



UNIVERSIDAD NACIONAL
AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

Introducción al método Monte Carlo Hamiltoniano

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Actuario

PRESENTA:

Granados Carmona Saúl Abraham

TUTOR

Mtro. Ramírez Bañales Isaías Manuel

CIUDAD DE MÉXICO, 2021





Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

El peor laberinto no es esa forma intrincada que puede atraparnos para siempre, sino una línea recta única y precisa.

Jorge Luis Borges.

*A mis padres, cada letra en esta tesis es también
un “gracias”.*

Agradecimientos

A Dios, por permitirme llegar a este punto tan importante en mi vida, por brindarme salud, amor, fuerza y felicidad en cada momento.

A mis padres, no alcanzan las palabras para agradecerles todo lo que han hecho por mi. La parte más complicada de este logro la hicieron ustedes; no lo hubiera conseguido sin su apoyo, amor y todo lo que me brindan. LOS AMO, GRACIAAAS.

A mis hermanos, gracias por siempre apoyarme y hacer más alegres mis días. A Betos, por ser un ejemplo a seguir y sus buenos consejos. A la Bere, por su alegría y porque estuvo sentada junto a mi casi en cada hoja de este trabajo, haciendo más ameno el camino. A mi hermano y colega Isra, que siempre adivinaba por mi estado de humor si iba bien en la tesis o estaba estancado, por su apoyo y ayuda siempre.

Al M. en C. Isaías Manuel Ramírez, por aceptar trabajar conmigo en este proyecto, por su tiempo, paciencia, ayuda, sus consejos, y por darme la oportunidad de cumplir un par de mis metas: dar clases en la facultad y tener un trabajo de investigación.

A los plausibles sinodales de este trabajo, por su tiempo, sus palabras entusiastas y por todas las correcciones y observaciones que sin duda mejoraron grandemente la tesis.

A mis tios y primos, por siempre contagiar su alegría, por sus consejos y por apoyarme en cada momento.

A mi abuela Ana, por su apoyo, cariño, alegría y su gran fotalenza. A mi abuela Malena, por su amor y porque todos los recuerdos que tengo a a su lado son de felicidad. A mis abuelos: Pancho y Enrique, por todo el cariño y apoyo que me brindaron.

A mis amigos: Angel Javier Cruz y Luis Alexis Trejo, por siempre escucharme y apoyarme en esta década de amistad, por sus ánimos y palabras de aliento cuando por miedo o inseguridad pensaba en desistir de este proyecto, por todas las alegrías, aventuras y risas que hemos compartido.

A mis amigos y colegas: Zyanya Pérez, Edgar Marbán, Fernando Gaspar y Francisco Olivares, porque desde el primer semestre hicieron que cada día en la facultad fuera divertido. Les agradezco por su amistad, confianza, por todas las desveladas, los momentos de estrés, estudio y diversión que nos tocó vivir, y por ser el mejor equipo de tareas y trabajos en cada una de las materias en que coincidíamos. Los admiro muchísimo.

A Gloria Cervantes, Adriana Corona y Luis Ángel Calderón, por su apoyo y por ser unas personas increíbles.

A la UNAM, mi *alma mater*, por todo lo que he aprendido en sus aulas y por todas las oportunidades que me ha brindado.

Índice general

1. Introducción	1
2. Fundamentos de Mecánica Analítica	3
2.1. Mecánica Analítica	3
2.1.1. Conceptos básicos	3
2.1.2. El principio de mínima acción	4
2.1.3. Ecuaciones de Euler- Lagrange	5
2.1.4. Ecuaciones de Hamilton	9
2.1.5. Propiedades del Hamiltoniano	18
2.1.5.1. Reversibilidad	18
2.1.5.2. Conservación	19
2.1.5.3. Preservacion del volumen	20
2.1.5.4. Dinámicas simplécticas	21
3. Algoritmos y temas precedentes	23
3.1. Métodos para solución de ecuaciones diferenciales ordinarias	23
3.1.1. Método de Euler	24
3.1.2. Método de Euler Simpléctico	25
3.1.3. Método Leapfrog	27
3.1.4. Comparación	28
3.2. Estadística bayesiana	34
3.3. Métodos de cadenas de Markov Monte Carlo	35
3.3.1. Metropolis-Hastings	36
3.3.2. Random Walk Metropolis-Hastings	37
3.3.2.1. Escalamiento óptimo en RWMH	38
3.3.2.2. Burn-in	41
3.3.2.3. Lag	45
3.3.3. T-walk	47

