

---

# La información de Fisher: teoría y aplicación en las distribuciones beta y gamma

---

María José Gil Herrera





UNIVERSIDAD DEL VALLE DE GUATEMALA  
Facultad de Ciencias y Humanidades

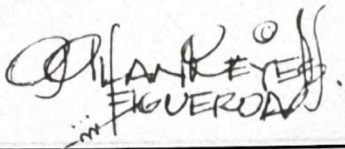


## La información de Fisher: teoría y aplicación en las distribuciones beta y gamma

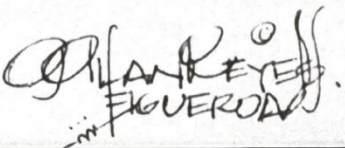
Trabajo de graduación en modalidad de tesis presentado por  
María José Gil Herrera  
para optar al grado académico de Licenciada en Matemática Aplicada


Guatemala, 2025

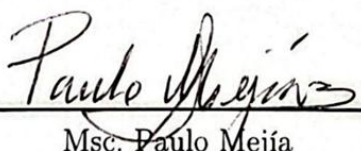
Vo.Bo.:

(f)   
Phd. Alan Reyes

Tribunal Examinador:

(f)   
Phd. Alan Reyes

(f)   
Msc. Dorval Carías

(f)   
Msc. Paulo Mejía

Fecha de aprobación: Guatemala, 03 de julio de 2025.

## Prefacio

Esta tesis nace de mi interés por entender cómo se cruzan la estadística y la geometría. Durante mis años de universidad me llamó la atención que los espacios de distribuciones de probabilidad pudieran concebirse como objetos geométricos con propiedades propias. Lo que al principio era una curiosidad matemática fue tomando forma y convirtiéndose en un proyecto que une herramientas de la inferencia estadística con ideas de la geometría diferencial.

El camino que me llevó hasta aquí no fue directo. He pasado temporadas estudiando probabilidad, estadística y geometría, y cada descubrimiento en uno de estos campos me daba nuevas pistas sobre los otros. Con el tiempo, estos avances se fueron complementando y dieron lugar al enfoque que presento en estas páginas.

Quisiera que esta tesis sirva de agradecimiento a quienes han abierto este camino —C.R. Rao, N.N. Chentsov y Shun-ichi Amari, entre otros— y que ayude a quienes se acerquen a estas ideas desde diversos contextos. Procuré detallar resultados que a menudo se omiten en la literatura, con la intención de tender un puente entre la estadística clásica y la geometría moderna.

Este trabajo no pretende cerrar el tema, sino invitar a otros a explorar las conexiones geométricas en los fundamentos de la inferencia estadística y a encontrar nuevas perspectivas en problemas clásicos y actuales del análisis de datos.

---

## Agradecimientos

---

A los que hicieron que las desveladas no fueran tanto sufrimiento. A mis amigos de Debate, particularmente a Lisa, Elisa y Juanpi; a Óscar y a Joshua, y a Luis que me insistió tanto en que la terminara.

<b>Prefacio</b>	III
<b>Agradecimientos</b>	IV
<b>Lista de figuras</b>	VIII
<b>Lista de cuadros</b>	IX
<b>Resumen</b>	X
<b>1. Introducción</b>	1
<b>2. Objetivos</b>	2
<b>3. Justificación</b>	3
<b>4. Marco Teórico</b>	4
4.1. Probabilidad e inferencia estadística . . . . .	4
4.2. Geometría de la información . . . . .	4
4.3. Aplicaciones modernas . . . . .	5
<b>5. Antecedentes</b>	6
<b>6. Alcance</b>	7
6.1. Delimitación del estudio . . . . .	7
6.2. Limitaciones . . . . .	7
6.3. Contribuciones principales . . . . .	8
<b>7. Conceptos de probabilidad y estadística</b>	9
7.1. Espacios de probabilidad . . . . .	9
7.1.1. Espacios de probabilidad completos . . . . .	11
7.1.2. Propiedades de los espacios de probabilidad . . . . .	11
7.2. Variables aleatorias . . . . .	12
7.2.1. Función de masa de probabilidad . . . . .	12
7.2.2. Función de distribución acumulada . . . . .	13
7.2.3. Función de densidad de probabilidad . . . . .	14
7.2.4. Valor esperado . . . . .	16
7.3. Estimación de parámetros . . . . .	16

7.3.1. Máxima verosimilitud . . . . .	16
7.3.2. Función de score y de información . . . . .	19
<b>8. Conceptos de geometría diferencial</b> . . . . .	<b>24</b>
8.1. Conceptos topológicos . . . . .	24
8.1.1. Funciones continuas . . . . .	25
8.1.2. Espacios 2º enumerables . . . . .	25
8.1.3. Localmente euclídiano . . . . .	26
8.2. Variedades diferenciables . . . . .	26
8.2.1. <i>Manifolds</i> o variedades topológicas . . . . .	27
8.2.2. Cartas locales y atlas . . . . .	27
8.2.3. Variedades diferenciables . . . . .	29
8.2.4. Vectores tangentes y espacio tangente . . . . .	30
8.3. Formas fundamentales . . . . .	33
8.3.1. Primera forma fundamental . . . . .	33
8.3.2. Segunda forma fundamental . . . . .	34
8.3.3. Símbolos de Christoffel . . . . .	35
8.4. Geodésicas . . . . .	36
8.5. Métrica de Riemann . . . . .	38
8.5.1. Ejemplos de métricas de Riemann . . . . .	39
8.5.2. Propiedades . . . . .	40
8.5.3. Geodésicas . . . . .	40
<b>9. La información de Fisher</b> . . . . .	<b>42</b>
9.1. Conceptos generales . . . . .	42
9.2. La métrica de Fisher . . . . .	43
9.2.1. Propiedades de la matriz de información de Fisher . . . . .	43
9.2.2. La información de Fisher como métrica de Riemann . . . . .	44
9.2.3. Desigualdad de Cramér–Rao . . . . .	45
9.3. Geometría de la información . . . . .	45
9.3.1. Variedades estadísticas . . . . .	45
9.3.2. Geodésicas . . . . .	45
9.3.3. Curvatura . . . . .	46
<b>10. Estudio de casos: aplicación de la métrica de Fisher y geodésicas</b> . . . . .	<b>47</b>
10.1. Distribución normal univariada . . . . .	47
10.1.1. Cálculo de la información de Fisher . . . . .	47
10.1.2. Métrica de Fisher–Rao . . . . .	48
10.1.3. Transformación al modelo del semiplano hiperbólico . . . . .	49
10.1.4. Cálculo de la curvatura gaussiana . . . . .	49
10.1.5. Geodésicas en el espacio normal . . . . .	50
10.1.6. Interpretación geométrica . . . . .	50
10.2. Distribución normal multivariada . . . . .	51
10.2.1. Cálculo de la información de Fisher . . . . .	51
10.2.2. Métrica de Fisher–Rao y estructura geométrica . . . . .	52
10.2.3. Geodésicas . . . . .	52
10.2.4. Interpretación geométrica . . . . .	53
10.3. Distribución beta . . . . .	54
10.3.1. Cálculo de la información de Fisher . . . . .	54
10.3.2. Métrica de Fisher–Rao . . . . .	55
10.3.3. Propiedades geométricas . . . . .	55
10.3.4. Geodésicas . . . . .	56
10.3.5. Interpretación geométrica . . . . .	56
10.4. Distribución gamma . . . . .	57

10.4.1. Cálculo de la información de Fisher . . . . .	58
10.4.2. Métrica de Fisher–Rao . . . . .	59
10.4.3. Propiedades geométricas . . . . .	59
10.4.4. Geodésicas . . . . .	60
10.4.5. Interpretación geométrica . . . . .	60
10.4.6. Esquemas de las geodésicas . . . . .	61
<b>11. Conclusiones</b>	<b>64</b>
<b>12. Recomendaciones</b>	<b>65</b>
12.1. Aplicaciones en imágenes médicas . . . . .	65
12.2. Agrupamiento estadístico mediante información de Fisher . . . . .	65
<b>Bibliografía</b>	<b>68</b>
<b>Lista de Símbolos</b>	<b>69</b>

7.1. Distribución binomial con $n = 15$ y $p = \frac{1}{13}$ . . . . .	13
7.2. Ejemplos de funciones de distribución acumulada . . . . .	14
7.3. Función de densidad de una distribución normal . . . . .	14
7.4. Área bajo una función de densidad $f$ en un intervalo $[a, b]$ . . . . .	15
8.1. Ejemplo de una carta . . . . .	27
8.2. Un mapeo de transición [19] . . . . .	28
8.3. Función de Weierstrass con $a = 0.5$ , $b = 5$ y $0 \leq n \leq 50$ . . . . .	30
8.4. Tira de Möbius con $R = 1$ . . . . .	31
8.5. Mapeo diferenciable entre variedades [7] . . . . .	32
8.6. Ejemplificación del vector tangente en una variedad diferenciable [7] . . . . .	32
8.7. Ejemplo de una superficie regular [6] . . . . .	33
8.8. Curvatura normal de una superficie [6] . . . . .	35
8.9. La derivada covariante [6] . . . . .	37
8.10. Geodésicas sobre un cilindro [6] . . . . .	38
10.1. Comportamiento de las geodésicas de las distribuciones normal, gamma y beta . . . . .	62
10.2. Comportamiento de las geodésicas de la distribución normal multivariada . . . . .	62

---

Lista de cuadros

---

7.1. Muestra de varias alturas	17
--------------------------------	----

Este trabajo presenta un estudio sobre la geometría de la información, un área que vincula la estadística paramétrica con la geometría diferencial. A partir de fundamentos de probabilidad se establece un marco que interpreta familias paramétricas de distribuciones como variedades diferenciables equipadas con la métrica de Fisher–Rao.

La tesis se organiza en tres bloques: (1) conceptos de probabilidad y estadística, donde se introduce la información de Fisher; (2) elementos de la geometría diferencial, centrados en variedades, formas fundamentales y geodésicas; y (3) la síntesis de ambos campos en la geometría de la información, analizando la estructura métrica inducida por la información de Fisher y sus implicaciones.

El trabajo concluye con un análisis de cuatro familias de distribuciones —normal univariada y multivariada, beta y gamma— derivando la métrica de Fisher para cada caso, estudiando propiedades como la curvatura y analizando las geodésicas de sus espacios paramétricos. En estos modelos se observa curvatura negativa con consecuencias estadísticas.

Este estudio reúne teoría y ejemplos para mostrar cómo el enfoque geométrico aporta una perspectiva común a problemas de inferencia, estimación y análisis de datos.

# CAPÍTULO 1

---

## Introducción

---

La geometría de la información surge como intersección entre la estadística matemática y la geometría diferencial, disciplinas vinculadas por sus estructuras subyacentes. Formalizada a finales del siglo XX por C.R. Rao y Shun-ichi Amari, proporciona un enfoque geométrico para abordar problemas de inferencia estadística.

El estudio de familias paramétricas de distribuciones de probabilidad constituye un elemento central en la inferencia; estos modelos se analizan desde perspectivas algebraicas o analíticas, atendiendo a propiedades como consistencia, eficiencia y suficiencia. Al interpretar un modelo estadístico como una variedad diferenciable donde cada punto representa una distribución, surgen estructuras geométricas que ayudan a describir su comportamiento.

La métrica de Fisher–Rao, derivada de la información de Fisher, dota a estos espacios de una estructura Riemanniana: las distancias miden diferencias estadísticas, las geodésicas describen trayectorias entre distribuciones y la curvatura refleja propiedades del modelo. Este enfoque también proporciona herramientas matemáticas para problemas de estimación, inferencia y análisis de datos.

La presente tesis introduce los conceptos básicos de la geometría de la información y aporta una exposición detallada de sus fundamentos. Se incluyen derivaciones completas de los resultados y se analizan casos en los que la literatura suele omitir pasos intermedios.

### Objetivo general

Exponer la teoría estadística de la información de Fisher, incluyendo los conceptos, teoría y principales resultados, y algunas aplicaciones principales. Mostrar mediante aplicaciones y casos concretos, las bondades y usos de la información de Fisher, principalmente para que otros estudiantes en las áreas de ciencia e ingeniería puedan sacar provecho de su aplicación.

### Objetivos específicos

- Presentar los fundamentos teóricos de la geometría de la información y cómo estos fundamentan la información de Fisher.
- Mostrar aplicaciones y cálculos asociados a la información de Fisher mediante casos de uso en distribuciones 2-paramétricas, específicamente la distribución beta y la distribución gamma.
- Explicar el uso y el impacto de la información de Fisher en aplicaciones prácticas dentro de las áreas de ciencia e ingeniería, destacando su utilidad y efectividad.

El desarrollo de esta tesis se motiva por la escasez de herramientas matemáticas adecuadas para comparar espacios de distribuciones de probabilidad. La comparación de modelos probabilísticos es central para el avance del aprendizaje automático y la estadística; disponer de métricas más ajustadas facilita el diseño de métodos de estimación y contraste con mayor solidez, útiles en áreas como ciencia de datos, reconocimiento de imágenes y análisis estadístico. La geometría de la información aporta métricas derivadas de la información de Fisher que son invariantes bajo reparametrizaciones suaves y poseen consistencia estadística, lo cual las hace especialmente atractivas para aplicaciones en análisis de señales, procesamiento de imágenes, optimización bayesiana y aprendizaje automático.

A nivel práctico, estas técnicas ya se utilizan en contextos como la clasificación de histogramas médicos, la geometría computacional y la optimización de modelos estadísticos. Sin embargo, la literatura especializada suele presentar la teoría de manera fragmentaria o abreviada, dificultando su comprensión por parte de estudiantes y profesionales fuera del ámbito matemático. Esta tesis busca subsanar ese vacío al exponer la teoría con detalle y ejemplos concretos, de modo que estudiantes de pregrado y el público interesado puedan comprender los fundamentos de las métricas probabilísticas.

Al documentar paso a paso derivaciones y aplicaciones, se amplía el alcance de estas herramientas y se facilita su adopción en proyectos académicos y profesionales. En suma, la justificación de este trabajo se apoya tanto en la necesidad de contar con métricas robustas para comparar distribuciones como en la oportunidad de difundir una perspectiva geométrica sobre problemas estadísticos, enriqueciendo así la interacción entre estadística, geometría y ciencia de datos.

La información de Fisher se introdujo en estadística para cuantificar cuánta información proporcionan los datos sobre un parámetro. Además de su papel en métodos de estimación, su interpretación geométrica conecta la probabilidad y la inferencia estadística con la geometría diferencial. En la geometría de la información, la información de Fisher se interpreta como una métrica Riemanniana sobre variedades de distribuciones de probabilidad [1, 4, 16].

## 4.1. Probabilidad e inferencia estadística

Los conceptos de espacio de probabilidad y de variable aleatoria constituyen la base de la inferencia estadística [12]. Una variable aleatoria  $X$  definida sobre un espacio de probabilidad  $(\Omega, \mathcal{F}, P)$  se describe mediante una función de densidad  $f(x | \theta)$  en el caso continuo o una función de masa en el caso discreto [32]. La información de Fisher cuantifica la precisión con que puede estimarse el parámetro  $\theta$  a partir de los datos. Bajo condiciones de regularidad, existen dos expresiones equivalentes: una define la información como la varianza del score (derivada de la log-verosimilitud), y la otra como el valor esperado de la segunda derivada de la log-verosimilitud [22, 13]. Para una variable aleatoria continua con densidad  $f(X | \theta)$ , ambas expresiones se escriben como

$$I(\theta) = \mathbb{E} \left[ \left( \frac{\partial}{\partial \theta} \log f(X | \theta) \right)^2 \right] = -\mathbb{E} \left[ \frac{\partial^2}{\partial \theta^2} \log f(X | \theta) \right]. \quad (4.1)$$

Esta cantidad se convierte en una métrica natural sobre el espacio de parámetros, lo que da origen a la geometría de la información [1, 4, 26].

## 4.2. Geometría de la información

La geometría de la información interpreta conjuntos de distribuciones de probabilidad como variedades diferenciables. En este marco, la información de Fisher define una métrica de Riemann que permite calcular distancias geodésicas, curvaturas y volúmenes [1, 10, 19]. Una métrica Riemanniana asigna a cada punto de la variedad un producto interno definido positivo en su espacio tangente [6], induciendo una geometría intrínseca sobre el conjunto de distribuciones.

Distribuciones de dos parámetros como las familias beta y gamma ilustran estas ideas. En la distribución beta, el espacio de parámetros presenta curvatura negativa constante, comparable a un plano hiperbólico [14], mientras que en la gamma la curvatura varía con el parámetro [28]. Estas características han motivado aplicaciones en clasificación de histogramas médicos, geometría computacional y optimización de modelos estadísticos [16, 28].

### 4.3. Aplicaciones modernas

Gracias al desarrollo de la geometría de la información, es posible emplear métricas estadísticas para comparar distribuciones en contextos como análisis de señales, procesamiento de imágenes, optimización bayesiana y aprendizaje automático [25, 22, 15]. Las métricas derivadas de la información de Fisher son invariantes bajo reparametrizaciones suaves y presentan consistencia estadística [1]. La información de Fisher une la estadística, la geometría y la ciencia de datos; su interpretación como métrica en el espacio de modelos estadísticos la convierte en una herramienta para la inferencia, visualización y comparación de distribuciones [4, 1, 32].

En los siguientes capítulos se desarrollan con detalle las propiedades geométricas de esta métrica y su aplicación a las distribuciones normal, beta y gamma, con énfasis en cálculos analíticos y su interpretación geométrica.

La información de Fisher fue introducida por Ronald Fisher en 1922 en su artículo de las *Philosophical Transactions of the Royal Society of London*; en aquella época se planteó como solución a problemas de estimación y no recibió todavía el nombre actual de “información de Fisher” [32]. Antes de ese trabajo, el método de máxima verosimilitud —desarrollado por Fisher entre 1912 y 1922— era la principal herramienta de inferencia estadística [31]. A partir de estas publicaciones, el concepto se difundió gradualmente y sirvió de base para el desarrollo de nuevas teorías y ramas matemáticas.

Tras la Segunda Guerra Mundial, en la década de 1940, surgió la *teoría de información* para abordar problemas de transmisión eficiente de datos; este campo se vio impulsado por avances de Ralph Hartley, Alan Turing y Claude E. Shannon, considerado “el padre de la teoría de la información”. La información de Fisher, al proporcionar una medida cuantitativa sobre la precisión de los estimadores, también se consideró relevante para el perfeccionamiento del método de máxima verosimilitud [32]. Uno de los resultados más conocidos de este desarrollo es la *cota de Cramér–Rao*, que establece un límite inferior para la varianza del error cuadrático medio de estimadores insesgados [14].

C. R. Rao fue uno de los primeros matemáticos en emplear la información de Fisher como métrica Riemanniana, lo que dio origen a la rama de la *geometría de la información*. Esta disciplina conecta la teoría estadística con la teoría geométrica mediante la métrica de Fisher y ha permitido la aplicación de estos conceptos en ámbitos más allá de la estadística, como la genética y la programación.

En años recientes, investigadores como Alice Le Brigant, Mohamed Sahbani y otros han aplicado la información de Fisher a tareas de análisis de imágenes, estudio de sistemas dinámicos y clasificación de histogramas médicos [14].

## 6.1. Delimitación del estudio

La presente investigación se sitúa dentro de la geometría de la información y se concentra en la estructura geométrica de familias paramétricas de distribuciones de probabilidad. El alcance se delimita considerando los siguientes aspectos:

- **Modelos paramétricos con condiciones de regularidad:** se analizan familias paramétricas que satisfacen condiciones de regularidad suficientes para admitir una estructura diferenciable y que poseen una métrica de Fisher bien definida en todo el espacio de parámetros.
- **Métrica de Fisher–Rao:** aunque existen otras métricas en espacios de distribuciones (como las divergencias alfa, beta o gamma de Amari), este trabajo se centra en la métrica de Fisher–Rao derivada de la información de Fisher.
- **Tratamiento teórico:** el enfoque es básicamente teórico, desarrollando la estructura formal que sustenta la interpretación geométrica de modelos estadísticos y priorizando derivaciones completas frente a implementaciones computacionales.
- **Familias específicas:** el análisis detallado se limita a cuatro familias de distribuciones —normal univariada, normal multivariada, beta y gamma— elegidas por su importancia teórica y práctica y porque representan distintos comportamientos geométricos.

## 6.2. Limitaciones

Las principales limitaciones de este estudio son:

- No se abordan algoritmos ni aspectos computacionales para el cálculo eficiente de geodésicas u otras magnitudes en espacios de alta dimensión.
- No se incluyen demostraciones completas de todos los resultados generales de geometría diferencial, para los cuales se remite a la literatura especializada.

- No se exploran extensiones recientes como la geometría de la información cuántica, las divergencias  $\alpha$ -conexiones generalizadas ni sus aplicaciones al aprendizaje profundo.
- No se desarrollan aplicaciones a problemas de inferencia estadística aplicada o análisis de datos reales, por lo que el enfoque se mantiene en el ámbito teórico.

### 6.3. Contribuciones principales

Las contribuciones de esta investigación son las siguientes:

- Se presenta un desarrollo integrado de conceptos de probabilidad, estadística y geometría diferencial, para construir un puente accesible entre estas disciplinas.
- Se proporcionan derivaciones detalladas de las estructuras geométricas (métricas, curvaturas, geodésicas) de las distribuciones estudiadas, incluyendo pasos intermedios que a menudo se omiten en la literatura.
- Se realiza un análisis comparativo de las propiedades geométricas de diferentes familias paramétricas y se discute la interpretación estadística de fenómenos como la curvatura negativa.
- Se examina la relación entre distintas parametrizaciones (natural, de esperanza, canónica) y sus implicaciones en la estructura geométrica resultante.

Este trabajo pretende ser una referencia didáctica para quienes se inician en el campo y, al mismo tiempo, una contribución que aporte claridad conceptual sobre los fundamentos de la geometría de la información.

## 7.1. Espacios de probabilidad

**Definición 7.1.1.** Un espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$  es un espacio medible compuesto por:

- **Espacio muestral**  $\Omega$ : el conjunto de todos los resultados posibles.
- **$\sigma$ -álgebra**  $\mathcal{F}$ : una colección de subconjuntos de  $\Omega$  que contiene el conjunto vacío, es cerrada bajo complementos y uniones numerables; es decir,
  - $\Omega \in \mathcal{F}$ .
  - Si  $A \in \mathcal{F}$ , entonces  $A^c \in \mathcal{F}$ .
  - Si  $A_1, A_2, \dots$  son subconjuntos de  $\Omega$  en  $\mathcal{F}$ , entonces  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ .
- **Medida de probabilidad**  $\mathbb{P}$ : una función  $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$  que asigna a cada evento  $A \in \mathcal{F}$  un número real y cumple:
  - *No negatividad*:  $\mathbb{P}(A) \geq 0$  para todo  $A \in \mathcal{F}$ .
  - *Aditividad numerable*: si  $\{A_i\}$  son eventos disjuntos, entonces

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

- *Normalización*:  $\mathbb{P}(\Omega) = 1$ .

Como se menciona en [21], la elección del espacio de probabilidad depende de los objetivos del modelo, comenzando por la selección de  $\Omega$ . Por ejemplo, al lanzar un dado una vez, el espacio muestral se modela usualmente por seis elementos  $\{1, 2, 3, 4, 5, 6\}$ , ignorando factores como la cara inicial o la posibilidad de que el dado quede sobre un borde para simplificar el análisis.

En general se distinguen dos tipos de espacios muestrales: *discretos*, cuando el conjunto de resultados posibles es finito o numerable, y *continuos*, cuando admite una cantidad no numerable de

resultados. Cabe señalar que estos modelos son ideales, ya que cualquier aparato de medición tiene precisión limitada, pero los resultados teóricos suelen aplicarse en la práctica [12].

Usualmente, cuando se trabaja en un marco teórico, se considera que el espacio muestral  $\Omega$  y la medida  $\mathbb{P}$  son fijos, y sólo debe determinarse la probabilidad de cada evento. Si  $A \subset \Omega$  es un evento, la probabilidad de  $A$  se define como la suma de las probabilidades elementales  $p_i$  asociadas a los puntos  $i \in A$ ; esto es,

$$\mathbb{P}(A) = \sum_{i \in A} p_i. \quad (7.1)$$

Para  $A \subset \Omega$ , el *complemento*  $A^c$  es el conjunto de los resultados que no pertenecen a  $A$ . Dado que la suma de probabilidades sobre  $\Omega$  es uno y los elementos de  $A$  y  $A^c$  son disjuntos, se tiene

$$\sum_{i \in A} p_i + \sum_{i \in A^c} p_i = \sum_{i \in \Omega} p_i = \mathbb{P}(\Omega) = 1,$$

de modo que

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c). \quad (7.2)$$

**Ejemplo 7.1.1.** Considérese una baraja estándar de 52 cartas de póquer y seleccionemos una carta al azar. El espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$  se define de la siguiente manera:

1. *Espacio muestral*  $\Omega$ : conjunto de todas las cartas. En este caso,

$$\Omega = \{A\heartsuit, A\diamondsuit, A\clubsuit, A\spadesuit, \dots, K\heartsuit, K\diamondsuit, K\clubsuit, K\spadesuit\},$$

donde  $|\Omega| = 52$ .

2.  $\sigma$ -álgebra  $\mathcal{F}$ : al ser  $\Omega$  finito,  $\mathcal{F}$  coincide con el conjunto de todas las partes de  $\Omega$ , es decir,  $\mathcal{F} = \mathcal{P}(\Omega)$ .
3. *Medida de probabilidad*  $\mathbb{P}$ : asigna a cada evento  $A \subseteq \Omega$  la proporción  $\mathbb{P}(A) = |A|/52$ . Por ejemplo, la probabilidad de extraer una carta concreta es  $1/52$ ; la probabilidad de obtener un cuatro es  $4/52$ ; y la probabilidad de sacar un corazón es  $13/52 = 1/4$ .

