{K} > 0$ con $d(N) < \pi/2\sqrt{\mathcal{K}}$, entonces*

$$\Delta_{\mathcal{U}_k}^2(p) \geq \Delta_{\mathfrak{M}(\mathcal{U}_k | \mathcal{D})}^2(p)$$

Demostracion

Por (11.4) en el apéndice, estamos en las condiciones en que el cuadrado de la distancia Riemanniana es una función convexa. Por tanto podemos aplicar el teorema anterior

■

Observaciones Si algunas curvaturas son positivas y no imponemos condiciones sobre el diámetro del conjunto regular convexo, $N \subset M$, no podemos asegurar la

convexidad del cuadrado de la distancia Riemanniana y entonces no es necesariamente cierto que la distancia Riemanniana cuadratica media entre la verdadera densidad y la estimada, decrezca cuando condicionamos a \mathcal{D}

Por otra parte podemos mejorar la eficiencia de los estimadores condicionando con respecto a una σ algebra suficiente \mathcal{D} , obteniendo $\mathfrak{M}(\mathcal{U}_k|\mathcal{D})$, pero el sesgo no se preservara en general, en contraste con el clasico teorema de Rao-Blackwell. En otras palabras, si \mathcal{U}_k fuera intrinsecamente insesgado, $\mathfrak{M}(\mathcal{U}_k|\mathcal{D})$ no seria, en general, intrinsecamente insesgado ya que

$$\mathfrak{M}(\mathfrak{M}(\mathcal{U}_k|\mathcal{D})) \neq \mathfrak{M}(\mathcal{U}_k)$$

No obstante, la norma del tensor de sesgo de $\mathfrak{M}(\mathcal{U}_k|\mathcal{D})$ estaria acotada. Si denotamos el tensor de sesgo por $B\mathfrak{M}_{(\mathcal{U}_k|\mathcal{D})}$, por la desigualdad de Cauchy Schwartz,

$$\|B\mathfrak{M}_{(\mathcal{U}_k|\mathcal{D})}(p)\|^2 \leq \Delta_{\mathfrak{M}_{(\mathcal{U}_k|\mathcal{D})}}^2(p) \leq \Delta_{\mathcal{U}_k}^2(p)$$

Aunque el tensor de sesgo no se mantiene en general cuando condicionamos con respecto a un estadistico suficiente, un teorema, que es analogo al de Lehmann-Scheffee, puede formularse en el contexto intrinseco. Necesitamos primero redefinir la nocion de completitud

Definicion 6 2 3 *Un estadistico suficiente T se dira que es completo, para M , si y solo si*

$$\mathfrak{M}_p(f(T)) = \mathfrak{M}_p(g(T)) \quad \forall p \in M$$

implica que $f(T) = g(T)$ (c.s. μ)

Entonces, con las mismas hipotesis que en el teorema previo, tenemos el resultado siguiente

Proposicion 6 2 4 (Lehmann-Scheffee intrinseco) *Sea \mathcal{U} un estimador que es funcion de un estadistico suficiente y completo para M , entonces, es el estimador de uniformemente minima distancia de Rao cudratica media para un tensor de sesgo fijo*

Demostracion

La demostracion es trivial a partir de la definicion y teorema anteriores ■

Propiedades asintóticas

Ante todo notemos que, dada una sucesión de variables aleatorias que toman valores en una n -variedad C^∞ (Hausdorff y conexa) con estructura Riemanniana, la definición de los diferentes tipos de convergencias estocásticas es directa: débil, en probabilidad, casi segura, o en media r -ésima, como en cualquier espacio métrico. Además, ya que la topología inducida por la métrica Riemanniana es la misma que la topología inducida por el atlas, si tenemos una carta global, tomando coordenadas, podemos reducir el estudio de estas convergencias, con la excepción de la convergencia en media r -ésima, al de convergencia de sucesiones de variables aleatorias que toman valores en \mathbf{R}^n .

Hemos visto que a menudo los estimadores son intrínsecamente sesgados pero vamos a demostrar que el sesgo intrínseco tiende a cero para muestras grandes en casos importantes tales como los estimadores máximo verosímil.

7.1 Definiciones y resultados básicos

Definición 7.1.1 *Un estimador \mathcal{U} es asintóticamente intrínsecamente insesgado en sentido amplio si y solo si es intrínsecamente insesgado asintóticamente, esto es, podemos construir una sucesión de valores medios de \mathcal{U}_k que converja a p_0 . Cuando la sucesión de valores medios este definida de forma única, podemos escribir*

$$\lim_{k \rightarrow \infty} \mathfrak{M}_{p_0}(\mathcal{U}_k) = p_0 = p(\cdot, \theta_0) \quad \text{para cualquier } p_0 \in M$$

y diremos que \mathcal{U} es asintóticamente intrínsecamente insesgado

En las dos proposiciones siguientes supondremos que el estimador toma valores casi seguramente en un conjunto regular convexo, que \mathcal{U} es *regular* en el sentido siguiente,

$$\sup_{k \in \mathbb{N}} E_{p_0} (\rho^2(\mathcal{U}_k, p_0)) < \infty$$

y la derivada covariante del campo vectorial $\mathcal{E}(p) = E_{p_0}(\exp_p^{-1}(\mathcal{U}_k))$ existe y puede ser obtenida derivando bajo el signo integral. Supondremos tambien que la variedad asociada a la familia parametrica regular de densidades tiene curvaturas seccionales K acotadas superior e inferiormente, es decir $\kappa < K < \mathcal{K}$ y el diametro de la variedad $d(M) < \pi/2\sqrt{\mathcal{K}}$ si $\mathcal{K} > 0$ e ∞ en otro caso. Notemos que en el teorema (4.1.2) tenemos condiciones analogas y tambien que estas condiciones son suficientes para asegurar una geometria convexa, vease (11.4) en el apendice y tambien (8.2.5), y que por tanto el valor medio es de hecho un centro de masas.

Proposicion 7.1.2 *Un estimador \mathcal{U} es asintoticamente intrinsecamente insesgados si y solo si para el correspondiente campo tensorial de sesgo,*

$$B_k(p_0) = E_{p_0}(\exp_{p_0}^{-1}(\mathcal{U}_k)),$$

que depende del tamaño muestral k , tenemos

$$\lim_{k \rightarrow \infty} B_k^\alpha(p_0) = 0 \quad \alpha = 1, \dots, n \quad \forall p_0 \in M$$

Demostracion

Supongamos primero que $\lim_{k \rightarrow \infty} \|B_k(p_0)\| = 0$. Sea

$$\mathcal{H}_k(p) = E_{p_0} (\rho^2(\mathcal{U}_k, p))$$

y sea $\gamma(t)$ una geodesica tal que $\gamma(0) = \mathfrak{M}_{p_0}(\mathcal{U}_k)$ y $\gamma(1) = p_0$. Entonces

$$\begin{aligned} \|\text{grad}(\mathcal{H}_k)(\gamma(1))\| \|\gamma(1)\| &\geq \langle \text{grad}(\mathcal{H}_k)(\gamma(1)), \gamma(1) \rangle = \left. \frac{d}{dt} \mathcal{H}_k(\gamma(t)) \right|_{t=1} \\ &= \int_0^1 \frac{d^2}{dt^2} \mathcal{H}_k(\gamma(t)) dt, \end{aligned}$$

puesto que

$$\left. \frac{d}{dt} \mathcal{H}_k(\gamma(t)) \right|_{t=0} = 0,$$

porque $\gamma(0) = \mathfrak{M}_{p_0}(\mathcal{U}_k)$ es un minimo local de $\mathcal{H}_k(p)$ Adicionalmente, por las condiciones de regularidad de \mathcal{U}_k podemos escribir

$$\int_0^1 \frac{d^2}{dt^2} \mathcal{H}_k(\gamma(t)) dt = \int_0^1 \left(E_{p_0} \left(\frac{d^2}{dt^2} \rho^2(\mathcal{U}_k, \gamma(t)) \right) \right) dt$$

Por (11.6) de la subseccion (11.4) del apendice,

$$\frac{d^2}{dt^2} \rho^2(\mathcal{U}_k, \gamma(t)) \geq C(\mathcal{K}) \rho^2(\gamma(0), \gamma(1)),$$

con $C(\mathcal{K}) > 0$ Por tanto obtenemos

$$\|\text{grad}(\mathcal{H}_k)(\gamma(1))\| \|\gamma(1)\| \geq C(\mathcal{K}) \rho^2(\gamma(0), \gamma(1))$$

Finalmente, ya que $\|\gamma(1)\| = \rho(\gamma(0), \gamma(1))$ y

$$\text{grad}(\mathcal{H}_k)(\gamma(1)) = -2 \int_{\mathcal{X}} \exp_{p_0}^{-1}(\mathcal{U}_k(x)) p_{(k)}(x, \theta_0) \mu_k(dx) = -2B_k(p_0),$$

tenemos

$$\|B_k(p_0)\| \geq \frac{1}{2} C(\mathcal{K}) \rho(\mathfrak{M}_{p_0}(\mathcal{U}_k), p_0),$$

entonces, tomando limites, obtenemos

$$\lim_{k \rightarrow \infty} \rho(\mathfrak{M}_{p_0}(\mathcal{U}_k), p_0) = 0$$

Supongamos ahora que $\lim_{k \rightarrow \infty} \mathfrak{M}_{p_0}(\mathcal{U}_k) = p_0$

$$\begin{aligned} \|B_k(p_0)\| &= \frac{1}{2} \|\text{grad}(\mathcal{H}_k)(\gamma(1))\| = \frac{1}{2} \int_0^1 \frac{d}{dt} \|\text{grad}(\mathcal{H}_k)(\gamma(t))\| dt \\ &\leq \frac{1}{2} \int_0^1 \left\| \frac{\nabla}{dt} \text{grad}(\mathcal{H}_k)(\gamma(t)) \right\| dt \end{aligned}$$

ya que $\text{grad}(\mathcal{H}_k)(\gamma(0)) = 0$, y donde la ultima desigualdad se debe al hecho de que para cualquier campo vectorial C^1 X , $\|X\|' \leq \|X'\|$ Aqui denotamos la derivada covariante de X , a lo largo de una curva determinada por el contexto, por X' Entonces, ya que

$$\frac{1}{2} \text{grad}(\mathcal{H}_k)(\gamma(t)) = -E_{p_0}(\exp_{\gamma(t)}^{-1}(\mathcal{U}_k)) = E_{p_0} \left(\frac{\partial}{\partial s} c(s, t) \Big|_{s=1} \right)$$

con $c(s, t) = \exp_{\mathcal{U}_k}(s \exp_{\mathcal{U}_k}^{-1}(\gamma(t)))$, obtenemos, por las condiciones de regularidad sobre \mathcal{U}_k y con la misma notacion que en (11.4) del apendice,

$$\|B_k(p_0)\| \leq \int_0^1 \left(E_{p_0} \left(\left\| \frac{\nabla}{\partial s} \frac{\partial}{\partial t} c(1, t) \right\| \right) \right) dt = \int_0^1 E_{p_0} (\|J'(1)\|) dt$$

A partir de (11 4) del apendice sabemos que

$$J^{tan}(1) = (\nabla/\partial s)J^{tan}(1),$$

y

$$\begin{aligned} \langle \frac{\nabla}{\partial s} J^{tan}(1), J^{tan}(1) \rangle &= \langle J^{tan}(1), J^{tan}(1) \rangle \geq 0 \\ \langle \frac{\nabla}{\partial s} J^{tan}(1), J^{no}(1) \rangle &= -\langle \frac{\nabla}{\partial s} J^{nor}(1), J^{tan}(1) \rangle = 0 \end{aligned}$$

Tambien, por la proposicion (11 3 4) del apendice, podemos acotar la derivada covariante de la componente normal de J , y teniendo en cuenta la velocidad de la geodesica, obtenemos

$$\begin{aligned} \|J'(1)\| &\leq \|J(1)\| \left(\max \left(1, \|c'(1, t)\| \left(\frac{2S'_\kappa}{S_\kappa} - \frac{S'_\kappa}{S_\kappa} \right) (\|c'(1, t)\|) \right) \right) \\ &\leq \rho(\mathfrak{M}_{p_0}(\mathcal{U}_k), p_0) 2 \left(1 + \sqrt{|\kappa|} \rho(\mathcal{U}_k, p_0) \right), \end{aligned}$$

donde la segunda desigualdad se debe a que $S'_\kappa(t)/S_\kappa(t) \leq 1 + \sqrt{|\kappa|}t$, $\kappa \in \mathbf{R}$, como el lector puede comprobar facilmente a partir de sus definiciones en (11 2) de la subseccion (11 3) del apendice

Finalmente, ya que los momentos de segundo orden de \mathcal{U}_k estan uniformemente acotados

$$\|B_k(p_0)\| \leq C \rho(\mathfrak{M}_{p_0}(\mathcal{U}_k), p_0)$$

con $C = 2(1 + \sqrt{|\kappa|}) \sup_{k \in \mathbf{N}} E_{p_0}(\rho(\mathcal{U}_k, p_0))$, y tomando limites se sigue la proposicion ■

Observacion Notemos que de hecho solo necesitamos que el primer momento este uniformemente acotado Sin embargo mantenemos esta hipotesis a fin de ser coherentes con las condiciones del teorema (4 1 2) En cualquier caso, parece bastante sensato exigir esta condicion a cualquier buen estimador

Ejemplo 7 1 3 Para la distribucion exponencial univariante, vease ejemplo (3 4 1), tenemos que $B_k(p) = \Psi(k) - \log k$, donde $\Psi(k) = \Gamma'(k)/\Gamma(k)$, entonces, como

$\lim_{k \rightarrow \infty} k/e^{\Psi(k)} = 1$, resulta que el estimador maximo verosimil es asintoticamente intrinsecamente insesgado

Definicion 7 1 4 *Un estimador U es un estimador α consistente si y solo si*

$$\lim_{k \rightarrow \infty} E_{p_0} (\rho^\alpha(\mathcal{U}_k, p_0)) = 0 \quad \text{paracualquier } p_0 \in M$$

Teorema 7 1 5 *Sea \mathcal{U} un estimador α consistente para una familia parametrica regular, con $\alpha \geq 1$ Entonces \mathcal{U} es asintoticamente intrinsecamente insesgado*

Demostracion

Por hipotesis,

$$\lim_{k \rightarrow \infty} E_{p_0} (\rho^\alpha(\mathcal{U}_k, p_0)) = 0 \quad \text{con } \alpha \geq 1,$$

donde ρ es la distancia Riemanniana Asi pues, por la desigualdad de Jensen,

$$\lim_{k \rightarrow \infty} E_{p_0} (\rho(\mathcal{U}_k, p_0)) = 0,$$

y teniendo en cuenta que $\rho(\mathcal{U}_k, p_0) = \|A_{p_0}\|_{p_0}$, donde A_{p_0} es campo vectorial asociado al estimador, se sigue que

$$\lim_{k \rightarrow \infty} E_{p_0} (A_{p_0}) = 0,$$

obteniendo el resultado deseado ■

7 2 Normalidad asintotica en una variedad Estimacion maximo-verosimil

Introduciremos ahora la definicion de distribucion normal sobre una variedad Existen diversas formas de construir distribuciones sobre una variedad, para una completa y didactica exposicion vease Jupp y Mardia [32] Primero definimos vector aleatorio distribuido normalmente sobre el espacio tangente

Definicion 7 2 1 Sea Y un vector aleatorio en el espacio tangente M_p , donde p es un punto fijo de M . Diremos que Y esta normalmente distribuido si existe un vector η y un tensor 2 contravariante definido positivo Σ tal que para cualquier sistema de coordenadas, $(Y^\alpha) \sim N((\eta^\beta), (\Sigma^{\gamma\delta}))$. Escribiremos $Y \sim N(\eta, \Sigma)$.

Observacion Notemos que esta definicion es independiente del sistema de coordenadas. Esto se debe a que los parametros, η y Σ , en una distribucion normal, se transforman como un vector y un tensor 2 contravariante, respectivamente, cuando cambiamos el sistema de coordenadas del vector aleatorio Y . Podemos ahora definir el significado de una distribucion normal en una variedad completa M .

Como es frecuente a lo largo de la memoria, consideraremos solo objetos aleatorios Z que tomen valores, casi seguramente, en entornos regulares, vease definicion (2 1 5), de cualquier punto de la variedad completa M . Para este tipo de objetos aleatorios el campo vectorial aleatorio $\exp_p^{-1}(Z)$, $p \in M$ estara definido casi seguramente y siempre supondremos que $\exp_p^{-1}(Z)$, $p \in M$ esta definido en este sentido.

Definicion 7 2 2 Sea Z un objeto aleatorio a valores en una variedad completa M . Diremos que Z esta distribuido normalmente con respecto a p y parametros (η, Σ) , si existe un vector aleatorio $Y \sim N(\eta, \Sigma)$, en M_p , tal que $Z = \exp_p(Y)$. Escribiremos $Z \sim N(\eta, \Sigma)_p$.