4. Monte Carlo Hamiltoniano	51
4.1. Construcción del algoritmo HMC	51
4.2. Resultados teóricos	58
4.2.1. Tasa de aceptación	58
4.2.2. Ecuaciones de balance detallado	59
4.2.3. Distribución estacionaria de la cadena	61
4.2.4. Ergodicidad de la cadena	62
4.3. Ejemplos y comparación entre HMC y RWMH	64
4.3.1. Simulación de la distribución Gamma(5,1)	64
4.3.2. Normal bivariada	68
4.3.3. Mezcla gaussiana	70
4.3.4. Distribución normal 150-dimensional	74
5. Aplicación de HMC en un modelo normal jerárquico	77
5.1. Modelo de las 8 escuelas	77
6. Conclusiones	87
Apéndice A. Breve introducción a la mecánica clásica	89
A.0.1. Leyes de Newton	90
A.0.1.1. Primera Ley de Newton:	90
A.0.1.2. Segunda ley de Newton	91
A.0.1.3. Tercera ley de Newton	92
A.0.2. Trabajo y energía cinética	94
A.0.3. Campos de fuerzas conservativas y energía potencial	96
A.0.4. Conservación de la energía mecánica	98

Capítulo 1

Introducción

Los métodos de cadenas de Markov Monte Carlo (MCMC, por sus siglas en inglés), son una herramienta útil que permite simular de distribuciones de probabilidad, cuando se conoce su función de densidad salvo constantes.

En la actualidad, el uso de los métodos MCMC se ha incrementado notablemente debido al desarrollo computacional, al costo decreciente de almacenamiento y recolección de datos, y a su utilidad en estadística. En particular, en estadística bayesiana los métodos MCMC resultan de gran utilidad ya que en esta área las densidades *a posteriori*, generalmente, no son conocidas analíticamente, por lo que para estimar cantidades de interés es necesario simular.

El objetivo de la presente tesis es dar una introducción, mostrar aplicaciones y presentar los códigos de programación en Python del algoritmo Monte Carlo Hamiltoniano (HMC, por sus siglas en inglés), un método MCMC que permite simular de distribuciones de probabilidad de una manera eficiente empleando dinámicas hamiltonianas. Estas dinámicas se rigen por las ecuaciones de Hamilton y son un área de estudio de la física, concretamente de la mecánica hamiltoniana.

Este trabajo está dirigido a alumnos de últimos semestres de actuaría con conocimientos de simulación, probabilidad y estadística. La parte referente a la mecánica hamiltoniana se plantea de manera lacónica y de forma estructurada, comenzando con ideas simples de mecánica clásica que se pueden consultar en el apéndice A, de tal forma que no se requieren conocimientos específicos de física o mecánica, sin embargo, se espera que para entender el desarrollo el alumno tenga conocimientos de cálculo diferencial e integral.

En el capítulo 2, se introducen los resultados y conceptos de mecánica hamiltoniana necesarios para la construcción del método HMC, además, se presentan las ideas de mecánica analítica que resultan útiles para una comprensión más clara de la mecánica hamiltoniana.

El capítulo 3 presenta los algoritmos Leapfrog, Euler simpléctico, y el método de Euler, enfocados a aproximar la solución de las ecuaciones de Hamilton. Por otro lado, se exponen los métodos de cadenas de Markov Monte Carlo, profundizando en el algoritmo *Random Walk Metropolis Hastings* (RWMH, por sus siglas en inglés), y se da bibliografía sugerida en caso de que el lector quiera profundizar en el tema. Se introduce también el algoritmo T-walk y se presenta una breve introducción a la estadística bayesiana.