**Ejemplo 7.1.2.** Supongamos ahora que en la baraja anterior hay 4 cartas ligeramente más pequeñas, lo que modifica la probabilidad de ser seleccionadas. El espacio muestral  $\Omega$  es el mismo, pero la medida ya no es uniforme. Se asignan probabilidades  $p$  a cada una de las 48 cartas de tamaño normal y  $q$  a cada una de las 4 cartas más pequeñas, donde  $q < p$ . Como  $\mathbb{P}(\Omega) = 1$ , se cumple

$$48p + 4q = 1, \quad q = \frac{1}{4} - 12p, \quad 0 \leq p \leq \frac{1}{48}.$$

Cuando las cartas son casi del mismo tamaño se espera que  $p$  y  $q$  estén próximos a  $1/52$ . Podemos expresar

$$p = \frac{1}{52} + \theta, \quad q = \frac{1}{52} - 12\theta,$$

donde  $-\frac{1}{52} \leq \theta \leq \frac{1}{624}$ . En este modelo,

$$p_1 = \dots = p_{48} = \frac{1}{52} + \theta, \quad p_{49} = \dots = p_{52} = \frac{1}{4} - 12\theta,$$

y  $\theta$  resume la diferencia en tamaño. Cuando  $\theta = 0$  se recupera la baraja estándar. Determinar  $\theta$  a partir de mediciones físicas suele ser difícil, por lo que se trata como un parámetro desconocido a estimar mediante repeticiones del experimento [12].

En ocasiones no es posible repetir indefinidamente el experimento para estimar el parámetro. Por ejemplo, si deseamos determinar la proporción de fósforos defectuosos en una fábrica encendiéndolos

uno a uno, el número de intentos necesarios antes de observar el primer defecto es aleatorio. En un modelo idealizado se asume que podemos encender fósforos hasta la aparición del primero defectuoso. El espacio muestral en este caso es el conjunto de enteros no negativos,

$$\Omega = \{0, 1, 2, \dots\},$$

donde cada entero representa el número de fósforos probados antes de encontrar el primero defectuoso.

### 7.1.1. Espacios de probabilidad completos

Un espacio de probabilidad se llama *completo* si cualquier subconjunto de un evento de probabilidad cero es medible y, por ende, también tiene probabilidad cero. Formalmente, si  $A \in \mathcal{F}$  y  $\mathbb{P}(A) = 0$ , entonces para todo  $B \subseteq A$  se tiene  $B \in \mathcal{F}$  y  $\mathbb{P}(B) = 0$ . Todo espacio de probabilidad puede completarse añadiendo a  $\mathcal{F}$  todos los subconjuntos de los eventos nulos.

**Ejemplo 7.1.3.** La medida de Lebesgue sobre el intervalo  $[0, 1]$  es completa, ya que cualquier subconjunto de un conjunto de medida cero es medible.

### 7.1.2. Propiedades de los espacios de probabilidad

**Definición 7.1.2.** Para una sucesión de eventos  $\{A_n\}$  en un espacio de probabilidad definimos los límites inferior y superior como

$$\liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m, \quad \limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m.$$

#### Propiedades de la medida de probabilidad

Enumeramos a continuación algunas propiedades básicas de la medida de probabilidad  $\mathbb{P}$  sobre  $(\Omega, \mathcal{F}, \mathbb{P})$ , que se utilizarán más adelante:

#### Propiedades básicas:

- *Probabilidad del conjunto vacío:*  $\mathbb{P}(\emptyset) = 0$ .
- *Probabilidad del complemento:*  $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ .
- *Monotonía:* si  $A \subseteq B$ , entonces  $\mathbb{P}(A) \leq \mathbb{P}(B)$ .

#### Aditividad y relaciones entre eventos:

- *Unión de eventos disjuntos:* si  $A \cap B = \emptyset$ , entonces  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ .
- *Unión de eventos no disjuntos:*  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ .
- *Subaditividad:*  $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$ .
- *Aditividad numerable:* si  $\{A_i\}_{i=1}^{\infty}$  es una sucesión de eventos disjuntos, entonces  $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ .

**Propiedades de continuidad y límites:**

- *Continuidad desde abajo:* si  $A_1 \subseteq A_2 \subseteq \dots$ , entonces  $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n)$ .
- *Continuidad desde arriba:* si  $A_1 \supseteq A_2 \supseteq \dots$  y  $\bigcap_i A_i = A$ , entonces  $\mathbb{P}(\bigcap_{i=1}^{\infty} A_i) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n)$ .
- *Relaciones entre  $\liminf$  y  $\limsup$ :*  $\mathbb{P}(\liminf A_n) \leq \liminf \mathbb{P}(A_n) \leq \limsup \mathbb{P}(A_n) \leq \mathbb{P}(\limsup A_n)$ .
- *Convergencia de probabilidades:* si  $\lim A_n$  existe, entonces  $\mathbb{P}(\lim A_n) = \lim \mathbb{P}(A_n)$ .

**Espacio muestral producto:** si se realizan dos experimentos con espacios muestrales  $\Omega_1$  y  $\Omega_2$ , el espacio muestral conjunto se define como el producto cartesiano

$$\Omega = \Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}.$$

## 7.2. Variables aleatorias

**Definición 7.2.1.** Una *variable aleatoria*  $X$  es una **función medible** definida sobre el espacio muestral  $\Omega$ ; a cada punto  $\omega \in \Omega$  le asigna un valor  $X(\omega) \in \mathbb{R}$ . Las variables aleatorias se clasifican en dos grupos principales [12]:

- **Discretas:** variables que sólo pueden tomar una cantidad finita o numerable de valores [12].
- **Continuas:** variables que pueden tomar cualquier valor dentro de un intervalo [12].

El concepto de variable aleatoria es más general que el de evento; por ejemplo, a cada evento podemos asociarle una variable indicadora que toma el valor 1 si el evento ocurre y 0 en caso contrario [21].

**Ejemplo 7.2.1.** Retomando el ejemplo del mazo de cartas, una **variable aleatoria discreta** podría ser la carta obtenida al seleccionar una al azar. Esta variable sólo asume uno de los 52 valores correspondientes a cada carta; no existe un valor intermedio como “4.5 de tréboles”. Por otro lado, una **variable aleatoria continua** podría ser, por ejemplo, una medida física asociada a una carta (como el área a la que puede reducirse); dicha variable podría tomar cualquier valor dentro de un intervalo continuo, por ejemplo entre 0 y  $9 \times 13$  cm.

### 7.2.1. Función de masa de probabilidad

Para variables discretas, la probabilidad de los distintos valores se describe mediante una *función de masa de probabilidad* (f.m.p.) [8, 12]. Formalmente:

**Definición 7.2.2.** Sea  $X$  una variable aleatoria definida sobre un espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$  con soporte finito o numerable  $S_X \subset \mathbb{R}$ . La **función de masa de probabilidad** de  $X$  es la función

$$f(x) = \mathbb{P}(X = x), \quad x \in S_X.$$

Debe satisfacer las siguientes propiedades [8, 12]:

- **No negatividad:**  $f(x) \geq 0$  para todo  $x \in S_X$ .
- **Normalización:**  $\sum_{x \in S_X} f(x) = 1$ .

- **Aditividad:** para  $A \subseteq S_X$ ,  $\mathbb{P}(X \in A) = \sum_{x \in A} f(x)$ .

El resultado de un experimento está modelado de manera que pertenece a  $\Omega$ ; un evento fuera del espacio muestral tiene probabilidad cero. La aditividad se cumple cuando los eventos considerados son disjuntos [12, 8].

La función de probabilidad puede describirse de diversas formas: mediante una tabla (cuando el número de valores es pequeño), mediante una fórmula algebraica o con una representación gráfica.

**Ejemplo 7.2.2.** Consideremos un juego en el que se intenta adivinar el número de la carta extraída de una baraja. Sea  $X$  la cantidad de aciertos en  $n$  intentos independientes. El espacio de valores es  $S_X = \{0, 1, 2, \dots, n\}$  y  $X$  sigue una distribución binomial:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x \in S_X,$$

donde  $p$  es la probabilidad de acierto en un intento [8]. Por ejemplo, si se realizan  $n = 15$  intentos y  $p = 1/13$ , entonces

$$f(x) = \binom{15}{x} \left(\frac{1}{13}\right)^x \left(1 - \frac{1}{13}\right)^{15-x}, \quad x \in S_X,$$

cuya gráfica se muestra en la Figura 7.1

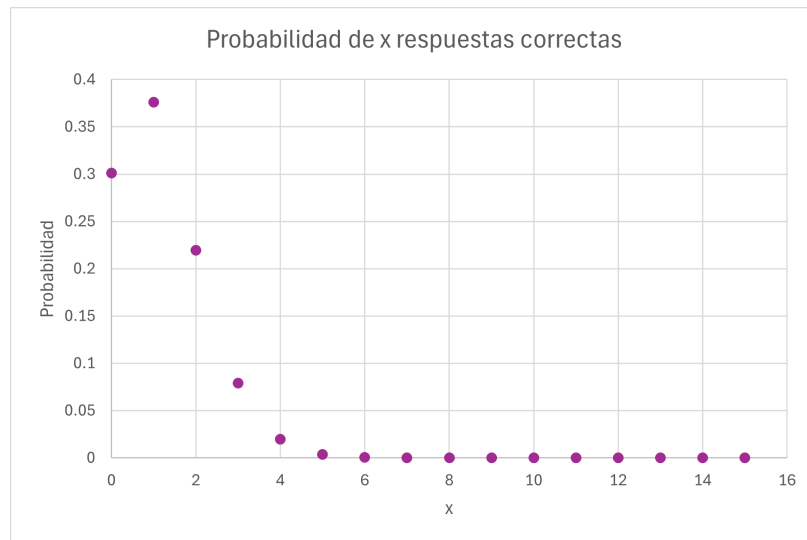


Figura 7.1: Distribución binomial con  $n = 15$  y  $p = \frac{1}{13}$

### 7.2.2. Función de distribución acumulada

**Definición 7.2.3.** Sea  $X$  una variable aleatoria en  $(\Omega, \mathcal{F}, \mathbb{P})$ . La **función de distribución acumulada** (F.D.A.) de  $X$  es la función

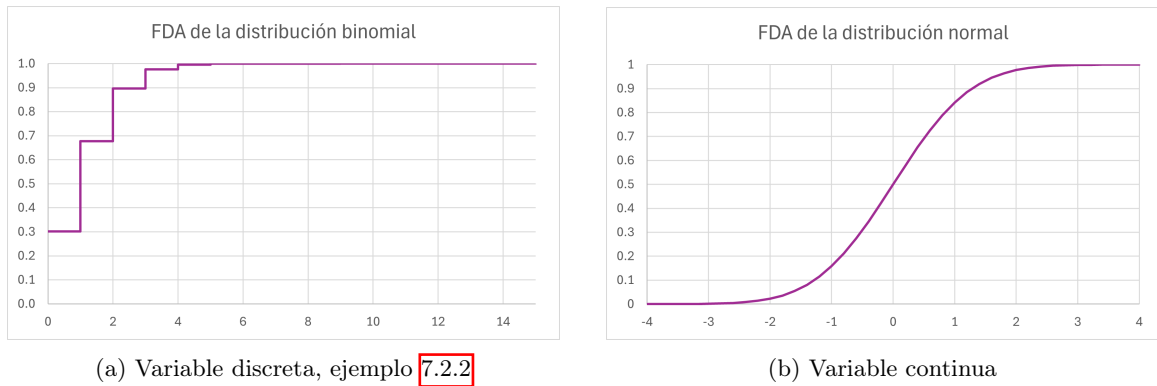
$$F(t) = \mathbb{P}(X \leq t), \quad -\infty < t < \infty,$$

que para variables discretas puede escribirse como

$$F(t) = \sum_{\substack{x \in S_X \\ x \leq t}} f(x).$$

La F.D.A. describe la probabilidad acumulada de que  $X$  tome un valor menor o igual a un umbral. Satisface propiedades básicas: es monótonamente no decreciente; cumple  $F(-\infty) = 0$  y  $F(\infty) = 1$ ; y es continua por la derecha. Por convención se utilizan letras mayúsculas (como  $F$ ) para la F.D.A. y minúsculas (como  $f$ ) para la función de masa o densidad.

En variables discretas la F.D.A. es una función escalonada (Figura 7.2a), mientras que para variables continuas es continua (Figura 7.2b). Puesto que la F.D.A. determina completamente la distribución, otras funciones como la f.m.p. o la función de densidad pueden derivarse a partir de ella [8, 12, 21].



(a) Variable discreta, ejemplo 7.2.2

(b) Variable continua

Figura 7.2: Ejemplos de funciones de distribución acumulada

### 7.2.3. Función de densidad de probabilidad

Para variables continuas se utiliza la **función de densidad de probabilidad** (F.D.P.), que describe cómo se distribuye la probabilidad a lo largo de la recta real [12].

**Definición 7.2.4.** Se dice que una variable aleatoria  $X$  es *continua* si su función de distribución acumulada  $F$  es continua y derivable, o si el conjunto de discontinuidades de  $F$  tiene medida cero [12]. En tal caso, la **función de densidad**  $f$  es la derivada de  $F$ :

$$f(x) = \frac{d}{dx}F(x).$$

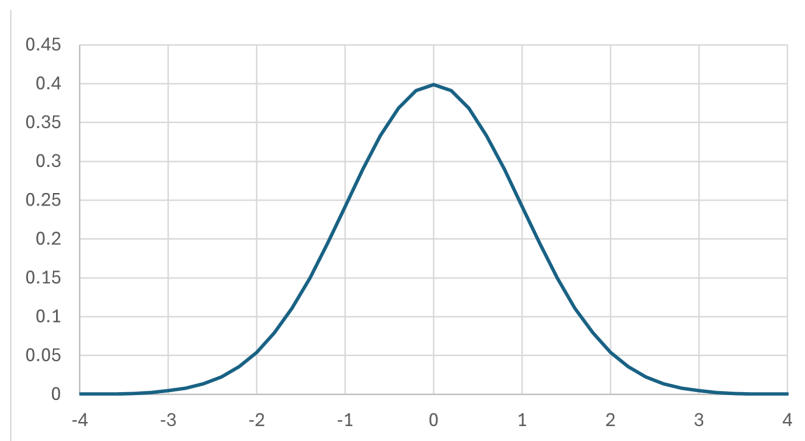


Figura 7.3: Función de densidad de una distribución normal

Dado que  $F$  es no decreciente y  $f(x) \geq 0$ , tomando  $F(-\infty) = 0$  se obtiene

$$F(x) = \int_{-\infty}^x f(t) dt, \tag{7.3}$$

y como  $F(\infty) = 1$ , también

$$\int_{-\infty}^{\infty} f(t) dt = 1. \tag{7.4}$$

La probabilidad de que  $X$  se encuentre en un intervalo  $[a, b]$  es el área bajo la curva de la F.D.P. en ese intervalo:

$$\mathbb{P}(a < X \leq b) = F(b) - F(a) = \int_a^b f(x) dx.$$

Por tanto, la probabilidad de un valor exacto es cero y, si el intervalo se reduce, la probabilidad tiende a cero [8, 12]. Más concretamente,

$$\mathbb{P}(a - \varepsilon \leq X \leq a + \varepsilon) = \int_{a-\varepsilon}^{a+\varepsilon} f(x) dx, \quad \mathbb{P}(X = a) = 0.$$

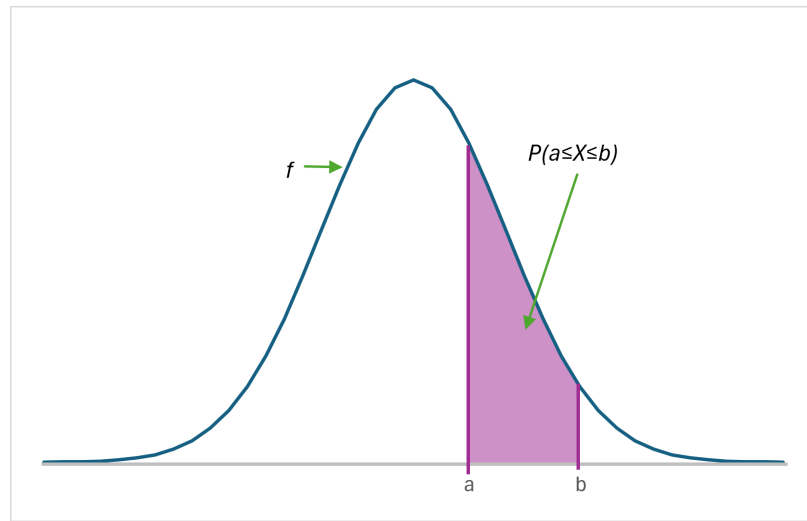


Figura 7.4: Área bajo una función de densidad  $f$  en un intervalo  $[a, b]$

**Ejemplo 7.2.3.** Supóngase que las alturas de un grupo de adultos siguen una distribución normal  $X \sim \mathcal{N}(170, 10^2)$ . La F.D.P. se usa para calcular, por ejemplo:

1. Probabilidad de que la altura esté entre 160 y 180 cm:

$$\mathbb{P}(160 \leq X \leq 180) = \int_{160}^{180} \frac{1}{\sqrt{2\pi \cdot 10^2}} e^{-\frac{(x-170)^2}{2 \cdot 10^2}} dx \approx 0.6827.$$

2. Probabilidad de medir menos de 155 cm:

$$\mathbb{P}(X \leq 155) = \int_{-\infty}^{155} \frac{1}{\sqrt{2\pi \cdot 10^2}} e^{-\frac{(x-170)^2}{2 \cdot 10^2}} dx \approx 0.0668.$$

3. Probabilidad de medir exactamente 170 cm:

$$\mathbb{P}(X = 170) = 0.$$

Esta última probabilidad es nula porque, incluso con una regla cada vez más precisa, siempre puede encontrarse una diferencia infinitesimal respecto a 170, de modo que la probabilidad de un valor exacto es cero.

### 7.2.4. Valor esperado

El **valor esperado** de una variable aleatoria  $X$  representa el promedio teórico que se obtendría al repetir el experimento muchas veces [8]. Para una variable discreta,

$$\mathbb{E}[X] = \sum_{x \in S_X} x \mathbb{P}(X = x),$$

y para una variable continua,

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f(x) dx.$$

En el caso de una variable aleatoria matricial  $\{X_{i,j}\}$ , el valor esperado se toma componente a componente:

$$(\mathbb{E}[X])_{i,j} = \mathbb{E}(X_{i,j}).$$

## 7.3. Estimación de parámetros

En muchas aplicaciones no se conocen la distribución subyacente ni los valores reales de sus parámetros. Por ello se recurre a *estimadores* obtenidos mediante distintos métodos, evaluando después si dichas estimaciones reflejan adecuadamente el comportamiento de los datos [20]. A menudo la estimación se formula como un problema de optimización en el que se minimiza una función de pérdida para obtener la mejor aproximación posible [25].

**Definición 7.3.1.** Un **modelo estadístico** es una familia de distribuciones parametrizada por  $\theta$ ; se describe mediante una función  $f(x | \theta)$  que indica cómo el parámetro  $\theta$  influye en los posibles valores  $x$  de una variable aleatoria  $X$  [23].

Un modelo puede depender de más de una variable aleatoria  $X_i$  o más de un parámetro  $\theta$ . De acuerdo con la naturaleza de  $X$ , la función  $f$  puede interpretarse como función de masa o densidad; se acostumbra escribir  $p_X(x | \theta) = \mathbb{P}(X = x | \theta) = f(x | \theta)$ .

**Ejemplo 7.3.1.** Retomando el ejemplo [7.2.3] supongamos que la distribución de las alturas es normal. El modelo estadístico queda descrito por

$$f(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right),$$

donde  $x$  denota la altura observada,  $\mu$  la media poblacional y  $\sigma^2$  la varianza. En este caso los parámetros  $\mu$  y  $\sigma^2$  son desconocidos y deberán estimarse para describir adecuadamente la variable  $X$ .

En general, dado un modelo con un parámetro desconocido  $\theta$  y una muestra de  $n$  observaciones independientes  $\mathbf{x} = (x_1, \dots, x_n)$ , se busca un *estimador* que proporcione el valor más plausible de  $\theta$ . Se suele suponer que las observaciones  $x_i$  son independientes e idénticamente distribuidas (i.i.d.) según la distribución de  $X$ ; bajo esta hipótesis, la función de verosimilitud conjunta se obtiene como el producto de las funciones de densidad o masa de las observaciones.

### 7.3.1. Máxima verosimilitud

La estimación por **máxima verosimilitud** (MLE) es el método más común para estimar parámetros. Consiste en seleccionar el valor de  $\theta$  que asigna la mayor probabilidad a los datos obtenidos.

Supongamos que, tras realizar un experimento, podemos agrupar los resultados en el evento  $A$  dentro del espacio muestral  $\Omega$  de nuestro modelo estadístico. Denotemos por

$$A = \{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\},$$

De este modo, la función de verosimilitud puede escribirse como

$$L(\theta | x_1, \dots, x_n) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | \theta) = P(A | \theta).$$

La MLE intenta maximizar  $P(A | \theta)$  [13].

**Definición 7.3.2.** La función de **verosimilitud** de  $\theta$  se define de la siguiente manera:

$$L(\theta) = cP(A | \theta), \tag{7.5}$$

donde  $c$  es una constante positiva cualquiera [13]. Si nuestra distribución es discreta, podemos escribir la función de verosimilitud así:

$$L(\theta) = p_\theta(x_1) p_\theta(x_2) p_\theta(x_3) \cdots p_\theta(x_n), \tag{7.6}$$

donde  $p_\theta$  denota la función de masa de probabilidad (f.m.p.). Por el contrario, si nuestra distribución es continua, con función de densidad  $f_\theta$ , entonces

$$L(\theta) = f_\theta(x_1) f_\theta(x_2) f_\theta(x_3) \cdots f_\theta(x_n). \tag{7.7}$$

**Definición 7.3.3.** La **log-verosimilitud** es la función obtenida al aplicar logaritmo natural a la función de verosimilitud [13]:

$$\ell(\theta) = \log L(\theta) = c' + \log P(A | \theta). \tag{7.8}$$

**Definición 7.3.4.** El **estimador de máxima verosimilitud**  $\hat{\theta}$  es aquella cantidad aleatoria que maximiza  $L(\theta)$  (o, de forma equivalente,  $\ell(\theta)$ ) [32, 13, 8]:

$$\hat{\theta} = \arg \max_{\theta \in \Omega} L(\theta) = \arg \max_{\theta \in \Omega} \ell(\theta). \tag{7.9}$$

En otras palabras, el MLE es el valor de  $\theta$  que maximiza la función de verosimilitud (7.5) (o, de manera equivalente, la log-verosimilitud (7.8)). Las fórmulas (7.6) y (7.7) corresponden a los casos discreto y continuo, respectivamente.

**Ejemplo 7.3.2.** Retomando el ejemplo [7.2.3], supongamos que tenemos la siguiente muestra: Supon-

Alturas (cm)	170	172	168	174	169
--------------	-----	-----	-----	-----	-----

Tabla 7.1: Muestra de varias alturas

gamos que nuestra distribución es normal. Queremos entonces estimar los parámetros desconocidos  $\mu$  y  $\sigma^2$ . Tomemos la F.D.P. de la distribución normal:

$$f(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right).$$

Usando la ecuación (7.7) para la verosimilitud, tenemos

$$L(\mu, \sigma^2) = \prod_{i=1}^n f(x_i | \mu, \sigma^2).$$

En este caso  $n = 5$ . Si expandimos  $L(\mu, \sigma^2)$ , obtenemos:

$$L(\mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}\right).$$

si vemos, la ecuación anterior integrada con nuestro modelo, es sumamente complicada, consiguiendo haremos uso de la log-verosimilitud para simplificar el cálculo, de modo que obtenemos:

$$\begin{aligned} \ell(\mu, \sigma^2) &= \log L(\mu, \sigma^2) = \log\left[(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right)\right] \\ &= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \\ &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2. \end{aligned}$$

En particular, para  $n = 5$ :

$$\ell(\mu, \sigma^2) = -\frac{5}{2} \log(2\pi) - \frac{5}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^5 (x_i - \mu)^2.$$

Para encontrar nuestro MLE debemos maximizar la ecuación anterior. En este caso, podemos derivar dicha ecuación y encontrar los intersejos para encontrar así los valores máximos y mínimos posibles, de modo que, obteniendo las derivadas respecto a  $\mu$  y a  $\sigma^2$ , y resolviendo para 0, tenemos que:

$$\begin{aligned} \frac{\partial \ell}{\partial \mu} &= \frac{\partial}{\partial \mu} \left[ -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right] \\ &= -\frac{1}{2\sigma^2} \frac{\partial}{\partial \mu} \left[ \sum_{i=1}^n (x_i - \mu)^2 \right] = -\frac{1}{2\sigma^2} \sum_{i=1}^n 2(x_i - \mu)(-1) \\ &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu). \end{aligned}$$

Igualando a cero:

$$\frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \implies \sum_{i=1}^n x_i - n\mu = 0 \implies \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Para nuestra muestra  $\{170, 172, 168, 174, 169\}$ ,

$$\sum_{i=1}^5 x_i = 170 + 172 + 168 + 174 + 169 = 853, \quad n = 5,$$

por lo que

$$\hat{\mu} = \frac{853}{5} = 170.6.$$

Ahora

$$\begin{aligned}\frac{\partial \ell}{\partial \sigma^2} &= \frac{\partial}{\partial \sigma^2} \left[ -\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right] \\ &= -\frac{n}{2} \frac{1}{\sigma^2} + \left(-\frac{1}{2}\right) (-\sigma^{-4}) \sum_{i=1}^n (x_i - \mu)^2 \\ &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (x_i - \mu)^2.\end{aligned}$$

Igualando a cero:

$$-\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (x_i - \mu)^2 = 0 \implies -n\sigma^2 + \sum_{i=1}^n (x_i - \mu)^2 = 0 \implies \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2.$$

Entonces:

$$\begin{aligned}(170 - 170.6)^2 &= (-0.6)^2 = 0.36, \\ (172 - 170.6)^2 &= (1.4)^2 = 1.96, \\ (168 - 170.6)^2 &= (-2.6)^2 = 6.76, \\ (174 - 170.6)^2 &= (3.4)^2 = 11.56, \\ (169 - 170.6)^2 &= (-1.6)^2 = 2.56.\end{aligned}$$

Sumando:

$$\sum_{i=1}^5 (x_i - 170.6)^2 = 0.36 + 1.96 + 6.76 + 11.56 + 2.56 = 23.20.$$

Por lo tanto,

$$\hat{\sigma}^2 = \frac{23.20}{5} = 4.64.$$

Finalmente, usando estos resultados, obtenemos:

$$\hat{\mu} = 170.6, \quad \hat{\sigma}^2 = 4.64.$$

Veamos que este caso es relativamente sencillo, con una respuesta directa. Sin embargo, hay que tomar en consideración que no siempre es posible resolver para un valor en específico. Dependiendo de la complejidad de nuestro modelo, es posible tener múltiples soluciones ante la derivada obtenida. Es necesario entonces comprobar si el resultado que se obtiene es solamente un máximo local o el verdadero MLE. Los métodos para lograr esto se explicarán en los capítulos más adelante.