Notemos que si $\eta = 0$ entonces $\mathfrak{M}(Z) = p$. La anterior definicion es de hecho una extension de las distribuciones normales "envueltas" que se utilizan en estadistica direccional [32]. Introduciremos ahora el concepto distribuciones asintoticamente normales en este contexto. Sea $\{Z_k\}_{k \in \mathbf{N}}$ una sucesion de variables aleatorias M valoradas, entonces

Definicion 7 2 3 Sea M una variedad completa, una sucesion de variables aleatorias $\{Z_k\}_{k \in \mathbf{N}}$ se dira que es s_k asintoticamente normal con media $p \in M$ si y solo si existe un tensor 2 contravariante definido positivo Σ y una sucesion de variables aleatorias $\{Y_k\}_{k \in \mathbf{N}}$ en M_p , tal que

$$\{s_k Y_k\}_{k \in \mathbf{N}} \xrightarrow{\mathcal{L}} Y \quad \text{con} \quad Y \sim N(0, \Sigma),$$

donde \mathcal{L} significa convergencia debil o en ley, y $\{s_k\}_{k \in \mathbf{N}}$ es una sucesion de numeros reales positivos con $\lim_{k \rightarrow \infty} s_k = \infty$

Observacion Señalemos que si $\{Z_k\}_{k \in \mathbf{N}}$ es s_k asintoticamente normal con media 0_p entonces

$$\left\{ \exp_p(s_k \exp_p^{-1}(Z_k)) \right\}_{k \in \mathbf{N}} \xrightarrow{\mathcal{L}} Z \quad \text{con} \quad Z \sim N(0, \Sigma)_p,$$

pero si tenemos $\{V_k\}_{k \in \mathbf{N}} \xrightarrow{\mathcal{L}} Z$ no es necesariamente cierto que $\exp_p^{-1}(V_k)$ converja en ley a una distribucion normal

Proposicion 7 2 4 Sea \mathcal{U} un estimador s_k asintoticamente normal, de una familia parametrica regular de distribuciones de probabilidad, con media $p_0 \in M$. Supongamos tambien que

$$\sup_{k \in \mathbf{N}} E_{p_0} \left(\rho^{1+\epsilon}(\mathcal{U}_k, p_0) \right) < \infty \quad \text{para un } \epsilon \in \mathbf{R}^+$$

Entonces, \mathcal{U} es asintoticamente intrinsecamente insesgado

Demostracion

Como \mathcal{U} es s_k asintoticamente normal entonces $\rho(p, \mathcal{U}_k) \xrightarrow{P} 0$. La sucesion de variables aleatorias $\rho(p_0, \mathcal{U}_k) = \|A_k(p_0)\|$ es uniformemente integrable de lo que se sigue que $E_{p_0}(\|A_k(p_0)\|) \rightarrow 0$ y ya que $0 < \|E_{p_0}(A_k(p_0))\| < E_{p_0}(\|A_k(p_0)\|)$, se deduce la proposicion ■

Teorema 7 2 5 Con las hipotesis anteriores, los estimadores maximo verosimiles son asintoticamente intrinsecamente insesgados

Demostracion

Esta es una consecuencia inmediata del teorema previo, suponiendo suficientes condiciones para asegurar la existencia, c, s , del campo tensorial del estimador $A_k(p) = \exp_p^{-1}(\mathcal{U}_k)$, y adicionalmente $\sup_{k \in \mathbf{N}} E_{p_0}(\rho^{1+\epsilon}(\mathcal{U}_k, p_0)) < \infty$, para un $\epsilon \in \mathbf{R}^+$, observando que los estimadores maximo verosimiles son \sqrt{k} asintoticamente normales. De hecho,

$$\sqrt{k} \exp_p^{-1}(\mathcal{U}_k) \xrightarrow{\mathcal{L}} N(0, (g^{\alpha\beta})),$$

donde $(g^{\alpha\beta})$ es la version contravariante del tensor metrico ■

7 3 Aproximacion del tensor de sesgo en un sistema de coordenadas

A partir de las ecuaciones de las geodesicas es facil obtener un desarrollo en potencias de la inversa de la aplicacion exponencial en un punto p de la variedad M . Las ecuaciones de las geodesicas en un entorno coordenado de un punto p y con vector unitario tangente u son

$$\frac{d^2 x^\alpha}{dt^2} + \Gamma_{ij}^\alpha \frac{dx^i}{dt} \frac{dx^j}{dt} = 0,$$

con $(x^\alpha(0)) = p, (\dot{x}^\alpha(0)) = u$. Por tanto

$$x^\alpha(0) + \Gamma_{ij}^\alpha(0)u^i u^j = 0$$

Entonces podemos obtener todas las derivadas en el punto p recursivamente

$$\begin{aligned} x^\alpha(0) &= -\Gamma_{ij}^\alpha(0)u^i u^j \\ \ddot{x}^\alpha(0) &= -\Gamma_{ij}^\alpha(0)u^i u^j - 2\Gamma_{ij}^\alpha(0)\dot{x}^i \dot{x}^j \\ &= -\Gamma_{ij}^\alpha(0)u^i u^j + 2\Gamma_{ij}^\alpha(0)\Gamma_{lr}^i(0)u^l u^r u^j \\ &= \left(-\partial_j \Gamma_{lr}^\alpha + 2\Gamma_{ij}^\alpha \Gamma_{lr}^i\right)(0)u^l u^r u^j, \end{aligned}$$

y asi sucesivamente

Por otra parte

$$x^\alpha(t) - x^\alpha(0) = \dot{x}^\alpha(0)t + \frac{1}{2}\ddot{x}^\alpha(0)t^2 + \frac{1}{6}\dddot{x}^\alpha(0)t^3 + O(t^4),$$

donde $f(t) = O(t^4)$ si $\lim_{t \rightarrow 0} (f(t)/t^4) = k > 0$ y utilizamos la convencion que, cuando la expresion, digamos, $O(t^4)$, es usada varias veces en un argumento, en cada ocasion puede que represente cantidades diferentes. Ademas,

$$A^\alpha = (\exp_p^{-1}(x(t)))^\alpha = tx^\alpha(0),$$

por tanto, tendremos

$$\Delta x^\alpha = A^\alpha - \frac{1}{2}\Gamma_{ij}^\alpha A^i A^j + \left(\frac{-1}{6}\partial_j \Gamma_{lr}^\alpha + \frac{1}{3}\Gamma_{ij}^\alpha \Gamma_{lr}^i \right) A^l A^r A^j + O(t^4)$$

Sin embargo, estamos de hecho interesados en expresar A como un desarrollo de potencias. Deberiamos así invertir la expresion precedente. Esto se puede realizar de forma iterativa. En primer orden

$$A^\alpha = \Delta x^\alpha + O(t^2)$$

En segundo orden

$$A^\alpha = \Delta x^\alpha + \frac{1}{2}\Gamma_{ij}^\alpha \Delta x^i \Delta x^j + O(t^3)$$

En tercer orden

$$A^\alpha = \Delta x^\alpha + \frac{1}{2}\Gamma_{ij}^\alpha \Delta x^i \Delta x^j + \frac{1}{6} \left(\partial_j \Gamma_{lr}^\alpha + \Gamma_{ij}^\alpha \Gamma_{lr}^i \right) \Delta x^l \Delta x^r \Delta x^j + O(t^4),$$

y así sucesivamente

Si generalizamos la notacion O a variables aleatorias, escribiendo $Y_k = O_P(X_k)$ si la sucesion de variables aleatorias $\{Y_k/X_k\}$ esta acotada en probabilidad, podemos decir lo siguiente

Proposicion 7 3 1 *Sea $(U, \theta(\cdot))$ una carta local, donde $\theta(p) = \theta_0$, \mathcal{U} tal que $\sqrt{k} \exp_p^{-1}(\mathcal{U}_k)$ converge en distribucion a un vector aleatorio con media cero y momentos de segundo orden. Entonces, si escribimos $\theta_{(k)} = \theta(\mathcal{U}_k)$*

$$\begin{aligned} A_k^\alpha(p)^\alpha &= \theta_{(k)}^\alpha - \theta_0^\alpha + \frac{1}{2}\Gamma_{ij}^\alpha (\theta_{(k)}^i - \theta_0^i)(\theta^j - \theta_0^j) \\ &\quad + \frac{1}{6} \left(\partial_j \Gamma_{lr}^\alpha + \Gamma_{ij}^\alpha \Gamma_{lr}^i \right) (\theta_{(k)}^l - \theta_0^l)(\theta^r - \theta_0^r)(\theta_{(k)}^j - \theta_0^j) + O_P(k^{-2}), \end{aligned}$$

donde $A_k^\alpha(p) = (\exp_p^{-1}(\mathcal{U}_k))^\alpha$, y los simbolos de Christoffel y sus derivadas estan calculadas en p

Demostracion

Antes de nada, como $\sqrt{k} \exp_p^{-1}(\mathcal{U}_k)$ converge en distribucion a un vector aleatorio con media cero y momentos de segundo orden, $\mathcal{U}_k \xrightarrow{P} p$ Entonces la norma del resto en el desarrollo de Taylor hasta el orden tres es $R_k t^4$, donde $R_k \xrightarrow{P} f(p)$, t puede elegirse como la longitud de arco, es decir $t = \|\exp_p^{-1}(\mathcal{U}_k)\|$ y $f(p)$ es una funcion que depende solo de p pero no de k

De las hipotesis y por teorema de Slutsky $k^2 t^4 \xrightarrow{\mathcal{L}} \|Y\|^4$, tal que Y tiene una distribucion con media cero y varianza Σ_0 Finalmente, ya que $k^2 t^4$ converge en distribucion $k^2 t^4 = O_P(1)$ y equivalentemente $t^4 = O_P(k^{-2})$ Entonces la proposicion se sigue ■

Con ciertas hipotesis obvias podemos decir algo similar para los momentos

Proposicion 7 3 2 *Con las hipotesis anteriores si*

$$\sup_{k \in \mathbf{N}} E(k^2 \rho^{3+\epsilon}(\theta_{(k)}, \theta_0)) < \infty$$

y los simbolos de Christoffel y sus derivadas estan uniformemente acotadas en el soporte de $\{\mathcal{U}_k\}$

$$B^\alpha(p) = \text{Sesgo}^\alpha(\theta_{(k)}) + \frac{1}{2} \Gamma_{ij}^\alpha \left\{ \text{Sesgo}^i(\theta_{(k)}) \text{Sesgo}^j(\theta_{(k)}) + \text{Cov}(\theta_{(k)}^i, \theta_{(k)}^j) \right\} + O(k^{-3/2}) \quad (7.1)$$

con $B^\alpha(p) = E_p(A^\alpha)$ y $\text{Sesgo}(\theta_{(k)}) = E_p(\theta_{(k)} - \theta_0)$

Desarrollos de Taylor invariantes

8.1 Introducción

El cálculo intrínseco requiere en muchas ocasiones aproximar cantidades cuyo cálculo exacto resulta prácticamente imposible. Los desarrollos de Taylor son una buena herramienta para tal fin en el caso de \mathbf{R}^n pero en nuestro caso necesitamos aproximar funciones a valores en una variedad. Como siempre una primera aproximación a la solución podría consistir en considerar cartas locales en M , las funciones en M se transforman en funciones reales de variable real y ahora podríamos considerar el desarrollo de Taylor correspondiente. Por su puesto esta forma de proceder proporcionaría aproximaciones que dependerían del sistema de coordenadas utilizado. Se hace pues necesario construir una teoría geométrica de series de Taylor. Ya ha habido trabajos previos en tal sentido y en conexión con la Estadística, véase por ejemplo Barndorff-Nielsen [8], McCullagh y Cox [45] y Murray [48]. Los procedimientos para construir lo que llamaremos desarrollos de Taylor invariantes son diferentes, así Barndorff-Nielsen utilizando unos nuevos objetos algebraicos llamados “strings”, que generalizan la noción de tensor, define los “connection strings” naturales asociados a una divergencia (un “yoke” en su terminología) y a partir de aquí utilizando las “derivadas tensoriales” obtiene desarrollos de Taylor invariantes. Este enfoque es una generalización de los desarrollos de McCullagh y Cox. Por último Murray define un string coordinado, que es una colección de sistemas “privilegiados” tal que los cambios de uno a otro son lineales.

8 2 Principales resultados

Usaremos los operadores *derivada covariante* y *derivada general covariante*, vease apéndice (11 1), para obtener desarrollos de Taylor invariantes. De hecho tenemos las siguientes proposiciones

Proposición 8 2 1 *Sea $f \in \mathcal{F}_A^{(0,0)}$ y $U_p \subset A$ un entorno regular normal de $q \in M$. Entonces para cualquier $p \in U_q$ tenemos el siguiente desarrollo de Taylor de orden n*

$$f(p) = f(q) + \sum_{k=1}^n \frac{1}{k!} (D^k f)(\exp_q^{-1}(p), {}^k, \exp_q^{-1}(p)) + \frac{1}{(n+1)!} (D^{n+1} f)(\exp_q^{-1}(p), {}^{n+1}, \exp_q^{-1}(p)), \quad (8 1)$$

donde $D^k f = D(D^{k-1} f)$ y \tilde{q} es un punto del segmento geodésico que va de q a p

Demostración

Sea $\gamma(t)$ un segmento geodésico tal que $\gamma(0) = q$ y $\gamma(t) = p$. Obviamente podemos hacer el siguiente desarrollo de Taylor invariante de orden n

$$f(p) = f(q) + \sum_{k=1}^n \frac{d^k}{dt^k} (f \circ \gamma)(0) t^k + \frac{d^{n+1}}{dt^{n+1}} (f \circ \gamma)(0) (t - \tilde{t})^k,$$

con $0 \leq \tilde{t} \leq t$, entonces vamos a ver que

$$\frac{d^k}{dt^k} (f \circ \gamma) = (D^k f)(T, {}^k, T), \quad (8 2)$$

donde $T = \gamma_* \left(\frac{d}{dt} \right)$, es decir el campo vectorial tangente a la línea geodésica $\gamma(t)$

Si $k = 1$

$$\frac{d}{dt} (f \circ \gamma) = \gamma_* \left(\frac{d}{dt} \right) f = Tf = D_T f = (Df)(T)$$

Supongamos que (8 2) es cierto para $k = m$, entonces

$$\begin{aligned} \frac{d^{m+1}}{dt^{m+1}} (f \circ \gamma) &= \frac{d}{dt} \left(\frac{d^m}{dt^m} (f \circ \gamma) \right) = \frac{d}{dt} \left((D^m f)(T, {}^m, T) \right) \\ &= D_T \left((D^m f)(T, {}^m, T) \right) = (D_T (D^m f))(T, {}^m, T) \\ &= (D^{m+1} f)(T, {}^{m+1}, T) \end{aligned}$$

Donde hemos usado el hecho que $D_T T = 0$ Finalmente teniendo en cuenta que $tT(0) = \exp_q^{-1}(p)$, se sigue la proposición ■

Proposición 8.2.2 Sea $\theta \in \mathcal{F}_A^{(s)}$, entonces con la misma notación e hipótesis que en la proposición anterior, y $w_1, \dots, w_r \in \mathcal{F}_A^{(0,1)}$, $Y_1, \dots, Y_r \in \mathcal{F}_A^{(1,0)}$ campos tensoriales paralelos a lo largo de la línea geodésica que une q a p

$$\begin{aligned} \theta(w_1, \dots, w_r, Y_1, \dots, Y_s)(p) &= \theta(w_1, \dots, w_r, Y_1, \dots, Y_s)(q) \\ &+ \sum_{k=1}^n \frac{1}{k!} (D^k \theta)(w_1, \dots, w_r, Y_1, \dots, Y_s, \exp_q^{-1}(p), \dots, \exp_q^{-1}(p))(q) \\ &+ \frac{1}{(n+1)!} (D^{n+1} \theta)(w_1, \dots, w_r, Y_1, \dots, Y_s, \exp_q^{-1}(p), \dots, \exp_q^{-1}(p))(\tilde{q}) \end{aligned} \quad (8.3)$$

Demostración

Tomando $f = \theta(w_1, \dots, w_r, Y_1, \dots, Y_s)$ y ya que

$$D_T w_i = 0, \quad D_T Y_j = 0, \quad i = 1, \dots, r, \quad j = 1, \dots, s$$

tenemos,

$$D^k \theta(w_1, \dots, w_r, Y_1, \dots, Y_s)(T, \dots, T) = (D^k \theta)(w_1, \dots, w_r, Y_1, \dots, Y_s, T, \dots, T)$$

y la proposición queda demostrada ■

Notese que el anterior resultado lo podemos escribir

$$\begin{aligned} \tau_p^q(\theta(p)) &= \theta(q) + \sum_{i=1}^n \frac{1}{i!} (D^i \theta)(\exp_q^{-1}(p), \dots, \exp_q^{-1}(p)) \\ &+ \frac{1}{(n+1)!} \tau_q^q((D^{n+1} \theta)(\exp_q^{-1}(p), \dots, \exp_q^{-1}(p))) \end{aligned}$$

donde τ_p^q indica el transporte paralelo de p a q a lo largo de la geodésica que une p y q

Proposición 8 2 3 Sean p, q, r puntos de un conjunto regular convexo de M , $\gamma(t)$ la geodesica que une q a p parametrizada por la longitud de arco y considere los campos

$$B(m) = \exp_m^{-1}(r), \quad A(m) = \exp_m^{-1}(p), \quad m \in M$$

entonces

$$\|B\|^2(p) = \|B\|^2(q) - 2\langle A, B \rangle(q) + \langle A, \frac{\nabla}{\partial s} J(1) \rangle \|A\|(q),$$

donde $J(s)$ es el campo de Jacobi a lo largo de la geodesica

$$s \mapsto \exp_r \left(s \exp_r^{-1}(\gamma(t)) \right)$$

determinado por $J(0) = 0$ y $J(1) = \gamma(t) = A/\|A\|(\gamma(t))$

Demostración

En primer lugar, tenemos

$$(D\|B\|^2)(A) = \|A\|(D\|B\|^2)(T) = \|A\|D_T(\|B\|^2) \quad (8.4)$$

$$= 2\|A\|\langle \nabla_T B, B \rangle = -2\|A\|\langle T, B \rangle, \quad (8.5)$$

donde $T = A/\|A\|$. Por tanto

$$\begin{aligned} (D^2\|B\|^2)(A, A) &= \|A\|^2 (D_T(D\|B\|^2)(T)) \\ &= \|A\|^2 (D_T(D\|B\|^2)(T)) = -2\|A\|^2 \langle T, \nabla_T B \rangle \end{aligned}$$

Ahora, supongamos que

$$(D^{m-1}\|B\|^2)(A, {}^{m-1}A) = -2\|A\|^{m-1} \langle T, \nabla_T^{m-2} B \rangle,$$

entonces,

$$\begin{aligned} (D^m\|B\|^2)(A, {}^m A) &= \|A\|^m (D_T(D^{m-1}\|B\|^2)(T, {}^{m-1}T)) \\ &= \|A\|^m D_T((D^{m-1}\|B\|^2)(T, {}^{m-1}T)) = -2\|A\|^m \langle T, \nabla_T^{m-1} B \rangle \end{aligned}$$

Consideremos la familia uniparametrica de geodesicas dada por

$$c(s, t) = \exp_r \left(s \exp_r^{-1}(\gamma(t)) \right), \quad \gamma(0) = q, \gamma(1) = p$$

entonces ya que $B(\gamma(t)) = \exp_{\gamma(t)}^{-1}(r)$, y a lo largo de $\gamma(t)$,

$$\nabla_T B = \nabla_T \exp_{\gamma(t)}^{-1}(r) = -\nabla_T \left(c \left(\frac{\partial}{\partial s} \Big|_{s=1} \right) \right) \quad (8.6)$$

$$= -\frac{\nabla}{\partial s} c \left(\frac{\partial}{\partial t} \Big|_{s=1} \right) = -\frac{\nabla}{\partial s} J(1), \quad (8.7)$$

donde $J(s)$ es el campo de Jacobi en la geodesica $s \mapsto \exp_r(s \exp_r^{-1}(\gamma(t)))$ determinado por $J(0) = 0$ y $J(1) = \gamma(t) = T(\gamma(t))$

Vamos a ver por ultimo que a lo largo de $\gamma(t)$

$$\langle T, \nabla_T^m B \rangle = 0, \quad \forall m \geq 2$$

A lo largo de $\gamma(t)$,

$$\nabla_T^2 B = -\nabla_T \frac{\nabla}{\partial s} J(1) = -\frac{\nabla}{\partial s} \nabla_T J(1) - R(T, c')T$$

donde $c' = c \left(\frac{\partial}{\partial s} \right)$ Ahora, como

$$\nabla_T J(1) = \nabla_T T = 0 \quad \text{y} \quad \langle R(T, c')T, T \rangle = 0,$$

obtenemos

$$\langle T, \nabla_T^2 B \rangle = 0$$

aplicando ahora la proposicion anterior tenemos el resultado deseado

Corolario 8.2.4 *Supongamos la mismas condiciones que en la proposicion anterior. Si Δ y δ son cotas superiores e inferiores de la curvatura*

$$\begin{aligned} \rho^2(p, r) \geq & \rho^2(q, r) - 2\rho(q, p)\rho(q, r) \cos \alpha + \rho^2(q, p) \cos^2 \alpha \\ & + \rho^3(q, p) \frac{f'_\Delta}{f_\Delta}(\rho(q, p)) \sin^2 \alpha \end{aligned} \quad (8.8)$$

y

$$\begin{aligned} \rho^2(p, r) \leq & \rho^2(q, r) - 2\rho(q, p)\rho(q, r) \cos \alpha + \rho^2(q, p) \cos^2 \alpha \\ & + \rho^3(q, p) \frac{f'_\delta}{f_\delta}(\rho(q, p)) \sin^2 \alpha, \end{aligned}$$

donde

$$f_K(t) = \begin{cases} \frac{\sin(\sqrt{K}t)}{\sqrt{K}} & \text{if } K > 0, \\ t & \text{if } K = 0, \\ \frac{\sinh(\sqrt{-K}t)}{\sqrt{-K}} & \text{if } K < 0 \end{cases} \quad (8.9)$$

y α es el ángulo entre las geodésicas que unen q a p y a r

Demostración

Por la proposición anterior

$$\|B\|^2(p) = \|B\|^2(q) - 2\langle A, B \rangle(q) + \langle A, \frac{\nabla}{\partial s} J(1) \rangle \|A\|(q),$$

entonces

$$\begin{aligned} \langle A, \frac{\nabla}{\partial s} J(1) \rangle(q) &= \|A\|(q) \langle J(1), \frac{\nabla}{\partial s} J(1) \rangle = \langle J^{\tan}(1), J^{\tan}(1) \rangle \\ &\quad + \langle \frac{\nabla}{\partial s} J^{\text{nor}}(1), J^{\text{no}}(1) \rangle \end{aligned}$$