Dentro del capítulo 4, se introduce el método HMC, se detalla el planteamiento y algunos resultados teóricos, como los que sustentan su funcionamiento. Por último, se realizan ejemplos y comparaciones con el algoritmo RWMH, con el objetivo de resaltar sus principales diferencias y analizar su desempeño.

En el capítulo 5, se presenta una ejemplo del método HMC para hacer inferencia bayesiana en un modelo gaussiano jerárquico, y se compara su efectividad con los algoritmos RWMH y T-walk para el mismo fin.

En el capítulo 6, se presenta una conclusión de los principales resultados obtenidos en este trabajo sobre el método HMC, y su comparación con RWMH y T-walk.

Los códigos en Python utilizados en este trabajo, se pueden consultar en el repositorio de Github <https://git.io/JtA70>.

Capítulo 2

Fundamentos de Mecánica Analítica

En este capítulo se introduce y se presentan conceptos, fundamentos y resultados de mecánica hamiltoniana importantes para la construcción del método Monte Carlo Hamiltoniano.

2.1. Mecánica Analítica

La mecánica analítica es una formulación abstracta de la mecánica newtoniana desarrollada en los siglos XVIII y XIX, principalmente por Joseph-Louis de Lagrange¹ y William Rowan Hamilton².

Las ecuaciones del movimiento bajo este enfoque se derivan de funciones escalares, a diferencia de la mecánica newtoniana donde es necesario definir las posiciones de las partículas y las fuerzas que actúan sobre ellas en espacios tridimensionales. Además, su formulación matemática es independiente de cualquier cambio de coordenadas, facilitando así su aplicación a diversos problemas como se explica en [13].

2.1.1. Conceptos básicos

Definimos como grados de libertad de un sistema a la cantidad mínima de valores escalares independientes, necesarios para determinar la posición de las partículas en el espacio. En mecánica newtoniana para conocer la posición de n partículas que se mueven sin restricción alguna en el espacio, son necesarios n vectores de posición, o bien, $3n$ grados de

¹(1736-1813), físico, matemático y astrónomo italiano, desarrolló la mecánica lagrangiana y tuvo una importante contribución en astronomía.

²(1805-1865), matemático, físico y astrónomo irlandés, realizó importantes contribuciones al desarrollo de la óptica, dinámica, álgebra y formuló la mecánica hamiltoniana.

libertad.

Las coordenadas generalizadas son un conjunto de parámetros q_1, q_2, \dots, q_s que permiten determinar de manera unívoca la posición de las partículas en el espacio, generalmente estos parámetros son magnitudes físicas, es decir propiedades medibles, como la posición en un eje de coordenadas, ángulos, distancias, etc.

En un sistema de partículas regido por s coordenadas generalizadas, denotaremos por $q : \mathbb{R} \rightarrow \mathbb{R}^s$ a la función que a un tiempo t devuelve los puntos de las coordenadas generalizadas en los que se encuentra el sistema de partículas, es decir, $q(t) = (q_1(t), q_2(t), \dots, q_s(t))$, donde $q_i(t)$ es la posición en la i -ésima coordenada al tiempo t .

El objetivo de usar las coordenadas generalizadas es obtener las ecuaciones de movimiento de un sistema de una manera más eficiente. Se pretende que al elegir estas coordenadas se tenga una función $\phi_i : \mathbb{R}^{s+1} \rightarrow \mathbb{R}^{3n}$ tal que

$$r_i(t) = \phi_i(q(t), t),$$

donde $r_i(t) \in \mathbb{R}^3$ representa la posición de la i -ésima partícula en coordenadas cartesianas, de esta manera con $q(t)$ puede conocerse también la posición completa del sistema en el espacio tridimensional.