**Lemma 7.3.1. Principio de la no varianza:** Si  $\hat{\theta}$  es el MLE de un parámetro  $\theta$  y  $g(\theta)$  es una función invertible de  $\theta$ , entonces  $g(\hat{\theta})$  es el MLE para  $g(\theta)$  [8, 32].

### 7.3.2. Función de score y de información

#### Función de score

Veamos que para evaluar  $\hat{\theta}$  debemos encontrar el valor máximo de  $l(\theta)$  sobre todos los posibles valores de  $\theta$ , como se mencionó anteriormente. Sin embargo, es posible que al resolver las diferentes ecuaciones, obtengamos una multiplicidad de valores, un mínimo, o un punto de inflexión (también llamado punto silla). Consiguientemente, es importante verificar que el valor encontrado es el máximo valor posible. Una de las maneras de hacer esto es utilizando la segunda derivada. [13]

**Definición 7.3.5.** La **función de score** se define como la derivada de la log-verosimilitud respecto a  $\theta$  [13]

$$S(\theta) = l'(\theta) = \frac{dl(\theta)}{d\theta}. \quad (7.10)$$

En otras palabras, la función de score nos indica la pendiente de la función de log-verosimilitud, de modo que donde la función de score sea  $S(\theta) = 0$  significa que nos encontramos ante un máximo de la log-verosimilitud, de modo que  $\theta$  es un posible candidato para ser el MLE. Si nos referimos al ejemplo 7.3.2, vemos que aplicamos la función de score al derivar nuestra ecuación para maximizarla, y encontrar los posibles candidatos para el MLE. Sin embargo, veamos que aún no se resuelve el problema planteado con anterioridad, por lo que introduciremos un nuevo concepto, la *información de Fisher*.

### Información de Fisher

Continuando con lo mencionado anteriormente, la información de Fisher, o la función de información, hace uso de la segunda derivada para obtener la optimización de nuestro MLE. Veamos la definición de una manera más formal.

**Definición 7.3.6.** La **función de información** se puede definir como la segunda derivada negativa de la log-verosimilitud, respecto a un parámetro  $\theta$

$$\mathcal{I}(\theta) = -l''(\theta) = -S'(\theta) = -\frac{d^2l(\theta)}{d\theta^2}.$$

De modo que podemos verificar si el valor obtenido al utilizar la función de score para maximizar nuestro MLE, es de hecho, un máximo. Sin embargo, esta no es la única definición que se puede dar de la información de Fisher. Veamos que la información de Fisher se utiliza también para cuantificar la cantidad de información contiene nuestro parámetro  $\theta$  dentro de nuestro modelo. En este sentido, la función de score nos describe qué tan sensible es el modelo a cambios  $\theta$ , en un  $\theta$  particular. Mientras que la información de Fisher mide la sensibilidad de la relación entre  $f$  y  $\theta$  al evaluar la sensibilidad en cada posible resultado de  $x$ , respecto a la probabilidad definida [23]. Para esto, definimos la información de Fisher de la siguiente manera:

**Definición 7.3.7.** La **información de Fisher**  $I_X(\theta)$  de una variable aleatoria  $X$  sobre  $\theta$  se define como

$$I_X(\theta) = \begin{cases} \sum_{x \in \Omega} \left( \frac{d}{d\theta} \log f(x | \theta) \right)^2 p_\theta(x) & \text{si } X \text{ es discreta,} \\ \int_{\Omega} \left( \frac{d}{d\theta} \log f(x | \theta) \right)^2 f_\theta(x) dx & \text{si } X \text{ es continua.} \end{cases} \quad (7.11)$$

Cabe mencionar que tanto la información de Fisher, como la función de Fisher suelen ser escritas con  $\frac{d}{d\theta}$  cuando  $\theta$  es una sola variable y  $\frac{\partial}{\partial \theta}$ , en el caso en que  $\theta$  es un valor de parámetros. A continuación utilizaremos esta segunda, para realizar pruebas generalizadas. Naturalmente es confuso el evaluar ambas definiciones, ya que la equivalencia es un poco contraintuitiva. Antes de demostrar esta equivalencia, entendamos la diferencia de uso. La primera definición de la función de información mide la curvatura de la función de log-verosimilitud, mientras que la segunda definición sirve para medir la variabilidad de la función de score. Veamos entonces la demostración de la equivalencia entre ambas. Para esto, debemos establecer algunas condiciones bajo las cuales se cumple la igualdad.

**Lemma 7.3.2.** La función de información y la información de Fisher son matemáticamente equivalentes si se cumplen las siguientes condiciones [23]:

1.  $f(x|\theta)$  es dos veces diferenciable respecto a  $\theta$ , y podemos libremente intercambiar la derivada con la integral.
2. El modelo es *suficiente* respecto al parámetro  $\theta$ , es decir la función  $f(x|\theta)$  no depende de  $\theta$  en conjuntos con medida 0.
3. El valor esperado existe y es finito,

$$\mathbb{E} \left[ \left| \frac{\partial^2}{\partial \theta^2} \log f(X|\theta) \right| \right] \leq \infty.$$

Con las condiciones de regularidad establecidas, procedamos a la demostración:

*Demostración.* Sea  $X$  una variable aleatoria con función de densidad (o masa)  $f(x|\theta)$ , donde  $\theta$  es un parámetro real. Bajo las condiciones de regularidad (en particular, que podemos intercambiar la derivada y la integral), definimos la función *score* como

$$\ell'(\theta) = \frac{\partial}{\partial \theta} \log f(X|\theta) = \frac{1}{f(X|\theta)} \frac{\partial}{\partial \theta} f(X|\theta).$$

Primeramente, comprobaremos que  $\mathbb{E}[\ell'(\theta)] = 0$ . En efecto,

$$\mathbb{E}[\ell'(\theta)] = \mathbb{E} \left[ \frac{\partial}{\partial \theta} \log f(X|\theta) \right] = \int_{\mathcal{X}} f(x|\theta) \frac{1}{f(x|\theta)} \frac{\partial}{\partial \theta} f(x|\theta) dx = \int_{\mathcal{X}} \frac{\partial}{\partial \theta} f(x|\theta) dx.$$

Ahora bien, por las condiciones de regularidad podemos intercambiar la derivada con la integral:

$$\int_{\mathcal{X}} \frac{\partial}{\partial \theta} f(x|\theta) dx = \frac{\partial}{\partial \theta} \underbrace{\int_{\mathcal{X}} f(x|\theta) dx}_{=1} = \frac{\partial}{\partial \theta} (1) = 0.$$

De este modo queda demostrado que

$$\mathbb{E}[\ell'(\theta)] = 0.$$

A continuación, consideremos la segunda derivada de la log-verosimilitud,

$$\ell''(\theta) = \frac{\partial^2}{\partial \theta^2} \log f(X|\theta).$$

Recordemos que, por definición, la información de Fisher se define como

$$I(\theta) = -\mathbb{E}[\ell''(\theta)].$$

Procedemos a calcular  $\ell''(\theta)$  usando la regla del producto sobre  $\ell'(\theta) = \partial_{\theta} \log f$ . Dado que  $\ell'(\theta) = (\partial_{\theta} f)/f$ , tenemos

$$\ell''(\theta) = \frac{\partial}{\partial \theta} \left[ \frac{1}{f(X|\theta)} \partial_{\theta} f(X|\theta) \right] = -\frac{1}{(f(X|\theta))^2} (\partial_{\theta} f(X|\theta))^2 + \frac{1}{f(X|\theta)} \partial_{\theta}^2 f(X|\theta).$$

En otras palabras,

$$\ell''(\theta) = -\left( \partial_{\theta} \log f(X|\theta) \right)^2 + \frac{1}{f(X|\theta)} \frac{\partial^2}{\partial \theta^2} f(X|\theta).$$

Tomamos entonces esperanza en ambos lados:

$$\mathbb{E}[\ell''(\theta)] = -\mathbb{E} \left[ \left( \partial_{\theta} \log f(X|\theta) \right)^2 \right] + \mathbb{E} \left[ \frac{1}{f(X|\theta)} \partial_{\theta}^2 f(X|\theta) \right].$$

Mostremos ahora que  $\mathbb{E}[f(X | \theta)^{-1} \partial_\theta^2 f(X | \theta)] = 0$ . Observemos:

$$\begin{aligned} \mathbb{E}\left[\frac{1}{f(X | \theta)} \partial_\theta^2 f(X | \theta)\right] &= \int_{\mathcal{X}} f(x | \theta) \frac{1}{f(x | \theta)} \frac{\partial^2}{\partial \theta^2} f(x | \theta) dx \\ &= \int_{\mathcal{X}} \frac{\partial^2}{\partial \theta^2} f(x | \theta) dx \\ &= \frac{\partial^2}{\partial \theta^2} \underbrace{\int_{\mathcal{X}} f(x | \theta) dx}_{=1} \\ &= \frac{\partial^2}{\partial \theta^2} (1) \\ &= 0. \end{aligned}$$

Aquí hemos aplicado la misma técnica que en la prueba de  $\mathbb{E}[\ell'(\theta)] = 0$ : intercambiamos derivada e integral sobre la función de densidad, cuya integral es siempre 1. Por lo tanto, se anula el segundo término en la esperanza de  $\ell''(\theta)$ , y obtenemos

$$\mathbb{E}[\ell''(\theta)] = -\mathbb{E}\left[(\partial_\theta \log f(X | \theta))^2\right].$$

Finalmente, usando la definición  $I(\theta) = -\mathbb{E}[\ell''(\theta)]$ , deducimos

$$I(\theta) = -[-\mathbb{E}[(\partial_\theta \log f(X | \theta))^2]] = \mathbb{E}\left[(\partial_\theta \log f(X | \theta))^2\right].$$

Esto concluye la prueba de que bajo las condiciones de regularidad la información de Fisher puede expresarse como

$$I(\theta) = \mathbb{E}\left[(\partial_\theta \log f(X | \theta))^2\right].$$

□

Veamos entonces que la información de Fisher es una herramienta extremadamente útil para varias aplicaciones dentro de la estadística. Para terminar de clarificar este concepto, veamos un ejemplo de la aplicación de la información de Fisher sobre la distribución normal. En los siguientes capítulos, utilizaremos principalmente la segunda definición.

**Ejemplo 7.3.3.** Supongamos una variable aleatoria  $X$  que sigue una distribución normal, de modo que  $X \sim \mathcal{N}(\mu, \sigma^2)$ , y tenemos  $\mu$  conocida, con parámetro desconocido  $\theta = \sigma^2$ , de modo que

$$\begin{aligned} f(x|\theta) &= \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{(x-\mu)^2}{2\theta}\right) \\ l(\theta) &= -\frac{1}{2} \log \theta - \frac{(x-\mu)^2}{2\theta} + c \\ l'(\theta) &= -\frac{1}{2\theta} + \frac{(x-\mu)^2}{2\theta^2} \\ -\mathbb{E}l''(\theta) &= -\mathbb{E}\left[\frac{1}{2\theta^2} - \frac{(x-\mu)^2}{\theta^3}\right] = -\frac{1}{2\theta^2} + \frac{1}{\theta^2} = \frac{1}{2\theta^2}. \end{aligned}$$

Por lo que

$$I(\theta) = \frac{n}{2\theta^2} = \frac{n}{2\sigma^4}.$$

Finalmente, cabe resaltar que si tenemos un modelo con múltiples variables desconocidas de modo que  $\theta = [\theta_1, \theta_2, \dots, \theta_n]$  es un vector, tendremos como resultado una *matriz* de información de Fisher, la cual se muestra a continuación.

**Definición 7.3.8.** Para cualesquier modelo con múltiples parámetros desconocidos, la información de Fisher se expresa como

$$I(\boldsymbol{\theta}) = \begin{bmatrix} \mathbb{E} \left[ \left( \frac{\partial}{\partial \theta_1} \log f(X | \boldsymbol{\theta}) \right)^2 \right] & \cdots & \mathbb{E} \left[ \frac{\partial}{\partial \theta_1} \log f(X | \boldsymbol{\theta}) \cdot \frac{\partial}{\partial \theta_k} \log f(X | \boldsymbol{\theta}) \right] \\ \vdots & \ddots & \vdots \\ \mathbb{E} \left[ \frac{\partial}{\partial \theta_k} \log f(X | \boldsymbol{\theta}) \cdot \frac{\partial}{\partial \theta_1} \log f(X | \boldsymbol{\theta}) \right] & \cdots & \mathbb{E} \left[ \left( \frac{\partial}{\partial \theta_k} \log f(X | \boldsymbol{\theta}) \right)^2 \right] \end{bmatrix} \quad (7.12)$$

Veamos un ejemplo utilizando la distribución normal.

**Ejemplo 7.3.4.** Sea una variable aleatoria  $X$  que sigue una distribución normal, tal que  $X \sim \mathcal{N}(\mu, \sigma^2)$ , donde tanto  $\mu$  y  $\sigma^2$  son desconocidos. Utilizando el ejemplo [7.3.2](#) la log-verosimilitud es

$$l(\mu, \sigma^2) = -\frac{1}{2} \log(\sigma^2) - \frac{(x - \mu)^2}{2\sigma^2} + c$$

A continuación obtenemos las derivadas parciales

$$\frac{\partial^2 l}{\partial \mu^2} = -\frac{1}{\sigma^2}, \quad \frac{\partial^2 l}{\partial \mu \partial \sigma^2} = -2 \frac{x - \mu}{\sigma^4}, \quad \frac{\partial^2 l}{\partial (\sigma^2)^2} = \frac{(x - \mu)^2}{2\sigma^4} - \frac{1}{2\sigma^2}, \quad \frac{\partial^2 l}{\partial \sigma^2 \partial \mu} = -2 \frac{x - \mu}{\sigma^4}.$$

Y al aplicar el valor esperado, obtenemos entonces que

$$I(\mu, \sigma^2) = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{bmatrix}.$$

Para un análisis más profundo y generalizado, con un nivel mayor de abstracción, sobre la información de Fisher, se recomienda consultar el libro *Theory of Statistics* de Schervish, el cual ofrece una visión rigurosa y completa de los fundamentos teóricos de la inferencia estadística [32](#).

## 8.1. Conceptos topológicos

Para entender la conexión entre la geometría diferencial y la teoría de la información es necesario presentar algunos fundamentos de topología. Las nociones de diferenciabilidad y las propiedades de las curvas dependen de conceptos topológicos básicos, que aquí se introducen suponiendo un conocimiento previo de topología general [27].

**Definición 8.1.1.** Un espacio topológico  $\mathcal{T}$  se llama de **Hausdorff** si para cada par de puntos  $\tau_1, \tau_2 \in \mathcal{T}$  con  $\tau_1 \neq \tau_2$  existen vecindades abiertas  $\mathcal{V}_1$  de  $\tau_1$  y  $\mathcal{V}_2$  de  $\tau_2$  tales que  $\mathcal{V}_1 \cap \mathcal{V}_2 = \emptyset$  [24].

Estos espacios poseen propiedades que no se garantizan en un espacio topológico general. Por ejemplo, todo espacio métrico es de Hausdorff. En particular,  $\mathbb{R}^n$  lo es: dados dos puntos distintos se pueden encontrar bolas abiertas centradas en cada uno que no se intersectan. Presentamos algunos resultados básicos.

**Teorema 8.1.1.** En un espacio de Hausdorff  $\mathcal{T}$ , todo conjunto finito es cerrado.

*Demostración.* Basta demostrar que un conjunto unitario  $\{\tau_0\}$  es cerrado. Sea  $\tau \neq \tau_0$ ; por la propiedad de Hausdorff existen vecindades abiertas disjuntas de  $\tau_0$  y  $\tau$ , de modo que  $\tau \notin \overline{\{\tau_0\}}$ . Por lo tanto, la clausura de  $\{\tau_0\}$  coincide con el conjunto mismo, que resulta ser cerrado [24].  $\square$

**Teorema 8.1.2.** Si  $\mathcal{T}$  es de Hausdorff, toda sucesión de puntos en  $\mathcal{T}$  converge, a lo sumo, a un único límite [24].

**Teorema 8.1.3.** Todo conjunto totalmente ordenado, provisto de la topología ordenada, es de Hausdorff. El producto de dos espacios de Hausdorff también es de Hausdorff. Además, cualquier subespacio de un espacio de Hausdorff es nuevamente de Hausdorff.

Las demostraciones de estos resultados siguen ideas similares a las del teorema 8.1.1 y pueden consultarse en la obra *Topology* de Munkres [24].

### 8.1.1. Funciones continuas

**Definición 8.1.2.** Sean  $\mathcal{T}$  y  $\mathcal{S}$  espacios topológicos. Una función  $f : \mathcal{T} \rightarrow \mathcal{S}$  es *continua* si para todo abierto  $\mathcal{V} \subset \mathcal{S}$  la preimagen  $f^{-1}(\mathcal{V})$  es abierto en  $\mathcal{T}$  [24].

La continuidad depende de la función y de las topologías asignadas al dominio y al codominio [24]. Este concepto es esencial porque la diferenciabilidad presupone continuidad, y servirá de base para las definiciones de variedades diferenciables.

**Teorema 8.1.4.** Sean  $\mathcal{T}$  y  $\mathcal{S}$  espacios topológicos y  $f : \mathcal{T} \rightarrow \mathcal{S}$ . Las siguientes condiciones son equivalentes:

1.  $f$  es continua.
2. Para todo  $A \subseteq \mathcal{T}$ , se tiene  $f(\overline{A}) \subseteq \overline{f(A)}$ .
3. Para cada cerrado  $B \subseteq \mathcal{S}$ , el preimagen  $f^{-1}(B)$  es cerrado en  $\mathcal{T}$ .
4. Para cada punto  $x \in \mathcal{T}$  y cada vecindad  $\mathcal{V}$  de  $f(x)$  existe una vecindad  $\mathcal{U}$  de  $x$  tal que  $f(\mathcal{U}) \subseteq \mathcal{V}$ .

Cumplir la última condición en un punto  $x$  se denomina *continuidad puntual* en  $x$ .

**Definición 8.1.3.** Sean  $\mathcal{T}$  y  $\mathcal{S}$  espacios topológicos y  $f : \mathcal{T} \rightarrow \mathcal{S}$  una función biyectiva. Si  $f$  y su inversa  $f^{-1} : \mathcal{S} \rightarrow \mathcal{T}$  son continuas, entonces  $f$  es un **homeomorfismo** [24].

Más adelante se verá la utilidad de los homeomorfismos en el desarrollo de la geometría de la información.

### 8.1.2. Espacios 2º enumerables

En topología se introducen los llamados axiomas de enumerabilidad, que permiten clasificar espacios según la existencia de bases contables. Existen dos niveles: espacios 1º enumerables y 2º enumerables. Nos interesan principalmente los segundos, aunque recordamos primero la definición de los primeros.

**Definición 8.1.4.** Un espacio topológico  $\mathcal{T}$  tiene una **base contable** en un punto  $x$  si existe una colección de vecindades  $\mathfrak{B}$  de  $x$  tal que toda vecindad de  $x$  contiene al menos un elemento de  $\mathfrak{B}$ . Si cada punto posee una base contable,  $\mathcal{T}$  se llama **1º enumerable** [24].

**Definición 8.1.5.** Un espacio topológico  $\mathcal{T}$  se llama **2º enumerable** si la topología tiene una base contable; es decir, existe una familia contable de abiertos tal que todo abierto es unión de elementos de esa familia [24].

Los espacios de Hausdorff garantizan la existencia de suficientes abiertos, mientras que la 2º enumerabilidad asegura que la colección de abiertos se pueda describir mediante una base contable [18]. Dado que la 2º enumerabilidad es una condición más fuerte, implica la 1º enumerabilidad [24, 18]. Esta propiedad permite trabajar con una base numerable de abiertos al estudiar coberturas [18].

**Definición 8.1.6.** Sea  $\mathcal{T}$  un espacio topológico. Una colección  $\mathcal{U}$  de subconjuntos de  $\mathcal{T}$  **cubre** a  $\mathcal{T}$  (o es una **cubierta**) si cada punto de  $\mathcal{T}$  pertenece a al menos un conjunto de  $\mathcal{U}$  [18].

**Definición 8.1.7.** Una **cubierta abierta** es una colección  $\mathcal{U}$  de abiertos que cubre a  $\mathcal{T}$  [18].

**Definición 8.1.8.** Dada una cubierta  $\mathcal{U}$ , una **subcubierta** es una subcolección de  $\mathcal{U}$  que sigue cubriendo a  $\mathcal{T}$  [18].

**Lemma 8.1.5.** Si  $\mathcal{T}$  es un espacio  $2^\circ$  enumerable, entonces toda cubierta abierta de  $\mathcal{T}$  tiene una subcubierta contable [17].

*Demostración.* Sea  $\mathfrak{B}$  una base contable para  $\mathcal{T}$  y  $\mathcal{U}$  una cubierta abierta. Sea  $\mathfrak{B}' \subseteq \mathfrak{B}$  el subconjunto de bases que están contenidas en algún elemento de  $\mathcal{U}$ ; como  $\mathfrak{B}$  es contable,  $\mathfrak{B}'$  también lo es. Para cada  $B \in \mathfrak{B}'$  elija  $U_B \in \mathcal{U}$  con  $B \subseteq U_B$  y defina  $\mathcal{U}' = \{U_B : B \in \mathfrak{B}'\}$ . La colección  $\mathcal{U}'$  es contable y contiene a cada punto  $x \in \mathcal{T}$ : dada  $x$ , existe  $U_0 \in \mathcal{U}$  con  $x \in U_0$ , y porque  $\mathfrak{B}$  es base existe  $B \in \mathfrak{B}$  con  $x \in B \subseteq U_0$ . Entonces  $B \in \mathfrak{B}'$  y  $x \in U_B \in \mathcal{U}'$ . Por lo tanto  $\mathcal{U}'$  es una subcubierta contable.  $\square$

La mayor parte de los espacios topológicos utilizados en análisis y geometría, como  $\mathbb{R}^n$ , son  $2^\circ$  enumerables. Además, si  $\mathcal{T}$  es  $2^\circ$  enumerable, cualquier abierto  $U \subseteq \mathcal{T}$  también lo es [18].

### 8.1.3. Localmente euclidiano

**Definición 8.1.9.** Un espacio topológico  $\mathcal{T}$  es **localmente  $n$ -euclidiano** si, para cada  $x \in \mathcal{T}$ , existe una vecindad de  $x$  homeomorfa a un abierto de  $\mathbb{R}^n$  [24].

En otras palabras,  $\mathcal{T}$  es localmente euclidiano de dimensión  $n$  si cada punto de  $\mathcal{T}$  posee una vecindad topológicamente equivalente a una bola abierta en  $\mathbb{R}^n$  [18]. Cabe señalar que esta propiedad no implica que el espacio sea necesariamente de Hausdorff.

**Lemma 8.1.6.** Un espacio topológico  $\mathcal{T}$  es localmente euclidiano de dimensión  $n$  si y sólo si se cumple alguna de las siguientes condiciones [18]:

1. Cada punto de  $\mathcal{T}$  tiene una vecindad homeomorfa a una bola abierta en  $\mathbb{R}^n$ .
2. Cada punto de  $\mathcal{T}$  tiene una vecindad homeomorfa a  $\mathbb{R}^n$ .

Esta definición también es válida para  $n = 0$  [18]. Además, un homeomorfismo entre un subconjunto  $U \subset \mathcal{T}$  y un abierto de  $\mathbb{R}^n$  se denomina *carta* sobre  $U$ ; este concepto se desarrollará en secciones posteriores [18].

## 8.2. Variedades diferenciables

La diferenciabilidad es necesaria para las estructuras que se abordarán en los capítulos posteriores. Para aplicar los principios de la geometría de la información es preciso introducir elementos de geometría diferencial. Los conceptos de continuidad y diferenciabilidad determinan los objetos sobre los que pueden definirse técnicas de cálculo y estadística de forma coherente.

**Definición 8.2.1.** Sea  $U \subset \mathbb{R}^n$  un abierto y  $f : U \rightarrow \mathbb{R}$ . Decimos que  $f$  es de clase  $C^k$  en  $U$ , con  $k \geq 0$ , si todas las derivadas parciales  $\frac{\partial^\alpha f}{\partial x^\alpha}$  existen y son continuas en  $U$  para  $|\alpha| \leq k$ . Particularmente  $f$  es  $C^0$  si es continua. Si  $f : U \rightarrow \mathbb{R}^m$ , decimos que  $f$  es  $C^k$  si cada componente  $f_i = \pi_i \circ f$  es  $C^k$ . Decimos que  $f$  es  $C^\infty$  si es  $C^k$  para todo  $k \geq 0$  [33].