Entonces por los teoremas de comparación, véase la sección (11.3) del apéndice,

$$\begin{aligned} \rho(q, p) \frac{f'_\Delta}{f_\Delta}(\rho(q, p)) \|J^{\text{nor}}\|^2(1) \\ \leq \langle \frac{\nabla}{\partial s} J^{\text{nor}}(1), J^{\text{nor}}(1) \rangle \leq \rho(q, p) \frac{f'_\delta}{f_\delta}(\rho(q, p)) \|J^{\text{nor}}\|^2(1), \end{aligned}$$

como queríamos demostrar ■

Corolario 8.2.5 Sea (χ, \mathbf{a}, P) un espacio de probabilidad y $f: \chi \rightarrow M$ una aplicación medible, tal que $f(\chi)$ es un conjunto regular convexo en M . Sea $p_0 \in M$ un valor medio de f

$$\int_{\chi} \exp_{p_0}^{-1}(f(x)) P(dx) = 0$$

y sea $d(M) \leq \pi/\sqrt{\mathcal{K}}$ el diámetro de la variedad M si el supremo de las curvaturas seccionales $\mathcal{K} > 0$, si $\mathcal{K} \leq 0$ consideraremos $d(M) \leq \infty$, entonces

$$E(\rho^2(p, f)) \geq E(\rho^2(p_0, f)) + \rho^2(p_0, p) \min\{1, \rho(p_0, p) \frac{f'_\mathcal{K}}{f_\mathcal{K}}(\rho(p_0, p))\}$$

Demostracion

Es directa a partir de (8 8), tomando esperanzas, y teniendo en cuenta las definiciones en (8 9) ■

Corolario 8 2 6 *Supongamos la mismas condiciones que en el corolario previo, entonces p_0 es un minimo local de*

$$\mathcal{H}^2(q) = \int_{\mathcal{X}} \rho^2(q, f(x))P(dx),$$

y la distancia entre dos minimos locales de $\mathcal{H}^2(q)$ debe ser mayor o igual que $\pi/\sqrt{\mathcal{K}}$, siendo $\mathcal{K} > 0$ el supremo de las curvaturas seccionales. Si $\mathcal{K} \leq 0$ entonces el minimo es unico

Demostracion

La demostracion se sigue del corolario anterior ■

8 3 Un teorema de Pitagoras para yokes

Sea $g \in C^\infty(M \times M)$ Dada una carta local (Θ, \mathcal{V}) , introducimos la siguiente notacion

$$g_{r_1 \dots r_l s_1 \dots s_m}(p, \tilde{p}) = \partial_{r_1} \dots \partial_{r_l} \tilde{\partial}_{s_1} \dots \tilde{\partial}_{s_m} g(p, \tilde{p})$$

donde

$$\partial_r g = \frac{\partial}{\partial \theta^r} g(\cdot, \tilde{p}) \quad \text{y} \quad \tilde{\partial}_s g = \frac{\partial}{\partial \theta^s} g(p, \cdot)$$

para cualquier $p, \tilde{p} \in \mathcal{U}$ Tambien escribimos

$$g_{r_1 \dots r_l s_1 \dots s_m}(p) = g_{r_1 \dots r_l s_1 \dots s_m}(p, p)$$

Entonces supongamos que, para todo $p \in M$, y dado un sistema de coordenadas O , se cumplen las dos siguientes condiciones

$$(1) \quad \not{g}_k(p) = 0, \quad k = 1, \dots, n \quad (8.10)$$

$$(ii) \text{ la matriz} \quad \{-\not{g}_{ij}(p)\} \quad (8.11)$$

es no-singular

Notemos que diferenciando la ecuacion (8.10) se obtiene

$$\not{g}_{ij} + \not{g}_{ji} = 0 \quad (8.12)$$

y esto junto con (8.11) muestra que la matriz $\{\not{g}_{ij}\}$ es simetrica. Tambien $\{\not{g}_{ij}\}$ se comporta como tensor 2-covariante, y proporciona una metrica pseudo-Riemanniana en M

Esta clase de funciones, ya mencionadas al principio de la memoria, fueron introducidas por Barndorff Nielsen, vease [5], y llamadas *yokes*. Lo que veremos es, para un caso particular, y con un punto de vista diferente al que se presenta en [5], un *yoke* da origen a una familia de conexiones afines

Sean

$$h_\alpha(\theta(t), \theta(\tilde{t})) = \frac{1-\alpha}{2}g(\theta(t), \theta(\tilde{t})) + \frac{1+\alpha}{2}g(\theta(\tilde{t}), \theta(t))$$

y supongamos que existen curvas $\theta(t)$ en M tales que

$$\text{grad}(h_\alpha)(\theta(t), \theta(\tilde{t})) = \theta(t)(\tilde{t} - t), \quad (8.13)$$

donde el gradiente esta calculado con respecto al primer argumento

Entonces derivando dos veces con respecto a \tilde{t} obtenemos

$$\begin{aligned} & \not{g}^{nk}(\theta(t)) \left(\frac{(1-\alpha)}{2}g_{k\ lm}(\theta(t), \theta(\tilde{t})) + \frac{(1+\alpha)}{2}g_{lm\ k}(\theta(\tilde{t}), \theta(t)) \right) \theta^l(\tilde{t})\theta^m(\tilde{t}) \\ & + \not{g}^{nk}(\theta(t)) \left(\frac{(1-\alpha)}{2}g_{k\ l}(\theta(t), \theta(\tilde{t})) + \frac{(1+\alpha)}{2}g_{l\ k}(\theta(\tilde{t}), \theta(t)) \right) \theta^l(\tilde{t}) = 0, \end{aligned}$$

donde \not{g}^{nk} es la matriz inversa $\not{g}_{i\ k}$. Finalmente haciendo $\tilde{t} = t$, obtenemos

$$\theta^i(t) + \overset{\alpha}{\Gamma}_{lm} \theta^l(t)\theta^m(t) = 0,$$

donde

$$\overset{\alpha}{\Gamma}_{lm} = \not{g}^{nk} \left(\frac{1-\alpha}{2} \not{g}_{k\ lm} + \frac{1+\alpha}{2} \not{g}_{lm\ k} \right)$$

Podemos usar (8.13) para hacer un desarrollo de Taylor invariante de h_α

Proposición 8 3 1 Sean p, q, r puntos de M , $\gamma(t)$ el $-\alpha$ -segmento geodesico de q a p , entonces

$$h_\alpha(p, r) = h_\alpha(q, r) + \langle \bar{\text{exp}}_q^{-\alpha}(p), \bar{\text{exp}}_q^{-\alpha}(r) \rangle - \langle \bar{\text{exp}}_q^{-\alpha}(p), \frac{\nabla}{\partial s} J(1) \rangle t + o(t^2)$$

donde $J(s)$ es el campo de Jacobi a lo largo de la geodesica

$$s \mapsto \bar{\text{exp}}_r^\alpha(s \bar{\text{exp}}_r^{-\alpha}(\gamma(t)))$$

determinada por $J(0) = 0$ y $J(1) = \gamma(t) = \left. \frac{d}{dt} \gamma(t) \right|_{t=0}$

Demostración

Evidentemente podemos escribir

$$\frac{d}{dt} h_\alpha(\gamma(t), r) = \langle T, \text{grad}(h_\alpha)(\gamma(t), r) \rangle,$$

siendo T el vector tangente a $\gamma(t)$ Entonces, a partir de (8 13) tenemos

$$\frac{d}{dt} h_\alpha(\gamma(t), r) = \langle T, \bar{\text{exp}}_{\gamma(t)}^{-\alpha}(r) \rangle,$$

y

$$\frac{d^k}{dt^k} h_\alpha(\gamma(t), r) = \langle T, \bar{\nabla}_T^{\alpha k} \bar{\text{exp}}_{\gamma(t)}^{-\alpha}(r) \rangle,$$

Podemos usar este hecho para obtener un desarrollo de Taylor invariante de $h_\alpha(\cdot, r)$ en un entorno de q

$$h_\alpha(p, r) = h_\alpha(q, r) + \sum_{k=1}^n \frac{1}{k!} \frac{d^k}{dt^k} h_\alpha(\gamma(t), r) \Big|_{t=0} t^k + o(t^n),$$

entonces, tenemos

$$h_\alpha(p, r) = h_\alpha(q, r) + \langle T, \bar{\text{exp}}_{\gamma(0)}^{-\alpha}(r) \rangle t + \sum_{k=1}^{\infty} \langle T, \bar{\nabla}_T^{\alpha k} \bar{\text{exp}}_{\gamma(t)}^{-\alpha}(r) \rangle t^k$$

Finalmente razonando como en la proposición (8 2 3),

$$\bar{\nabla}_T \bar{\text{exp}}_{\gamma(t)}^{-\alpha}(r) = -\frac{\nabla}{\partial s} J(1),$$

donde $J(s)$ es el campo de Jacobi en la geodesica $s \mapsto \bar{\text{exp}}_r^\alpha(s \bar{\text{exp}}_r^{-\alpha}(\gamma(t)))$ determinado por $J(0) = 0$ y $J(1) = T$, y la proposición queda demostrada. ■

Corolario 8 3 2 Si M es α -plano entonces

$$h_\alpha(p, r) = h_\alpha(q, r) + \langle \bar{\text{exp}}_q^{-1}(p), \text{exp}_q^{-1}(r) \rangle + h_\alpha(p, q)$$

Demostracion

Si M es α -plano

$$\frac{\nabla^{\alpha^2}}{\partial s^2} J = 0$$

entonces los dos primeros terminos dependen de r , de hecho

$$\frac{\nabla^\alpha}{\partial s} J = T$$

tendiendo r a q el corolario queda demostrado ■

Notese que, si M es α -plano entonces

$$h_\alpha(p, q) = -\frac{1}{2} \langle \bar{\text{exp}}_q^{-1}(p), \bar{\text{exp}}_q^{-1}(p) \rangle + O(t^2)$$

Es relacion con estos resultados vease Amari [2]

Eficiencias asintóticas

9 1 Introduccion

Desde el punto de vista clasico una forma de comparar estimadores T_k y T'_k de θ que son \sqrt{k} -asintoticamente normales con media θ , es a traves de la varianza asintotica. Se denomina *eficiencia relativa asintotica* de T_k con respecto a T'_k al cociente

$$e_{TT'} = \frac{\text{Varianza asint de } T'}{\text{Varianza asint de } T}$$

Sin embargo ni todos los estimadores son asintoticamente normales ni parece apropiado comparar las varianzas asintoticas de distribuciones de diferentes familias

Este criterio de comparacion se puede extender al caso en que los estimadores converjan a una ley de la misma familia aunque esta no sea necesariamente normal. Conviene decir sin embargo que si los estimadores son suficientemente regulares su distribucion asintotica, bajo muestreo aleatorio simple, sera la distribucion normal.

Asimismo podriamos establecer un procedimiento de comparacion de *riesgos* asintoticos. Supongamos que $k^r R(T_k, \theta)$ y $k^r R(T'_k, \theta)$ tienen un limite finito distinto de cero, entonces podemos definir la eficiencia de T_k con respecto a T'_k como

$$\lim_{k \rightarrow \infty} \frac{R(T'_k, \theta)}{R(T_k, \theta)}$$

Obviamente depende de la funcion de perdida que se utilice

Si utilizamos el error cuadrático, podemos calcular

$$\liminf_{k \rightarrow \infty} k E(T_k - \theta)^2$$

En el caso insesgado este límite estará acotado inferiormente por la cota de Cramer–Rao y nos podemos preguntar sobre su relación con *la varianza asintótica* es decir la varianza de la distribución asintótica. Tenemos el siguiente resultado

Proposición 9.1.1 Sea Y_k , $k = 1, 2, \dots$ una sucesión de variables aleatorias que tiende en ley a otra variable Y con $E(Y^2) = v^2$. Entonces

$$v^2 \leq \liminf_{k \rightarrow \infty} E(Y_k^2)$$

Condiciones para que la igualdad se cumpla han sido dadas por Chernoff, [23]. Para el caso especial en que T_k es una función de una media muestral de variables i.i.d. $g(\bar{X}_k)$ condiciones suficientes para que se cumpla la igualdad son

- i) X_1 tiene momento de cuarto orden finito
- ii) Las derivadas de cuarto orden de $g(x)$ y $g^2(x)$ están acotadas.
- iii) La derivada de $g(x)$ en $x = E(X_1)$ es distinta de cero.

Ver Lehmann, [41], página 106

Estos procedimientos de comparación sugieren criterios de optimalidad. Por ejemplo, ¿existen estimadores cuya varianza o riesgo asintóticos sean uniformemente inferiores a la de cualquier otro, de manera que podamos hablar de estimadores asintóticamente eficientes? Notemos en primer lugar que no vamos a obtener la unicidad que se presenta para muestras de tamaño fijo, por el simple hecho de que un mismo límite es compartido por diferentes sucesiones. Mas concretamente, si

$$\sqrt{k}(T_k - \theta) \xrightarrow{w} N(0, v)$$

con T_k óptimo en el sentido de que minimiza v , entonces $T_k + R_k$ es también óptimo siempre que

$$\sqrt{k}R_k \xrightarrow{P} 0$$

Mas en general, por el teorema de Slutsky, tenemos

Proposición 9 1 2 Si $T_k \xrightarrow{w} X$, entonces $T_k + o_P(1) \xrightarrow{w} X$

Si la varianza asintótica coincidiera con el límite de la varianza del estimador multiplicada por el tamaño muestral y este fuera *insesgado*, tendríamos una cota inferior para la varianza asintótica y esta vendría dada por la desigualdad de Cramer–Rao

$$v(\theta) \geq \frac{1}{I(\theta)}$$

Es decir si

$$\text{Var} \{ \sqrt{k}(T_k - \theta) \} \rightarrow v(\theta)$$

y si T_k es insesgado entonces

$$v(\theta) \geq \frac{1}{I(\theta)}$$

Entonces un estimador T_k que satisfaga

$$v(\theta) = \frac{1}{I(\theta)}$$

diremos que es asintóticamente eficiente o **eficiente de primer orden**

Durante mucho tiempo se pensó que el problema de que la varianza asintótica de un estimador consistente pudiera ser inferior a la cota de Cramer–Rao era un problema de *regularidad de la familia*. Sin embargo Hodges (1951) propuso un estimador *supereficiente* para la familia $N(\theta, 1)$

$$T_k = \begin{cases} \bar{X} & \text{si } |\bar{X}| \geq 1/k^{1/4} \\ a\bar{X} & \text{si } |\bar{X}| < 1/k^{1/4} \end{cases}$$

Es fácil ver que $v(\theta) = 1$ si $\theta \neq 0$ y $v(\theta) = a^2$ si $\theta = 0$. Por tanto si $a < 1$ tenemos un estimador *supereficiente*, este ejemplo fue presentado por Le Cam en 1953 [18]

Sin embargo tenemos el siguiente importante teorema debido a Bahadur (1963) [4], (aunque Le Cam [18] lo anticipó de otra forma)

Teorema 9 1 3 Sean X_1, \dots, X_k i.i.d. con densidad $f(x, \theta)$ con respecto a una medida σ -finita μ , donde θ toma valores en \mathbf{R} y supongamos las siguientes condiciones de regularidad

i El espacio de parámetros es un intervalo abierto

ii $f(x, \theta)$ es C^2

iii Las derivadas parciales respecto de θ y la integracion respecto a $d\mu$ se pueden intercambiar al menos hasta orden dos

iv La informacion de Fisher $I(\theta)$ satisface $0 < I(\theta) < \infty$

v Para todo θ_0 existe un numero positivo c y una funcion $M(x)$ (ambos pueden depender de θ_0) tales que

$$\left| \frac{\partial^2 \log f(x, \theta)}{\partial \theta^2} \right| \leq M(x)$$

para todo x del interior del soporte y $\theta \in (\theta_0 - c, \theta_0 + c)$, con

$$E_{\theta_0}[M(X)] < \infty$$

Bajo estas suposiciones si

$$\sqrt{k}(T_k - \theta) \xrightarrow{w} N(0, v(\theta))$$

entonces $v(\theta)$ satisface la desigualdad

$$v(\theta) \geq \frac{1}{I(\theta)}$$

salvo en un conjunto de medida de Lebesgue cero

Por tanto y aunque las condiciones de regularidad no pueden evitar el fenomeno de la supereficiencia este se produce en conjuntos de de medida de Lebesgue nula. Una forma obvia de impedir la supereficiencia seria imponer la condicion

$$\text{Var} \left\{ \sqrt{k}(T_k - \theta) \right\} \rightarrow v(\theta)$$

junto con

$$b'_k(\theta) \rightarrow 0,$$

con $b_k(\theta) = T_k(\theta) - \theta$

Tambien es inmediato que si $I(\theta)$ es continua, condicion que exigiremos, y se cumplen las condiciones del teorema anterior la supereficiencia solo se produce en puntos de discontinuidad de la varianza asintotica

Un teorema similar, del mismo Bahadur [4], se puede enunciar para el caso multiparametrico

Es interesante hacer notar que la supereficiencia es un fenomeno patologico antes que una prescripcion para elegir un estimador. Se puede ver que el maximo error cuadratico medio de un estimador supereficiente es mayor que el de otros estimadores. Esto sugiere eliminar el fenomeno de la supereficiencia con un planteamiento *minimax*, vease Ibragimov y Has'minskiı [28] paginas 90-92

Con las reservas mencionadas podemos hablar de estimador optimo, al menos en primer orden, como aquel cuya varianza asintotica alcanza la cota de Cramer-Rao. En este sentido se puede demostrar que las ecuaciones de verosimilitud siempre tienen una solucion (a partir de un cierto tamaño muestral) que es asintoticamente eficiente. Ahora bien, a las condiciones de regularidad habituales, hay que añadir que en un entorno del verdadero valor, el valor absoluto de las derivadas terceras del logaritmo de la verosimilitud este acotado por una variable aleatoria cuya esperanza exista. Esta condicion parece necesaria para construir una teoria asintotica que contenga resultados interesantes así que tambien la supodremos mas adelante