Definimos como velocidades generalizadas a un tiempo t al vector de derivadas de las coordenadas generalizadas $\frac{dq(t)}{dt}$ y llamaremos espacio fase de velocidades, denotado por M , al subconjunto de \mathbb{R}^{2s} tal que para todo tiempo t

$$\left(q_1(t), q_2(t), \dots, q_s(t), \frac{dq_1(t)}{dt}, \frac{dq_2(t)}{dt}, \dots, \frac{dq_s(t)}{dt} \right) \in M.$$

2.1.2. El principio de mínima acción

La reformulación de la mecánica newtoniana está basada en el principio de mínima acción, también conocido como principio de Hamilton, a partir del cual pueden deducirse las ecuaciones que describen el movimiento del sistema de partículas en el espacio fase de velocidades. Para comprender este principio, introduciremos primero qué es el lagrangiano y la acción.

El lagrangiano de un sistema de partículas es una función $L : \mathbb{R}^{2s+1} \rightarrow \mathbb{R}$, dependiente del tiempo, de las coordenadas y las velocidades generalizadas y puede tener diversas reglas de correspondencia como se explica en la sección 2.1.3.

La acción A es una función, cuyo argumento es el lagrangiano en un intervalo de tiempo $[t_0, t_1]$, se define como

$$A[L] = \int_{t_0}^{t_1} L \left(q(t), \frac{dq(t)}{dt}, t \right) dt. \quad (2.1)$$

El principio de Hamilton, como se afirma en [24], menciona que todo sistema de partículas está caracterizado por un lagrangiano y que de todas las trayectorias posibles que puede tomar el sistema en un intervalo de tiempo $[t_0, t_1]$, la única que sigue es aquella que hace a la acción (2.1), un valor mínimo o un punto silla.

2.1.3. Ecuaciones de Euler- Lagrange

En esta sección deduciremos las ecuaciones Euler-Lagrange, que para un sistema de partículas con n grados de libertad son un conjunto de n ecuaciones diferenciales de segundo orden, cuyas incógnitas son las coordenadas generalizadas q_1, q_2, \dots, q_n , las cuales determinan completamente el movimiento del sistema de partículas.

Sea $q(t) = (q_1(t), q_2(t), \dots, q_n(t))$ la trayectoria que sigue el sistema de partículas en un intervalo de tiempo $[t_0, t_1]$, y definamos un conjunto de curvas parametrizadas como

$$q^\alpha(t) = q(t) + \alpha \eta(t),$$

donde $\alpha \in \mathbb{R}$ y $\eta : \mathbb{R} \rightarrow \mathbb{R}^n$ es una función arbitraria que cumple que $\eta(t_0) = \eta(t_1) = 0$, de modo que los puntos extremos de $q^\alpha(t)$ coincidan con los de $q(t)$. De manera intuitiva, las curvas q^α representan todas las trayectorias posibles entre los puntos $q(t_0)$ y $q(t_1)$.

Notemos además que para todo $i \in \{1, 2, \dots, n\}$ se cumplen las siguientes igualdades, que serán de utilidad para deducir las ecuaciones de Euler- Lagrange:

$$\frac{\partial q_i^\alpha}{\partial t} = \frac{dq_i}{dt} + \alpha \frac{d\eta_i}{dt}, \quad (2.2)$$

$$\frac{\partial q_i^\alpha}{\partial \alpha} = \eta_i, \quad (2.3)$$

$$\frac{\partial}{\partial \alpha} \left(\frac{\partial q_i^\alpha}{\partial t} \right) = \frac{d\eta_i}{dt}. \quad (2.4)$$

Tengamos presente además, que al tomar $\alpha = 0$, la trayectoria q^α será la que sigue el sistema de partículas, pues $q^0(t) = q(t)$, en este caso por el principio de Hamilton debe cumplirse que la acción evaluada en el lagrangiano L^α , asociado a la trayectoria q^α sea un punto crítico, es decir,

$$\left. \frac{\partial A(L^\alpha)}{\partial \alpha} \right|_{\alpha=0} = 0. \quad (2.5)$$