**Nota.** Un homeomorfismo que además es de clase  $C^\infty$  se denomina *difeomorfismo*.

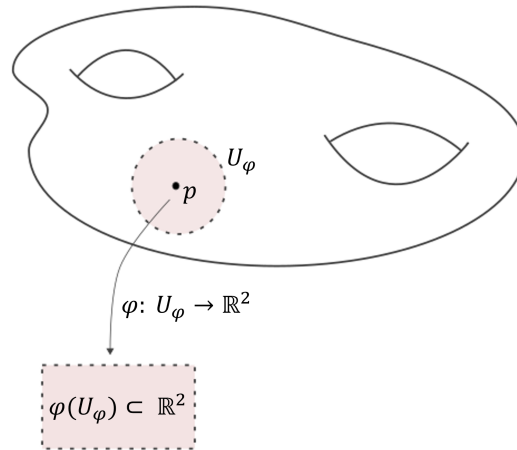


Figura 8.1: Ejemplo de una carta

### 8.2.1. *Manifolds* o variedades topológicas

**Definición 8.2.2.** Una **variedad topológica de dimensión  $n$**  es un espacio topológico Hausdorff,  $2^\circ$  enumerable, y localmente  $n$ -euclidiano [18].

Si  $n = 2$ , la variedad se denomina *superficie*. El ejemplo básico de una  $n$ -variedad es  $\mathbb{R}^n$ , utilizando la identidad como homeomorfismo [18].

**Lemma 8.2.1.** Todo subconjunto abierto de una  $n$ -variedad es nuevamente una  $n$ -variedad.

*Demostración.* Sea  $M$  una  $n$ -variedad y  $V \subset M$  un abierto. Cada punto  $p \in V$  tiene una vecindad en  $M$  homeomorfa a un abierto de  $\mathbb{R}^n$ . La intersección de esa vecindad con  $V$  es abierta en  $V$  y sigue siendo homeomorfa a un abierto de  $\mathbb{R}^n$ , por lo que  $V$  es localmente euclidiana. Los abiertos de un espacio Hausdorff son Hausdorff y los subespacios abiertos de un espacio  $2^\circ$  enumerable son  $2^\circ$  enumerables, luego  $V$  es una  $n$ -variedad [18].  $\square$

En adelante, cuando la dimensión sea clara, se empleará simplemente *variedad* en lugar de  $n$ -variedad. Cada variedad tiene una dimensión bien determinada; por lo tanto, objetos de distintas dimensiones no son homeomorfos bajo esta definición.

**Teorema 8.2.2.** Si  $M$  es una variedad topológica no vacía de dimensión  $n$ , no puede ser homeomorfa a una variedad topológica no vacía de dimensión  $m$  a menos que  $n = m$  [19].

La demostración de este teorema está fuera del alcance de este trabajo debido a su complejidad técnica. Para el lector interesado, puede consultarse el capítulo 2 de la segunda edición de *Introduction to Smooth Manifolds* de John M. Lee [19].

### 8.2.2. Cartas locales y atlas

**Definición 8.2.3.** Sea  $M$  una variedad topológica. Una **carta** sobre  $M$  es un par  $(U, \varphi)$  formado por un abierto  $U \subset M$  y un homeomorfismo  $\varphi: U \rightarrow \hat{U}$  con  $\hat{U} = \varphi(U) \subset \mathbb{R}^n$  abierto [19, 7] (véase la Figura 8.1).

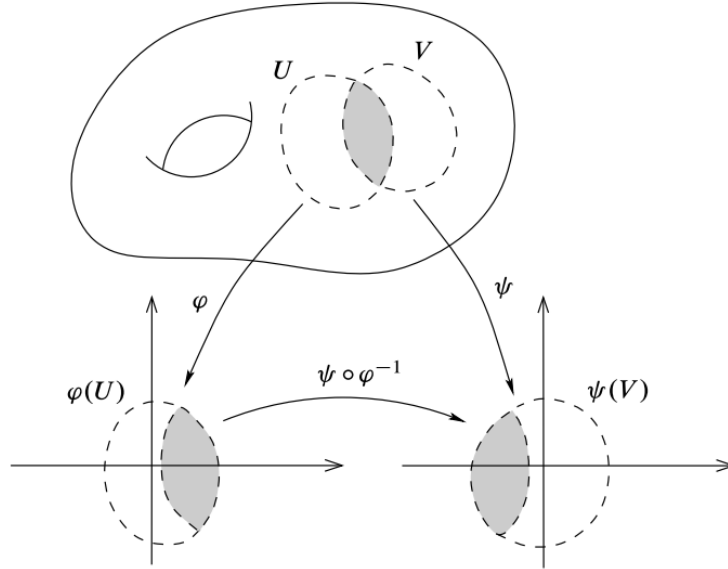


Figura 8.2: Un mapeo de transición [19]

Por la definición de variedad topológica [8.2.2] cada  $x \in M$  pertenece al dominio de alguna carta  $(U, \varphi)$ . Si  $\varphi(x) = 0$ , se dice que la carta está *centrada* en  $x$ . A partir de una carta que contiene a  $x$  puede construirse otra centrada en  $x$  restando el vector constante  $\varphi(x)$  [19].

**Definición 8.2.4.** Dada una carta  $(U, \varphi)$ , se denomina a  $U$  **vecindad dominio** o **dominio coordinado** de sus puntos [19].

En algunos casos se emplean nombres específicos: una *bola coordinada* cuando  $\varphi(U)$  es una bola abierta de  $\mathbb{R}^n$  y un *cubo coordinado* cuando es un cubo abierto. El mapeo  $\varphi$  se llama **mapeo coordinado local**. Si escribimos  $\varphi(x) = (\varphi_1(x), \dots, \varphi_n(x))$ , las funciones  $\varphi_i$  reciben el nombre de *coordenadas locales* en  $U$  [19].

Para que las propiedades que definamos no dependan de la carta concreta, se restringe el análisis a *cartas suaves*, es decir, aquellas cuyo mapeo de transición con otras cartas es de clase  $C^k$  (ver definición [8.2.1]).

**Definición 8.2.5.** Un **atlas**  $\mathcal{A}$  de  $M$  es una colección de cartas cuyos dominios cubren  $M$ . El atlas es **suave** si cualquier par de cartas de  $\mathcal{A}$  son suavemente compatibles [19].

**Definición 8.2.6.** Sea  $M$  una variedad topológica. Si  $(U, \varphi)$  y  $(V, \psi)$  son cartas con  $U \cap V \neq \emptyset$ , el mapeo compuesto  $\psi \circ \varphi^{-1} : \varphi(U \cap V) \rightarrow \psi(U \cap V)$  se llama **mapeo de transición** de  $\varphi$  a  $\psi$  [19] (véase la Figura [8.2]).

**Definición 8.2.7.** Dos cartas  $(U, \varphi)$  y  $(V, \psi)$  son **suavemente compatibles** si  $U \cap V = \emptyset$  o si el mapeo de transición  $\psi \circ \varphi^{-1}$  es un difeomorfismo [19].

**Ejemplo 8.2.1.** Considérese la esfera unitaria  $S^2 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$ . Definimos dos cartas mediante proyecciones estereográficas. Sea  $N = (0, 0, 1)$  el polo norte y  $S = (0, 0, -1)$  el polo sur. Tomamos los abiertos  $U_N = S^2 \setminus \{N\}$  y  $U_S = S^2 \setminus \{S\}$ . Considerando  $\mathbb{R}^2$  en el plano  $z = 0$ , definimos

$$\varphi : U_N \rightarrow \mathbb{R}^2, \quad \varphi(x, y, z) = \left( \frac{x}{1-z}, \frac{y}{1-z} \right), \quad \psi : U_S \rightarrow \mathbb{R}^2, \quad \psi(x, y, z) = \left( \frac{x}{1+z}, \frac{y}{1+z} \right).$$

Cada mapeo es un homeomorfismo con inversa

$$\varphi^{-1}(u, v) = \left( \frac{2u}{u^2 + v^2 + 1}, \frac{2v}{u^2 + v^2 + 1}, \frac{u^2 + v^2 - 1}{u^2 + v^2 + 1} \right),$$

y análogamente para  $\psi^{-1}$ . En la intersección  $U_N \cap U_S = S^2 \setminus \{N, S\}$ , el mapeo de transición es

$$\psi \circ \varphi^{-1}(u, v) = \left( \frac{u}{u^2 + v^2}, \frac{v}{u^2 + v^2} \right) \quad \text{para } (u, v) \neq (0, 0).$$

Este mapeo es  $C^\infty$ , por lo que  $\mathcal{A} = \{(U_N, \varphi), (U_S, \psi)\}$  constituye un *atlas suave* sobre  $S^2$ .

Aunque hemos definido cartas, atlas y compatibilidad suave, para desarrollar estructuras más profundas en geometría diferencial no basta con un atlas concreto: diferentes atlas pueden definir la misma estructura diferenciable en una variedad. Para formalizar esto introducimos la noción de *estructura diferenciable*.

**Definición 8.2.8.** Una **estructura diferenciable** de clase  $C^k$ , con  $1 \leq k \leq \infty$ , en un espacio localmente euclidiano  $M$  es un atlas  $\{(U_\alpha, \varphi_\alpha)\}_{\alpha \in A}$  tal que:

1.  $\bigcup_{\alpha \in A} U_\alpha = M$ ,
2. Para todo  $\alpha, \beta \in A$ , las transiciones  $\varphi_\beta \circ \varphi_\alpha^{-1}$  son de clase  $C^k$  donde están definidas,
3. El atlas es **máximo** con respecto a estas propiedades, es decir, no está contenido en ningún atlas diferenciable más grande (tal atlas se dice *completo*) [19, 33].

No todas las variedades admiten estructura diferenciable; sin embargo, sin ella no es posible hablar de variedades diferenciables.

**Lemma 8.2.3.** Dos atlas diferenciables sobre  $M$  inducen la misma estructura diferenciable si y sólo si su unión es un atlas diferenciable [19].

### 8.2.3. Variedades diferenciables

Una vez definida la noción de variedad topológica, introducimos las *variedades diferenciables* (o *smooth manifolds*). Intuitivamente, son espacios que localmente se parecen a  $\mathbb{R}^n$ , lo que permite aplicar herramientas del cálculo [19]. Ejemplos que no son el plano incluyen esferas, toros y otras superficies suaves [19].

**Definición 8.2.9.** Una **variedad diferenciable** es un par  $(M, \mathcal{A})$  en el que  $M$  es una variedad topológica y  $\mathcal{A}$  una estructura diferenciable sobre  $M$  [19].

El interés de las variedades diferenciables radica en que permiten integrar y derivar sobre estos objetos, lo que es fundamental en muchas aplicaciones de la geometría diferencial [19]. La condición de *suavidad* exige que la variedad no presente cambios bruscos: en términos de cálculo, que exista una pendiente (tangente) bien definida en cada punto. Cabe recordar que la continuidad y la suavidad son conceptos distintos; una función puede ser continua sin ser suave.

**Ejemplo 8.2.2.** Una función continua pero no suave es la **función de Weierstrass**,

$$W(x, y) = \sum_{n=0}^{\infty} a^n \cos(b^n x) \cos(b^n y),$$

donde  $0 < a < 1$ ,  $b$  es un entero impar y  $ab > 1 + \frac{3\pi}{2}$ . Su representación gráfica muestra que, al aumentar  $n$ , la función adquiere una superficie con picos irregulares: no admite tangentes en ningún punto aunque sigue siendo continua.

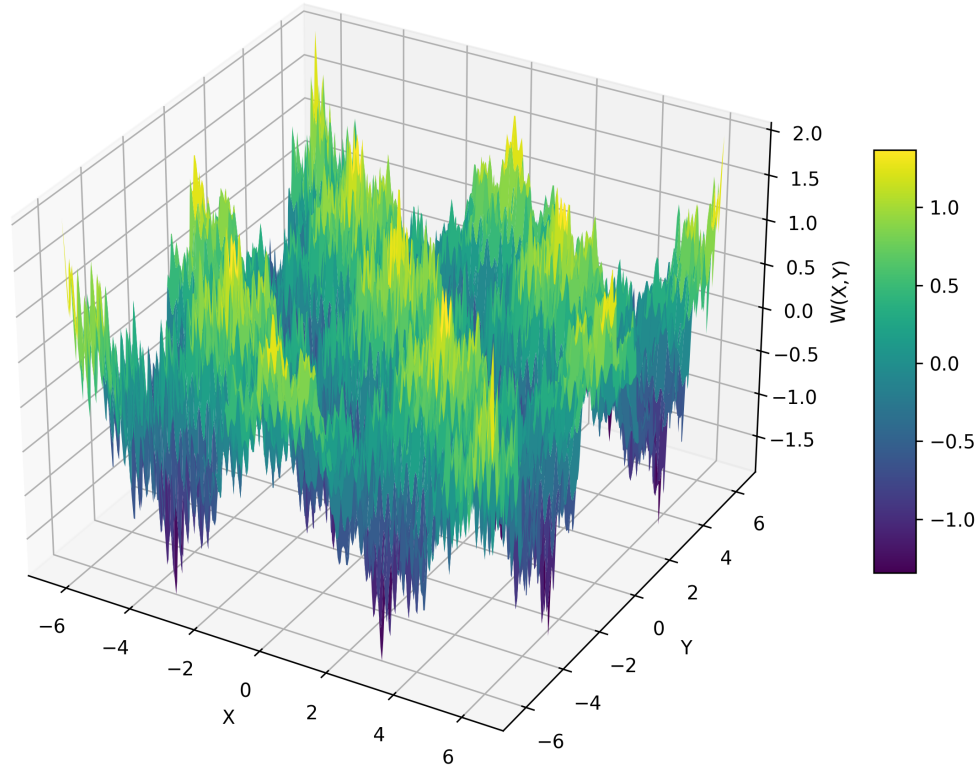


Figura 8.3: Función de Weierstrass con  $a = 0.5$ ,  $b = 5$  y  $0 \leq n \leq 50$

**Ejemplo 8.2.3.** Un ejemplo no orientable de variedad diferenciable es la *tira de Möbius*, parametrizada por

$$\begin{aligned} x(t, s) &= (R + s \cos(\frac{1}{2}t)) \cos t, \\ y(t, s) &= (R + s \cos(\frac{1}{2}t)) \sin t, \quad 0 < s < 1, \quad 0 < t < 2\pi, \\ z(t, s) &= s \sin(\frac{1}{2}t). \end{aligned}$$

Aunque globalmente no es orientable, localmente se comporta como  $\mathbb{R}^2$  y admite una tangente en cada punto.

En adelante consideraremos solo variedades diferenciables y los objetos necesarios para su estudio; muchas de las definiciones anteriores se adaptan con pequeñas modificaciones.

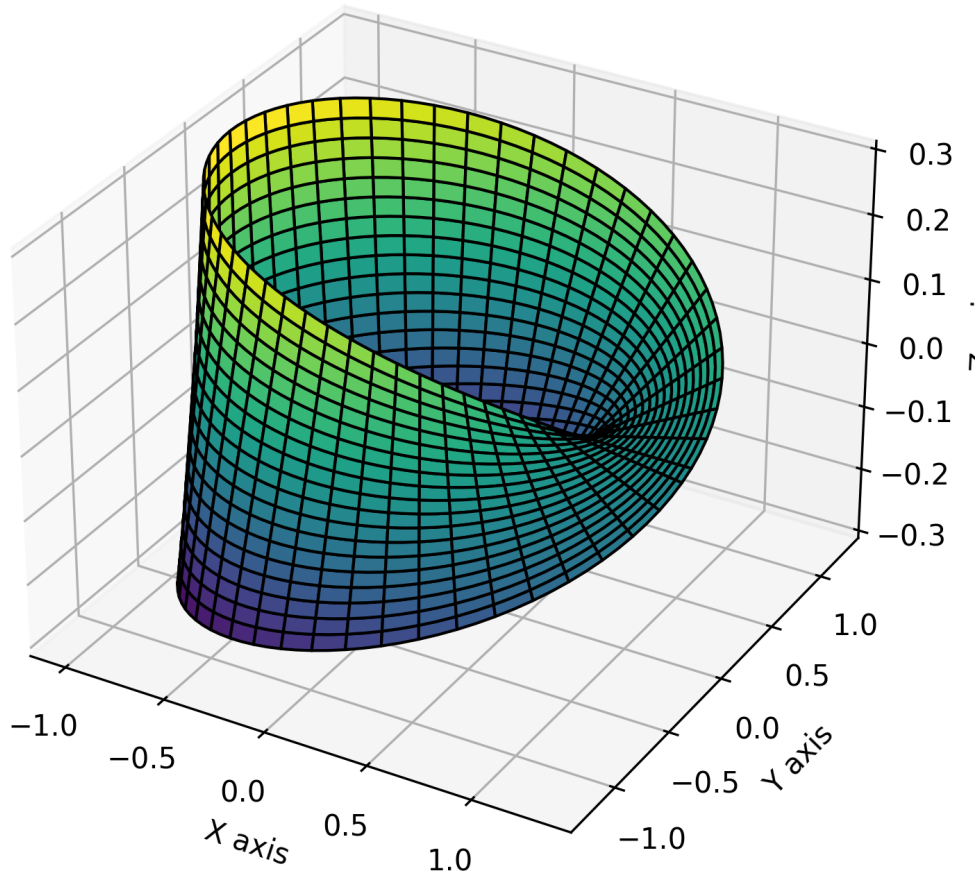
**Definición 8.2.10.** Sean  $M_1$  una  $n$ -variedad y  $M_2$  una  $m$ -variedad diferenciables. Un mapeo  $\varphi : M_1 \rightarrow M_2$  es **diferenciable en**  $p \in M_1$  si, dadas cartas  $\rho : V \rightarrow M_2$  en  $\varphi(p)$  y  $\varrho : U \rightarrow M_1$  en  $p$  con  $\varphi(\varrho(U)) \subset \rho(V)$ , el compuesto

$$\rho^{-1} \circ \varphi \circ \varrho : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$$

es diferenciable en  $\varrho^{-1}(p)$  [7].

### 8.2.4. Vectores tangentes y espacio tangente

**Definición 8.2.11.** Sea  $M$  una variedad diferenciable. Una **curva diferenciable** en  $M$  es una función diferenciable  $\alpha : (-\varepsilon, \varepsilon) \rightarrow M$  con  $\alpha(0) = p$  para algún  $p \in M$ , donde  $\varepsilon > 0$  es pequeño [7].

Figura 8.4: Tira de Möbius con  $R = 1$ 

**Definición 8.2.12.** Sea  $M$  una variedad diferenciable y sea  $\mathcal{D}$  el conjunto de funciones diferenciables definidas en un entorno de  $p \in M$ . El **vector tangente** a la curva  $\alpha$  en  $t = 0$  es el operador

$$\alpha'(0) : \mathcal{D} \rightarrow \mathbb{R}, \quad \alpha'(0)(f) = \left. \frac{d}{dt}(f \circ \alpha) \right|_{t=0},$$

para toda  $f \in \mathcal{D}$  [7].

El vector tangente representa la derivada direccional de las funciones diferenciables a lo largo de  $\alpha$  en  $p$ .

**Definición 8.2.13.** El **espacio tangente**  $T_p M$  de  $M$  en  $p$  es el conjunto de todos los vectores tangentes en  $p$ . Con las operaciones naturales de suma y multiplicación por escalares,  $T_p M$  es un espacio vectorial real de dimensión  $n$ , donde  $n$  es la dimensión de  $M$  [7].

Si  $\varphi : U \rightarrow M$  es una carta con  $\varphi(0) = p$ , un vector tangente puede escribirse como

$$\alpha'(0) = \sum_{i=1}^n x'_i(0) \left( \frac{\partial}{\partial x_i} \right)_p,$$

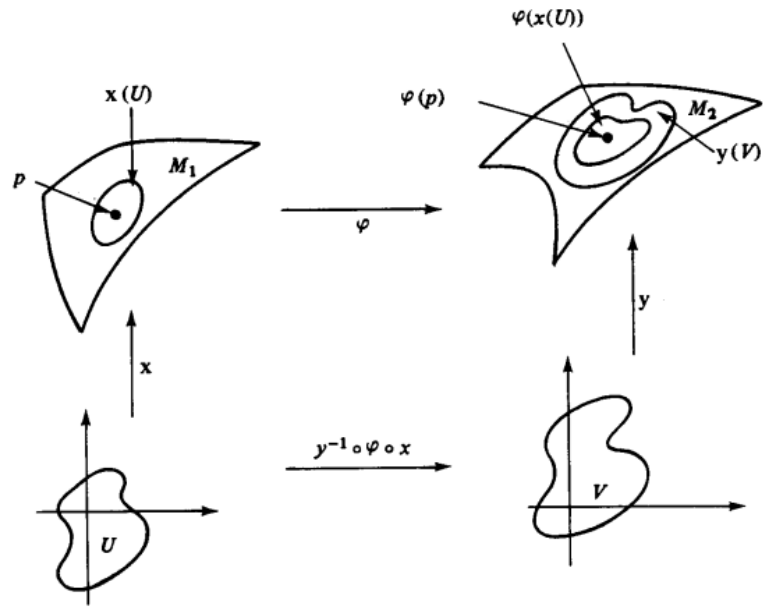


Figura 8.5: Mapeo diferenciable entre variedades [7]

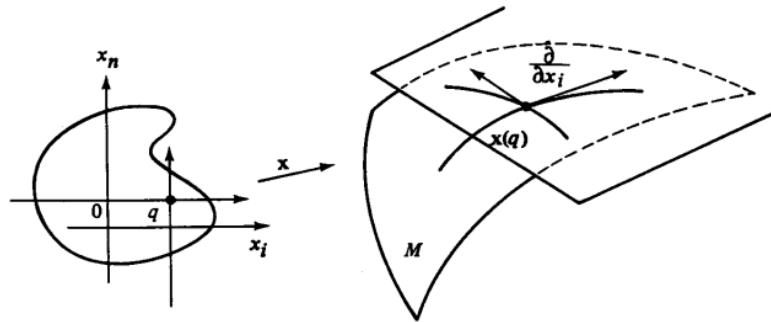


Figura 8.6: Ejemplificación del vector tangente en una variedad diferenciable [7]

donde  $\left(\frac{\partial}{\partial x_i}\right)_p$  son los operadores derivadas parciales en  $p$  respecto a las coordenadas locales inducidas por  $\varphi$ . Una base natural de  $T_p M$  es  $\left\{\left(\frac{\partial}{\partial x_1}\right)_p, \dots, \left(\frac{\partial}{\partial x_n}\right)_p\right\}$  [7].

**Definición 8.2.14.** Sean  $M_1$  y  $M_2$  variedades diferenciables y  $\varphi : M_1 \rightarrow M_2$  un mapeo diferenciable. El **diferencial de  $\varphi$  en  $p \in M_1$** , denotado  $d\varphi_p$ , es la transformación lineal

$$d\varphi_p : T_p M_1 \rightarrow T_{\varphi(p)} M_2$$

definida por

$$d\varphi_p(v) = \left. \frac{d}{dt} \varphi(\alpha(t)) \right|_{t=0},$$

donde  $\alpha : (-\varepsilon, \varepsilon) \rightarrow M_1$  es una curva diferenciable con  $\alpha(0) = p$  y  $\alpha'(0) = v$  [7].

La intuición detrás del diferencial es que nos permite linealizar el comportamiento local de  $\varphi$  alrededor del punto  $p$ , de forma análoga a la derivada usual en  $\mathbb{R}^n$ .

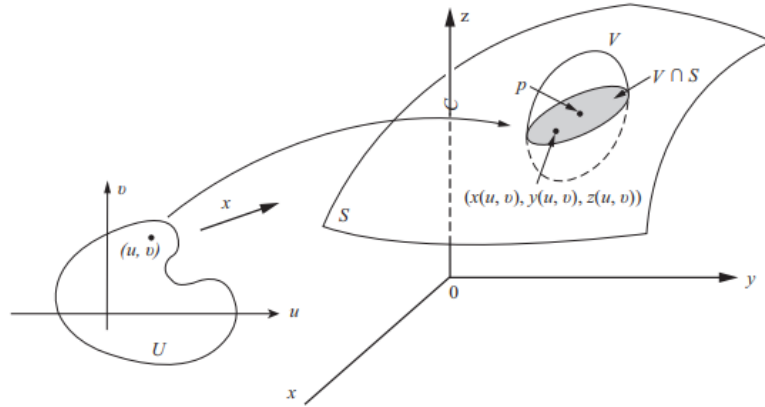


Figura 8.7: Ejemplo de una superficie regular [6]

### 8.3. Formas fundamentales

#### 8.3.1. Primera forma fundamental

Hasta ahora hemos descrito las variedades desde el punto de vista diferenciable. En esta sección introducimos las estructuras geométricas asociadas a nuestras variedades [6]. Para simplificar, nos centramos en las variedades de  $\mathbb{R}^3$ , es decir, superficies.

**Definición 8.3.1.** Un subconjunto  $S \subset \mathbb{R}^3$  es una **superficie regular** si, para cada  $p \in S$ , existen un abierto  $U \subset \mathbb{R}^2$ , un abierto  $V \subset \mathbb{R}^3$ , y un mapa  $\varphi : U \rightarrow V \cap S$  tales que:

1.  $\varphi$  es  $C^\infty$ ;
2.  $\varphi$  es un homeomorfismo sobre su imagen  $V \cap S$ ;
3. para cada  $q \in U$ , el diferencial  $d\varphi_q : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  es inyectivo.

Sea  $S$  una superficie y  $T_p S$  su plano tangente en  $p \in S$ . El producto interno de  $\mathbb{R}^3$  induce un producto interno en  $T_p S$ , denotado  $\langle \cdot, \cdot \rangle_p$  [6].