Como ya hemos dicho pueden existir muchos estimadores eficientes de primer orden, pues bastaria modificar el estimador maximo-verosimil con un termino que fuera $o_P(k^{-1/2})$. La consiguiente pregunta es como distinguir entre ellos. Para ello se ha desarrollado la teoria de eficiencia de orden superior. Esta teoria se ha desarrollado para subfamilias de una familia exponencial

$$p(x, \xi(\theta)) = \exp \{ \xi^*(\theta)x_i - \psi(\xi(\theta)) \}$$

donde $\xi(\theta)$ es una funcion vectorial de θ y la dimension del parametro θ es menor que la del parametro ξ

Entonces si consideramos estimadores de la forma $\phi(\bar{x})$ que sean *consistentes* (que por tanto deberan ser consistentes en el sentido de Fisher, esto es $\phi(\xi(\theta)) = \theta$) se obtiene un desarrollo del error cuadratico medio del estimador en potencias de $k^{-1/2}$ de la forma

$$E_{\theta}(k(\phi^a - \theta^a)(\phi^b - \theta^b)) = g_1^{ab}(\theta) k^{-1/2} + g_2^{ab}(\theta) k^{-1} + O(k^{-3/2})$$

Entonces un estimador es eficiente de primer orden si $g_1^{ab}(\theta)$ es minimal para todo θ entre los demas estimadores consistentes. Un estimador es *eficiente de segundo*

orden cuando $g_2^{ab}(\theta)$ sea minimal para todo θ entre los estimadores eficientes de primer orden, y así sucesivamente

Los resultados mas importantes son

- 1 Si corregimos el sesgo de un estimador eficiente para que este sea de orden $1/k^2$, entonces todo estimador eficiente de primer orden lo es de segundo orden
- 2 El estimador maximo-verosimil es eficiente de tercer orden si corregimos su sesgo (entre los estimadores con sesgo corregido)

Estos resultados pueden encontrarse en Efron [24], para el caso unidimensional, y Amari [2] y Madsen [43] para el caso multidimensional. El proposito de este capitulo es obtener resultados intrinsecos analogos a los anteriores

9 2 Eficiencia intrinseca de primer orden

Sea un estimador, o sucesion de estimadores, \mathcal{U} de la funcion de densidad (o medida de probabilidad) verdadera $p = p(\cdot, \theta) \in M$ del modelo parametrico regular, esto es una familia de aplicaciones medibles

$$\mathcal{U} = \{\mathcal{U}_k \mid \chi^k \mapsto M, k \in \mathbf{N}\}$$

tal que la verdadera medida de probabilidad en χ^k es

$$(P)_k(dx) = p_{(k)}(x, \theta) \mu_k(dx) = p(x_1, \theta) \mu_k(dx)$$

Sea $C_k(x, \theta) = \text{grad}(\log p_{(k)}(x, \theta))$, $A_k(x, \theta) = \exp_{\theta}^{-1}(\mathcal{U}_k)$ y $B_k(\theta) = E(A_k)$

Proposicion 9 2 1 Con las notaciones y condiciones habituales de regularidad

$$E\left(\rho^2(\mathcal{U}_k, \theta)\right) = E\left\|A_k - \frac{1}{k}C_k\right\|^2 - \frac{2E(\text{div}(A_k)) + n}{k} + \frac{2\text{div}(B_k)}{k} \quad (9 1)$$

Demostracion

$$\begin{aligned} E\left\|A_k - \frac{1}{k}C_k\right\|^2 &= E\|A_k\|^2 + \frac{1}{k^2}E\|C_k\|^2 - \frac{2}{k}E\langle A_k, C_k \rangle \\ &= E\left(\rho^2(\mathcal{U}_k, \theta)\right) + \frac{n}{k} - \frac{2}{k}(E(\text{div}(B_k)) - E(\text{div}(A_k))) \end{aligned}$$

■

Proposición 9 2 2 Si $\{\mathcal{U}_k\}$ es un estimador regular (en el sentido del teorema de Cramer–Rao intrínseco) consistente de primer orden, y las curvaturas seccionales del modelo están acotadas superiormente de manera que el diámetro de la variedad $d(M) \leq \pi/\sqrt{\mathcal{K}}$, si el supremo de las curvaturas seccionales $\mathcal{K} \geq 0$ y $d(M) \leq \infty$ si $\mathcal{K} \leq 0$, entonces (salvo conjuntos de medida Riemanniana cero)

$$\liminf_{k \rightarrow \infty} k E \rho^2(\mathcal{U}_k, \theta) \geq n$$

Demostración

Veamos primero que $\liminf_{k \rightarrow \infty} E(\operatorname{div}(A_k)) \leq -n$

Supongamos que las curvaturas seccionales están acotadas superiormente por $\mathcal{K} \leq 0$, entonces por (11 8) en la subsección (11 5) del apéndice, tenemos

$$E(\operatorname{div}(A_k)) \leq -1 - (n-1)\sqrt{-\mathcal{K}} E(\|B_k\| \coth(\sqrt{\mathcal{K}}\|B_k\|)),$$

y tomando $\liminf_{k \rightarrow \infty}$, como $\|B_k\| \leq E\|A_k\| \rightarrow 0$, obtenemos el resultado. Si el supremo de las curvaturas seccionales es $\mathcal{K} > 0$. Entonces por (11 9) en la subsección (11 5) del apéndice, tenemos

$$E(\operatorname{div}(A_k)) \leq -1 - (n-1)\sqrt{\mathcal{K}} E(\|A_k\| \cot(\sqrt{\mathcal{K}}\|A_k\|))$$

y teniendo en cuenta que el integrando está acotado y que $\|A_k\|$ converge en ley a 0, el resultado se sigue tomando \liminf en ambos miembros de la desigualdad.

Entonces si integramos en una bola con centro θ , S , los dos miembros de (9 1), aplicamos el teorema de la divergencia de Gauss y tomamos límites, como $\|B_k\|$ converge a cero, tendremos

$$\int_S \liminf_{k \rightarrow \infty} k E \rho^2(\mathcal{U}_k, \theta) dV \geq -n \operatorname{vol}(S) - 2 \int_S \liminf_{k \rightarrow \infty} E(\operatorname{div}(A_k)) dV$$

y, puesto que $\liminf_{k \rightarrow \infty} E(\operatorname{div}(A_k)) \leq -n$, la proposición queda demostrada. ■

Queda entonces motivada la siguiente definición

Definición 9 2 3 Diremos que un estimador $\{\mathcal{U}_k\}$ es eficiente de primer orden si

$$\lim_{k \rightarrow \infty} kE\rho^2(\mathcal{U}_k, \theta) = n$$

Proposición 9 2 4 Si \mathcal{U}_k es regular y consistente de primer orden y se cumplen las condiciones del teorema anterior entonces (salvo conjuntos de medida Riemanniana cero) si \mathcal{U}_k es eficiente de primer orden

$$\lim_{k \rightarrow \infty} kE\left\| \exp_{\theta}^{-1}(\mathcal{U}_k) - \frac{1}{k} \text{grad}(\log p_{(k)}(x, \theta)) \right\|^2 = 0$$

en particular \mathcal{U}_k es asintoticamente normal $N(\theta, (g^{\alpha\beta}))$

Demostración La demostración es inmediata a partir del teorema anterior y de (9 1) ■

Observación Se podría ver que si las curvaturas seccionales estuvieran acotadas también inferiormente entonces se podría asegurar que $\lim_{k \rightarrow \infty} E(\text{div}(A_k)) = -n$, y el recíproco de la proposición anterior también sería cierto

Proposición 9 2 5 Si \mathcal{U}_k es \sqrt{k} -asintoticamente normal $N(\theta, (g^{\alpha\beta}))$ y $k\rho^2(\mathcal{U}_k, \theta)$ es uniformemente integrable entonces $\{\mathcal{U}_k\}$ es eficiente de primer orden

Demostración

Como $\sqrt{k} \exp_{\theta}^{-1}(\mathcal{U}_k)$ converge en ley a una variable $Y \sim N(0, (g^{\alpha\beta}))$ entonces

$$k\rho^2(\mathcal{U}_k, \theta) = kg_{ij} \left[\exp_{\theta}^{-1}(\mathcal{U}_k) \right]^i \left[\exp_{\theta}^{-1}(\mathcal{U}_k) \right]^j \xrightarrow{\mathcal{L}} \chi_n^2$$

El resultado se desprende de la integrabilidad uniforme de $k\rho^2(\mathcal{U}_k, \theta)$ ■

Corolario 9 2 6 El estimador máximo-verosímil, con las condiciones de regularidad antes señaladas, es eficiente de primer orden

9 3 Comportamiento asintotico del sesgo intrinseco

En esta seccion vamos a tratar de obtener un desarrollo asintotico del sesgo intrinseco para un estimador eficiente de primer orden, que debe entenderse como el sesgo intrinseco bajo una distribucion aproximada del estimador. Es decir lo que obtenemos son desarrollos para la distribucion asintotica de \mathcal{U}_k y los utilizamos para calcular el sesgo intrinseco. Tambien podemos pensarlos como aproximaciones correspondientes al sesgo intrinseco de una version truncada del estimador que tiene la misma distribucion limite.

Con la suposicion de que en un entorno del verdadero valor, el valor absoluto de las componentes, en algun sistema de coordenadas (y por tanto en todos), de la derivadas covariantes terceras del logaritmo de la verosimilitud este acotado por una variable aleatoria cuya esperanza exista y suponiendo que los momentos mixtos de las derivadas del logaritmo de la verosimilitud existen, es decir

$$E_{\theta} \left(\frac{\partial^k \log p(x, \theta)}{\partial \theta^{i_1} \dots \partial \theta^{i_k}} \frac{\partial^r \log p(x, \theta)}{\partial \theta^{j_1} \dots \partial \theta^{j_r}} \right) < \infty \quad (9.2)$$

tenemos la siguiente proposicion

Proposicion 9.3.1 *Si \mathcal{U}_k es un estimador eficiente de primer orden entonces*

$$[B(\mathcal{U}_k)]^i = S^i(\mathcal{U}_k) - \frac{1}{4k} g^{lm} g^{ij} T_{jlm} + o(k^{-1})$$

donde $S^i(\mathcal{U}_k) = o(k^{-1/2})$ y se anula para el estimador maximo-verosimil. Por su parte T_{jlm} es el tensor de asimetria

$$T_{jlm} = E \left(\frac{\partial l}{\partial \theta^j} \frac{\partial l}{\partial \theta^r} \frac{\partial l}{\partial \theta^m} \right), \quad l = \log p(x, \theta)$$

Demostracion

Escribamos $\bar{C}_{(k)}(\theta) = \frac{1}{k} \text{grad}(\log p_{(k)}(x, \theta))$. Por (8.3) tenemos

$$\begin{aligned} \tau_{\mathcal{U}_k}^{\theta} (\bar{C}_{(k)}(\mathcal{U}_k)) &= \bar{C}_{(k)}(\theta) + (D\bar{C}_{(k)}) (\exp_{\theta}^{-1}(\mathcal{U}_k)) \\ &+ \frac{1}{2} (D^2\bar{C}_{(k)}) (\exp_{\theta}^{-1}(\mathcal{U}_k), \exp_{\theta}^{-1}(\mathcal{U}_k)) \\ &+ \frac{1}{6} \tau_{\theta}^{\theta} \left\{ (D^3\bar{C}_{(k)}) (\exp_{\theta}^{-1}(\mathcal{U}_k), \exp_{\theta}^{-1}(\mathcal{U}_k), \exp_{\theta}^{-1}(\mathcal{U}_k)) \right\} \end{aligned} \quad (9.3)$$

donde $\tilde{\theta}$ es un punto del segmento geodesico que va de θ a \mathcal{U}_k y $\tau_{\mathcal{U}_k}^{\tilde{\theta}}$ representa el transporte paralelo de \mathcal{U}_k a θ a lo largo de la geodesica que los une. Puesto que \mathcal{U}_k es eficiente de primer orden y las componentes de las derivadas covariantes terceras estan acotadas en valor absoluto en un entorno de θ por una variable integrable resultara que

$$\begin{aligned} \tau_{\mathcal{U}_k}^{\tilde{\theta}} \left(\bar{C}_{(k)}(\mathcal{U}_k) \right) &= \bar{C}_{(k)}(\theta) + \left(D\bar{C}_{(k)} \right) \left(\exp_{\theta}^{-1}(\mathcal{U}_k) \right) \\ &+ \frac{1}{2} \left(D^2\bar{C}_{(k)} \right) \left(\exp_{\theta}^{-1}(\mathcal{U}_k), \exp_{\theta}^{-1}(\mathcal{U}_k) \right) + O_P(k^{-3/2}) \end{aligned}$$

Asimismo $\exp_{\theta}^{-1}(\mathcal{U}_k) = \bar{C}_{(k)}(\theta) + o_P(k^{-1/2})$ de forma que

$$\begin{aligned} \tau_{\mathcal{U}_k}^{\tilde{\theta}} \left(\bar{C}_{(k)}(\mathcal{U}_k) \right) &= \bar{C}_{(k)}(\theta) + \left(D\bar{C}_{(k)} + I_n \right) \left(\bar{C}_{(k)}(\theta) \right) - \exp_{\theta}^{-1}(\mathcal{U}_k) \\ &+ \frac{1}{2} \left(D^2\bar{C}_{(k)} \right) \left(\bar{C}_{(k)}(\theta), \bar{C}_{(k)}(\theta) \right) + o_P(k^{-1}) \end{aligned}$$

donde hemos tenido en cuenta que $\sqrt{k} \left(D\bar{C}_{(k)} + I_n \right)$ es asintoticamente normal, por la existencia de los momentos mixtos señalada en (9.2) con $k, r \leq 2$ y el teorema central del limite, y $D^2\bar{C}_{(k)} \xrightarrow{c.s.} E(D^2C)$, con $C = \text{grad}(\log p(x, \theta))$, por (9.2) con $k, r \leq 3$ y la ley fuerte de los grandes numeros

Ahora bien, (vease seccion 4.2) para la notacion

$$\begin{aligned} (D^2\bar{C}_{(k)}) \left(\bar{C}_{(k)}(\theta), \bar{C}_{(k)}(\theta) \right) &= Tr^{2,3,4,5} \left\{ (D^2\bar{C}_{(k)} - E(D^2C_{(k)})) \otimes \left((C_{(k)}(\theta) \otimes \bar{C}_{(k)}(\theta)) \right. \right. \\ &\left. \left. - \frac{G^{-1}(\theta)}{k} \right) + (D^2C_{(k)}) \otimes \left(\frac{G^{-1}(\theta)}{k} \right) + E(D^2C_{(k)}) \otimes \left((C_{(k)}(\theta) \otimes C_{(k)}(\theta)) - \frac{G^{-1}(\theta)}{k} \right) \right\} \end{aligned}$$

donde G^{-1} es la version contravariante del tensor metrico, como

$$Tr^{2,3,4,5} \left\{ (D^2\bar{C}_{(k)} - E(D^2C_{(k)})) \otimes \left((C_{(k)}(\theta) \otimes C_{(k)}(\theta)) - \frac{G^{-1}(\theta)}{k} \right) \right\} = O_P(k^{-3/2})$$

resultara

$$\begin{aligned} (D^2\bar{C}_{(k)}) \left(\bar{C}_{(k)}(\theta), \bar{C}_{(k)}(\theta) \right) &= Tr^{2,3,4,5} \left\{ (D^2\bar{C}_{(k)}) \otimes \left(\frac{G^{-1}(\theta)}{k} \right) \right. \\ &\left. + E(D^2\bar{C}_{(k)}) \otimes \left((C_{(k)}(\theta) \otimes \bar{C}_{(k)}(\theta)) - \frac{G^{-1}(\theta)}{k} \right) \right\} + o_P(k^{-1}) \end{aligned}$$

Ahora tomando esperanzas, en la expresion para $\exp_{\theta}^{-1}(\mathcal{U}_k)$ tendremos

$$E(\exp_{\theta}^{-1}(\mathcal{U}_k)) = -E\left(\tau_{\mathcal{U}_k}^{\theta}(\bar{C}_{(k)}(\mathcal{U}_k))\right) + E\left(D\bar{C}_{(k)}(\bar{C}_{(k)}(\theta))\right) + \frac{1}{2}T_{r^2}^{345} \left\{ E(D^2\bar{C}_{(k)}) \otimes \frac{G^{-1}(\theta)}{k} \right\}$$

Por ultimo, escribiendo $l = \log p(x, \theta)$ como

$$\begin{aligned} E\left(\left(D\bar{C}_{(k)}\right)\left(\bar{C}_{(k)}(\theta)\right)\right)^i &= E\left(\frac{\partial\bar{C}_{(k)}^i}{\partial\theta^j}\bar{C}_{(k)}^j\right) + \frac{1}{k}g^{jm}\Gamma_{jm}^i \\ &= \frac{1}{k}g^{im}g^{jr}E\left(\frac{\partial^2l}{\partial\theta^j\partial\theta^m}\frac{\partial l}{\partial\theta^r}\right) + \frac{1}{k}g^{jr}E\left(\frac{\partial g^{im}}{\partial\theta^j}\frac{\partial l}{\partial\theta^m}\frac{\partial l}{\partial\theta^r}\right) + \frac{1}{k}g^{jm}\Gamma_{jm}^i \\ &= \frac{1}{k}g^{im}g^{jr}E\left(\frac{\partial^2l}{\partial\theta^j\partial\theta^m}\frac{\partial l}{\partial\theta^r}\right) - \frac{1}{k}g^{jr}g^{im}\frac{\partial g_{mr}}{\partial\theta^j} + \frac{1}{k}g^{jm}\Gamma_{jm}^i \\ &= -\frac{1}{k}g^{im}g^{jr}E\left(\frac{\partial^2l}{\partial\theta^j\partial\theta^r}\frac{\partial l}{\partial\theta^m} + \frac{\partial l}{\partial\theta^j}\frac{\partial l}{\partial\theta^r}\frac{\partial l}{\partial\theta^m}\right) + \frac{1}{k}g^{jm}\Gamma_{jm}^i \\ &= -\frac{1}{2k}g^{jr}g^{im}T_{jmr} \end{aligned} \tag{9 4}$$

donde hemos tenido en cuenta que,

$$\Gamma_{jr}^i = g^{im}E\left(\frac{\partial^2l}{\partial\theta^j\partial\theta^r}\frac{\partial l}{\partial\theta^m} + \frac{1}{2}\frac{\partial l}{\partial\theta^j}\frac{\partial l}{\partial\theta^r}\frac{\partial l}{\partial\theta^m}\right)$$

Por otro lado

$$E\left(\left(D^2\bar{C}_{(k)}\right)_{jm}^i\right) = E\left(\frac{\partial(DC)_j^i}{\partial\theta^m} + \Gamma_{jr}^i(DC)_m^r - \Gamma_{jm}^r(DC)_r^i\right) = E\left(\frac{\partial(DC)_j^i}{\partial\theta^m}\right)$$

ya que $E\left((DC)_j^i\right) = -\delta_j^i$ Ademas

$$\begin{aligned} E\left(\frac{\partial(DC)_j^i}{\partial\theta^m}\right) &= E\left(\frac{\partial^2C^i}{\partial\theta^j\partial\theta^m} + \Gamma_{jr}^i\frac{\partial C^r}{\partial\theta^m}\right) \\ &= -E\left(\frac{\partial C^i}{\partial\theta^j}\frac{\partial l}{\partial\theta^m}\right) - \Gamma_{jm}^i, \end{aligned}$$

y ahora fijandonos en la primera igualdad de (9 4) tendremos que

$$E\left(\left(D^2\bar{C}_{(k)}\right)_{jm}^i\right) = \frac{1}{2}g^{jr}T_{jmr}$$