Obteniendo la derivada de la acción respecto al parámetro α , obtenemos que

$$\begin{aligned} \frac{\partial A(L^\alpha)}{\partial \alpha} &= \frac{d}{d\alpha} \int_{t_0}^{t_1} L \left(q^\alpha(t), \frac{d(q^\alpha(t))}{dt}, t \right) dt \\ &= \frac{\partial}{\partial \alpha} \int_{t_0}^{t_1} L \left(q_1^\alpha(t), q_2^\alpha(t), \dots, q_n^\alpha(t), \frac{dq_1^\alpha(t)}{dt}, \frac{dq_2^\alpha(t)}{dt}, \dots, \frac{dq_n^\alpha(t)}{dt}, t \right) dt \\ &= \int_{t_0}^{t_1} \sum_{i=1}^n \left(\frac{\partial L}{\partial q_i^\alpha} \frac{\partial q_i^\alpha}{\partial \alpha} + \frac{\partial L}{\partial \left(\frac{dq_i^\alpha}{dt} \right)} \frac{\partial \left(\frac{dq_i^\alpha}{dt} \right)}{\partial \alpha} \right) + \frac{dt}{d\alpha} dt, \end{aligned} \quad (2.6)$$

sustituyendo (2.3) y (2.4) en (2.6),

$$\begin{aligned} \frac{\partial A(L^\alpha)}{\partial \alpha} &= \int_{t_0}^{t_1} \sum_{i=1}^n \left(\frac{\partial L}{\partial q_i^\alpha} \eta_i(t) + \frac{\partial L}{\partial \left(\frac{dq_i^\alpha}{dt} \right)} \frac{d\eta_i}{dt} \right) dt \\ &= \sum_{i=1}^n \int_{t_0}^{t_1} \frac{\partial L}{\partial q_i^\alpha} \eta_i(t) dt + \sum_{i=1}^n \int_{t_0}^{t_1} \frac{\partial L}{\partial \left(\frac{dq_i^\alpha}{dt} \right)} \frac{d\eta_i}{dt} dt. \end{aligned} \quad (2.7)$$

Integrando por partes el segundo sumando de (2.7), teniendo en cuenta que $\eta(t_0) = \eta(t_1) = 0$, obtenemos que

$$\begin{aligned} \int_{t_0}^{t_1} \frac{\partial L}{\partial \left(\frac{dq_i^\alpha}{dt} \right)} \frac{d\eta_i}{dt} dt &= \left. \frac{\partial L}{\partial \left(\frac{dq_i^\alpha}{dt} \right)} \eta_i(t) \right|_{t_0}^{t_1} \\ &\quad - \int_{t_0}^{t_1} \eta_i(t) \frac{d}{dt} \left(\frac{\partial L}{\partial \left(\frac{dq_i^\alpha}{dt} \right)} \right) dt \\ &= - \int_{t_0}^{t_1} \eta_i(t) \frac{d}{dt} \left(\frac{\partial L}{\partial \left(\frac{dq_i^\alpha}{dt} \right)} \right) dt. \end{aligned} \quad (2.8)$$

Sustituyendo el valor de la integral (2.8) en (2.7),

$$\begin{aligned} \frac{dA(L^\alpha)}{d\alpha} &= \sum_{i=1}^n \int_{t_0}^{t_1} \frac{\partial L}{\partial q_i^\alpha} \eta_i(t) dt - \sum_{i=1}^n \int_{t_0}^{t_1} \eta_i(t) \frac{d}{dt} \left(\frac{\partial L}{\partial \left(\frac{dq_i^\alpha}{dt} \right)} \right) dt \\ &= \int_{t_0}^{t_1} \sum_{i=1}^n \left[\frac{\partial L}{\partial q_i^\alpha} - \frac{d}{dt} \left(\frac{\partial L}{\partial \left(\frac{dq_i^\alpha}{dt} \right)} \right) \right] \eta_i(t) dt, \end{aligned}$$

tomando $\alpha = 0$ y por (2.5) se cumple que

$$\int_{t_0}^{t_1} \sum_{i=1}^n \left[\frac{\partial L}{\partial q_i} - \frac{d}{dt} \left(\frac{\partial L}{\partial \left(\frac{dq_i}{dt} \right)} \right) \right] \eta_i(t) dt = 0. \quad (2.9)$$