**Definición 8.3.2.** Para vectores tangentes  $w_1, w_2 \in T_p S$ , su **producto interno** se define por

$$\langle w_1, w_2 \rangle_p = \langle w_1, w_2 \rangle,$$

donde el producto de la derecha es el producto interno usual en  $\mathbb{R}^3$ . Este producto es una forma bilineal simétrica [30]; si  $a, b \in \mathbb{R}$  y  $w_i \in T_p S$ , se cumple:

$$\langle aw_1 + bw_2, w_3 \rangle_p = a\langle w_1, w_3 \rangle_p + b\langle w_2, w_3 \rangle_p, \quad \langle w_1, w_2 \rangle_p = \langle w_2, w_1 \rangle_p.$$

**Definición 8.3.3.** La función cuadrática  $I_p : T_p S \rightarrow \mathbb{R}$  dada por

$$I_p(w) = \langle w, w \rangle_p = \|w\|^2 \geq 0$$

se llama **primera forma fundamental** de la superficie regular  $S \subset \mathbb{R}^3$  en  $p$  [6].

Sea  $\varphi(u, v)$  una parametrización local de  $S$  en  $p$ . Un vector  $w \in T_p S$  puede escribirse como derivada de una curva  $\alpha(t) = \varphi(u(t), v(t))$  con  $\alpha(0) = p$ . Entonces

$$I_p(\alpha'(0)) = \langle \alpha'(0), \alpha'(0) \rangle_p = \langle \varphi_u u' + \varphi_v v', \varphi_u u' + \varphi_v v' \rangle_p,$$

lo que, al desarrollar, produce

$$I_p(\alpha'(0)) = \langle \varphi_u, \varphi_u \rangle_p u'^2 + 2\langle \varphi_u, \varphi_v \rangle_p u'v' + \langle \varphi_v, \varphi_v \rangle_p v'^2.$$

Definimos los coeficientes

$$E = \langle \varphi_u, \varphi_u \rangle_p, \quad F = \langle \varphi_u, \varphi_v \rangle_p, \quad G = \langle \varphi_v, \varphi_v \rangle_p,$$

que forman la matriz asociada a la primera forma fundamental en la base  $\{\varphi_u, \varphi_v\}$  de  $T_pS$  [6]. Con esta notación, la expresión anterior se escribe  $I_p(\alpha'(0)) = E u'^2 + 2F u'v' + G v'^2$ .

**Definición 8.3.4.** Sea  $R \subset S$  una región acotada contenida en la imagen de la carta  $\varphi : U \subset \mathbb{R}^2 \rightarrow S$ . El **área** de  $R$  se define como

$$A(R) = \iint_{\varphi^{-1}(R)} |\varphi_u \wedge \varphi_v| \, du \, dv, \quad \text{donde } \varphi^{-1}(R) \subset U.$$

Se observa que  $|\varphi_u \wedge \varphi_v|^2 + \langle \varphi_u, \varphi_v \rangle^2 = \|\varphi_u\|^2 \|\varphi_v\|^2$ ; por tanto, el integrando puede escribirse como  $|\varphi_u \wedge \varphi_v| = \sqrt{EG - F^2}$ . De esta manera, la fórmula de área queda:

$$A(R) = \iint_{\varphi^{-1}(R)} \sqrt{EG - F^2} \, du \, dv \quad [6].$$

### 8.3.2. Segunda forma fundamental

La primera forma fundamental permite obtener características intrínsecas de una superficie  $S$ , como distancias y áreas. Para estudiar cómo se curva  $S$  en  $\mathbb{R}^3$  introducimos la segunda forma fundamental, que recoge información extrínseca de la curvatura.

Consideraremos  $S$  como una superficie regular orientable con un campo de normales unitarias  $N$ ; diremos que  $S$  está orientada por  $N$  [6].

**Definición 8.3.5.** Sea  $S \subset \mathbb{R}^3$  una superficie orientada. El mapa  $N : S \rightarrow \mathbb{R}^3$  toma valores en la esfera unitaria

$$S^2 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}.$$

A  $N$  se le denomina **mapa de Gauss** de  $S$  [6].

**Definición 8.3.6.** Sea  $S$  una superficie orientada. La **segunda forma fundamental** de  $S$  en un punto  $p$  es la forma cuadrática

$$II_p : T_pS \rightarrow \mathbb{R}$$

definida por

$$II_p(\mathbf{v}) = -\langle dN_p(\mathbf{v}), \mathbf{v} \rangle. \quad (8.1)$$

Si  $\alpha : (-\varepsilon, \varepsilon) \rightarrow S$  es una curva con  $\alpha(0) = p$  y  $\alpha'(0) = \mathbf{v}$ , parametrizada por longitud de arco ( $|\alpha'(s)| = 1$ ), entonces  $\langle N(\alpha(s)), \alpha'(s) \rangle = 0$  para todo  $s$ . Derivando y evaluando en  $s = 0$  se obtiene

$$\langle N(p), \alpha''(0) \rangle = -\langle dN_p(\mathbf{v}), \mathbf{v} \rangle = II_p(\mathbf{v}),$$

y como  $\alpha''(0) = \kappa \mathbf{n}$  (donde  $\kappa$  es la curvatura de  $\alpha$  en  $p$  y  $\mathbf{n}$  su vector normal), se deduce

$$II_p(\mathbf{v}) = \kappa \langle N(p), \mathbf{n} \rangle, \quad (8.2)$$

donde  $\kappa$  es la curvatura de  $\alpha$  en  $s = 0$  [6].

**Definición 8.3.7.** Una **curva regular** es una función diferenciable  $\alpha : I \rightarrow \mathbb{R}^3$ , definida en un intervalo abierto  $I \subset \mathbb{R}$ , tal que  $\alpha'(t) \neq 0$  para todo  $t \in I$ . Si  $|\alpha'(t)| = 1$  para todo  $t$ , se dice que la curva está parametrizada por longitud de arco.

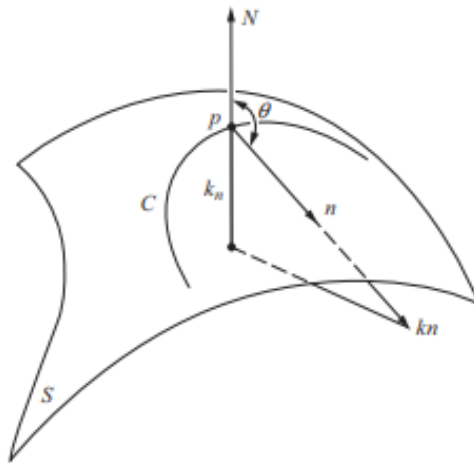


Figura 8.8: Curvatura normal de una superficie [6]

**Definición 8.3.8.** Sea  $\alpha : I \rightarrow \mathbb{R}^3$  una curva regular parametrizada por  $t$  y tal que  $|\alpha'(t)| = 1$ . La **curvatura** de  $\alpha$  en  $\alpha(t)$  se define por [6]

$$\kappa(t) = |\alpha''(t)|.$$

La curvatura mide la rapidez con que cambia la dirección del vector tangente unitario de la curva. Si  $\alpha$  no está parametrizada por longitud de arco, la curvatura viene dada por

$$\kappa(t) = \frac{|\alpha'(t) \times \alpha''(t)|}{|\alpha'(t)|^3}.$$

**Definición 8.3.9.** Sea  $C \subset S$  una curva regular que pasa por  $p \in S$ . Sean  $\kappa$  su curvatura en  $p$  y  $\cos \theta = \langle \mathbf{n}, N \rangle$ , donde  $\mathbf{n}$  es el vector normal de  $C$  y  $N$  el vector normal de  $S$  en  $p$ . La cantidad

$$\kappa_n = \kappa \cos \theta$$

se llama **curvatura normal** de  $C$  en  $p$  [6]. La curvatura normal es el componente normal de la curvatura de la curva y cambia de signo según la orientación de la superficie.

La curvatura normal  $\kappa_n$  es el valor absoluto de la proyección de  $\kappa \mathbf{n}$  sobre el vector normal a la superficie en  $p$ , con el signo determinado por la orientación. Cabe resaltar que, aunque  $\kappa_n$  no depende de la orientación de la curva, sí cambia de signo según la orientación de la superficie.

### 8.3.3. Símbolos de Christoffel

La primera y la segunda formas fundamentales permiten distinguir entre cantidades intrínsecas y extrínsecas en una superficie  $S$ . Una cantidad es *intrínseca* si se expresa mediante la primera forma fundamental y *extrínseca* si depende de la segunda. No siempre es fácil determinar a cuál tipo pertenece una característica.

Para una parametrización  $\varphi$  de  $S$ , los coeficientes

$$g_{ij} = \langle \varphi_i, \varphi_j \rangle, \quad h_{ij} = \langle N, \varphi_{ij} \rangle, \quad a_{ij} = \langle N_i, \varphi_j \rangle,$$

recogen información fundamental sobre la geometría de  $S$ . Aquí  $g_{ij}$  corresponden a la primera forma fundamental,  $h_{ij}$  a la segunda forma, y  $a_{ij}$  miden la variación de la normal  $N$  en las direcciones tangentes [29].

El diferencial del campo normal  $N$ , denotado  $DN$ , satisface las ecuaciones de Weingarten:

$$DN = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = - \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}^{-1}.$$

De este modo, el diferencial del mapa de Gauss se obtiene multiplicando la matriz de la segunda forma fundamental por el inverso de la matriz de la primera, con signo negativo.

Sea  $S$  una superficie regular orientada y  $\varphi : U \subset \mathbb{R}^2 \rightarrow S$  una parametrización compatible con la orientación. En cada punto de  $\varphi(U)$  se considera la base  $\{\varphi_u, \varphi_v, N\}$  [6]. Expresando las derivadas de  $\varphi_u, \varphi_v$  y  $N$  en dicha base, se obtienen las siguientes relaciones:

$$\begin{aligned} \varphi_{uu} &= \Gamma_{11}^1 \varphi_u + \Gamma_{11}^2 \varphi_v + h_{11} N, \\ \varphi_{uv} &= \Gamma_{12}^1 \varphi_u + \Gamma_{12}^2 \varphi_v + h_{12} N, \\ \varphi_{vu} &= \Gamma_{21}^1 \varphi_u + \Gamma_{21}^2 \varphi_v + h_{21} N, \\ \varphi_{vv} &= \Gamma_{22}^1 \varphi_u + \Gamma_{22}^2 \varphi_v + h_{22} N, \\ N_u &= a_{11} \varphi_u + a_{21} \varphi_v, \\ N_v &= a_{12} \varphi_u + a_{22} \varphi_v. \end{aligned}$$

Los coeficientes  $\Gamma_{ij}^k$ , con  $i, j, k = 1, 2$ , se denominan **símbolos de Christoffel** de la parametrización  $\varphi$  [6].

Los símbolos  $\Gamma_{ij}^k$  describen cómo varían las bases tangentes  $\{\varphi_u, \varphi_v\}$  al moverse sobre la superficie. Aunque no son tensores, contienen información geométrica intrínseca. Si  $g_{ij} = \langle \varphi_i, \varphi_j \rangle$  es la primera forma fundamental y  $(g^{ij})$  su inversa, entonces

$$\Gamma_{ij}^k = \frac{1}{2} \sum_{\ell=1}^2 g^{k\ell} \left( \frac{\partial g_{i\ell}}{\partial u^j} + \frac{\partial g_{j\ell}}{\partial u^i} - \frac{\partial g_{ij}}{\partial u^\ell} \right),$$

lo que muestra que dependen únicamente de la primera forma fundamental y de sus derivadas parciales [29, 6]. En consecuencia, son cantidades totalmente *intrínsecas*, en contraste con las magnitudes definidas por la segunda forma fundamental.

## 8.4. Geodésicas

Las geodésicas son una propiedad intrínseca de la superficie, ya que dependen únicamente de la primera forma fundamental. Para definir las utilizaremos los símbolos de Christoffel y conceptos relacionados. Intuitivamente, una geodésica es la curva más corta entre dos puntos de la superficie.

**Definición 8.4.1.** Sea  $w$  un campo vectorial diferenciable en un abierto  $U \subset S$  y sea  $p \in U$ . Tomemos  $y \in T_p S$  y una curva parametrizada  $\alpha : (-\varepsilon, \varepsilon) \rightarrow U$  con  $\alpha(0) = p$  y  $\alpha'(0) = y$ . Si  $w(t)$  denota el campo restringido a  $\alpha$ , el vector obtenido al proyectar ortogonalmente  $\frac{dw}{dt}(0)$  sobre  $T_p S$  se llama **derivada covariante** de  $w$  en  $p$  relativa a  $y$ . Esta derivada se denota  $(\frac{Dw}{dt})(0)$  o  $(D_y w)(p)$ .

La derivada covariante describe cómo cambia un campo vectorial en la superficie; puede verse como un análogo del gradiente, adaptado a la curvatura de  $S$ .

**Definición 8.4.2.** Una curva  $\alpha$  en  $S \subset \mathbb{R}^3$  es una **geodésica** si su aceleración  $\alpha''$  es normal a  $S$  [27]. De manera equivalente,  $\alpha$  es geodésica si su derivada covariante satisface  $D_{\alpha'} \alpha' = 0$ .

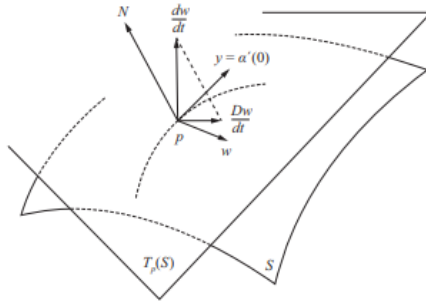


Figura 8.9: La derivada covariante [6]

La condición  $D_{\alpha'}\alpha' = 0$  expresa que el vector velocidad se mantiene paralelo a lo largo de la curva con respecto a la conexión de Levi-Civita. En consecuencia, una geodésica mantiene velocidad constante y describe la trayectoria más recta posible en la superficie.

**Definición 8.4.3.** Un **campo de triedros** sobre  $S$  es una asignación suave que asocia a cada  $p \in S$  una base ortonormal  $\{E_1(p), E_2(p)\}$  de  $T_pS$ . Este campo facilita la descripción local de velocidades y aceleraciones.

Si  $E_1, E_2$  forman un campo de triedros sobre  $S$ , la velocidad y aceleración de una curva  $\alpha$  pueden expresarse como [27]:

$$\alpha' = v_1E_1 + v_2E_2, \quad \alpha'' = A_1E_1 + A_2E_2.$$

La curva  $\alpha$  es geodésica si y sólo si  $A_1 = A_2 = 0$ , donde [27]:

$$A_1 = v_1' + v_2\omega_1^2(\alpha'), \quad A_2 = v_2' + v_1\omega_2^1(\alpha').$$

**Teorema 8.4.1.** Sea  $\varphi$  un sistema de coordenadas ortogonales en  $S$ . Una curva  $\alpha(t) = \varphi(\alpha_1(t), \alpha_2(t))$  es una **geodésica** si y sólo si satisface:

$$A_1 = \alpha_1'' + \frac{1}{2E} (E_u (\alpha_1')^2 + 2E_v \alpha_1' \alpha_2' - G_u (\alpha_2')^2) = 0,$$

$$A_2 = \alpha_2'' + \frac{1}{2G} (-E_v (\alpha_1')^2 + 2G_u \alpha_1' \alpha_2' + G_v (\alpha_2')^2) = 0,$$

donde  $E, G$  son los coeficientes de la primera forma fundamental y  $E_u, E_v, G_u, G_v$  sus derivadas parciales [27].

**Teorema 8.4.2.** Dado  $p \in S$  y un vector  $\mathbf{v} \in T_pS$ , existe una única geodésica  $\alpha$  definida en un intervalo  $I \ni 0$  tal que [27]:

$$\alpha(0) = p, \quad \alpha'(0) = \mathbf{v}.$$

Así, por cada posición inicial y dirección se determina una geodésica. No obstante, el intervalo  $I$  puede ser limitado. Si  $\gamma_1 : I_1 \rightarrow S$  y  $\gamma_2 : I_2 \rightarrow S$  son geodésicas con las mismas condiciones iniciales, el teorema de unicidad implica  $\gamma_1 = \gamma_2$  en  $I_1 \cap I_2$ . Se puede por tanto definir una **geodésica maximal**  $\gamma : I \rightarrow S$  como aquella que no admite extensión con las mismas condiciones iniciales [27].

**Definición 8.4.4.** Una superficie  $S$  es **completa** si toda geodésica maximal está definida en todo  $\mathbb{R}$  [27].

La completitud implica que cualquier geodésica no constante tiene longitud infinita. Por ejemplo, las geodésicas de una esfera o de un cilindro son completas.

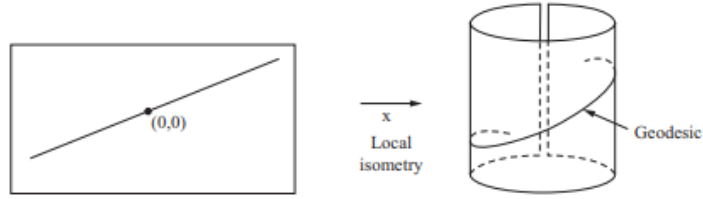


Figura 8.10: Geodésicas sobre un cilindro [6]

**Lemma 8.4.3.** Sea  $\{E_1, E_2\}$  un campo de triedros y sea  $\alpha$  una curva de velocidad constante tal que  $\alpha'$  y  $E_2$  nunca sean ortogonales. Si  $A_1 = 0$ , entonces  $A_2 = 0$ ; en consecuencia,  $\alpha$  es geodésica [27].

**Corolario 8.4.3.1.** Sea  $\beta$  una curva de velocidad unitaria en una región con un campo de triedros  $\{E_1, E_2\}$ . Si  $\varphi$  es el ángulo desde  $E_1$  hacia  $\beta'$  a lo largo de  $\beta$ , entonces

$$\kappa_g = \frac{d\varphi}{ds} + \omega_{12}(\beta'),$$

donde  $\kappa_g$  es la curvatura geodésica y  $\omega_{12}$  la forma de conexión correspondiente [27].

**Lemma 8.4.4.** Una curva regular  $\alpha$  en  $M$  es geodésica si y sólo si tiene velocidad constante y curvatura geodésica  $\kappa_g = 0$  [27].

## 8.5. Métrica de Riemann

Introducimos ahora estructuras métricas en variedades diferenciables, lo que permite definir distancias, ángulos, áreas y curvatura. Para ello se dota a la variedad de una estructura adicional: la **métrica de Riemann**, que extiende el producto interno usual de  $\mathbb{R}^n$  a contextos más generales.

**Definición 8.5.1.** Una **métrica Riemanniana** en una variedad diferenciable  $M$  es una familia de productos internos en los espacios tangentes  $T_pM$  que varía suavemente con el punto  $p$ . Una **variedad Riemanniana** es una variedad diferenciable provista de una métrica Riemanniana [11].

Sea  $x = (x^1, \dots, x^d)$  un sistema de coordenadas locales sobre  $M$ . En estas coordenadas, una métrica Riemanniana se representa mediante una matriz simétrica y definida positiva:

$$(g_{ij}(x))_{i,j=1,\dots,d},$$

con  $g_{ij} = g_{ji}$  para todo  $i, j$  y

$$\sum_{i=1}^d \sum_{j=1}^d g_{ij}(x) \xi^i \xi^j > 0 \quad \text{para todo } \xi = (\xi^1, \dots, \xi^d) \neq 0,$$

siendo  $g_{ij}$  funciones diferenciables de  $x$  [11].

Para  $v, w \in T_pM$ , expresados en coordenadas como

$$v = \sum_{i=1}^d v^i \frac{\partial}{\partial x^i}, \quad w = \sum_{j=1}^d w^j \frac{\partial}{\partial x^j},$$

el producto interno inducido por la métrica se define por

$$\langle v, w \rangle = \sum_{i=1}^d \sum_{j=1}^d g_{ij}(x(p)) v^i w^j. \quad (8.3)$$

En particular, las bases coordenadas satisfacen

$$\left\langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right\rangle = g_{ij} \quad \square \square \square.$$

La norma de  $v \in T_p M$  viene dada por

$$\|v\| = \sqrt{\langle v, v \rangle} = \sqrt{\sum_{i=1}^d \sum_{j=1}^d g_{ij}(x(p)) v^i v^j}.$$

Consideremos un cambio de coordenadas  $y = f(x)$ . En el nuevo sistema, los vectores se escriben como

$$v = \sum_{i=1}^d \tilde{v}^i \frac{\partial}{\partial y^i}, \quad w = \sum_{j=1}^d \tilde{w}^j \frac{\partial}{\partial y^j},$$

donde

$$\tilde{v}^i = \sum_{k=1}^d v^k \frac{\partial f^i}{\partial x^k}, \quad \tilde{w}^j = \sum_{l=1}^d w^l \frac{\partial f^j}{\partial x^l}.$$

Si  $(h_{kl})$  son los coeficientes de la métrica en estas coordenadas, entonces

$$\sum_{k=1}^d \sum_{l=1}^d h_{kl}(f(x)) \tilde{v}^k \tilde{w}^l = \sum_{i=1}^d \sum_{j=1}^d g_{ij}(x) v^i w^j.$$

La condición de invariancia de la métrica se traduce en

$$\sum_{k=1}^d \sum_{l=1}^d h_{kl}(f(x)) \frac{\partial f^k}{\partial x^i} \frac{\partial f^l}{\partial x^j} = g_{ij}(x), \quad (8.4)$$

lo que asegura que la métrica es intrínseca a la variedad.

**Teorema 8.5.1.** Toda variedad diferenciable admite una métrica Riemanniana  $\square \square \square$ .

### 8.5.1. Ejemplos de métricas de Riemann

- **Espacio euclidiano.** En  $\mathbb{R}^n$  se utiliza el producto interno estándar:

$$\langle v, w \rangle = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n.$$

- **Subvariedades de  $\mathbb{R}^n$ .** Una superficie regular  $S \subset \mathbb{R}^3$  hereda la métrica natural inducida por el producto interno de  $\mathbb{R}^3$ .
- **Superficie esférica.** En  $S^2 \subset \mathbb{R}^3$  la métrica inducida coincide con la restricción del producto interno de  $\mathbb{R}^3$ .

### 8.5.2. Propiedades

**Definición 8.5.2.** Sea  $[a, b] \subset \mathbb{R}$  un intervalo cerrado y  $\gamma : [a, b] \rightarrow M$  una curva de clase  $C^\infty$ . La **longitud** de  $\gamma$  se define por

$$L(\gamma) = \int_a^b \left\| \frac{d\gamma}{dt}(t) \right\| dt, \quad (8.5)$$

y la **energía** de  $\gamma$  como  $\square\square$

$$E(\gamma) = \frac{1}{2} \int_a^b \left\| \frac{d\gamma}{dt}(t) \right\|^2 dt. \quad (8.6)$$

Estas integrales pueden calcularse en coordenadas locales. Si  $x^i(\gamma(t))$  son las coordenadas de  $\gamma$  y  $\dot{x}^i(t)$  sus derivadas, entonces  $\square\square$

$$L(\gamma) = \int_a^b \sqrt{g_{ij}(x(\gamma(t))) \dot{x}^i(t) \dot{x}^j(t)} dt,$$

$$E(\gamma) = \frac{1}{2} \int_a^b g_{ij}(x(\gamma(t))) \dot{x}^i(t) \dot{x}^j(t) dt.$$

**Definición 8.5.3.** En una variedad Riemanniana  $M$ , la **distancia** entre dos puntos  $p$  y  $q$  se define por  $\square\square$

$$d(p, q) = \inf \{ L(\gamma) \mid \gamma : [a, b] \rightarrow M \text{ curva suave por tramos con } \gamma(a) = p, \gamma(b) = q \}. \quad (8.7)$$

**Lemma 8.5.2.** La función  $d(p, q)$  definida en  $\square(8.7)$  cumple las propiedades de un espacio métrico  $\square\square$ :

- **No-negatividad:**  $d(p, q) \geq 0$ , y  $d(p, q) = 0$  solo si  $p = q$ .
- **Simetría:**  $d(p, q) = d(q, p)$ .
- **Desigualdad triangular:**  $d(p, q) \leq d(p, r) + d(r, q)$ .

**Corolario 8.5.2.1.** La topología generada por  $d$  coincide con la topología original de la variedad  $M$   $\square\square$ .

**Lemma 8.5.3.** Sea  $\gamma : [a, b] \rightarrow M$  una curva de clase  $C^\infty$ . Se cumple

$$L(\gamma)^2 \leq 2(b-a) E(\gamma), \quad (8.8)$$

con igualdad si y solo si  $\|\gamma'(t)\|$  es constante  $\square\square$ .

**Lemma 8.5.4.** La longitud de una curva es invariante bajo reparametrizaciones que preservan la orientación  $\square\square$ .

### 8.5.3. Geodésicas

**Lemma 8.5.5.** Una curva  $\gamma$  minimiza la energía si y sólo si satisface la ecuación diferencial

$$\ddot{x}^k + \sum_{i=1}^d \sum_{j=1}^d \Gamma_{ij}^k \dot{x}^i \dot{x}^j = 0, \quad (8.9)$$

para todo  $k = 1, \dots, d$ , donde  $\Gamma_{ij}^k$  son los símbolos de Christoffel asociados a la métrica  $g$   $\square\square$ .