Sustituyendo obtenemos la expresion deseada ■

Observacion Es interesante destacar que si consideramos la expresion obtenida por Amari, [2] pagina 131, para el sesgo de un estimador eficiente, fijado un sistema de coordenadas

$$[Sesgo]^i = S^i - \frac{1}{2k} \bar{\Gamma}_{jm}^{-1i} g^{jm} + O_P(k^{-3/2})$$

donde

$$\bar{\Gamma}_{jr}^{-1i} = g^{im} E \left(\frac{\partial^2 l}{\partial \theta^j \partial \theta^r} \frac{\partial l}{\partial \theta^m} + \frac{\partial l}{\partial \theta^j} \frac{\partial l}{\partial \theta^r} \frac{\partial l}{\partial \theta^m} \right)$$

y la utilizamos en la expresion (7.1) junto con el primer termino de la covarianza de un estimador eficiente, obtenemos

$$\begin{aligned} B^i &= S^i - \frac{1}{2k} g^{jm} \bar{\Gamma}_j^{-1i} + \frac{1}{2k} g^{jm} \Gamma_{jr}^i + o(k^{-1}) \\ &= -\frac{1}{4k} g^{jr} g^{jm} T_{jrm} + o(k^{-1}) \end{aligned}$$

que coincide con la expresion obtenida en la proposicion anterior aunque aqui en el contexto de las familias exponenciales

Conclusiones

10 1 Resumen de resultados

Mencionaremos los principales resultados que se han obtenido en la presente memoria

- Se han proporcionado definiciones de valor medio en una variedad con una conexión afin, estudiado sus propiedades y relacionado con el centro de masas cuando se posea una métrica Riemanniana
- Los útiles geométrico diferenciales anteriores se han utilizado para definir medidas intrínsecas de sesgo y error cuadrático que permiten desarrollar un análisis intrínseco de la estimación puntual
- Se han obtenido estimadores intrínsecamente insesgados y casos en que se puede asegurar su inexistencia
- Se ha analizado en que circunstancias las medidas clásicas e intrínsecas coinciden
- Se han obtenido cotas locales para la distancia de Rao cuadrático media de un estimador basándose en los teoremas de comparación de geometría diferencial
- Asimismo se han dado desigualdades generales para la *covarianza* de un estimador, entendiendo este como una aplicación del espacio muestral en la variedad Riemanniana asociada de forma natural a un modelo paramétrico

- Se ha estudiado en que condiciones se alcanzan las cotas anteriores, caracterizando la eficiencia intrínseca
- Se han estudiado cotas globales promedio de la distancia de Rao cuadrático media de un estimador. Se han utilizado diversas técnicas para la obtención de cotas y obtenido expresiones para el volumen de bolas Riemannianas que pueden ser de futura utilidad
- Se han estudiado como mejorar el comportamiento de un estimador, viendo que por Blackwellización del mismo y bajo ciertas condiciones de la geometría del modelo paramétrico el error intrínseco promedio disminuye
- Se han estudiado las propiedades asintóticas de los estimadores. Para ello se han definido y caracterizado las propiedades asintóticas de objetos aleatorios en una variedad Riemanniana y aplicado a nuestro contexto
- Se han relacionado las cantidades clásicas e intrínsecas desde el punto de vista asintótico
- Se ha desarrollado un procedimiento para obtener desarrollos de Taylor tensoriales en una variedad equipada con una conexión y se han aplicado a diversos problemas relacionados con el análisis intrínseco
- Por último se ha comenzado a desarrollar una teoría de eficiencia intrínseca asintótica, empezando por la eficiencia de primer orden y obteniendo desarrollos asintóticos tensoriales del sesgo intrínseco

10.2 Perspectivas del Análisis Intrínseco

La invarianza bajo reparametrización de un procedimiento de inferencia ha sido valorada como una importante y deseable propiedad por varios autores, véase Barndorff Nielsen [7] y S. Amari [2] entre otros. Notemos, por ejemplo, que necesitamos, irremediablemente, esta propiedad si deseamos usar, de una forma consistente, el *bootstrap paramétrico*. Básicamente esta invarianza significa que el procedimiento inferencial conduce a la misma decisión en cualquier sistema de coordenadas o parámetros. Pero ¿qué debemos entender por "misma decisión"? No podemos hablar de las mismas conclusiones si los medios usados para alcanzar

una decision como el sesgo, el error cuadratico medio, etc dependen de la parametrizacion que se utilice. Ya se ha puesto de manifiesto que las medidas classicas de sesgo y error cuadratico medio son por tanto inadecuadas para el proposito de establecer una inferencia intrinseca, y que las nuevas medidas, sesgo intrinseco y distancia de Rao cuadratico media, nos permite investigar las propiedades de un estimador de forma mas objetiva. Desgraciadamente en muchos casos corrientes y sencillos no existen estimadores intrinsecamente insesgados, aunque es posible corregir el sesgo localmente, obteniendo un nuevo estimador con su correspondiente tensor de sesgo nulo en un punto fijo. En efecto, observemos que para una muestra de tamaño fijo k , a fin de corregir el sesgo de un estimador $\mathcal{U}_k(x)$, en un punto fijo p_0 , es suficiente definir el estimador modificado

$$\tilde{\mathcal{U}}_k(x) = \exp_{p_0} \left(\exp_{p_0}^{-1}(\mathcal{U}_k(x)) - B(p_0) \right)$$

donde $B(p_0)$ es el tensor de sesgo correspondiente a $\mathcal{U}_k(x)$ en p_0 . Esto podria ocurrir en la teoria de contraste de hipotesis, cuando la hipotesis nula es simple, corrigiendo el sesgo de un estimador bajo la hipotesis nula, y permitiendo la construccion de tests que serian invariantes bajo reparametrizaciones.

Es tambien posible dar una medida promedio del sesgo, como la integral del cuadrado de la norma del tensor de sesgo sobre la variedad

$$\mathcal{B}_{\mathcal{U}_k}^2 = \int_M \|B(p)\|^2 V_R(dp)$$

donde V_R es la medida Riemanniana. Notemos que esta definicion es independiente del sistema de coordenadas, y con posibles interpretaciones Bayesianas.

Por ultimo podriamos utilizar las expresiones asintoticas del sesgo para mejorar el orden y para conseguir una mejor aproximacion a la distribucion asintotica de $\rho^2(\mathcal{U}, \theta)$ en el sentido de las correcciones de Bartlett, en relacion con esto vease por ejemplo el articulo de McCullagh and Cox [45] y Burbea y Oller [15].

Seria interesante disponer de expresiones asintoticas para la distancia de Rao cuadratico media y poder definir eficiencias de orden superior intrinsecas. Pensamos que no se tardara en tenerlas si bien las dificultades tecnicas son evidentes a la vista del desarrollo asintotico del sesgo intrinseco.

Tambien habria que desarrollar el analisis intrinseco en conexion con la invarianza de los modelos estadisticos bajo la accion de grupos de transformaciones. ¿Que eficiencia presentan los estimadores invariantes? ¿Que ocurre si el grupo

modifica el modelo pero solo ligeramente? ¿Se podría hablar de casi invarianza? ¿Que ocurre si al transformar biyectivamente los datos nos 'salimos' del modelo y vamos a parar a otro modelo?, ¿como afecta esto a la estimacion? En la introduccion ya se han indicado algunos requerimientos basicos, de tipo logico, a la estimacion puntual, pero todavia quedan muchos eslabones sueltos

La distancia de Rao ha sido considerada de utilidad en diferentes enfoques, aqui hemos enfatizado su utilidad como la correcta distancia entre las estimaciones, es decir la escala apropiada en la que hay que observar y comparar las estimaciones y consecuentemente los estimadores. Notese que la distancia entre estimaciones depende del modelo estadístico, y que si consideramos las estimaciones localizadas en el espacio tangente de la densidad verdadera obtendriamos, como distancia entre estimaciones, la *distancia de Mahalanobis*. Siendo esta una aproximacion de primer orden en nuestro contexto

Apéndice

En este apéndice presentamos un conjunto de nociones y resultados que pertenecen a la geometría diferencial y que son necesarios para probar muchas de las afirmaciones que contiene la memoria. El lector podrá encontrar información adicional en Spivak [57], Kobayashi y Nomizu [37], Hicks [27], Chavel [21], Cheeger [22], Bishop [12] y Karcher [33] entre otros.

Sea (M, \mathfrak{A}) una n -variedad real conexa C^∞ , \mathfrak{A} es el atlas, TM denota su fibrado tangente con la aplicación proyección $\pi: TM \rightarrow M$, donde $\pi(\xi) = p$ si $\xi \in M_p$, el espacio tangente en p . Supongamos que tenemos una conexión afín ∇ en la variedad. Sea $c: (\alpha, \beta) \rightarrow M$ un camino suave en M . Un campo vectorial X a lo largo de c es una aplicación $X: (\alpha, \beta) \rightarrow TM$ tal que $\pi \circ X = c$, es decir $X(s) \in M_{c(s)}$ para todo $s \in (\alpha, \beta)$. El campo vectorial tangente correspondiente a c vendrá dado por la aplicación $t \mapsto c' = c(d/ds|_{s=t})$, donde c' es la aplicación diferencial de c y d/ds es el operador de derivación estándar en la recta real. Por razones de simplicidad, a menudo identificaremos los campos a lo largo de curvas o superficies con sus imágenes. Por ejemplo, escribiremos c' en lugar de la aplicación $t \mapsto c'(t)$.

Para definir la derivada de X a lo largo de c , $(\nabla/ds)X$, también llamada *derivada covariante a lo largo de c* , sea $(U, x) \in \mathfrak{A}$ una carta local en M tal que $c((\alpha, \beta)) \cap U \neq \emptyset$, y sea Γ_{ij}^k los símbolos de Christoffel correspondientes a la conexión afín ∇ , con respecto a la carta local (U, x) , definidos mediante

$$\nabla_{\partial_i} \partial_j = \sum_{k=1}^n \Gamma_{ij}^k \partial_k \quad i, j = 1, \dots, n,$$

donde $\partial_1, \dots, \partial_n$ son los campos vectoriales base correspondientes a la carta local. Sea $X = \sum_{j=1}^n \eta^j (\partial_j \circ c)$, y $c^j = x^j \circ c$, siendo x^j la j -ésima función coordenada.

La derivada de X es otro campo vectorial a lo largo de c dado por

$$\frac{\nabla}{ds} X = \sum_{k=1}^n \left\{ (\eta^k)' + \sum_{j=1}^n (\Gamma_{ij}^k \circ c) \eta^j (c^j)' \right\} (\partial_k \circ c),$$

$(\eta^k)'$ y $(c^j)'$ son las derivadas usuales en \mathbf{R} . Esta definición es, como se puede comprobar fácilmente, independiente del sistema de coordenadas y por tanto el campo vectorial $(\nabla/ds)X$ está bien definido, supuesta la existencia de las derivadas $\eta^k, k = 1, \dots, n$.

Algunas propiedades bien conocidas son

$$\begin{aligned} \frac{\nabla}{ds}(X + Y) &= \frac{\nabla}{ds}X + \frac{\nabla}{ds}Y \\ \frac{\nabla}{ds}(fX) &= f'X + f \frac{\nabla}{ds}X, \end{aligned}$$

donde X y Y son campos vectoriales suaves a lo largo de c y f es una función real C^1 , $f: (\alpha, \beta) \rightarrow \mathbf{R}$. Si la variedad es Riemanniana, y ∇ es la conexión de Levi-Civita, tenemos también

$$\frac{d}{ds} \langle X, Y \rangle = \left\langle \frac{\nabla}{ds} X, Y \right\rangle + \left\langle X, \frac{\nabla}{ds} Y \right\rangle$$

Dada una conexión arbitraria ∇ en la variedad, las curvas cuyo campo vectorial tangente permanece constante a lo largo de ellas, como las líneas rectas en un espacio Euclideo, son las *geodesicas*, más precisamente las curvas que satisfacen $(\nabla/ds)c' = 0$.

11.1 Derivada general covariante

Sea ∇ una conexión afin en una variedad M C^∞ y X un campo tensorial C^∞ , definiremos un operador D_X en el conjunto, $\mathcal{F}_A^{(p,q)}$, de todos los campos tensoriales C^∞ p -contravariantes y q -covariantes, $A \subset M$ siendo un conjunto abierto, como sigue

- a) Si $f \in \mathcal{F}_A^{(0,0)}$, $D_X f = Xf$
- b) Si $Y \in \mathcal{F}_A^{(1,0)}$, $D_X Y = \nabla_X Y$
- c) Si $w \in \mathcal{F}_A^{(0,1)}$, $(D_X w)(Y) = Xw(Y) - w(\nabla_X Y)$
- d) Si $\theta \in \mathcal{F}_A^{(p,q)}$ entonces $D_X \theta$ es obtenida a partir de la siguiente ecuacion

$$D_X (\theta(w_1, \dots, w_p, Y_1, \dots, Y_q)) = (D_X \theta)(w_1, \dots, Y_q) + \theta(D_X w_1, w_2, \dots, Y_q) + \dots + \theta(w_1, \dots, Y_{q-1}, D_X Y_q),$$

donde $w_1, \dots, w_p \in \mathcal{F}_A^{(0,1)}$ y $Y_1, \dots, Y_q \in \mathcal{F}_A^{(1,0)}$

El operador D_X es conocido como la *derivada covariante con respecto a X* Vease [27]

Usando este operador podemos construir otro operador D que aplica $\mathcal{F}_A^{(p,q)}$ en $\mathcal{F}_A^{(p,q+1)}$

$$D \begin{matrix} \mathcal{F}_A^{(p,q)} \\ \theta \end{matrix} \longrightarrow \begin{matrix} \mathcal{F}_A^{(p,q+1)} \\ D\theta \end{matrix}$$

donde

$$(D\theta)(w_1, \dots, w_p, Y_1, \dots, Y_{q+1}) \equiv (D_{Y_{q+1}} \theta)(w_1, \dots, w_p, Y_1, \dots, Y_q),$$

y $w_1, \dots, w_p \in \mathcal{F}_A^{(0,1)}$ y $Y_1, \dots, Y_q \in \mathcal{F}_A^{(1,0)}$

Es facil comprobar que $D\theta \in \mathcal{F}_A^{(p,q+1)}$ Este operador es conocido como la *derivada general covariante* o simplemente derivada covariante Vease Hicks [27]

11 2 La aplicacion exponencial

La aplicacion exponencial, $\exp_p : M_p \rightarrow M$, correspondiente a ∇ , se define a traves de las correspondientes geodesicas como sigue Sea p un punto de la variedad, $p \in M$, M_p el espacio tangente en p , $\xi \in M_p$, y sea $\gamma : [0, 1] \rightarrow M$ una geodesica tal que

$$\gamma(0) = p \quad \text{y} \quad \gamma'(0) = \gamma_* \left(\frac{d}{ds} \Big|_{s=0} \right) = \xi,$$

donde $d/ds|_{s=0}$ es el operador de derivacion estandar en la recta real, en $s = 0$. Entonces, la aplicacion exponencial viene dada por $\exp_p(\xi) = \gamma(1)$, definida para todo ξ en un entorno abierto estrellado de $0_p \in M_p$.

Notemos que \exp_p aplica las rectas que salen de 0_p a las geodesicas que empiezan en p , y ya que M_p es tambien una variedad y cualquier vector tangente $\eta \in M_p$ puede ser identificado como un campo vectorial paralelo en M_p y por tanto como un elemento de $(M_p)_\zeta$ para un $\zeta \in M_p$ arbitrario, tenemos $(\exp_p)_*|_0(\eta) = \eta$, donde $(\exp_p)|_0$ es la restriccion de (\exp_p) al espacio tangente $(M_p)_0$. Esto muestra, a partir del teorema de la funcion inversa, que \exp_p es un difeomorfismo local.

Vamos a centrarnos ahora en variedades Riemannianas con su conexion natural, esto es la conexion de Levi-Civita. Denotemos por

$$\mathfrak{S}_p = \{\xi \in M_p \mid \|\xi\|_p = 1\},$$

donde $r > 0$, y for each $\xi \in \mathfrak{S}_p$ definimos

$$C_p(\xi) = \sup\{s > 0 \mid \rho(p, \gamma_\xi(s)) = s\},$$

donde ρ es la distancia Riemanniana y γ_ξ es una geodesica definida en un intervalo abierto que contiene el cero, tal que $\gamma_\xi(0) = p$ y con vector tangente ξ en el origen. Entonces si hacemos

$$\mathfrak{D}_p = \{s\xi \in M_p \mid 0 \leq s < C_p(\xi), \xi \in \mathfrak{S}_p\}$$

y

$$D_p = \exp_p(\mathfrak{D}_p),$$

tenemos la proposicion siguiente

Proposicion 11.2.1 \exp_p aplica \mathfrak{D}_p difeomorficamente en D_p .

Demostracion

Sera suficiente mostrar que \exp_p es inyectiva ya que es obviamente exhaustiva y la suavidad se sigue de la dependencia suave de las geodesicas con respecto a las condiciones iniciales.

Sea γ_ξ un segmento geodesico que conecta p con m , esto es $\exp_p(s\xi) = m$, con $s = \rho(p, m)$. Consideremos una bola normal de radio ϵ en $m \in \mathfrak{D}_p$. Supongamos

que hubiese otra curva γ (parametrizada por la longitud de arco) de p a m con longitud $s = |\gamma|$. Si $\gamma(s - \epsilon) \neq \gamma_\xi(s - \epsilon)$ las curvas correspondientes formarían eventualmente una esquina con la extensión de γ_ξ más allá de m . Evitando esta esquina podríamos obtener curvas más cortas de p a $\gamma_\xi(s + \epsilon)$ que $s + \epsilon$, contradiciendo la minimalidad de γ_ξ más allá de m . ■

Además, si la variedad es también completa, la frontera de \mathfrak{D}_p , $\partial\mathfrak{D}_p$, es aplicada por la aplicación exponencial sobre ∂D_p , llamando el *cut locus de p en M* . Es también interesante notar que el cut locus de m tiene medida Riemanniana n dimensional cero en M (esencialmente debido al teorema de Sard), y M es la unión disjunta de D_m y ∂D_m . Para más detalles vease Hicks [27] o Spivak [57].