Notemos que para que se satisfaga (2.9) para cualquier función $\eta_i(t)$ debe suceder que el valor entre corchetes se anule, es decir, debe suceder que

$$\frac{\partial L}{\partial q_i} = \frac{d}{dt} \left(\frac{\partial L}{\partial \left(\frac{dq_i}{dt} \right)} \right) \quad \text{para todo } i \in \{1, 2, \dots, n\}. \quad (2.10)$$

Las ecuaciones (2.10) son las ecuaciones de Euler-Lagrange y junto con sus condiciones iniciales, describen completamente el movimiento del sistema de partículas.

La función lagrangiana no es única, como se explica en [10, p.21] existen infinitas funciones de las que se puede deducir las ecuaciones de movimiento. Generalmente, se trabaja con el lagrangiano estándar, que sólo depende del tiempo a través de las coordenadas y velocidades generalizadas, y se define como la diferencia entre la energía cinética y la potencial, es decir,

$$L \left(q(t), \frac{dq(t)}{dt} \right) = T \left(\frac{dq(t)}{dt} \right) - U(q(t)). \quad (2.11)$$

Una demostración que explica por qué puede considerarse así al lagrangiano, partiendo del principio de D'Alambert, que postula que las fuerzas externas que actúan sobre un cuerpo y las fuerzas de inercia están en equilibrio, se puede consultar en [17, pp. 267-27].

En el siguiente ejemplo se muestra el cálculo del lagrangiano estándar de un sistema y a partir de éste se deduce la trayectoria de la partícula.

Ejemplo 1. Consideremos un cuerpo de masa constante m , unido a un resorte colocado de manera horizontal como se muestra en la imagen 2.1 (generada en la página web <https://www.edrawsoft.com/es/>), supongamos además que la única fuerza que existe en

el ejemplo es la del resorte denotada por F . El objetivo es encontrar una función que al tiempo t devuelva la posición del cuerpo.

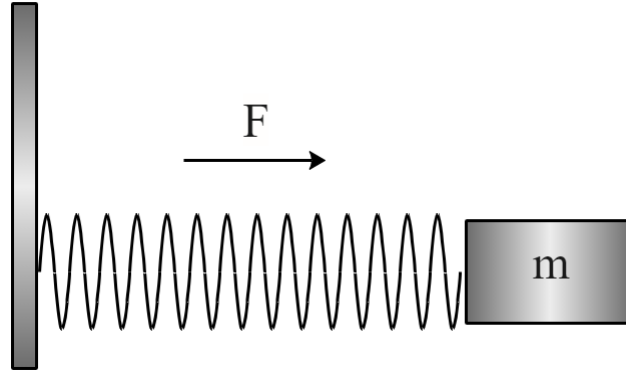


Figura 2.1: Sistema masa-resorte.

El lagrangiano estándar definido en (2.11), está dado por

$$\begin{aligned} L\left(q(t), \frac{dq(t)}{dt}\right) &= T\left(\frac{dq(t)}{dt}\right) - U(q(t)) \\ &= \frac{1}{2}m\left(\frac{dq(t)}{dt}\right)^2 - \frac{1}{2}kq(t)^2. \end{aligned} \quad (2.12)$$

La energía cinética y potencial de la ecuación (2.12) se calcularon en el apéndice A.0.4.