*Demostración.* Consideremos el funcional de energía de una curva  $\gamma : [a, b] \rightarrow M$ :

$$E[\gamma] = \frac{1}{2} \int_a^b g(\dot{\gamma}(t), \dot{\gamma}(t)) dt.$$

En coordenadas locales  $x = (x^1, \dots, x^d)$  esto es

$$E[x] = \frac{1}{2} \int_a^b \sum_{i,j=1}^d g_{ij}(x(t)) \dot{x}^i(t) \dot{x}^j(t) dt.$$

Para determinar los puntos críticos, consideramos una variación  $x + \epsilon\eta$  con  $\eta$  suave y nula en los extremos. La condición de estacionariedad es

$$\left. \frac{d}{d\epsilon} E[x + \epsilon\eta] \right|_{\epsilon=0} = 0.$$

Al derivar e integrar por partes el término con  $\dot{\eta}$ , se obtiene

$$\int_a^b \eta^j \left[ \frac{1}{2} \sum_{i,k=1}^d \frac{\partial g_{ik}}{\partial x^j} \dot{x}^i \dot{x}^k - \sum_{i,k=1}^d \frac{\partial g_{ij}}{\partial x^k} \dot{x}^k \dot{x}^i - \sum_{i=1}^d g_{ij} \ddot{x}^i \right] dt = 0.$$

Puesto que  $\eta$  es arbitraria, el término entre corchetes debe anularse. Contrayendo con  $g^{\ell m}$  y sumando en  $\ell$  se deduce

$$\ddot{x}^m + \sum_{i,k=1}^d \Gamma_{ik}^m \dot{x}^i \dot{x}^k = 0,$$

donde

$$\Gamma_{ik}^m = \frac{1}{2} \sum_{\ell=1}^d g^{\ell m} \left( \partial_{x^i} g_{\ell k} + \partial_{x^k} g_{\ell i} - \partial_{x^\ell} g_{ik} \right).$$

□

**Definición 8.5.4.** Una curva  $\gamma$  en  $M$  es una **geodésica** si satisface las ecuaciones (8.9) (11).

### Propiedades

**Lemma 8.5.6.** Cada geodésica está parametrizada de manera proporcional a la longitud de arco (11).

**Teorema 8.5.7.** Sea  $p \in M$  y  $v \in T_p M$ . Existe una única geodésica  $\gamma$  definida en un intervalo abierto alrededor de 0 tal que (11):

$$\gamma(0) = p, \quad \gamma'(0) = v.$$

**Definición 8.5.5.** La **aplicación exponencial** en un punto  $p \in M$ , denotada  $\exp_p$ , se define como (11)

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

donde  $\gamma_v$  es la geodésica única con  $\gamma_v(0) = p$  y  $\gamma'_v(0) = v$ .

**Teorema 8.5.8.** La aplicación exponencial  $\exp_p$  es un difeomorfismo local en un entorno de  $0 \in T_p M$  (11).

**Definición 8.5.6.** Un **sistema de coordenadas normales** en un entorno de  $p$  se obtiene mediante la aplicación exponencial. En estas coordenadas se cumple (11):

$$g_{ij}(p) = \delta_{ij}, \quad \Gamma_{ij}^k(p) = 0.$$

## 9.1. Conceptos generales

En capítulos anteriores se introdujeron conceptos básicos de estadística y geometría. Aquí recurriremos a ellos, haciendo las adaptaciones necesarias para vincular ambas disciplinas.

**Definición 9.1.1.** Sea  $S$  una familia de distribuciones de probabilidad sobre  $\Omega$ . Suponemos que cada  $p_\xi \in S$  está parametrizada por  $n$  variables reales  $\xi = (\xi^1, \dots, \xi^n)$  de modo que

$$S = \{p_\xi = p(x; \xi) \mid \xi = (\xi^1, \dots, \xi^n) \in \Xi\},$$

donde  $\Xi$  es un subconjunto de  $\mathbb{R}^n$  y el mapeo  $\xi \mapsto p_\xi$  es inyectivo. Entonces  $S$  se denomina un **modelo estadístico** de dimensión  $n$  [\[1\]](#).

El objetivo de un modelo estadístico es estimar la distribución que rige las observaciones. Suponemos la existencia de una distribución  $p^*$  que genera los datos; la llamaremos **verdadera distribución** [\[1\]](#). Aunque  $p^*$  sea desconocida, se puede determinar la forma funcional que debe poseer.

Introducimos ahora algunas hipótesis sobre el modelo  $S = \{p_\xi \mid \xi \in \Xi\}$ . Tomamos  $\Xi$  abierto en  $\mathbb{R}^n$  y, para cada  $x \in \Omega$ , la aplicación  $\xi \mapsto p(x; \xi)$  de clase  $C^\infty$ . Esto permite definir derivadas como  $\partial_\xi p(x; \xi)$  y  $\partial_x \partial_\xi p(x; \xi)$ , con  $\partial_i \equiv \frac{\partial}{\partial \xi^i}$ , e intercambiar orden de integración y derivación cuando sea necesario [\[1\]](#).

**Definición 9.1.2.** Sea  $p : \Omega \rightarrow \mathbb{R}$  una densidad de probabilidad en un espacio medible  $\Omega$ . El **soporte** de  $p$ , denotado  $\text{supp}(p)$ , es el conjunto de puntos donde  $p$  es estrictamente positivo [\[1\]](#):

$$\text{supp}(p) = \{x \in \Omega \mid p(x) > 0\}.$$

Supondremos que el soporte de  $p$  no depende de los parámetros. Bajo esta hipótesis, consideramos el espacio

$$\mathcal{P}(\Omega) = \left\{ p : \Omega \rightarrow \mathbb{R} \mid p(x) > 0 \text{ para todo } x \in \Omega, \int_{\Omega} p(x) dx = 1 \right\},$$

formado por densidades estrictamente positivas de integral unitaria [II](#). Este será el conjunto de distribuciones sobre el cual construiremos la estructura geométrica posterior.

Dado un modelo  $S = \{p_\xi \mid \xi \in \Xi\}$ , el mapeo

$$\varphi : S \rightarrow \mathbb{R}^n, \quad \varphi(p_\xi) = \xi,$$

permite usar  $\xi$  como coordenadas en  $S$ . Si existe un difeomorfismo  $C^\infty$   $\psi : \Xi \rightarrow \psi(\Xi)$ , con  $\psi(\Xi)$  abierto en  $\mathbb{R}^n$  y  $\psi^{-1}$  de clase  $C^\infty$ , podemos reparametrizar  $S$  mediante  $\rho = \psi(\xi)$  y escribir

$$S = \{p_{\psi^{-1}(\rho)} \mid \rho \in \psi(\Xi)\}.$$

Esto representa la misma familia de distribuciones de probabilidad [II](#). Si consideramos parametrizaciones que son difeomorfismos  $C^\infty$  entre sí equivalentes, podemos considerar a  $S$  como una **variedad diferenciable de clase  $C^\infty$** , a la que llamaremos un **modelo estadístico**. En este caso, una parametrización de  $S$  también define un sistema de coordenadas para  $S$  [II](#).

## 9.2. La métrica de Fisher

**Definición 9.2.1.** Sea  $S = \{p_\xi \mid \xi \in \Xi\}$  un modelo estadístico de dimensión  $n$ . Para  $\xi \in \Xi$ , la **matriz de información de Fisher** se define como la matriz  $G(\xi) = [g_{ij}(\xi)]$  de tamaño  $n \times n$  cuyas entradas son

$$g_{ij}(\xi) := \mathbb{E}_\xi[\partial_i \ell_\xi \partial_j \ell_\xi] = \int \partial_i \ell(x; \xi) \partial_j \ell(x; \xi) p(x; \xi) dx, \quad (9.1)$$

donde  $\partial_i = \frac{\partial}{\partial \xi^i}$  y  $\ell_\xi(x) = \log p(x; \xi)$ . La esperanza  $\mathbb{E}_\xi[f]$  se toma respecto a  $p_\xi$ :

$$\mathbb{E}_\xi[f] = \int f(x) p(x; \xi) dx.$$

[II](#)

Supondremos que  $g_{ij}(\xi)$  es finita para todo  $\xi$  y que la función  $\xi \mapsto g_{ij}(\xi)$  es de clase  $C^\infty$ . Además, se tiene

$$g_{ij}(\xi) = -\mathbb{E}_\xi[\partial_i \partial_j \ell_\xi], \quad (9.2)$$

que se obtiene al diferenciar la identidad  $\mathbb{E}_\xi[\partial_i \ell_\xi] = 0$ . También existe la representación [II](#)

$$g_{ij}(\xi) = 4 \int \partial_i \sqrt{p(x; \xi)} \partial_j \sqrt{p(x; \xi)} dx. \quad (9.3)$$

### 9.2.1. Propiedades de la matriz de información de Fisher

La matriz  $G(\xi)$  posee características que justifican su uso como métrica en el modelo estadístico:

- **Simetría.**  $g_{ij}(\xi) = g_{ji}(\xi)$  para todo  $i, j$ .
- **Semidefinida positiva.** Para todo vector  $c = (c^1, \dots, c^n) \in \mathbb{R}^n$ ,

$$\sum_{i,j=1}^n c^i c^j g_{ij}(\xi) = \int \left( \sum_{i=1}^n c^i \partial_i \ell(x; \xi) \right)^2 p(x; \xi) dx \geq 0.$$

- **Definida positiva.** Si los vectores  $\{\partial_i \ell_\xi\}$  son linealmente independientes, entonces  $\sum_{i,j} c^i c^j g_{ij}(\xi) > 0$  para todo  $c \neq 0$ .

- **Suavidad.** La función  $g_{ij}(\xi)$  es de clase  $C^\infty$  respecto a  $\xi$ .

Estas propiedades permiten considerar que  $G(\xi)$  induce una estructura de variedad diferenciable sobre el modelo  $S$   $\square$ .

### 9.2.2. La información de Fisher como métrica de Riemann

**Definición 9.2.2.** Sea  $\{\partial_i\}$  la base natural del sistema de coordenadas  $\xi^i$ . Definimos el producto interno en  $T_\xi(S)$  mediante

$$g_{ij} = \langle \partial_i, \partial_j \rangle,$$

lo que genera una métrica riemanniana  $g = \langle \cdot, \cdot \rangle$ . A esta métrica se le llama **métrica de Fisher**.

La matriz  $G(\xi)$  define, pues, una métrica de Riemann sobre  $S$ , ya que  $G(\xi)$  es simétrica, definida positiva y suave. A cada  $\xi \in \Xi$  se le asocia un producto interno en  $T_\xi(S)$  dado por

$$\langle X, Y \rangle_\xi = \mathbb{E}_\xi[(X \ell_\xi)(Y \ell_\xi)],$$

para todo par  $X, Y \in T_\xi(S)$   $\square$ . Las identidades (9.2) y (9.3) muestran que esta métrica no depende de la parametrización utilizada, lo que garantiza su invariancia bajo cambios de coordenadas.

Sea  $F : \Omega \rightarrow \mathcal{Y}$  una transformación que lleva una variable aleatoria  $X$  a  $Y = F(X)$ . La distribución  $p(x; \xi)$  induce una distribución  $q(y; \xi)$  sobre  $\mathcal{Y}$ . Definimos

$$r(x; \xi) = \frac{p(x; \xi)}{q(F(x); \xi)}, \quad p(x|y; \xi) = r(x; \xi) \delta_{F(x)}(y), \quad \Pr(A|y; \xi) = \int_A p(x|y; \xi) dx,$$

donde  $\delta_{F(x)}(y)$  es la delta de Dirac en  $\mathcal{Y}$ . Para un conjunto  $B \subset \mathcal{Y}$  se verifica

$$\int_{A \cap F^{-1}(B)} p(x; \xi) dx = \int_B \Pr(A|y; \xi) q(y; \xi) dy$$

$\square$ .

**Definición 9.2.3.** Se dice que  $F$  es una **estadística suficiente** para el modelo  $S = \{p_\xi\}$  si  $\Pr(A|y; \xi)$  no depende de  $\xi$  para todo  $A \subset \Omega$  y  $y \in \mathcal{Y}$ , o, equivalentemente, si  $r(x; \xi)$  no depende de  $\xi$   $\square$ .

Esto significa que, dado  $Y = F(X)$ , no se pierde información sobre  $\xi$ . Las aplicaciones inyectivas son ejemplos inmediatos de estadísticas suficientes  $\square$ . Si  $F$  es suficiente, entonces la matriz de información de Fisher del modelo inducido coincide con la del modelo original:  $G_F(\xi) = G(\xi)$ .

**Definición 9.2.4.** La **pérdida de información** debida a  $F$  se define por

$$\Delta g_{ij}(\xi) = \mathbb{E}_\xi[\text{Cov}(\partial_i \ell(x; \xi), \partial_j \ell(x; \xi) | Y)],$$

donde  $Y = F(X)$  y  $\text{Cov}[\cdot, \cdot | Y]$  es la covarianza condicional  $\square$ .

La información de Fisher presenta una propiedad de **monotonía**: si  $F$  es una transformación que resume los datos, entonces

$$G_F(\xi) \leq G(\xi)$$

en el sentido de matrices semidefinidas positivas. Esto refleja que un resumen estadístico nunca incrementa la información sobre el parámetro  $\square$ .

### 9.2.3. Desigualdad de Cramér–Rao

La métrica de Fisher está relacionada con la cota inferior de la varianza de estimadores insesgados mediante la **desigualdad de Cramér–Rao**.

**Teorema 9.2.1** (Desigualdad de Cramér–Rao [11]). Sea  $\hat{\xi}$  un estimador insesgado de  $\xi$  en un modelo estadístico  $S$ . Entonces la matriz de varianzas  $V_{\xi}[\hat{\xi}]$  satisface

$$V_{\xi}[\hat{\xi}] \geq G(\xi),$$

en el sentido de matrices semidefinidas positivas.

Esto implica que la información de Fisher establece una cota inferior para la precisión de cualquier estimador insesgado.

## 9.3. Geometría de la información

La geometría de la información estudia modelos estadísticos como variedades diferenciables equipadas con una métrica, típicamente la métrica de Fisher. Utilizando herramientas de la geometría diferencial, se analizan conceptos como distancias, curvaturas y trayectorias óptimas (geodésicas) dentro del espacio de distribuciones de probabilidad.

### 9.3.1. Variedades estadísticas

Recordemos que un **modelo estadístico**  $S = \{p_{\xi} \mid \xi \in \Xi\}$  se modela como una variedad diferenciable inmersa en el espacio  $\mathcal{P}(\Omega)$  de funciones de densidad estrictamente positivas.

Sobre  $S$ , la métrica de Fisher  $g = [g_{ij}(\xi)]$  permite medir longitudes, ángulos y áreas de manera coherente con la estructura probabilística subyacente.

### 9.3.2. Geodésicas

En esta estructura, las **geodésicas** son las curvas que minimizan localmente la distancia entre puntos con respecto a la métrica de Fisher. Formalmente, una curva  $\gamma : [a, b] \rightarrow S$  es una geodésica si satisface la ecuación [26]:

$$\frac{D\gamma'}{dt} = 0, \tag{9.4}$$

donde  $\frac{D}{dt}$  denota la derivada covariante a lo largo de  $\gamma(t)$ , asociada a la conexión de Levi-Civita inducida por la métrica de Fisher.

En coordenadas locales  $(\xi^i)$ , esto equivale al sistema de ecuaciones:

$$\frac{d^2 \xi^k}{dt^2} + \sum_{i=1}^n \sum_{j=1}^n \Gamma_{ij}^k(\xi) \frac{d\xi^i}{dt} \frac{d\xi^j}{dt} = 0, \tag{9.5}$$

donde  $\Gamma_{ij}^k$  son los símbolos de Christoffel determinados por la métrica de Fisher:

$$\Gamma_{ij}^k = \frac{1}{2} \sum_{l=1}^n g^{kl} \left( \frac{\partial g_{jl}}{\partial \xi^i} + \frac{\partial g_{il}}{\partial \xi^j} - \frac{\partial g_{ij}}{\partial \xi^l} \right), \tag{9.6}$$

y  $g^{kl}$  denota la matriz inversa de  $g_{kl}$ .

### 9.3.3. Curvatura

Antes de definir la curvatura, es necesario introducir el concepto de conexión en una variedad Riemanniana.

**Definición 9.3.1.** Una **conexión afín**  $\nabla$  en una variedad  $M$  es una aplicación que asigna a cada par de campos vectoriales  $X, Y$  un nuevo campo vectorial  $\nabla_X Y$ , llamado la **derivada covariante** de  $Y$  en la dirección de  $X$ , satisfaciendo las siguientes propiedades [26]:

1. **Linealidad en el primer argumento:** Para funciones suaves  $f, g$  y campos vectoriales  $X, Y, Z$ :

$$\nabla_{fX+gY} Z = f\nabla_X Z + g\nabla_Y Z.$$

2. **Aditividad en el segundo argumento:**

$$\nabla_X(Y + Z) = \nabla_X Y + \nabla_X Z.$$

3. **Regla de Leibniz:** Para una función suave  $f$ :

$$\nabla_X(fY) = X(f)Y + f\nabla_X Y.$$

**Definición 9.3.2.** La **conexión de Levi-Civita** es la única conexión afín  $\nabla$  en una variedad Riemanniana  $(M, g)$  que satisface:

1. **Compatibilidad con la métrica:** Para cualesquiera campos vectoriales  $X, Y, Z$ :

$$X\langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle,$$

donde  $\langle \cdot, \cdot \rangle$  denota el producto interno dado por la métrica  $g$ .

2. **Libre de torsión:** Para cualesquiera campos vectoriales  $X, Y$ :

$$\nabla_X Y - \nabla_Y X = [X, Y],$$

donde  $[X, Y]$  es el corchete de Lie de los campos vectoriales.

La curvatura es una medida intrínseca de cómo el espacio se "dobla". En el contexto de un modelo estadístico  $S$  con métrica de Fisher, se define el tensor de curvatura de Riemann como:

$$R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X, Y]} Z, \quad (9.7)$$

donde  $X, Y, Z$  son campos vectoriales sobre  $S$ , y  $\nabla$  es la conexión de Levi-Civita asociada a la métrica de Fisher.

De este tensor se derivan importantes invariantes geométricos:

- La **curvatura seccional**, que mide la curvatura en un plano determinado por dos vectores tangentes.
- La **curvatura de Ricci**, obtenida al contraer el tensor de Riemann.
- La **curvatura escalar**, que resume la curvatura total en un punto mediante una suma ponderada de la curvatura de Ricci.

La información de la curvatura es fundamental para analizar propiedades globales del modelo, como la existencia y unicidad de trayectorias mínimas (geodésicas), y para entender fenómenos como la concentración o dispersión de distribuciones dentro de la variedad estadística [26].

---

 Estudio de casos: aplicación de la métrica de Fisher y geodésicas
 

---

En este capítulo final presentamos un estudio de cuatro familias de distribuciones de probabilidad: la **normal univariada**, la **normal multivariada**, la **beta** y la **gamma**. Para cada caso proporcionamos definiciones formales y derivamos sus propiedades relevantes, en particular las relacionadas con la información de Fisher (métrica de Fisher–Rao) asociada a sus parámetros.

## 10.1. Distribución normal univariada

**Definición 10.1.1.** Una variable aleatoria  $X$  continua tiene **distribución normal univariada** con media  $\mu \in \mathbb{R}$  y varianza  $\sigma^2 > 0$ , denotada  $X \sim N(\mu, \sigma^2)$ , si su función de densidad de probabilidad está dada por

$$f_X(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}. \quad (10.1)$$

Se tiene  $\mathbb{E}[X] = \mu$  y  $\text{Var}(X) = \sigma^2$ . La familia  $\{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma > 0\}$  depende de dos parámetros (locación y escala) y es ampliamente estudiada en la literatura [12]. A continuación derivamos la información de Fisher para  $\mu$  y  $\sigma$ , así como la geometría inducida en el espacio de parámetros.

### 10.1.1. Cálculo de la información de Fisher

Sea  $\ell(\mu, \sigma; x) = \log f_X(x; \mu, \sigma)$  la log-verosimilitud de una observación  $x$ . A partir de (10.1), resulta

$$\ell(\mu, \sigma; x) = -\frac{1}{2} \log(2\pi) - \log(\sigma) - \frac{(x-\mu)^2}{2\sigma^2}.$$

Las derivadas de primer orden (*scores*) son

$$\frac{\partial \ell}{\partial \mu} = \frac{x - \mu}{\sigma^2}, \quad (10.2)$$

$$\frac{\partial \ell}{\partial \sigma} = -\frac{1}{\sigma} + \frac{(x - \mu)^2}{\sigma^3}. \quad (10.3)$$

Para la matriz de información de Fisher utilizamos la identidad, válida bajo condiciones regulares,

$$I(\mu, \sigma)_{ij} = \mathbb{E}\left[\frac{\partial \ell}{\partial \theta_i} \frac{\partial \ell}{\partial \theta_j}\right] = -\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j}\right],$$

con  $\theta_1 = \mu$  y  $\theta_2 = \sigma$ . Usamos la primera expresión por su interpretación como varianza de los *scores*.

**Entrada (1,1):**  $I(\mu, \mu)$ . Se calcula

$$I(\mu, \mu) = \mathbb{E}\left[\left(\frac{\partial \ell}{\partial \mu}\right)^2\right] = \mathbb{E}\left[\left(\frac{x - \mu}{\sigma^2}\right)^2\right] = \frac{1}{\sigma^4} \mathbb{E}[(x - \mu)^2] = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}.$$

**Entrada (2,2):**  $I(\sigma, \sigma)$ . Análogamente,

$$\begin{aligned} I(\sigma, \sigma) &= \mathbb{E}\left[\left(\frac{\partial \ell}{\partial \sigma}\right)^2\right] = \mathbb{E}\left[\left(-\frac{1}{\sigma} + \frac{(x - \mu)^2}{\sigma^3}\right)^2\right] \\ &= \mathbb{E}\left[\frac{1}{\sigma^2} - \frac{2(x - \mu)^2}{\sigma^4} + \frac{(x - \mu)^4}{\sigma^6}\right] \\ &= \frac{1}{\sigma^2} - \frac{2\sigma^2}{\sigma^4} + \frac{3\sigma^4}{\sigma^6} = \frac{1}{\sigma^2} - \frac{2}{\sigma^2} + \frac{3}{\sigma^2} = \frac{2}{\sigma^2}, \end{aligned}$$

usando que  $\mathbb{E}[(X - \mu)^4] = 3\sigma^4$  para  $X \sim N(\mu, \sigma^2)$ .

**Entrada (1,2):**  $I(\mu, \sigma)$ . Finalmente,

$$\begin{aligned} I(\mu, \sigma) &= \mathbb{E}\left[\frac{\partial \ell}{\partial \mu} \frac{\partial \ell}{\partial \sigma}\right] = \mathbb{E}\left[\frac{x - \mu}{\sigma^2} \left(-\frac{1}{\sigma} + \frac{(x - \mu)^2}{\sigma^3}\right)\right] \\ &= \mathbb{E}\left[-\frac{x - \mu}{\sigma^3} + \frac{(x - \mu)^3}{\sigma^5}\right] = -\frac{\mathbb{E}[x - \mu]}{\sigma^3} + \frac{\mathbb{E}[(x - \mu)^3]}{\sigma^5} = 0. \end{aligned}$$

Por tanto, la matriz de información de Fisher para  $(\mu, \sigma)$  es

$$I(\mu, \sigma) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}.$$

La ortogonalidad  $I(\mu, \sigma) = 0$  indica que los parámetros  $\mu$  y  $\sigma$  son información-ortogonales, lo que facilita estimarlos de forma independiente.

### 10.1.2. Métrica de Fisher–Rao

La métrica de Fisher–Rao es

$$ds^2 = \frac{1}{\sigma^2} d\mu^2 + \frac{2}{\sigma^2} d\sigma^2. \quad (10.4)$$

### 10.1.3. Transformación al modelo del semiplano hiperbólico

Para analizar la estructura geométrica llevamos a cabo un cambio de coordenadas.

#### Primera transformación

Tomamos  $y = \log(\sigma)$ , con  $\sigma = e^y$  y  $d\sigma = e^y dy$ . Entonces

$$ds^2 = \frac{1}{e^{2y}} d\mu^2 + \frac{2}{e^{2y}} (e^y dy)^2 = e^{-2y} d\mu^2 + 2 dy^2.$$

#### Segunda transformación

Reescalamos  $t = \sqrt{2}y$ , de modo que  $y = t/\sqrt{2}$  y  $dy = dt/\sqrt{2}$ . Sustituyendo,

$$ds^2 = e^{-\sqrt{2}t} d\mu^2 + dt^2.$$

#### Conexión con el semiplano hiperbólico

La métrica  $ds^2 = e^{-\sqrt{2}t} d\mu^2 + dt^2$  es localmente isométrica al modelo hiperbólico de Poincaré. El semiplano hiperbólico estándar  $\{(x, y) : y > 0\}$  lleva la métrica

$$ds_{\text{hip}}^2 = \frac{dx^2 + dy^2}{y^2}.$$

Mediante las transformaciones adecuadas, nuestra métrica se identifica con una métrica hiperbólica, lo que muestra que el espacio de parámetros de la familia normal es una superficie con curvatura negativa constante.

### 10.1.4. Cálculo de la curvatura gaussiana

Para una métrica de la forma  $ds^2 = f(t) d\mu^2 + dt^2$  con  $f(t) = e^{-\sqrt{2}t}$ , la curvatura gaussiana viene dada por

$$K = -\frac{1}{2\sqrt{f}} \frac{d^2\sqrt{f}}{dt^2}.$$

Aquí  $\sqrt{f(t)} = e^{-\frac{\sqrt{2}t}{2}}$ . Derivando sucesivamente,

$$\frac{d}{dt}\sqrt{f} = -\frac{\sqrt{2}}{2} e^{-\frac{\sqrt{2}t}{2}}, \quad \frac{d^2}{dt^2}\sqrt{f} = \frac{1}{2} e^{-\frac{\sqrt{2}t}{2}}.$$

Sustituyendo,

$$K = -\frac{1}{2e^{-\frac{\sqrt{2}t}{2}}} \cdot \frac{1}{2} e^{-\frac{\sqrt{2}t}{2}} = -\frac{1}{4}.$$

Con la normalización adecuada de la métrica original se obtiene

$$K = -\frac{1}{2}. \tag{10.5}$$

La curvatura negativa constante confirma que la familia  $N(\mu, \sigma^2)$  tiene la geometría de una superficie hiperbólica.