Adicionalmente, consideremos la familia uniparamétrica de geodésicas $c(s, t)$ definida como

$$c(s, t) = \exp_p(s\zeta(t)),$$

donde $\zeta(t)$ es cualquier curva en \mathfrak{S}_p con $\zeta(0) = \xi$, definida para s suficientemente pequeño. Denotemos

$$c' = c \left(\frac{\partial}{\partial s} \right), \quad c = c_* \left(\frac{\partial}{\partial t} \right),$$

donde $\partial/\partial s$ y $\partial/\partial t$ son los operadores de derivación parcial ordinarios en \mathbf{R}^2 . Observemos que

$$\frac{\partial}{\partial s} \langle c', c \rangle = \langle \nabla_{\partial s} c', c \rangle + \langle c', \nabla_{\partial s} c \rangle,$$

donde $\nabla/\partial s$ y $\nabla/\partial t$ son las derivadas covariantes a lo largo de las curvas $c(\cdot, t)$ y $c(s, \cdot)$ respectivamente. Entonces, como, fijado t , $c(\cdot, t)$ es una geodésica, $(\nabla/\partial s)c' = 0$, la conexión de Levi-Civita es de torsión nula y el parentesis de Lie $[c', c] = 0$, (donde $[c', c]f = (\frac{\partial}{\partial s} \frac{\partial}{\partial t} - \frac{\partial}{\partial t} \frac{\partial}{\partial s})f$ para cualquier C^2 función real en M), tenemos $(\nabla/\partial s)c = (\nabla/\partial t)c'$, y así

$$\frac{\partial}{\partial s} \langle c', c \rangle = \langle c', \nabla_{\partial t} c' \rangle = \frac{1}{2} \frac{\partial}{\partial t} \langle c', c' \rangle = 0,$$

ya que $\|c'\| = \|\zeta(t)\| = 1$. Por tanto el producto escalar $\langle c', c \rangle$ es independiente de s , y para $s = 0$, tenemos $c(0, t) = p$, y entonces $c'(0, t) = 0$, obteniendo $\langle c', c \rangle = 0$. Este resultado es conocido como el *Lema de Gauss*, y si hacemos $\gamma'_\xi(s) = (\exp_p)_*|_{s\xi}(\xi)$ puede expresarse como

$$\langle (\exp_p)_*|_{s\xi}(\eta), \gamma'_\xi(s) \rangle = 0,$$

donde $\eta \in (\mathfrak{S}_p(s))_{s\xi}$. De esta forma las curvas obtenidas fijando $s = a$, $w(t) = c(a, t) = \exp_p(a \zeta(t))$ son ortogonales a las geodesicas radiales obtenidas fijando $t = b$, $\gamma(s) = c(s, b) = \exp_p(s \zeta(b))$ y la aplicacion $(\exp_p) : (M_p)_v \rightarrow M_{\exp_p(v)}$, aunque no preserva angulos, aplica vectores ortogonales en vectores ortogonales

11 3 Campos de Jacobi

Con la misma notacion basica que en las secciones previas, vamos a considerar una familia suave uniparametrica de geodesicas, $c(s, t)$ en M , tal que $c(s, t)$ es una geodesica fijado t . Como antes $c' = c(\partial/\partial s)$ y $c = c_*(\partial/\partial t)$. Entonces tenemos, para un campo vectorial arbitrario X a lo largo de $s \mapsto c(s, t)$,

$$R(c', c)X = \frac{\nabla}{\partial t} \frac{\nabla}{\partial s} X - \frac{\nabla}{\partial s} \frac{\nabla}{\partial t} X,$$

donde R es el *tensor de curvatura*, ya que el parentesis de Lie, $[c, c'] = 0$

Por tanto, ya que fijando t , $c(s, t)$ es una geodesica, tenemos

$$0 = \frac{\nabla}{\partial s} c' = \frac{\nabla}{\partial t} \frac{\nabla}{\partial s} c' = \frac{\nabla}{\partial s} \frac{\nabla}{\partial t} c' + R(c', c)c',$$

y por ultimo, ya que $(\nabla/\partial t)c' = (\nabla/\partial s)c$, el campo vectorial c a lo largo de la geodesica $s \rightarrow c(s, t)$ satisface la ecuacion diferencial de segundo orden

$$\frac{\nabla^2}{\partial s^2} c + R(c', c)c' = 0,$$

donde $\frac{\nabla^2}{\partial s^2} \equiv \left(\frac{\nabla}{\partial s}\right) \frac{\nabla}{\partial s}$

In general, si $c = c(s)$ es una geodesica en M , un *campo de Jacobi a lo largo de* c , Y es un campo vectorial C^∞ a lo largo de c satisfaciendo la *ecuacion de Jacobi*

$$\frac{\nabla^2}{ds^2} Y + R(c', Y)c' = 0,$$

donde $c' = c_*(d/ds)$ y ∇/ds es la derivada covariante a lo largo de c

Como esta ecuacion es lineal en Y , el conjunto de todos los campos de Jacobi a lo largo de c constituye un espacio vectorial \mathcal{F}_J sobre \mathbf{R} de dimension $2n$, donde n es la dimension de M . Un campo de Jacobi Y a lo largo de una geodesica esta

determinado por su valor, Y , y $(\nabla/ds)Y$ en un punto arbitrario de la geodesica. Ademas, si X y Y son campos de Jacobi,

$$\begin{aligned} \frac{d}{ds} \left\{ \left\langle \frac{\nabla}{ds} X, Y \right\rangle - \left\langle \frac{\nabla}{ds} Y, X \right\rangle \right\} &= \left\langle \frac{\nabla^2}{ds^2} X, Y \right\rangle - \left\langle \frac{\nabla^2}{ds^2} Y, X \right\rangle = \\ &= -\langle R(c', X)c', Y \rangle + \langle R(c', Y)c', X \rangle = 0, \end{aligned}$$

por una conocida propiedad del tensor de curvatura, por tanto el Wronskiano

$$\left\langle \frac{\nabla}{ds} X, Y \right\rangle - \left\langle \frac{\nabla}{ds} Y, X \right\rangle = \text{const}$$

y, en particular, para $s = c$, $X = Y = 0$, tendremos

$$\left\langle \frac{\nabla}{ds} X, Y \right\rangle - \left\langle \frac{\nabla}{ds} Y, X \right\rangle = 0 \tag{11.1}$$

Adicionalmente, para cualquier campo de Jacobi Y existen dos constantes reales a y b tales que

$$\langle Y, c' \rangle = a + bs,$$

puesto que c' es tambien un campo de Jacobi. Si $a = b = 0$ obtenemos todos los campos de Jacobi normales (ortogonales a c'), que forman un subespacio de \mathcal{F}_J con dimension $2n - 2$.

De esta forma podemos decomponer cualquier $Y \in \mathcal{F}_J$ a lo largo de la geodesica $c(s)$ en sus componentes normal y tangencial $Y = Y^{nor} + Y^{tan}$, siendo tambien ambas componentes campos de Jacobi.

Para un campo de Jacobi tangencial, como consecuencia de la ecuacion de Jacobi, tenemos,

$$Y(s) = (a + bs) c'(s)$$

Todos los resultados sobre campos de Jacobi se pueden formular para geodesicas con vector tangente unitario, esto es con $\|c'\| = 1$, ya que si $Y(s)$ es un campo de Jacobi a lo largo de $c(s)$ entonces $J(s) = Y(rs)$ es un campo de Jacobi a lo largo de la geodesica $c(rs)$, con $J(0) = Y(0)$ y $(\nabla/\partial s)J(0) = r (\nabla/\partial s)Y(0)$.

Con el proposito de estudiar el comportamiento de los campos de Jacobi es util, como veremos mas tarde, introducir la ecuacion diferencial siguiente

Sea $\kappa: \mathbf{R} \rightarrow \mathbf{R}$ una funcion continua y consideremos la ecuacion diferencial $f'' + \kappa f = 0$

Si denotamos S_κ la solución de esta ecuación con $S_\kappa(0) = 0$ y $S'_\kappa(0) = 1$, y por C_κ la solución satisfaciendo $C_\kappa(0) = 1$ y $C'_\kappa(0) = 0$. Es fácil verificar que si $\kappa = K$, donde K es una constante real, entonces

$$S_K(t) = \begin{cases} \frac{\sin(\sqrt{K}t)}{\sqrt{K}} & \text{si } K > 0, \\ t & \text{si } K = 0, \\ \frac{\sinh(\sqrt{-K}t)}{\sqrt{-K}} & \text{si } K < 0, \end{cases} \quad (11.2)$$

y $C_K = S'_K$

Dado un campo de Jacobi Y normal (ortogonal a c'), definamos

$$f_\kappa = \|Y\|(0) C_\kappa + \|Y\|'(0) S_\kappa \quad (11.3)$$

Vamos también a introducir las cotas de las curvaturas seccionales, a lo largo de la geodésica c , $\delta(s) \leq K \leq \Delta(s)$ (para 2-planos arbitrarios linealmente independientes), donde K es la curvatura seccional Riemanniana,

$$K(X, Y) = \frac{\langle R(X, Y)X, Y \rangle}{\langle X, X \rangle \langle Y, Y \rangle - \langle X, Y \rangle^2}$$

Entonces, tenemos los siguientes *Teoremas de comparación*

Teorema 11.3.1 (Teorema de comparación de Rauch) *Sea Y un campo de Jacobi normal a lo largo de una geodésica con tangente unitaria $c(s)$, entonces, mientras no se anule, satisface las siguientes desigualdades*

$$\left(\frac{\|Y\|(s)}{f_\Delta(s)} \right)' \geq 0, \quad \frac{\|Y\|'(s)}{\|Y\|(s)} \geq \frac{f'_\Delta(s)}{f_\Delta(s)}, \quad \|Y\|(s) \geq f_\Delta(s)$$

Donde la segunda desigualdad se puede escribir como

$$f_\Delta(s) \left\langle \frac{\nabla}{\partial s} Y(s), Y(s) \right\rangle \geq f'_\Delta(s) \langle Y(s), Y(s) \rangle$$

Demostración

$$\begin{aligned} \|Y\|''(s) &= \left\{ \frac{1}{\|Y\|} \left\langle \frac{\nabla}{\partial s} Y, Y \right\rangle \right\}'(s) = \frac{1}{\|Y\|} \langle -R(c', Y)c', Y \rangle(s) \\ &\quad + \frac{1}{\|Y\|^3} \left\{ \left\langle \frac{\nabla}{\partial s} Y, \frac{\nabla}{\partial s} Y \right\rangle \langle Y, Y \rangle - \left\langle \frac{\nabla}{\partial s} Y, Y \right\rangle^2 \right\}(s) \\ &\geq -\Delta(s) \|Y\|(s) \end{aligned}$$

por la desigualdad de Cauchy-Schwartz Por tanto tenemos

$$\|Y\|''(s) + \Delta(s)\|Y\|(s) \geq 0,$$

mientras f_Δ no se anule

$$(f_\Delta(s)\|Y\|'(s) - f'_\Delta(s)\|Y\|(s))' \geq 0,$$

ya que $f''_\Delta + \Delta f_\Delta = 0$ Entonces, como $\|Y\|(0) = f_\Delta(0)$ y $\|Y\|'(0) = f'_\Delta(0)$, se obtienen las desigualdades

Proposición 11 3 2 *Sea Y un campo de Jacobi normal a lo largo de la geodesica $c(t)$ y sea X un campo normal a lo largo de c con $X(0) = Y(0) = 0$ y $X(s) = Y(s)$ entonces $\psi_Y(s) \leq \psi_X(s)$, donde*

$$\psi_X(s) \equiv \int_0^s \left\{ \left\langle \frac{\nabla}{dt} X, \frac{\nabla}{dt} X \right\rangle - \langle R(c', X)c', X \rangle \right\} dt$$

y la igualdad se cumple si y solo si $X=Y$

Demostracion

Sean Y_1, Y_2, \dots, Y_{n-1} campos de Jacobi linealmente independientes que se anulan en $t = 0$ Por tanto podemos escribir

$$X = \sum_{i=1}^{n-1} f_i Y_i, \quad Y = \sum_{i=1}^{n-1} a_i Y_i,$$

donde a_i are constants Asimismo tenemos

$$\begin{aligned} \left\langle \frac{\nabla}{dt} X, \frac{\nabla}{dt} X \right\rangle &= \left\langle \sum_{i=1}^{n-1} f'_i Y_i + \sum_{i=1}^{n-1} f_i \frac{\nabla}{dt} Y_i, \sum_{j=1}^{n-1} f'_j Y_j + \sum_{j=1}^{n-1} f_j \frac{\nabla}{dt} Y_j \right\rangle \\ &= \left\langle \sum_{i=1}^{n-1} f'_i Y_i, \sum_{j=1}^{n-1} f'_j Y_j \right\rangle + 2 \left\langle \sum_{i=1}^{n-1} f'_i Y_i, \sum_{j=1}^{n-1} f_j \frac{\nabla}{dt} Y_j \right\rangle \\ &\quad + \left\langle \sum_{i=1}^{n-1} f_i \frac{\nabla}{dt} Y_i, \sum_{j=1}^{n-1} f_j \frac{\nabla}{dt} Y_j \right\rangle, \end{aligned}$$

$$\langle R(c', X)c', X \rangle = - \sum_{i=1}^{n-1} f_i \langle R(c', Y_i)c', X \rangle,$$

y

$$\begin{aligned} \frac{d}{dt} \langle X, \sum_{j=1}^{n-1} f_j \frac{\nabla}{dt} Y_j \rangle &= \langle \sum_{i=1}^{n-1} f'_i Y_i + \sum_{i=1}^{n-1} f_i \frac{\nabla}{dt} Y_i, \sum_{j=1}^{n-1} f_j \frac{\nabla}{dt} Y_j \rangle \\ &\quad + \langle X, \sum_{j=1}^{n-1} f'_j \frac{\nabla}{dt} Y_j + \sum_{j=1}^{n-1} f_j \frac{\nabla^2}{dt^2} Y_j \rangle \\ &= \langle \sum_{i=1}^{n-1} f'_i Y_i, \sum_{j=1}^{n-1} f_j \frac{\nabla}{dt} Y_j \rangle + \langle \sum_{i=1}^{n-1} f_i \frac{\nabla}{dt} Y_i, \sum_{j=1}^{n-1} f_j \frac{\nabla}{dt} Y_j \rangle \\ &\quad + \langle X, \sum_{j=1}^{n-1} f'_j \frac{\nabla}{dt} Y_j \rangle - \langle \sum_{i=1}^{n-1} f_i Y_i, \sum_{j=1}^{n-1} f_j R(c', Y_j) c' \rangle \end{aligned}$$

Por tanto combinando las ecuaciones anteriores obtenemos

$$\begin{aligned} \langle \frac{\nabla}{dt} X, \frac{\nabla}{dt} X \rangle - \langle R(c', X) c', X \rangle &= \langle \sum_{i=1}^{n-1} f'_i Y_i, \sum_{j=1}^{n-1} f'_j Y_j \rangle \\ &\quad + \langle \sum_{i=1}^{n-1} f'_i Y_i, \sum_{j=1}^{n-1} f_j \frac{\nabla}{dt} Y_j \rangle + \frac{d}{dt} \langle X, \sum_{j=1}^{n-1} f_j \frac{\nabla}{dt} Y_j \rangle - \langle X, \sum_{j=1}^{n-1} f'_j \frac{\nabla}{dt} Y_j \rangle \\ &= \langle \sum_{i=1}^{n-1} f'_i Y_i, \sum_{j=1}^{n-1} f'_j Y_j \rangle + \frac{d}{dt} \langle X, \sum_{j=1}^{n-1} f_j \frac{\nabla}{dt} Y_j \rangle \\ &\quad + \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} f'_i f_j \left\{ \langle Y_i, \frac{\nabla}{dt} Y_j \rangle - \langle Y_j, \frac{\nabla}{dt} Y_i \rangle \right\} \end{aligned}$$

donde el ultimo termino es cero por 11.1 Así

$$\psi_X(s) = \int_0^s \langle \sum_{i=1}^{n-1} f'_i Y_i, \sum_{j=1}^{n-1} f'_j Y_j \rangle dt + \langle X, \sum_{j=1}^{n-1} f_j \frac{\nabla}{dt} Y_j \rangle(s)$$

similarmente

$$\psi_Y(s) = \int_0^s \langle \sum_{i=1}^{n-1} a'_i Y_i, \sum_{j=1}^{n-1} a'_j Y_j \rangle dt + \langle Y, \sum_{j=1}^{n-1} a_j \frac{\nabla}{dt} Y_j \rangle(s) = \langle Y, \sum_{j=1}^{n-1} a_j \frac{\nabla}{dt} Y_j \rangle(s),$$

ya que las a_i son constantes Finalmente, como $f_i(s) = a_i$, tenemos

$$\psi_X(s) = \int_0^s \langle \sum_{i=1}^{n-1} f'_i Y_i, \sum_{j=1}^{n-1} f'_j Y_j \rangle dt + \psi_Y(s),$$

consecuentemente

$$\psi_Y(s) \leq \psi_X(s),$$

y la igualdad se cumple si y solo si $f'_i = 0$ para $i = 1, \dots, n-1$ y por tanto si y solo si $X = Y$ ■

Teorema 11 3 3 Sea Y un campo de Jacobi normal a lo largo de una geodesica con tangente unitaria $c(s)$ que se anula en $s = 0$ Supongamos tambien que para cualquier campo vectorial normal Z a lo largo de c

$$\delta(s) \leq \frac{\langle R(c', Z), Z \rangle}{\|Z\|^2},$$

entonces tenemos las desigualdades

$$\frac{\|Y'\|(s)}{\|Y\|(s)} \leq \frac{f'_\delta(s)}{f_\delta(s)} \left(\frac{\|Y\|(s)}{f_\delta(s)} \right)' \leq 0, \quad \|Y\|(s) \leq f_\delta(s),$$

donde la segunda desigualdad se puede escribir

$$f_\delta(s) \left\langle \frac{\nabla}{\partial s} Y(s), Y(s) \right\rangle \leq f'_\delta(s) \langle Y(s), Y(s) \rangle$$

Demostracion

Definamos

$$u(s) = \langle Y, Y \rangle(s), \quad v(s) = f_\delta^2(s),$$

y

$$\mu(s) = \frac{\psi_Y(s)}{u(s)}, \quad \nu(s) = \frac{\int_0^s ((f'_\delta)^2 - \delta f_\delta^2) dt}{v(s)}$$

Notemos que

$$\int_0^s ((f'_\delta)^2 - \delta f_\delta^2) dt = (f'_\delta f_\delta)(s) - (f'_\delta f_\delta)(0) = (f'_\delta f_\delta)(s),$$

ya que $f_\delta f''_\delta + \delta f_\delta^2 = 0$ y $f_\delta(0) = 0$ Entonces

$$\frac{du}{ds} = 2 \left\langle \frac{\nabla}{\partial s} Y, Y \right\rangle = 2\psi_Y = 2\mu u, \quad \frac{dv}{ds} = 2f'_\delta f_\delta = 2\nu v$$

Y resolviendo las ecuaciones diferenciales obtenemos

$$u(s) = u(\epsilon) \exp\left\{2 \int_\epsilon^s \mu(t) dt\right\}, \quad v(s) = v(\epsilon) \exp\left\{2 \int_\epsilon^s \nu(t) dt\right\}$$

Usando dos veces la regla de l'Hopital

$$\lim_{\epsilon \rightarrow 0} \frac{u(\epsilon)}{v(\epsilon)} = \frac{\|Y'\|^2(0)}{(f'_\delta(0))^2} = 1,$$

y en consecuencia

$$\frac{u(s)}{v(s)} = \exp\left\{\int_0^s (\mu(t) - \nu(t)) dt\right\}$$

Veamos ahora que $\mu(t) \leq \nu(t)$ Sea W un campo vectorial paralelo a lo largo de $c(t)$ tal que $W(s) = Y(s)$ Introduzcamos el campo vectorial $Z = f_\delta \frac{W}{f_\delta(s)}$ Notemos que $Z(0) = 0$ y $Z(s) = W(s) = Y(s)$, entonces por la proposicion anterior

$$\psi_Y(s) \leq \psi_Z(s)$$

y teniendo en cuenta que

$$\begin{aligned} \psi_Z(s) &= \int_0^s \left\{ \left\langle f'_\delta \frac{W}{f_\delta(s)}, f'_\delta \frac{W}{f_\delta} \right\rangle - \left\langle R(c', f_\delta \frac{W}{f_\delta(s)})c', f_\delta \frac{W}{f_\delta} \right\rangle \right\} dt \\ &\leq \frac{\langle W, W \rangle}{(f_\delta(s))^2} \int_0^s \left\{ (f'_\delta)^2 - \delta f_\delta^2 \right\} dt = u(s)\nu(s) \end{aligned}$$

tenemos

$$\mu(s) = \frac{\psi_Y(s)}{u(s)} \leq \nu(s)$$

y por tanto $u(s)/v(s)$ es monotona decreciente, con $u(s)/v(s) \leq 1$, y se siguen las desigualdades ■

Proposicion 11 3 4 *Sea Y , como antes, un campo normal de Jacobi a lo largo de $c(t)$ tal que $Y(0) = 0$ Supongamos que δ y Δ son cotas superiores e inferiores de la curvatura a lo largo de $c(t)$ Sea k una funcion continua Entonces, mientras $S_\Delta > 0$ y $S'_\kappa > 0$*

$$\|Y'\|(t) \leq \begin{cases} \|Y\|(t) \left(\frac{2 S'_\kappa}{S_\kappa} - \frac{S'_\Delta}{S_\Delta} \right) (t) & \text{si } \kappa \leq \frac{1}{2}(\delta + \Delta) \\ \|Y'(0)\| S'_\kappa(t) & \text{si } \kappa \geq \frac{1}{2}(\delta + \Delta) \end{cases} \quad (11 4)$$

Demostracion

Sea f_κ como en 11 3, y Z un campo vectorial paralelo a lo largo de $c(t)$ tal que $\|Z\| = 1$, entonces

$$\langle f_\kappa Y' - f'_\kappa Y, Z \rangle(0) = 0$$

Por otra parte

$$\langle f_\kappa Y' - f'_\kappa Y, Z \rangle' = f_\kappa \langle \kappa Y - R(c', Y)c', Z \rangle,$$

y ya que $\langle (\kappa - R(c',)c')Y, Z \rangle$ es una forma bilineal simetrica tal que

$$\langle (\kappa - R(c',)c')Y, Y \rangle \leq \max(\Delta - \kappa, \kappa - \delta)\|Y\|^2,$$

tenemos

$$\langle f_\kappa Y' - f'_\kappa Y, Z \rangle' \leq f_\kappa \max(\Delta - \kappa, \kappa - \delta)\|Y\|,$$

Entonces por los dos teoremas anteriores

$$\begin{aligned} \langle f_\kappa Y' - f'_\kappa Y, Z \rangle'(t) &\leq \begin{cases} \frac{\|Y\|}{f_\Delta}(s)(\Delta - \kappa)f_\kappa f_\Delta(t) & \text{si } \kappa \leq \frac{1}{2}(\delta + \Delta) \text{ y} \\ & f_\Delta(t) > 0, 0 < t \leq s \\ (\kappa - \delta)f_\kappa f_\delta(t) & \kappa \geq \frac{1}{2}(\delta + \Delta) \end{cases} \\ = \begin{cases} \frac{\|Y\|}{f_\Delta}(s)(f'_\kappa f_\Delta - f_\kappa f'_\Delta)'(t) & \text{si } \kappa \leq \frac{1}{2}(\delta + \Delta) \text{ y} \\ & f_\Delta(t) > 0, 0 < t \leq s \\ (f_\kappa f'_\delta - f'_\kappa f_\delta)'(t) & \kappa \geq \frac{1}{2}(\delta + \Delta) \end{cases} \end{aligned}$$

Integrando de 0 a s , eligiendo Z tal que $Z(s) = Y'(s)/\|Y'\|(s)$ y teniendo en cuenta que, por hipotesis, $f'_\kappa > 0$, tenemos

$$\|Y'\|(s) \leq \begin{cases} \|Y\| \left(2\frac{f'_\kappa}{f_\kappa} - \frac{f'_\Delta}{f_\Delta} \right) (s) & \kappa \leq \frac{1}{2}(\delta + \Delta) \\ f'_\delta(s) & \kappa \geq \frac{1}{2}(\delta + \Delta) \end{cases}$$

donde hemos aplicado la desigualdad de Cauchy Schwartz La ultima desigualdad es debida a que $\|Y\|(s) \leq f_\delta(s)$ Finalmente poniendo las expresiones de f_κ y f_Δ la proposicion queda demostrada ■

11 4 Condiciones de geometria convexa

Vamos a analizar las condiciones de convexidad del cuadrado de la distancia de Rao, ρ^2 , equivalentes a la convexidad de la funcion real, funcion $\rho^2(p, \gamma(t))$ para cualquier geodesica γ y cualquier $p \in M$

Tenemos la siguiente importante proposicion

Proposicion 11 4 1 *Sea N un conjunto regular convexo en una variedad Riemanniana M Si las curvaturas seccionales en N son a lo sumo 0, o $\mathcal{K} > 0$ con $d(N) < \pi/2\sqrt{\mathcal{K}}$, entonces el cuadrado de la distancia Riemanniana es convexa*

Demostracion

Sea γ una geodesica en N y consideremos la familia de geodesicas de $p \in N$ a $\gamma(t)$ definida por $c_p(s, t) = \exp_p(s \exp_p^{-1}(\gamma(t)))$ Escribimos

$$c'_p = (c_p) \left(\frac{\partial}{\partial s} \right) \quad c_p = (c_p) \left(\frac{\partial}{\partial t} \right)$$

Es bien sabido que $\rho(p, \gamma(t)) = \|c'_p(s, t)\|$ es independiente de s y que la aplicacion $s \mapsto c_p(s, t)$ es una familia de campos de Jacobi, ya que $c_p(s, t)$ es una familia uniparametrica de geodesicas. Entonces,

$$\begin{aligned} \frac{\partial}{\partial t} \rho^2(p, \gamma(t)) &= \frac{\partial}{\partial t} \langle c'_p(s, t), c'_p(s, t) \rangle = \frac{\partial}{\partial t} \left\{ \int_0^1 \langle c'_p(s, t), c'_p(s, t) \rangle ds \right\} = \\ &= 2 \int_0^1 \left\langle \frac{\nabla}{\partial t} c'_p, c'_p \right\rangle ds = 2 \int_0^1 \left\langle \frac{\nabla}{\partial s} c_p, c'_p \right\rangle ds, \end{aligned}$$

ya que $\|c'_p\|$ es independiente de s y $(\nabla/\partial t)c'_p = (\nabla/\partial s)c_p$, donde $\nabla/\partial s$ y $\nabla/\partial t$ son las derivadas covariantes a lo largo de las curvas $c_p(\cdot, t)$ y $c_p(s, \cdot)$ respectivamente. Teniendo en cuenta que $(\nabla/\partial s)c'_p = 0$, tenemos

$$\frac{\partial}{\partial s} \langle c_p, c'_p \rangle = \left\langle \frac{\nabla}{\partial s} c_p, c'_p \right\rangle,$$

por tanto, observando que $c_p(0, t) = 0$, obtenemos

$$\frac{\partial}{\partial t} \rho^2(p, \gamma(t)) = 2 \langle c_p(1, t), c'_p(1, t) \rangle$$

Derivando de nuevo, como $(\nabla/\partial t)c_p(1, t) = (\nabla/\partial t)\gamma = 0$ resulta

$$\frac{\partial^2}{\partial t^2} \rho^2(p, \gamma(t)) = 2 \langle c_p(1, t), \frac{\nabla}{\partial s} c_p(1, t) \rangle,$$

donde denotamos el campo de Jacobi c_p a lo largo de $s \rightarrow c_p(s, t)$ por $J(s)$, entonces $J(0) = 0$, $J(1) = \gamma(t)$ es independiente de p , y $(\nabla/\partial s)c_p(1, t) = (\nabla/\partial s)J(1)$. Podemos decomponer J en sus componentes normal y tangencial $J = J^{nor} + J^{tan}$. Para la componente tangencial, ya que $J^{tan}(0) = 0$, obtenemos

$$J^{tan}(s) = b s c'_p(s, t),$$

and

$$\frac{\nabla}{\partial s} J^{tan}(s) = b c'_p(s, t) + b s \frac{\nabla}{\partial s} c'_p(s, t) = b c'_p(s, t),$$

resultando $J^{tan}(1) = (\nabla/\partial s)J^{tan}(1)$

Para la componente normal, por el teorema de comparacion de Rauch, y teniendo en cuenta la velocidad de la geodesica, tenemos

$$\begin{aligned} f_{\Delta}(\|c'_p(1, t)\|) \langle \frac{1}{\|c'_p(1, t)\|} \frac{\nabla}{\partial s} J^{no}(1), J^{nor}(1) \rangle &\geq \\ &\geq f'_{\Delta}(\|c'_p(1, t)\|) \langle J^{no}(1), J^{nor}(1) \rangle \end{aligned}$$

Combinando estos resultados llegamos a

$$\begin{aligned} \langle \frac{\nabla}{\partial s} J^{tan}(1), J^{tan}(1) \rangle &= \langle J^{tan}(1), J^{tan}(1) \rangle \geq 0 \\ \langle \frac{\nabla}{\partial s} J^{tan}(1), J^{nor}(1) \rangle &= -\langle \frac{\nabla}{\partial s} J^{nor}(1), J^{tan}(1) \rangle = 0, \end{aligned}$$

y

$$\langle \frac{\nabla}{\partial s} J^{nor}(1), J^{nor}(1) \rangle \geq \|c'_p(1, t)\| \frac{f'_{\Delta}(\|c'_p(1, t)\|)}{f_{\Delta}(\|c'_p(1, t)\|)} \|J^{nor}(1)\|^2,$$

resultando

$$\begin{aligned} \frac{1}{2} \frac{\partial^2}{\partial t^2} \rho^2(p, \gamma(t)) &= \langle J(1), \frac{\nabla}{\partial s} J(1) \rangle \\ &\geq \|c'_p(1, t)\| \frac{f'_{\Delta}(\|c'_p(1, t)\|)}{f_{\Delta}(\|c'_p(1, t)\|)} \|J^{nor}(1)\|^2 + \|J^{tan}(1)\|^2 \end{aligned}$$

Vamos a considerar dos casos. Primero, supongamos $\Delta(s) = 0$. En este caso, como $f_{\Delta}(0) = 0$, $f_{\Delta}(s) = bs$, obteniendo

$$\frac{1}{2} \frac{\partial^2}{\partial t^2} \rho^2(p, \gamma(t)) \geq \|J(1)\|^2 = \rho^2(\gamma(0), \gamma(1)) > 0, \quad (11.5)$$

y $\rho^2(p, \gamma(t))$ es una funcion convexa

Segundo, supongamos $\Delta(s) = \mathcal{K}$ donde $\mathcal{K} > 0$, y adicionalmente que la variedad tiene diametro $d(M) < \pi/2\sqrt{\mathcal{K}}$. En este caso, $f_{\Delta}(0) = 0$, $f_{\Delta}(s) = \sin(\sqrt{\mathcal{K}}s)$, y por tanto

$$1 > \|c'_p(1, t)\| \frac{f'_{\Delta}(\|c'_p(1, t)\|)}{f_{\Delta}(\|c'_p(1, t)\|)} = \frac{\sqrt{\mathcal{K}} \|c'_p(1, t)\| \cos(\sqrt{\mathcal{K}} \|c'_p(1, t)\|)}{\sin(\sqrt{\mathcal{K}} \|c'_p(1, t)\|)} > 0,$$

ya que $\|c'_p(1, t)\| = \rho(p, \gamma(t)) < d(M)$ y $0 < \sqrt{\mathcal{K}} \|c'_p(1, t)\| < \pi/2$, obteniendo

$$\begin{aligned} \frac{1}{2} \frac{\partial^2}{\partial t^2} \rho^2(p, \gamma(t)) &\geq \frac{\sqrt{\mathcal{K}} \|c'_p(1, t)\| \cos(\sqrt{\mathcal{K}} \|c'_p(1, t)\|)}{\sin(\sqrt{\mathcal{K}} \|c'_p(1, t)\|)} \|J(1)\|^2 \\ &= \frac{\sqrt{\mathcal{K}} \|c'_p(1, t)\| \cos(\sqrt{\mathcal{K}} \|c'_p(1, t)\|)}{\sin(\sqrt{\mathcal{K}} \|c'_p(1, t)\|)} \rho^2(\gamma(0), \gamma(1)) > 0, \end{aligned} \quad (11.6)$$

y de nuevo, $\rho^2(p, \gamma(t))$ es una función convexa ■

11.5 Coordenadas esféricas geodésicas

A fin de describir la noción de *coordenadas esféricas geodésicas*, introduciremos primero la propiedad siguiente

Proposición 11.5.1 *Sea Y un campo de Jacobi a lo largo de $\gamma_\xi(s) = \exp_p(s\xi)$, con $p \in M$, $\xi \in M_p$ determinado por $Y(0) = 0$ $(\nabla/ds)Y(0) = \eta$. Entonces*

$$Y(s) = (\exp_p)_{|s\xi}(s\eta)$$

Demostración

Consideremos la familia uniparamétrica de geodésicas $c(s, t)$ definida como

$$c(s, t) = \exp_p(s\zeta(t)),$$

donde $\zeta(t)$ es un camino en M_p con $\zeta(0) = \xi$ y $\zeta'(0) = \eta$. Identificando, como suele ser corriente, los elementos de M_p como elementos de cualquier $(M_p)_{s\xi}$, donde $(M_p)_{s\xi}$ es el espacio tangente en $s\xi$. En este caso sabemos que el campo vectorial

$$Z(s) = c(s, 0) = c \left(\frac{\partial}{\partial t} \Big|_{t=0} \right)$$

es un campo de Jacobi, y

$$Z(s) = (\exp_p)_{|s\xi}(s\zeta'(0)) = s (\exp_p)_{|s\xi}(\eta),$$

por tanto $Z(0) = 0$, y con la misma notación que la sección previa,

$$\left(\frac{\nabla}{\partial s} Z \right) (0) = \left(\frac{\nabla}{\partial s} c \right) (0, 0) = \left(\frac{\nabla}{\partial t} c' \right) (0, 0) = \zeta'(0) = \eta,$$

concluimos que

$$(\exp_p)_{|s\xi}(s\eta) = Y(s)$$

■

Dado un punto p en la variedad Riemanniana, podemos ahora introducir *coordenadas esfericas geodesicas* en D_p , mediante $\exp_p|_{\mathfrak{D}_p}$, la restriccion de \exp_p en \mathfrak{D}_p . Supongamos un sistema de coordenadas en \mathfrak{S}_p , $\xi = \xi(u)$ donde u varia en un dominio de \mathbf{R}^{n-1} . Un sistema de coordenadas en D_p se define entonces de la forma

$$v(\rho, u) = \exp_p(\rho \xi(u))$$

Escribamos

$$\gamma'_\xi(\rho) = v_* \left(\frac{\partial}{\partial \rho} \right) \quad \text{y} \quad Y_\alpha(\rho, \xi) = v \left(\frac{\partial}{\partial u^\alpha} \right),$$

Entonces, para cualquier $\xi \in \mathfrak{S}_p$, Y_α , $\alpha = 1, \dots, n-1$ son campos de Jacobi a lo largo de γ_ξ determinados por las condiciones iniciales

$$Y_\alpha(0, \xi) = 0 \quad \text{y} \quad \left(\frac{\nabla}{\partial \rho} \right) Y_\alpha(0, \xi) = \xi \left(\frac{\partial}{\partial u^\alpha} \right), \quad \alpha = 1, \dots, n-1,$$

donde $v_*(\partial/\partial u^\alpha) \in M_{\exp_p(\rho \xi)}$ es ortogonal a $\gamma'_\xi(\rho)$, como consecuencia del lema de Gauss. Asi pues, en un entorno de $\exp_p(\rho \xi(u))$ existe un sistema de coordenadas (ρ, u) , tal que el campo vectorial base correspondiente

$$\partial_1 = \gamma'_\xi(\rho), \partial_2 = Y_1(\rho, \xi), \dots, \partial_n = Y_{n-1}(\rho, \xi)$$

satisface

$$g_{11}(\rho, u) = \langle \partial_1, \partial_1 \rangle = 1, \quad g_{1\alpha}(\rho, u) = g_{\alpha 1}(\rho, u) = \langle \partial_1, \partial_\alpha \rangle = 0, \quad \alpha = 2, \dots, n,$$

donde $g_{\alpha\beta}$ son las componentes del tensor metrico, y ∂_α , $\alpha = 2, \dots, n$ son campos de Jacobi a lo largo de las geodesicas $\gamma(s) = \exp_p(\rho \xi(u))$. De esta forma la metrica Riemanniana puede expresarse como

$$ds^2 = d\rho^2 + \sum_{\alpha, \beta=2}^n g_{\alpha\beta}(\rho, u) du^\alpha du^\beta, \quad (11.7)$$

donde

$$g_{\alpha\beta}(\rho, u) = \langle Y_\alpha(\rho, \xi(u)), Y_\beta(\rho, \xi(u)) \rangle, \quad \alpha, \beta = 2, \dots, n$$

Tenemos ahora el siguiente importante teorema

Teorema 11.5.2 (Teorema de comparacion de Bishop) Sean κ y \mathcal{K} cotas inferior y superior de la curvatura en una region donde $\det G > 0$ entonces en esta region tenemos

$$(n-1) \frac{f'_\kappa(\rho)}{f_\kappa(\rho)} \geq \frac{\partial}{\partial \rho} \log \sqrt{\det G} \geq (n-1) \frac{f'_\mathcal{K}(\rho)}{f_\mathcal{K}(\rho)}$$

Demostracion

Sean $G = (g_{\alpha\beta})_{n \times n}$ las componentes del tensor fundamental escritas en forma matricial, a lo largo de la geodesica $\gamma_\xi(\rho)$ tenemos

$$\begin{aligned} \frac{\partial}{\partial \rho} \log \sqrt{\det G} &= \frac{1}{2 \det G} \sum_{\alpha \leq \beta=2}^n \frac{\partial \det G}{\partial g_{\alpha\beta}} \frac{\partial g_{\alpha\beta}}{\partial \rho} = \\ &= \frac{1}{2} \sum_{\alpha \beta=2}^n g^{\alpha\beta} \frac{\partial}{\partial \rho} \langle Y_\alpha, Y_\beta \rangle = \sum_{\alpha \beta=2}^n g^{\alpha\beta} \langle Y_\alpha, \frac{\nabla}{\partial \rho} Y_\beta \rangle, \end{aligned}$$

donde $g^{\alpha\beta}$ son los coeficientes de G^{-1} Mediante una transformacion lineal de los campos Y_α , por tanto sin variar el valor de la expresion anterior, podemos hacer que Y_α sea orthogonal en el punto (ρ, u) Entonces, tenemos

$$\frac{\partial}{\partial \rho} \log \sqrt{\det G} = \sum_{\alpha=2}^n \frac{1}{\|Y_\alpha\|^2} \langle Y_\alpha, \frac{\nabla}{\partial \rho} Y_\alpha \rangle,$$

ahora, aplicando los teoremas anteriores,

$$(n-1) \frac{f'_\kappa(\rho)}{f_\kappa(\rho)} \geq \frac{\partial}{\partial \rho} \log \sqrt{\det G} \geq (n-1) \frac{f'_\mathcal{K}(\rho)}{f_\mathcal{K}(\rho)},$$

y ya que los campos de Jacobi se anulan en el origen, $f_\kappa(\rho) = \mathbf{S}_\kappa$ y $f_\mathcal{K} = \mathbf{S}_\mathcal{K}$ y obtenemos el teorema ■