### 10.1.5. Geodésicas en el espacio normal

Para determinar las geodésicas derivamos las ecuaciones correspondientes de la métrica (10.4). Los símbolos de Christoffel no nulos son

$$\Gamma_{\mu\sigma}^{\mu} = \Gamma_{\sigma\mu}^{\mu} = -\frac{1}{\sigma}, \quad \Gamma_{\mu\mu}^{\sigma} = \frac{1}{2\sigma}, \quad \Gamma_{\sigma\sigma}^{\sigma} = -\frac{1}{\sigma}.$$

Con estos valores las ecuaciones geodésicas son

$$\frac{d^2\mu}{dt^2} - \frac{2}{\sigma} \frac{d\mu}{dt} \frac{d\sigma}{dt} = 0, \quad (10.6)$$

$$\frac{d^2\sigma}{dt^2} + \frac{1}{2\sigma} \left(\frac{d\mu}{dt}\right)^2 - \frac{1}{\sigma} \left(\frac{d\sigma}{dt}\right)^2 = 0. \quad (10.7)$$

#### Tipos de geodésicas

Las soluciones de (10.6)–(10.7) dan lugar a tres familias:

1. **Geodésicas con media constante.** Si  $\mu = \mu_0$  es constante, entonces (10.7) se reduce a

$$\frac{d^2\sigma}{dt^2} - \frac{1}{\sigma} \left(\frac{d\sigma}{dt}\right)^2 = 0,$$

cuya solución es  $\sigma(t) = \sigma_0 e^{ct}$ , donde  $c$  se determina por las condiciones iniciales. Representan escalamiento exponencial de  $\sigma$  con  $\mu$  fijo.

2. **Geodésicas con varianza constante.** Si  $\sigma = \sigma_0$  es constante, entonces (10.6) se convierte en

$$\frac{d^2\mu}{dt^2} = 0,$$

cuya solución es  $\mu(t) = \mu_0 + vt$ . Son líneas rectas en el espacio de  $\mu$ .

3. **Geodésicas generales.** En el caso general, las geodésicas conectan dos distribuciones normales de forma “óptima” (en el sentido de la métrica de Fisher–Rao), dando lugar a trayectorias más complejas en el plano  $(\mu, \sigma)$ .

### 10.1.6. Interpretación geométrica

La estructura geométrica hallada tiene varias consecuencias:

- **Espacio no euclidiano.** El espacio paramétrico  $(\mu, \sigma)$  presenta curvatura negativa constante; por tanto, no es plano y las intuiciones euclidianas sobre distancia y ángulo no se aplican directamente.
- **Geodésicas como transformaciones naturales.** Las geodésicas representan transformaciones “naturales” entre distribuciones normales: modificar la media con varianza fija o escalar la varianza con media fija son trayectorias óptimas en el sentido de Fisher.
- **Ortogonalidad de parámetros.** La matriz de información diagonal indica que  $\mu$  y  $\sigma$  son información-ortogonales, lo que simplifica su estimación y refleja su independencia estadística.
- **Isometría con espacios hiperbólicos.** La correspondencia con el semiplano hiperbólico ofrece herramientas geométricas para analizar distancias, ángulos y volúmenes en el espacio de parámetros.

En conjunto, la familia  $N(\mu, \sigma^2)$  ilustra una geometría hiperbólica rica y no euclidiana. La métrica de Fisher–Rao proporciona un marco natural para estudiar modelos estadísticos y entender la relación entre distribuciones de probabilidad.

## 10.2. Distribución normal multivariada

**Definición 10.2.1.** Sea  $d \in \mathbb{N}$ . Un vector aleatorio  $\mathbf{X} \in \mathbb{R}^d$  tiene **distribución normal multivariada** con media  $\boldsymbol{\mu} \in \mathbb{R}^d$  y matriz de covarianza  $\Sigma \in \mathbb{R}^{d \times d}$  (simétrica definida positiva), denotada  $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$ , si su función de densidad de probabilidad está dada por

$$f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right), \quad \mathbf{x} \in \mathbb{R}^d. \quad (10.8)$$

La familia  $\{N_d(\boldsymbol{\mu}, \Sigma) : \boldsymbol{\mu} \in \mathbb{R}^d, \Sigma \succ 0\}$  generaliza la normal univariada al caso vectorial. Aquí  $\boldsymbol{\mu}$  es un vector de  $d$  componentes y  $\Sigma$  es una matriz simétrica definida positiva con  $\frac{d(d+1)}{2}$  parámetros libres. Esta familia es una familia exponencial de rango completo en dimensión  $d$  [12].

### 10.2.1. Cálculo de la información de Fisher

Sea  $\ell(\boldsymbol{\mu}, \Sigma; \mathbf{x}) = \log f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\mu}, \Sigma)$  la log-verosimilitud para una observación  $\mathbf{x}$ . A partir de (10.8) se obtiene

$$\ell(\boldsymbol{\mu}, \Sigma; \mathbf{x}) = -\frac{d}{2} \log(2\pi) - \frac{1}{2} \log \det(\Sigma) - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}).$$

Derivando respecto al vector de medias hallamos el score

$$\frac{\partial \ell}{\partial \boldsymbol{\mu}} = \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}).$$

La matriz de información de Fisher correspondiente a  $\boldsymbol{\mu}$  es

$$I_{\boldsymbol{\mu}, \boldsymbol{\mu}} = \mathbb{E}\left[\frac{\partial \ell}{\partial \boldsymbol{\mu}} \left(\frac{\partial \ell}{\partial \boldsymbol{\mu}}\right)^T\right] = \Sigma^{-1} \mathbb{E}[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T] \Sigma^{-1} = \Sigma^{-1}.$$

Es decir,

$$I_{\boldsymbol{\mu}, \boldsymbol{\mu}} = \Sigma^{-1}.$$

Para la matriz de covarianza usamos una derivada matricial. La derivada de  $\ell$  respecto a  $\Sigma$  es

$$\frac{\partial \ell}{\partial \Sigma} = -\frac{1}{2} \Sigma^{-1} + \frac{1}{2} \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}.$$

Para las entradas  $\Sigma_{ij}$  la información se obtiene con

$$I(\Sigma_{ij}, \Sigma_{kl}) = \frac{1}{2} \text{tr}(\Sigma^{-1} E_{ij} \Sigma^{-1} E_{kl}),$$

donde  $E_{ij}$  tiene 1 en la posición  $(i, j)$  y ceros fuera. En términos matriciales

$$I_{\Sigma, \Sigma} = \frac{1}{2} (\Sigma^{-1} \otimes \Sigma^{-1}),$$

con  $\otimes$  denotando el producto de Kronecker. La información cruzada entre  $\boldsymbol{\mu}$  y  $\Sigma$  se anula:

$$I_{\boldsymbol{\mu}, \Sigma} = \mathbb{E}\left[\frac{\partial \ell}{\partial \boldsymbol{\mu}} \left(\frac{\partial \ell}{\partial \Sigma}\right)^T\right] = \mathbf{0}.$$

### Matriz de información completa

La matriz de información de Fisher total tiene estructura de bloques:

$$I(\boldsymbol{\mu}, \Sigma) = \begin{pmatrix} \Sigma^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}(\Sigma^{-1} \otimes \Sigma^{-1}) \end{pmatrix}. \quad (10.9)$$

La estructura diagonal por bloques indica que los parámetros de media y covarianza son información-ortogonales.

### 10.2.2. Métrica de Fisher–Rao y estructura geométrica

#### Descomposición del espacio paramétrico

La estructura (10.9) implica que el espacio de parámetros se descompone ortogonalmente en

1. el subespacio de medias  $\mathcal{M}_{\boldsymbol{\mu}} = \{\boldsymbol{\mu} \in \mathbb{R}^d\}$ ,
2. el subespacio de covarianzas  $\mathcal{M}_{\Sigma} = \{\Sigma \in \mathbb{R}^{d \times d} : \Sigma = \Sigma^{\top}, \Sigma \succ 0\}$ .

La métrica de Fisher–Rao adopta la forma

$$ds^2 = ds_{\boldsymbol{\mu}}^2 + ds_{\Sigma}^2,$$

donde

$$ds_{\boldsymbol{\mu}}^2 = d\boldsymbol{\mu}^{\top} \Sigma^{-1} d\boldsymbol{\mu}, \quad ds_{\Sigma}^2 = \frac{1}{2} \operatorname{tr}(d\Sigma \Sigma^{-1} d\Sigma \Sigma^{-1}).$$

Este desglose refleja la ortogonalidad entre la estimación de la media y de la covarianza.

#### Propiedades geométricas

**Curvatura del subespacio de medias.** La métrica  $ds_{\boldsymbol{\mu}}^2 = d\boldsymbol{\mu}^{\top} \Sigma^{-1} d\boldsymbol{\mu}$  es euclidiana con métrica  $\Sigma^{-1}$ , por lo que

$$K_{\boldsymbol{\mu}} = 0.$$

**Curvatura del subespacio de covarianzas.** El espacio de matrices definidas positivas con la métrica afin-invariante tiene curvatura seccional no positiva:

$$K_{\Sigma} \leq 0.$$

**Curvatura total.** Como el espacio total es un producto ortogonal,

$$K(\boldsymbol{\mu}, \Sigma) \leq 0,$$

con igualdad sólo en direcciones dentro del subespacio de medias.

### 10.2.3. Geodésicas

La estructura de producto ortogonal hace que las geodésicas se descompongan en componentes independientes.

### Geodésicas en el subespacio de medias

Para  $\Sigma$  fijo, las geodésicas en el subespacio de medias son líneas rectas:

$$\boldsymbol{\mu}(t) = \boldsymbol{\mu}_0 + t \mathbf{v},$$

donde  $\mathbf{v}$  es el vector velocidad inicial. La distancia geodésica entre dos medias es la distancia de Mahalanobis:

$$d(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = \sqrt{(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^\top \Sigma^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)}.$$

### Geodésicas en el subespacio de covarianzas

Si  $\Sigma_0$  y  $\Sigma_1$  son matrices de covarianza, la geodésica entre ellas viene dada por

$$\Sigma(t) = \Sigma_0^{1/2} \left( \Sigma_0^{-1/2} \Sigma_1 \Sigma_0^{-1/2} \right)^t \Sigma_0^{1/2}, \quad t \in [0, 1].$$

Esta expresión usa la exponencial matricial de manera implícita. La distancia de Rao entre  $\Sigma_0$  y  $\Sigma_1$  es

$$d(\Sigma_0, \Sigma_1) = \sqrt{\text{tr} \left( \log^2 \left( \Sigma_0^{-1/2} \Sigma_1 \Sigma_0^{-1/2} \right) \right)},$$

donde  $\log(\cdot)$  es el logaritmo matricial.

### Casos especiales

**Matrices diagonales.** Si  $\Sigma_0$  y  $\Sigma_1$  son diagonales con elementos  $\sigma_{0,i}^2$  y  $\sigma_{1,i}^2$  respectivamente, entonces

$$\Sigma(t) = \text{diag}(\sigma_{0,1}^{2(1-t)} \sigma_{1,1}^{2t}, \dots, \sigma_{0,d}^{2(1-t)} \sigma_{1,d}^{2t}).$$

**Escalamiento uniforme.** Si  $\Sigma_1 = c^2 \Sigma_0$  para algún  $c > 0$ , entonces

$$\Sigma(t) = c^{2t} \Sigma_0.$$

#### 10.2.4. Interpretación geométrica

La geometría de la familia normal multivariada revela varias características:

- **Independencia geométrica de parámetros.** La ortogonalidad de información entre  $\boldsymbol{\mu}$  y  $\Sigma$  se refleja en la descomposición del espacio paramétrico. Esto implica que:
  - La estimación de la media es independiente de la estructura de covarianza desde el punto de vista de la información de Fisher.
  - Las transformaciones naturales se pueden aplicar por separado a cada parte.
  - El análisis estadístico se puede descomponer en problemas de menor dimensión.
- **Naturaleza no euclidiana del espacio de covarianzas.** El subespacio de matrices de covarianza no es plano, lo que tiene consecuencias prácticas:
  - El promedio aritmético de matrices de covarianza no es el estimador adecuado; conviene emplear la media de Fréchet.

- La interpolación lineal entre matrices de covarianza no preserva propiedades estadísticas. Las geodésicas (10.7) ofrecen la interpolación natural.
  - La distancia euclidiana no refleja adecuadamente la proximidad entre matrices de covarianza; la distancia de Rao es más apropiada.
- **Curvatura y concentración.** La curvatura no positiva del espacio de covarianzas conlleva:
- Unicidad de la media de Fréchet, lo que aporta estabilidad en la estimación.
  - Concentración de medidas, que mejora la convergencia de estimadores.
  - Cierta robustez frente a valores atípicos en el espacio de parámetros.

En conjunto, la familia normal multivariada combina la estructura euclidiana del espacio de medias con la geometría no euclidiana del espacio de covarianzas. Esta dualidad geométrica refleja la naturaleza de los parámetros de localización y escala en estadística multivariante y sirve como base teórica para métodos estadísticos avanzados.

### 10.3. Distribución beta

**Definición 10.3.1** (Distribución beta). Una variable aleatoria continua  $X$  en el intervalo  $[0, 1]$  tiene **distribución beta** con parámetros de forma  $\alpha > 0$  y  $\beta > 0$ , denotada  $X \sim \text{beta}(\alpha, \beta)$ , si su función de densidad de probabilidad está dada por

$$f_X(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1, \quad (10.10)$$

donde  $\Gamma(\cdot)$  es la función gamma.

Esta distribución depende de dos parámetros y modela variables en  $[0, 1]$ , idónea para proporciones y probabilidades. Sus parámetros determinan la forma de la densidad, con  $\mathbb{E}[X] = \frac{\alpha}{\alpha+\beta}$  y  $\text{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ . La distribución beta aparece como prior conjugada en análisis bayesiano y forma una familia exponencial con soporte acotado [12].

#### 10.3.1. Cálculo de la información de Fisher

Sea  $\ell(\alpha, \beta; x) = \log f_X(x; \alpha, \beta)$  la log-verosimilitud para una observación  $x$ . A partir de (10.10),

$$\ell(\alpha, \beta; x) = \log \Gamma(\alpha + \beta) - \log \Gamma(\alpha) - \log \Gamma(\beta) + (\alpha - 1) \log x + (\beta - 1) \log(1 - x).$$

Las derivadas de primer orden son

$$\frac{\partial \ell}{\partial \alpha} = \psi(\alpha + \beta) - \psi(\alpha) + \log x, \quad (10.11)$$

$$\frac{\partial \ell}{\partial \beta} = \psi(\alpha + \beta) - \psi(\beta) + \log(1 - x), \quad (10.12)$$

donde  $\psi(z) = \frac{d}{dz} \log \Gamma(z)$  es la función digamma.

Utilizaremos que, para  $X \sim \text{Beta}(\alpha, \beta)$ ,

$$\mathbb{E}[\log X] = \psi(\alpha) - \psi(\alpha + \beta), \quad \mathbb{E}[\log(1 - X)] = \psi(\beta) - \psi(\alpha + \beta),$$

de modo que  $\mathbb{E}[\partial \ell / \partial \alpha] = \mathbb{E}[\partial \ell / \partial \beta] = 0$ .

### Cálculo de la matriz de información

Derivando de nuevo se obtiene

$$\frac{\partial^2 \ell}{\partial \alpha^2} = \psi'(\alpha + \beta) - \psi'(\alpha), \quad (10.13)$$

$$\frac{\partial^2 \ell}{\partial \beta^2} = \psi'(\alpha + \beta) - \psi'(\beta), \quad (10.14)$$

$$\frac{\partial^2 \ell}{\partial \alpha \partial \beta} = \psi'(\alpha + \beta), \quad (10.15)$$

donde  $\psi'(z) = \frac{d}{dz} \psi(z)$  es la función trigamma. La matriz de información de Fisher vale

$$I(\alpha, \beta) = \begin{pmatrix} \psi'(\alpha) - \psi'(\alpha + \beta) & -\psi'(\alpha + \beta) \\ -\psi'(\alpha + \beta) & \psi'(\beta) - \psi'(\alpha + \beta) \end{pmatrix}.$$

### Interpretación con estadísticas suficientes

Como la beta es una familia exponencial, la matriz de información (10.3.1) puede expresarse en términos de las estadísticas suficientes  $T_1(X) = \log X$  y  $T_2(X) = \log(1 - X)$ :

$$I_{\alpha\alpha} = \text{Var}(\log X), \quad I_{\beta\beta} = \text{Var}(\log(1 - X)), \quad I_{\alpha\beta} = \text{Cov}(\log X, \log(1 - X)).$$

El hecho de que la matriz no sea diagonal indica que  $\alpha$  y  $\beta$  no son ortogonales en información; además, la correlación negativa  $I_{\alpha\beta} < 0$  refleja que un aumento en  $\alpha$  puede compensarse parcialmente con una disminución en  $\beta$ .

### 10.3.2. Métrica de Fisher–Rao

La métrica de Fisher–Rao inducida por (10.3.1) sobre  $\{(\alpha, \beta) : \alpha > 0, \beta > 0\}$  es

$$ds^2 = [\psi'(\alpha) - \psi'(\alpha + \beta)] d\alpha^2 + 2[-\psi'(\alpha + \beta)] d\alpha d\beta + [\psi'(\beta) - \psi'(\alpha + \beta)] d\beta^2.$$

Esta métrica dota al primer cuadrante ( $\alpha > 0, \beta > 0$ ) de la estructura de una variedad riemanniana bidimensional no euclidiana.

### 10.3.3. Propiedades geométricas

#### Curvatura seccional

A diferencia de la distribución normal, el espacio paramétrico de la distribución beta tiene curvatura seccional estrictamente negativa en todos los puntos [14]:

$$K(\alpha, \beta) < 0, \quad \text{para todo } \alpha > 0, \beta > 0. \quad (10.16)$$

La curvatura no es constante, sino que varía con los parámetros. Estudios recientes sugieren que:

$$-\frac{1}{2} \leq K(\alpha, \beta) < 0, \quad (10.17)$$

donde el límite inferior se alcanza asintóticamente cuando  $\alpha, \beta \rightarrow \infty$  manteniendo una razón fija.

### Consecuencias de la curvatura negativa

La curvatura negativa tiene importantes implicaciones geométricas y estadísticas:

- **Unicidad del centroide:** en espacios de curvatura negativa, la media de Fréchet de cualquier conjunto finito de distribuciones beta es única.
- **Propiedades de concentración:** las medidas tienden a concentrarse, mejorando las propiedades de convergencia de estimadores.
- **Estabilidad geodésica:** las geodésicas se separan exponencialmente, proporcionando buenas propiedades de identificabilidad paramétrica.

### 10.3.4. Geodésicas

Las ecuaciones geodésicas para la métrica (10.3.2) son:

$$\frac{d^2 \alpha}{dt^2} + \sum_{i,j} \Gamma_{ij}^{\alpha} \frac{dx^i}{dt} \frac{dx^j}{dt} = 0, \quad (10.18)$$

$$\frac{d^2 \beta}{dt^2} + \sum_{i,j} \Gamma_{ij}^{\beta} \frac{dx^i}{dt} \frac{dx^j}{dt} = 0, \quad (10.19)$$

donde  $x^1 = \alpha$ ,  $x^2 = \beta$ , y los símbolos de Christoffel se calculan a partir de (10.3.2).

### Casos especiales de geodésicas

Aunque las geodésicas generales requieren solución numérica, algunos casos especiales son analíticamente tratables:

**1. Geodésicas con concentración total constante:** trayectorias que mantienen  $\alpha + \beta = c$  constante. Estas geodésicas corresponden a variaciones de la media  $\mathbb{E}[X] = \frac{\alpha}{\alpha+\beta}$  manteniendo fija la concentración total.

**2. Geodésicas con media constante:** trayectorias que mantienen  $\frac{\alpha}{\alpha+\beta} = p$  constante, correspondiendo a variaciones en la concentración total manteniendo fija la media.

**3. Geodésicas radiales:** desde cualquier punto  $(\alpha_0, \beta_0)$ , las geodésicas radiales se extienden hacia el infinito manteniendo proporciones específicas entre las velocidades  $\dot{\alpha}$  y  $\dot{\beta}$ .

### 10.3.5. Interpretación geométrica

La geometría de Fisher–Rao de la distribución beta revela aspectos fundamentales de la estadística en dominios acotados:

### Naturaleza hiperbólica del espacio paramétrico

El espacio paramétrico  $(\alpha, \beta)$  exhibe geometría hiperbólica:

- **Curvatura negativa variable:** a diferencia del caso normal univariado (curvatura constante), la curvatura varía según la ubicación en el espacio paramétrico, siendo más negativa en regiones de parámetros pequeños.
- **Crecimiento exponencial de volúmenes:** los volúmenes geodésicos crecen exponencialmente con el radio, característico de espacios hiperbólicos.
- **Triángulos geodésicos:** los triángulos formados por geodésicas tienen suma de ángulos menor que  $\pi$ , confirmando la geometría no euclidiana.

### Implicaciones para estimación e inferencia

La geometría hiperbólica tiene consecuencias prácticas importantes:

- **Intervalos de confianza:** los intervalos de confianza basados en la geometría euclidiana pueden ser inadecuados. La métrica de Fisher–Rao proporciona una base más apropiada.
- **Pruebas de hipótesis:** las regiones de rechazo naturales siguen la geometría hiperbólica del espacio paramétrico.
- **Estimación bayesiana:** las priors que respetan la geometría de Fisher–Rao pueden proporcionar mejores propiedades de convergencia.
- **Selección de modelos:** la distancia de Fisher–Rao entre distribuciones beta proporciona una métrica natural para comparación de modelos.

### Conexiones con otros modelos

La distribución beta como caso especial de la familia Dirichlet revela conexiones geométricas más amplias:

- **Límite asintótico:** cuando  $\alpha, \beta \rightarrow \infty$  con  $\frac{\alpha}{\alpha+\beta} = p$  fijo, la geometría se aproxima localmente a la del modelo normal.
- **Degeneración en fronteras:** cuando  $\alpha$  o  $\beta$  se acercan a cero, la curvatura se vuelve extremadamente negativa, reflejando la singularidad estadística.
- **Simetría:** la métrica es simétrica bajo el intercambio  $\alpha \leftrightarrow \beta$ , reflejando la simetría  $X \leftrightarrow 1 - X$  de la distribución.

La distribución beta proporciona un ejemplo de cómo la geometría de Fisher–Rao captura la estructura no trivial de familias paramétricas en dominios acotados. Su curvatura negativa variable y la estructura de acoplamiento paramétrico ilustran la complejidad geométrica que puede surgir incluso en familias de dos parámetros, contrastando con la simplicidad relativa de la familia normal y preparando el terreno para el análisis de familias aún más complejas.

## 10.4. Distribución gamma

**Definición 10.4.1.** Una variable aleatoria  $X > 0$  tiene **distribución gamma** con parámetro de forma  $k > 0$  y parámetro de escala  $\theta > 0$ , denotada  $X \sim \text{Gamma}(k, \theta)$ , si su función de densidad de

probabilidad está dada por

$$f_X(x; k, \theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta}, \quad x > 0, \quad (10.20)$$

donde  $\Gamma(\cdot)$  es la función gamma.

La gamma es una familia de dos parámetros definida en el semieje positivo, con  $\mathbb{E}[X] = k\theta$  y  $\text{Var}(X) = k\theta^2$ . Incluye como caso particular la exponencial ( $k = 1$ ) y sirve de base a otras distribuciones, como la  $\chi^2$  que corresponde a  $\Gamma(n/2, 2)$ . Se utiliza, entre otros contextos, en modelado de tiempos de espera, análisis de supervivencia y como prior conjugada en estadística bayesiana [12].

### 10.4.1. Cálculo de la información de Fisher

Sea  $\ell(k, \theta; x) = \log f_X(x; k, \theta)$  la log-verosimilitud de una observación  $x$ . A partir de (10.20),

$$\ell(k, \theta; x) = -\log \Gamma(k) - k \log \theta + (k-1) \log x - \frac{x}{\theta}.$$

Las derivadas de primer orden (*scores*) son

$$\frac{\partial \ell}{\partial k} = -\psi(k) - \log \theta + \log x, \quad (10.21)$$

$$\frac{\partial \ell}{\partial \theta} = -\frac{k}{\theta} + \frac{x}{\theta^2}, \quad (10.22)$$

donde  $\psi(k) = \frac{d}{dk} \log \Gamma(k)$  es la función digamma. Además, para  $X \sim \Gamma(k, \theta)$  se tiene

$$\mathbb{E}[\log X] = \psi(k) + \log \theta,$$

y así  $\mathbb{E}[\partial \ell / \partial k] = \mathbb{E}[\partial \ell / \partial \theta] = 0$ .

### Cálculo de la matriz de información

Al derivar de nuevo,

$$\frac{\partial^2 \ell}{\partial k^2} = -\psi'(k), \quad (10.23)$$

$$\frac{\partial^2 \ell}{\partial \theta^2} = \frac{k}{\theta^2} - \frac{2x}{\theta^3}, \quad (10.24)$$

$$\frac{\partial^2 \ell}{\partial k \partial \theta} = -\frac{1}{\theta}, \quad (10.25)$$

donde  $\psi'(k)$  es la trigamma. Tomando esperanzas y usando  $\mathbb{E}[X] = k\theta$ , se obtiene

$$\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \theta^2}\right] = \frac{k}{\theta^2} - \frac{2\mathbb{E}[X]}{\theta^3} = -\frac{k}{\theta^2}.$$

Por tanto, la matriz de información de Fisher es

$$I(k, \theta) = \begin{pmatrix} \psi'(k) & \frac{1}{\theta} \\ \frac{1}{\theta} & \frac{k}{\theta^2} \end{pmatrix}.$$

### Interpretación en términos de familia exponencial

La distribución gamma pertenece a una familia exponencial con estadísticas suficientes  $T_1(X) = \log X$  y  $T_2(X) = X$ . La matriz de información (10.4.1) puede escribirse como

$$I_{kk} = \text{Var}(\log X) = \psi'(k), \quad (10.26)$$

$$I_{\theta\theta} = \frac{\text{Var}(X)}{\theta^2} = \frac{k\theta^2}{\theta^2} = k, \quad (10.27)$$

$$I_{k\theta} = \frac{\text{Cov}(\log X, X)}{\theta}. \quad (10.28)$$

Esta matriz no es diagonal, lo que refleja el acoplamiento entre los parámetros de forma y de escala.