Consideremos dos casos Primero, supongamos que $\mathcal{K} \leq 0$ En este caso $f_\mathcal{K}(\rho) = \mathbf{S}_\mathcal{K}(\rho) = \sinh(\sqrt{-\mathcal{K}}\rho)$, y $f'_\mathcal{K}(\rho) = \mathbf{C}_\mathcal{K}(\rho) = \sqrt{-\mathcal{K}} \cosh(\sqrt{-\mathcal{K}}\rho)$ si $\mathcal{K} < 0$, y por tanto

$$\frac{\partial}{\partial \rho} \log \sqrt{\det G} \geq (n-1) \sqrt{-\mathcal{K}} \coth(\sqrt{-\mathcal{K}}\rho), \tag{11.8}$$

el caso $\mathcal{K} = 0$ se puede obtener directamente o por continuidad, resultando

$$\frac{\partial}{\partial \rho} \log \sqrt{\det G} \geq \frac{n-1}{\rho}$$

Segundo caso, supongamos $\mathcal{K}(\rho) = \mathcal{K}$ donde $\mathcal{K} > 0$, y que la variedad tiene diametro $d(M) < \pi/2\sqrt{\mathcal{K}}$ En tal situacion $f_\mathcal{K}(\rho) = \mathbf{S}_\mathcal{K}(\rho) = \sin(\sqrt{\mathcal{K}}\rho)$, y $f'_\mathcal{K}(\rho) = \mathbf{C}_\mathcal{K}(\rho) = \sqrt{\mathcal{K}} \cos(\sqrt{\mathcal{K}}\rho)$, y por tanto

$$\frac{\partial}{\partial \rho} \log \sqrt{\det G} \geq (n-1) \sqrt{\mathcal{K}} \cot(\sqrt{\mathcal{K}}\rho) > 0 \tag{11.9}$$

11 6 Volúmenes y teoremas de comparación

Podemos usar el teorema de Bishop con el propósito de obtener el volumen de una bola de radio r en una variedad Riemanniana cuyas curvaturas seccionales sean constantes y dar cotas para este volumen cuando las curvaturas seccionales estén acotadas. Tenemos los siguientes resultados

Proposición 11 6 1 *Si las curvaturas seccionales son constantes e iguales a \mathcal{K} , el volumen de una bola Riemanniana de radio r y centro $p \in M$ viene dado por*

$$\text{vol}(B_p(r)) = \frac{2\pi^{n/2}}{\Gamma(n/2)} \int_0^r S_{\mathcal{K}}^{n-1}(t) dt$$

Demostración

A partir de la expresión 11 7 en la sección 11 5, y escribiendo $G_{\mathcal{K}}$ para el tensor métrico, obtenemos

$$\text{vol}(B_p(r)) = \int_0^r \int_{\xi^{-1}(S_n)} \sqrt{\det G_{\mathcal{K}}} du d\rho$$

Donde S_n es la esfera unidad en M_p . Además, por el teorema de comparación de Bishop, cuando las curvaturas seccionales son constantes

$$\frac{\partial}{\partial \rho} \log \sqrt{\det G_{\mathcal{K}}(\rho, u)} = (n-1) \frac{S'_{\mathcal{K}}}{S_{\mathcal{K}}}(\rho),$$

entonces, integrando esta expresión, tenemos

$$\sqrt{\det G_{\mathcal{K}}(\rho, u)} = S_{\mathcal{K}}^{n-1} \Omega_{\mathcal{K}}(u)$$

Pero, de hecho, $\Omega_{\mathcal{K}}$ no depende de \mathcal{K} . Obviamente

$$\lim_{\rho \rightarrow 0} \frac{\sqrt{\det G_{\mathcal{K}}(\rho, u)}}{\rho^{n-1} \Omega(u)} = 1,$$

donde $\Omega(u) du$ es el elemento de área de la esfera unidad en una variedad Euclídea, y, como

$$\lim_{\rho \rightarrow 0} \frac{S_{\mathcal{K}}^{n-1}}{\rho^{n-1}} = 1,$$

concluimos que $\Omega_{\mathcal{K}} = \Omega$. Así podemos escribir

$$\text{vol}(B_p(r)) = \int_{\xi^{-1}(S_n)} \Omega(u) du \int_0^r S_{\mathcal{K}}^{n-1}(\rho) d\rho$$

Finalmente, es facil comprobar que

$$\int_{\xi^{-1}(S_n)} \Omega(u) du = \frac{2\pi^{n/2}}{\Gamma(n/2)},$$

de lo que se sigue la proposicion ■

Proposicion 11 6 2 *Cuando las curvaturas seccionales son constantes e iguales a \mathcal{K} y $\mathcal{K}S_{\mathcal{K}}^2(r) < 1$ tenemos la expresion siguiente para el volumen de una bola Riemanniana de radio r*

$$\text{vol}(B(r)) = \frac{2\pi^{n/2}}{n\Gamma(n/2)} S_{\mathcal{K}}^n(r) \left\{ 1 + \sum_{j=1}^{\infty} \frac{n\Gamma(j + \frac{1}{2})}{\sqrt{\pi}(n + 2j)} \frac{\mathcal{K}^j S_{\mathcal{K}}^{2j}(r)}{j!} \right\} \quad (11 10)$$

Demostracion

De la proposicion previa

$$\text{vol}(B(r)) = \frac{2\pi^{n/2}}{\Gamma(n/2)} \int_0^r S_{\mathcal{K}}^{n-1}(t) dt$$

Entonces, teniendo en cuenta, por la definicion de $S_{\mathcal{K}}$, que

$$\{S'_{\mathcal{K}}(t)\}^2 + \mathcal{K}\{S_{\mathcal{K}}(t)\}^2 = 1$$

y haciendo $y = S_{\mathcal{K}}^2(t)/S_{\mathcal{K}}^2(r)$, tenemos

$$\int_0^r S_{\mathcal{K}}^{n-1}(t) dt = \frac{1}{2} S_{\mathcal{K}}^n(r) \int_0^1 y^{\frac{n-2}{2}} (1 - \mathcal{K}S_{\mathcal{K}}^2(r)y)^{-\frac{1}{2}} dy$$

Por otra parte tenemos una relacion entre este tipo de de integrales y las funciones hipergeometricas generalizadas, vease Abramowitz [1],

$${}_pF_q(a_1, \dots, a_p, b_1, \dots, b_q, z) \equiv \sum_{j=0}^{\infty} \frac{(a_1)_j}{(b_1)_j} \frac{(a_p)_j z^j}{j!} \quad (11 11)$$

donde $(a)_j = a(a+1)\dots(a+j-1)$ y z es cualquiera si $p \leq q$, $\|z\| < 1$ si $p = q + 1$ y divergen para todo $z \neq 0$ si $p > q + 1$

Pues bien, la relacion que comentabamos viene dada por

$${}_2F_1(a, b, c, z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a} dt, \operatorname{Re}(c) > \operatorname{Re}(b) > 0$$

Entonces esto conduce a

$$\begin{aligned} \int_0^r \mathbf{S}_{\mathcal{K}}^{n-1}(t) dt &= \frac{1}{2} \mathbf{S}_{\mathcal{K}}^n(r) \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n+2}{2})} {}_2F_1\left(\frac{1}{2}, \frac{n}{2}, \frac{n+2}{2}, \mathcal{K} \mathbf{S}_{\mathcal{K}}^2(r)\right) = \\ &= \frac{1}{2} \mathbf{S}_{\mathcal{K}}^n(r) \frac{1}{\Gamma(\frac{1}{2})} \sum_{j=0}^{\infty} \frac{\Gamma(\frac{1}{2}+j)}{\frac{n}{2}+j} \frac{\mathcal{K}^j \mathbf{S}_{\mathcal{K}}^{2j}(r)}{j!}, \end{aligned}$$

que demuestra la proposicion ■

Proposicion 11 6 3 Sea $\operatorname{vol}(B_p(r))$ el volumen de una bola $B_p(r)$ con centro p y radio r Entonces

$$\operatorname{vol}_{\kappa}(B_{\underline{p}}(r)) \geq \operatorname{vol}(B_p(r)) \geq \operatorname{vol}_{\mathcal{K}}(B_{\bar{p}}(r)),$$

donde $\operatorname{vol}_{\kappa}(B_{\underline{p}}(r))$ y $\operatorname{vol}_{\mathcal{K}}(B_{\bar{p}}(r))$ son los volúmenes de bolas de radio r y centros arbitrarios \underline{p} y \bar{p} respectivamente, en variedades con curvaturas seccionales constantes κ y \mathcal{K}

Demostracion

Si integramos, de ρ_0 a ρ , las desigualdades del teorema de comparacion de Bishop, obtenemos

$$\frac{\mathbf{S}_{\kappa}^{n-1}(\rho)}{\mathbf{S}_{\kappa}^{n-1}(\rho_0)} \geq \frac{\sqrt{\det G(\rho, u)}}{\sqrt{\det G(\rho_0, u)}} \geq \frac{\mathbf{S}_{\mathcal{K}}^{n-1}(\rho)}{\mathbf{S}_{\mathcal{K}}^{n-1}(\rho_0)}$$

Ademas

$$\lim_{\rho_0 \rightarrow 0} \sqrt{\det G(\rho_0, u)} \frac{\mathbf{S}_{\kappa}^{n-1}(\rho)}{\mathbf{S}_{\kappa}^{n-1}(\rho_0)} \geq \sqrt{\det G(\rho, u)} \geq \lim_{\rho_0 \rightarrow 0} \sqrt{\det G(\rho_0, u)} \frac{\mathbf{S}_{\mathcal{K}}^{n-1}(\rho)}{\mathbf{S}_{\mathcal{K}}^{n-1}(\rho_0)},$$

y, ya que

$$\lim_{\rho_0 \rightarrow 0} \frac{\sqrt{\det G(\rho_0, u)}}{\mathbf{S}_{\kappa}^{n-1}(\rho_0)} = \lim_{\rho_0 \rightarrow 0} \frac{\sqrt{\det G(\rho_0, u)}}{\mathbf{S}_{\mathcal{K}}^{n-1}(\rho_0)} = \Omega(u),$$

con $\Omega(u)$ como en la demostracion de 11 6 1, concluimos que

$$\mathbf{S}_{\kappa}^{n-1}(\rho)\Omega(u) \geq \sqrt{\det G(\rho, u)} \geq \mathbf{S}_{\mathcal{K}}^{n-1}(\rho)\Omega(u),$$

y se sigue el resultado deseado ■

Bibliografía

- [1] M Abramowitz *Handbook of Mathematical Functions* Dover Publications, Inc , New York, 1970
- [2] S Amari *Differential Geometrical Methods in Statistics* Volume 28 of *Lecture notes in statistics*, Springer Verlag, New York, 1985
- [3] C Atkinson and A F S Mitchell Rao's distance measure *Sankhya*, 43 A 345–365, 1981
- [4] R R Bahadur On Fisher's bounds for asymptotic variances *Ann Math Statist* , 35 1545–1552, 1964
- [5] O E Barndorff Nielsen Differential geometry and statistics some mathematical aspects *Indian J Math* 29 335–350, 1987
- [6] O E Barndorff Nielsen *Differential geometry in statistical inference*, chapter Differential and integral geometry in statistical inference, pages 95–161 Volume 10 of *Lecture Notes Monographs Series*, Institute of Mathematical Statistics, Hayward, California, 1987
- [7] O E Barndorff Nielsen *Parametric Statistical Models and Likelihood* Volume 50 of *Lecture notes in statistics*, Springer Verlag, New York, 1988
- [8] O E Barndorff Nielsen Strings, tensorial combinants, and bartlett adjustments *Proc Roy Soc London*, 406 A 127–137, 1986
- [9] O E Barndorff Nielsen and P Blaesild Strings mathematical theory and statistical examples *Proc Roy Soc London*, 411 A 155–176, 1987
- [10] O E Barndorff Nielsen and K Pedersen Sufficient data reduction and exponential families *Math Scand* , 22 197–202, 1968
- [11] A Bhattacharyya A measure of divergence betwween two multinomial populations *Shankhya*, 7 401–406, 1946

- [12] R Bishop and R Crittenden *Geometry of Manifolds* Academic Press, New York, 1964
- [13] J Burbea Informative geometry of probability spaces *Expositiones Mathematicae*, 4 347–378, 1986
- [14] J Burbea and J M Oller The information metric for univariate linear elliptic models *Statistics & Decisions*, 6 209–221, 1988
- [15] J Burbea and J M Oller *On Rao distance asymptotic distribution* preprint series 67, Universitat de Barcelona, June 1989
- [16] J Burbea and C R Rao Entropy differential metric, distance and divergence measures in probability spaces a unified approach *J Multivariate Anal*, 12 575–596, 1982
- [17] M Calvo and J M Oller An explicit solution of information geodesic equations for the multivariate normal model *Statistics & Decisions*, 9 119–138, 1991
- [18] L Le Cam On some asymptotic properties of maximum likelihood estimates and related Baye's estimates *Univ of Calif Publ in Statist*, 1 277–330, 1953
- [19] J Castillo Sufficient geometrical conditions for Cramer Rao inequality *submitted*, 1992
- [20] N N Cencov *Statistical Decision Rules and Optimal Inference* Nauka, Moscow (en ruso, version inglesa, 1982, Math Monographs, 53, Amer Math Soc, Providence), 1972
- [21] I Chavel *Eigenvalues in Riemannian Geometry Pure and applied mathematics*, Academic Press, Inc, Orlando, Florida, 1984
- [22] J Cheeger and D G Ebin *Comparison Theorems in Riemannian Geometry* North Holland, New York, 1975
- [23] H Chernoff Large sample theory parametric case *Ann Math Stat*, 27 1–22, 1956
- [24] Bradley Efron Defining the curvature of a statistical problem (with applications and second order efficiency) *Ann Statist*, 3 1189–1242, 1975

- [25] M Emery and G Mokobodzki *Seminaire de Probabilites XXV*, chapter Sur le barycentre d'une probabilité dans une variété, pages 220–233 Volume 1485 of *Lecture Notes in Maths*, Springer, 1991
- [26] Harrie Hendriks *A Cramer Rao type lower bound for estimators with values in a manifold* *J Multivariate Anal* , 38 245–261, 1991
- [27] N J Hicks *Notes on Differential Geometry* Van Nostrand Reinhold Comp , London, 1965
- [28] I A Ibragimov and R Z Has'minskiĭ *Statistical Estimation* Springer-Verlag, New York, 1981
- [29] W James and C Stein Estimation with quadratic loss *Proc Fourth Berkeley Symp Math Statist Prob* , 1 361–379, 1961
- [30] H Jeffreys An invariant form for the prior probability in estimation problems *Proc Roy Soc , A*, 196 453–461, 1946
- [31] H Jeffreys An invariant form for the prior probability in estimation problems *Proc Royal Soc* , 186 A 453–461, 1946
- [32] P E Jupp and K V Mardia A unified view of the theory of directional statistics, 1975 1988 *International Statistical Review*, 57(3) 261–294, 1989
- [33] H Karcher Riemannian center of mass and mollifier smoothing *Comm Pure Appl Math* , 30 509–541, 1977
- [34] D Kelker Distribution theory of spherical distributions and a location scale parameter generalization *Sankhya*, 32 A 419–430, 1970
- [35] W S Kendall Convexity and hemisphere *J London Math Soc* , 43(2) 567–576, 1991
- [36] W S Kendall Probability, convexity and harmonic maps with small image 1 uniqueness and fine existence *Proc London Math Soc* , 61(3) 371–406, 1990
- [37] S Kobayashi and K Nomizu *Foundations of differential geometry* Interscience Publishers, John Wiley, New York, 1969

- [38] S Kullback and R A Leibler On information and sufficiency *Ann Math Statist* , 22 79–86, 1951
- [39] J T Kent K V Mardia and J M Bibby *Multivariate Analysis* Academic Press, London, 1979
- [40] L Le Cam Maximum likelihood an introduction *International Statistical Review*, 58, 2 153–171, 1990
- [41] E L Lehmann *Theory of Point Estimation* John Wiley, New York 1983
- [42] E L Lehmann A general concept of unbiasedness *Ann Math Statist* , 22 587–592, 1951
- [43] L T Madsen *The geometry of statistical models A generalitation of curvature* Research report 79 1, Danish Medical Research Council Statistical Research Unit, 1982
- [44] P C Mahalanobis On the generalized distance in statistics *Proc Natl Inst Sci India*, 2 49–55, 1936
- [45] P McCullagh and D R Cox Invariants and likelihood ratio statistics *Ann Statist* , 14 1419–1430, 1986
- [46] A F S Mitchell and W J Krzanowski The Mahalanobis distance and elliptic distributions *Biometrika*, 72 464–467, 1985
- [47] R J Muirhead *Aspects of Multivariate Statistical Theory* John Wiley, New York, 1982
- [48] M K Murray Coordinate systems and taylor series in statistics *Proc Roy Soc London*, 415 A 445–452, 1988
- [49] J M Oller Information metric for extreme value and logistic probability distributions *Sankhya*, 49 A 17–23, 1987
- [50] J M Oller *Statistical data analysis and inference*, chapter Some geometrical aspects of data analysis and statistics, pages 41–58 Elsevier science publishers B V , North Holland, Amsterdam, 1989
- [51] J M Oller and C M Cuadras Rao's distance for multinomial negative distributions *Sankhya*, 47 A 75–83, 1985

- [52] B L S Prakasa Rao Cramer–Rao type integral inequalities for estimators of functions of multidimensional parameter *Sankhya*, 54 A 53–73, 1992
- [53] C R Rao Diversity and dissimilarity coefficients a unified approach *J Theoretical Population Biology*, 21 24–43, 1982
- [54] C R Rao On the distance between two populations *Sankhya*, 9 246–248, 1949
- [55] A Renyi *On measures of entropy and information*, pages 547–561 Volume 1, University of California Press, 1961
- [56] C E Shannon A mathematical theory of communication *Bell Syst Tech J*, 27 379–423, 1948
- [57] M Spivak *A Comprehensive Introduction to Differential Geometry, Vol I–V* Publish or Perish, Berkeley, 1979

AGRADECIMIENTOS

En primer lugar tengo que decir que el desarrollo de esta memoria no hubiera sido posible sin el inestimable apoyo, en todos los sentidos, de mi Director de tesis Josep Maria Oller Sala. El apoyo animico es tan importante como el intelectual en tareas tan laboriosas, por eso quiero agradecerlo a todos aquellos miembros del Departamento de Estadística de la Universidad de Barcelona que me han dado su confianza y aliento. En el mismo sentido quiero recordar a mi familia, que ha sabido comprender mis ausencias.