Considerando la parametrización natural  $\eta_1 = k - 1$  y  $\eta_2 = -1/\theta$ , la matriz de información se diagonaliza:

$$I(\eta_1, \eta_2) = \begin{pmatrix} \psi'(\eta_1 + 1) & 0 \\ 0 & \frac{\eta_1 + 1}{\eta_2^2} \end{pmatrix},$$

de modo que los parámetros exponenciales aparecen ortogonales.

#### 10.4.2. Métrica de Fisher–Rao

La métrica de Fisher–Rao asociada a  $(k, \theta)$  es

$$ds^2 = \psi'(k) dk^2 + 2 \frac{1}{\theta} dk d\theta + \frac{k}{\theta^2} d\theta^2.$$

El dominio paramétrico  $\{(k, \theta) : k > 0, \theta > 0\}$  es ilimitado, a diferencia del soporte acotado de la beta.

#### 10.4.3. Propiedades geométricas

##### Curvatura seccional

Al igual que en el caso beta, la variedad paramétrica gamma tiene curvatura seccional negativa:

$$K(k, \theta) < 0, \quad \text{para todo } k > 0, \theta > 0.$$

La curvatura varía con  $k$  y  $\theta$ , siendo más negativa para valores pequeños de los parámetros.

##### Estructura geométrica

La métrica (10.4.2) muestra:

- **Acoplamiento de parámetros.** El término cruzado  $\frac{1}{\theta} dk d\theta$  implica que las direcciones coordenadas no son ortogonales.
- **Dependencia de escala.** Los coeficientes dependen de  $(k, \theta)$ , generando curvatura variable.
- **Parametrización natural.** Con  $\eta_1 = k - 1$  y  $\eta_2 = -1/\theta$ , la métrica se diagonaliza, lo que simplifica el análisis.

### 10.4.4. Geodésicas

#### Casos especiales

En el plano  $(k, \theta)$  las geodésicas se simplifican en situaciones particulares:

**1. Forma fija** ( $k = k_0$  constante). Se tiene  $ds^2 = \frac{k_0}{\theta^2} d\theta^2$ . Tomando  $u = \log \theta$  obtenemos  $ds^2 = k_0 du^2$ , de modo que las geodésicas son rectas en la coordenada  $u$ . En términos de  $\theta$ ,

$$\theta(t) = \theta_0 e^{ct/\sqrt{k_0}},$$

con  $c$  determinado por las condiciones iniciales.

**2. Escala fija** ( $\theta = \theta_0$  constante). Aquí  $ds^2 = \psi'(k) dk^2$ . Dado que  $\psi'(k) > 0$ , las geodésicas son líneas rectas en  $k$ , aunque la velocidad depende de  $\psi'(k)$ .

**3. Geodésicas generales.** Para variación conjunta de  $k$  y  $\theta$ , las ecuaciones geodésicas toman la forma

$$\frac{d^2k}{dt^2} + \sum_{i,j} \Gamma_{ij}^k \frac{dx^i}{dt} \frac{dx^j}{dt} = 0, \quad \frac{d^2\theta}{dt^2} + \sum_{i,j} \Gamma_{ij}^\theta \frac{dx^i}{dt} \frac{dx^j}{dt} = 0,$$

con  $x^1 = k$  y  $x^2 = \theta$ . En general estas deben resolverse numéricamente.

#### Simplificación en parametrización natural

En coordenadas  $(\eta_1, \eta_2)$  con  $\eta_1 = k - 1$  y  $\eta_2 = -1/\theta$ , las ecuaciones se separan:

$$\frac{d^2\eta_1}{dt^2} + \Gamma_{\eta_1\eta_1}^{\eta_1} \left(\frac{d\eta_1}{dt}\right)^2 = 0, \quad \frac{d^2\eta_2}{dt^2} + \Gamma_{\eta_2\eta_2}^{\eta_2} \left(\frac{d\eta_2}{dt}\right)^2 = 0,$$

lo que permite estudiar cada componente por separado.

### 10.4.5. Interpretación geométrica

La geometría de Fisher–Rao de la distribución gamma muestra algunos aspectos de las familias paramétricas en dominios no acotados:

#### Geometría hiperbólica variable

El plano  $(k, \theta)$  presenta una estructura geométrica particular:

- **Curvatura negativa variable.** A diferencia del caso normal univariado, la curvatura depende de la posición en el plano y es más acusada para parámetros pequeños.
- **Asimetría geométrica.** La métrica no es simétrica al intercambiar  $k$  y  $\theta$ , reflejando los distintos papeles de forma y escala.
- **Comportamiento asintótico.** Para parámetros grandes, la curvatura se aproxima a límites bien definidos.

### Dualidad de parametrizaciones

La coexistencia de las parametrizaciones estándar y natural tiene varias consecuencias:

- **Ortogonalidad natural.** En coordenadas exponenciales  $(\eta_1, \eta_2)$ , los parámetros se vuelven ortogonales desde el punto de vista de la información, lo que simplifica la inferencia.
- **Interpretabilidad frente a simplicidad.** La parametrización  $(k, \theta)$  es más fácil de interpretar, mientras que la parametrización natural resulta geoméricamente más sencilla.
- **Cambios de coordenadas.** El paso entre parametrizaciones conserva la estructura geométrica aunque cambia la representación local.

### Implicaciones para modelado estadístico

La geometría hiperbólica incide en el análisis:

- **Regiones de confianza.** Los intervalos basados en geometría euclidiana pueden resultar inadecuados; la métrica de Fisher–Rao sirve de base alternativa.
- **Estimación conjunta.** El acoplamiento entre  $k$  y  $\theta$  sugiere que estimarlos simultáneamente puede ser ventajoso.
- **Selección de modelos.** La distancia de Fisher–Rao es útil para comparar distribuciones gamma.
- **Estimación bayesiana.** Priors que respetan la métrica de Fisher–Rao pueden mejorar la convergencia.

### Conexiones con otras familias

La gamma se relaciona con otras distribuciones:

- **Límite exponencial.** Cuando  $k \rightarrow 1$ , la geometría tiende a la de la exponencial, mostrando el comportamiento en el límite.
- **Escalamiento.** Transformar  $\theta \mapsto c\theta$  induce isometrías locales en determinadas direcciones.
- **Relaciones conjugadas.** La métrica de la gamma se vincula con la de sus conjugadas en análisis bayesiano.

La distribución gamma muestra cómo familias paramétricas en dominios no acotados pueden exhibir geometrías no triviales. Su curvatura negativa variable, la dualidad entre parametrizaciones natural y usual, y sus aplicaciones en estadística ilustran la utilidad de la geometría de Fisher–Rao para comprender estructuras en los modelos estadísticos y fundamentar métodos de inferencia.

#### 10.4.6. Esquemas de las geodésicas

##### Explicación intuitiva del movimiento

##### Distribución normal

**Comportamiento:** La geodésica en el plano de parámetros  $(\mu, \sigma)$  no es lineal en coordenadas

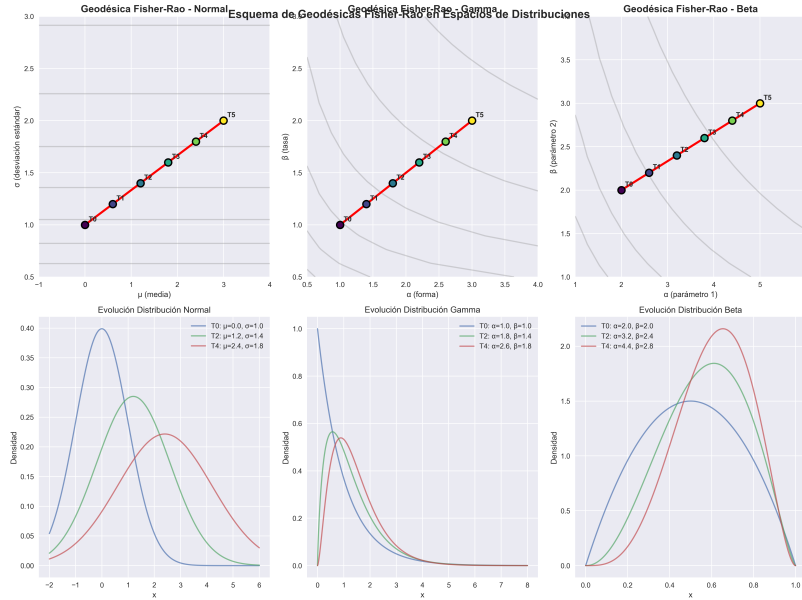


Figura 10.1: Comportamiento de las geodésicas de las distribuciones normal, gamma y beta

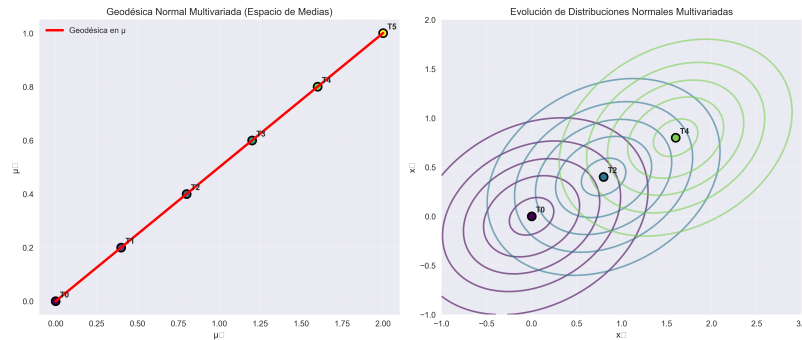


Figura 10.2: Comportamiento de las geodésicas de la distribución normal multivariada

naturales.

**Razón:** La métrica de Fisher para la normal es anisótropa:

- La información sobre  $\mu$  es proporcional a  $1/\sigma^2$ .
- La información sobre  $\sigma$  es proporcional a  $2/\sigma^2$ .

Por tanto, cambios en  $\sigma$  son más costosos informativamente y la geodésica se curva para reducir desplazamiento en esa dirección.

### Distribución gamma

**Comportamiento:** Las geodésicas se curvan con fuerza, especialmente para parámetros pequeños.

**Razón:** La métrica de Fisher depende de la trigamma (derivada de la digamma), lo que introduce una marcada no linealidad:

- Para  $\alpha$  pequeño, la curvatura del espacio es alta.
- Las geodésicas tienden a evitar regiones de baja información (alta curvatura).

**Distribución beta**

**Comportamiento:** Las geodésicas son muy curvas debido al acoplamiento entre parámetros.

**Razón:** En la métrica de Fisher para la beta:

- $\alpha$  y  $\beta$  influyen juntos en la curvatura a través de  $\psi'(\alpha + \beta)$ .
- Las geodésicas equilibran variaciones en ambos parámetros para reducir la divergencia.
- Evitan regiones cercanas a los bordes (valores pequeños) donde surgen singularidades.

**Distribución normal multivariada**

**Comportamiento:** En el subespacio de medias las geodésicas sí son rectas.

**Razón:** Con  $\Sigma$  fija, la información de Fisher respecto a  $\boldsymbol{\mu}$  es la matriz  $\Sigma^{-1}$ :

- No hay acoplamiento no lineal entre componentes de  $\boldsymbol{\mu}$ .
- La métrica inducida es euclidiana, de modo que las geodésicas son segmentos rectos.

**Principio general**

Las geodésicas de Fisher–Rao minimizan la divergencia estadística entre distribuciones. Su forma depende de la métrica:

- Son rectas cuando la métrica es plana.
- Se curvan cuando la variedad estadística tiene curvatura.
- Evitan zonas de baja información, donde la métrica es singular o degenerada.

La curvatura del espacio refleja cuánta información se pierde al desplazarse en distintas direcciones: cuanto mayor la curvatura, más delicado es el cambio paramétrico desde el punto de vista informativo.

Esta investigación ha analizado la geometría de la información y ha demostrado cómo cualquier familia paramétrica regular se puede interpretar como una variedad diferenciable dotada de la métrica de Fisher–Rao, donde los parámetros estadísticos actúan como coordenadas locales. Se ha visto que la información de Fisher cuantifica la precisión estadística a la vez que induce una estructura geométrica intrínseca e invariante bajo reparametrizaciones.

En las distribuciones estudiadas, la normal univariada presenta una geometría hiperbólica con curvatura constante negativa  $K = -\frac{1}{2}$ . La normal multivariada muestra una estructura en producto, plana en el subespacio de medias y de curvatura negativa variable en el subespacio de covarianzas. Las distribuciones beta y gamma tienen curvaturas negativas en todo su espacio paramétrico, con matices debidos a sus soportes y parametrizaciones. Se han descrito geodésicas en estos espacios, que actúan como trayectorias óptimas entre distribuciones bajo la métrica de Fisher.

Los resultados muestran que la curvatura negativa indica cómo la información se concentra de forma no lineal, afectando al comportamiento de estimadores. La ortogonalidad entre parámetros, como media y varianza en la normal, se interpreta geoméricamente como ortogonalidad de vectores tangentes. En familias exponenciales, parametrizaciones naturales diagonalizan la métrica, y la desigualdad de Cramér–Rao se entiende como una propiedad de longitud mínima en la variedad estadística.

Quedan abiertas varias líneas de trabajo futuro: desarrollar una teoría geométrica para modelos no paramétricos de dimensión infinita; implementar algoritmos eficientes para geodésicas en espacios estadísticos de alta dimensión; explorar conexiones con aprendizaje profundo; extender el análisis a modelos mixtos, jerárquicos o bayesianos; y estudiar otras métricas de información y su influencia en la inferencia.

En síntesis, la geometría de la información unifica conceptos estadísticos bajo una perspectiva geométrica coherente. Este enfoque ofrece un lenguaje natural para estudiar propiedades estadísticas y diseñar nuevos métodos. La presencia de curvaturas negativas sugiere que este rasgo podría ser característico de la información estadística, reflejando cómo la incertidumbre se organiza en los espacios paramétricos. En un contexto de datos y modelos cada vez más complejos, la geometría de la información aporta un marco prometedor para la ciencia de datos moderna.

## 12.1. Aplicaciones en imágenes médicas

Una línea de investigación es la aplicación de la geometría de Fisher–Rao al análisis de imágenes médicas. Características como histogramas de intensidad, textura o grosor cortical pueden modelarse mediante distribuciones beta o gamma, de modo que cada muestra se represente como un punto en una variedad diferenciable con métrica Riemanniana.

En este marco se han empleado histogramas ajustados con distribuciones beta para detectar hipertensión pulmonar y para apoyar el diagnóstico de Alzheimer mediante medidas de grosor cortical. La curvatura negativa de la variedad asegura unicidad de la media de Fréchet, útil en agrupamiento y comparación. Sería oportuno explorar este enfoque en otras modalidades como fMRI, PET o ecografía, así como en tareas de segmentación, normalización espacial y análisis longitudinal, donde la geometría estadística podría mejorar la detección de patologías.

## 12.2. Agrupamiento estadístico mediante información de Fisher

La métrica de Fisher en espacios de distribuciones como las familias beta y Dirichlet permite definir distancias geodésicas que reflejan la estructura probabilística de los datos. Esta curvatura negativa garantiza que algoritmos de centroides, como el  $K$ -means geométrico, tengan unicidad y estabilidad en los centroides.

Se recomienda desarrollar métodos numéricos eficientes para calcular distancias geodésicas y promedios en estos espacios, y combinar esta geometría con técnicas como *spectral clustering* o *t-SNE* en análisis de datos que consisten en colecciones de distribuciones (por ejemplo, histogramas de textura o distribuciones empíricas de sensores). La distancia de Fisher–Rao aporta una medida natural de disimilitud intrínseca, adecuada para estos problemas.

- [1] Amari, Shun ichi y Hiroshi Nagaoka: *Methods of Information Geometry*, volumen 191 de *Translations of Mathematical Monographs*. American Mathematical Society, 2000.
- [2] Analytics, IBM Cognos: *Statistical distribution*. <https://www.ibm.com/docs/en/cognos-analytics/12.0.0?topic=terms-statistical-distribution>, 2018. Accessed: 2024-09-25.
- [3] Analytics, IBM Cognos: *Normal distribution*. <https://www.ibm.com/docs/en/cognos-analytics/11.2.0?topic=terms-normal-distribution>, 2024. Accessed: 2024-09-25.
- [4] Ay, Nihat, Jürgen Jost, Hông Vân Lê y Lorenz Schwachhöfer: *Information Geometry and Sufficient Statistics*. Probability Theory and Related Fields, 162:327–364, 2015.
- [5] Carlson, Stephan C.: *Hausdorff space*, n.d. <https://www.britannica.com/science/Hausdorff-space>, Accessed April 24, 2025.
- [6] Carmo, Manfredo P. do: *Differential Geometry of Curves and Surfaces*. Dover Publications, Mineola, New York, second revised and updated edition edición, 2016, ISBN 978-0-486-80699-0.
- [7] Carmo, Manfredo Perdigão do: *Riemannian Geometry*. Mathematics: Theory & Applications. Birkhäuser, Boston, MA, 1992, ISBN 978-0-8176-3490-8.
- [8] Dekking, F.M., C. Kraaikamp, H.P. Lopuhaä y L.E. Meester: *A Modern Introduction to Probability and Statistics: Understanding Why and How*. Springer Texts in Statistics. Springer, London, 2005, ISBN 978-1-85233-896-1.
- [9] He, Wangzhe: *An introduction to continuous functions, metric space, manifolds, topological spaces and its properties*. En *Proceedings of the 2nd International Conference on Computing Innovation and Applied Physics (CONF-CIAP 2023)*. CONF-CIAP, 2023. <https://doi.org/10.54254/2753-8818/5/20230301>.
- [10] Jost, Jürgen: *Riemannian Geometry and Geometric Analysis*. Springer International Publishing, Cham, 7th edición, 2017, ISBN 978-3-319-61859-3. <https://link.springer.com/book/10.1007/978-3-319-61860-9>.
- [11] Jost, Jürgen: *Riemannian Geometry and Geometric Analysis*. Springer-Verlag Berlin Heidelberg, Germany, 4th edición, 2005, ISBN 978-3-540-25907-7.
- [12] Kalbfleisch, J.G.: *Probability and Statistical Inference, Volume 1: Probability*. Springer Texts in Statistics, 1, 1985.

- [13] Kalbfleisch, J.G.: *Probability and Statistical Inference, Volume 2: Statistical Inference*. Springer Texts in Statistics, 2, 1985.
- [14] Le Brigant, Alice, Nicolas Guigui, Sana Rebbah y Stéphane Puechmorel: *Classifying Histograms of Medical Data Using Information Geometry of Beta Distributions*. IFAC PapersOnLine, 54(9):514–520, 2021.
- [15] Le Brigant, Alice, Stephen C. Preston y Stéphane Puechmorel: *Fisher-Rao Geometry of Dirichlet Distributions*. arXiv preprint arXiv:2005.05608, 2020. <https://arxiv.org/abs/2005.05608>.
- [16] Le Brigant, Alice y Stéphane Puechmorel: *The Fisher-Rao Geometry of Beta Distributions Applied to the Study of Canonical Moments*. arXiv preprint arXiv:1904.08247, 2019. <https://arxiv.org/abs/1904.08247>.
- [17] Leclerc, Gaétan: *Riemannian Geometry and Geometric Analysis*. Internship Report, Université de Rennes and University of Freiburg, June–August 2019. [https://perso.imj-prg.fr/gaetan-leclerc/wp-content/uploads/sites/36/2021/08/Riemannian\\_Geometry\\_and\\_Geometric\\_Analysis.pdf](https://perso.imj-prg.fr/gaetan-leclerc/wp-content/uploads/sites/36/2021/08/Riemannian_Geometry_and_Geometric_Analysis.pdf).
- [18] Lee, John M.: *Introduction to Topological Manifolds*. Springer, New York, 1st edición, 2000.
- [19] Lee, John M.: *Introduction to Smooth Manifolds*. Springer, 2012.
- [20] Levy, Roger: *Probabilistic Models in the Study of Language*, 2012. [https://www.mit.edu/~rplevy/pmsl\\_textbook/book\\_draft.pdf](https://www.mit.edu/~rplevy/pmsl_textbook/book_draft.pdf), Unpublished textbook draft, accessed April 23, 2024.
- [21] Loève, Michel: *Probability Theory I*, volumen 45 de *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1977.
- [22] Ly, Alexander, Maarten Marsman, Josine Verhagen, Raoul Grasman y Eric Jan Wagenmakers: *A Tutorial on Fisher Information*. Journal of Mathematical Psychology, 80:40–55, 2017.
- [23] Ly, Alexander, Maarten Marsman, Josine Verhagen, Raoul Grasman y Eric Jan Wagenmakers: *A Tutorial on Fisher Information*. arXiv preprint arXiv:1705.01064, 2017. <https://arxiv.org/abs/1705.01064>.
- [24] Munkres, James R.: *Topology*. Pearson Education, Harlow, Essex, 2nd edición, 2014, ISBN 9780321986927.
- [25] Murphy, Kevin P.: *Probabilistic Machine Learning: An introduction*. MIT Press, 2022. <http://probml.github.io/book1>.
- [26] Nielsen, Frank: *An Elementary Introduction to Information Geometry*. arXiv preprint arXiv:1808.08271, 2020. Available at <https://arxiv.org/abs/1808.08271>.
- [27] O'Neill, Barrett: *Elementary Differential Geometry*. Elsevier Academic Press, Amsterdam; Boston; Heidelberg; London; New York; Oxford; Paris; San Diego; San Francisco; Singapore; Sydney; Tokyo, revised second edition edición, 2006, ISBN 978-0-12-088735-4.
- [28] Rebbah, Sana, Florence Nicol y Stéphane Puechmorel: *The Geometry of the Generalized Gamma Manifold and an Application to Medical Imaging*. Mathematics, 7:674, 2019.
- [29] Reyes-Figueroa, Alan: *Geometría Intrínseca de Superficies*. Clase del 23 de abril de 2024, 2024. <https://pfafer.github.io/gd2024/aulas/Aula29.pdf>, Notas de clase, Geometría Diferencial (Aula 29).
- [30] Reyes-Figueroa, Alan: *Primera Forma Fundamental*. Clase del 19 de marzo de 2024, 2024. <https://pfafer.github.io/gd2024/aulas/Aula19.pdf>, Notas de clase, Geometría Diferencial (Aula 19).

- [31] Savage, Leonard J.: *On Rereading R. A. Fisher*. The Annals of Statistics, 4(3):441–500, 1976.
- [32] Schervish, M.J.: *Theory of Statistics*. Springer Series in Statistics, 1995.
- [33] Warner, Frank W.: *Foundations of Differentiable Manifolds and Lie Groups*, volumen 94 de *Graduate Texts in Mathematics*. Springer-Verlag, Berlin Heidelberg, 1983.

$(\Omega, \mathcal{F}, \mathbb{P})$	Espacio de probabilidad
$(U, \varphi)$	Carta con dominio $U$
$\alpha$	Curva paramétrica
$\ell(\theta; x)$	Función de log-verosimilitud
$\Gamma(z)$	Función gamma
$\Gamma_{jk}^i$	Símbolos de Christoffel
$\hat{\theta}$	Estimador de máxima verosimilitud de $\theta$
$\mathbb{E}[X]$	Valor esperado de la variable aleatoria $X$
$\mathbb{P}$	Medida de probabilidad
$\mathcal{F}$	$\sigma$ -álgebra, colección de subconjuntos medibles de $\Omega$
Beta( $\alpha, \beta$ )	Distribución beta con parámetros de forma $\alpha$ y $\beta$
Gamma( $k, \theta$ )	Distribución gamma con parámetro de forma $k$ y escala $\theta$
$\nabla$	Conexión de Levi-Civita
$\Omega$	Espacio muestral
Var( $X$ )	Varianza de la variable aleatoria $X$
$\psi'(z)$	Función trigamma, derivada de la función digamma
$\psi(z)$	Función digamma, $\psi(z) = \frac{d}{dz} \ln \Gamma(z)$
$\varphi$	Carta local o parametrización
$\xi$	Parámetro en un modelo estadístico
$d(p, q)$	Distancia geodésica entre puntos $p$ y $q$
$F_X(x)$	Función de distribución acumulada de $X$
$f_X(x)$	Función de densidad de probabilidad de $X$
$g$	Métrica de Riemann

$G(\xi)$	Matriz de información de Fisher en $\xi$
$g_{ij}$	Componentes de la métrica en coordenadas locales
$g_{ij}(\xi)$	Componentes de la métrica de Fisher
$I(\theta)$	Información de Fisher
$K$	Curvatura seccional
$L(\theta)$	Función de verosimilitud
$M$	Variedad diferenciable
$N(\mu, \sigma^2)$	Distribución normal con media $\mu$ y varianza $\sigma^2$
$N_d(\boldsymbol{\mu}, \Sigma)$	Distribución normal multivariada de dimensión $d$
$p_\theta$	Distribución de probabilidad con parámetro $\theta$
$S$	Modelo estadístico (como variedad)
$S(\theta)$	Función de score o puntuación
$T_p M$	Espacio tangente a la variedad $M$ en el punto $p$
$X$	Variable aleatoria