Para obtener las ecuaciones de movimiento aplicando las ecuaciones de Euler-Lagrange, calculemos las siguientes expresiones:

$$\begin{aligned} \frac{\partial L}{\partial q(t)} &= \frac{\partial}{\partial q(t)} \left(\frac{1}{2}m\left(\frac{dq(t)}{dt}\right)^2 - \frac{1}{2}kq(t)^2 \right) \\ &= -kq(t), \end{aligned} \quad (2.13)$$

$$\begin{aligned} \frac{\partial L}{\partial \left(\frac{dq(t)}{dt}\right)} &= \frac{\partial}{\partial \left(\frac{dq(t)}{dt}\right)} \left(\frac{1}{2}m\left(\frac{dq(t)}{dt}\right)^2 - \frac{1}{2}kq(t)^2 \right) \\ &= m\left(\frac{dq(t)}{dt}\right), \end{aligned}$$

además, notemos que

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \left(\frac{dq(t)}{dt} \right)} \right) = m \frac{d^2 q(t)}{dt^2}. \quad (2.14)$$

Sustituyendo (2.13) y (2.14) en la ecuación de Euler-Lagrange $\frac{\partial L}{\partial q(t)} = \frac{d}{dt} \left(\frac{\partial L}{\partial \left(\frac{dq(t)}{dt} \right)} \right)$, obtenemos la siguiente ecuación diferencial

$$-kq(t) = m \frac{d^2 q(t)}{dt^2}, \quad (2.15)$$

que es la misma que surge al resolver este ejercicio con el enfoque newtoniano, como se presenta en el ejemplo 7. Su solución se presenta en ese misma ejemplo.

2.1.4. Ecuaciones de Hamilton

El enfoque hamiltoniano brinda una manera distinta de describir el movimiento en los sistemas de partículas y ha servido como base para la construcción de la mecánica cuántica como se afirma en [30].

Las ecuaciones de Hamilton son un sistema de $2n$ ecuaciones diferenciales de primer orden equivalentes a las n ecuaciones diferenciales de segundo orden de Euler Lagrange, presentadas en (2.10), y surgen al aplicar una transformada de Legendre al lagrangiano L . El concepto de ésta transformación se presenta en la definición 1.

Definición 1 (Transformada de Legendre). *Dada una función $f(x_1, x_2, \dots, x_n)$ cuyas parciales $\frac{\partial f}{\partial x_i}$ existen y son distintas de cero, se define la transformada de Legendre de f como la función*

$$g(\mathbf{s}, \mathbf{x}) = \sum_{i \in S} s_i x_i - f(x_1, x_2, \dots, x_n),$$

donde S es un subconjunto de $\{1, 2, \dots, n\}$, $s_i = \frac{\partial f}{\partial x_i}$, $\mathbf{s} = \{s_i : i \in S\}$ y $\mathbf{x} = \{x_i : i \notin S\}$.

De manera intuitiva, la transformada de Legendre de la función f es una función g que reemplaza un subconjunto de variables x_i con $i \in S$ del argumento de f , por las variables $s_i = \frac{\partial f}{\partial x_i}$; el ejemplo 2 ilustra esta idea.

Ejemplo 2. Transformada de Legendre. Supongamos que tenemos la función $f(x_1, x_2) = 2x_1^2 + x_2^2$ y que queremos cambiar la variable x_1 por s_1 , donde $s_1 = \frac{df}{dx_1} = 4x_1$, la transformada de Legendre en este caso está dada por

$$\begin{aligned} g(s_1, x_2) &= \sum_{i=1}^1 s_i x_i - f(x_1, x_2) \\ &= x_1 s_1 - f(x_1, x_2). \end{aligned} \quad (2.16)$$

Dado que $s_1 = 4x_1$ tenemos que $x_1 = \frac{s_1}{4}$, sustituyendo x_1 en (2.16) se obtiene que

$$\begin{aligned} g(s_1, x_2) &= \left(\frac{s_1}{4}\right) s_1 - f\left(\frac{s_1}{4}, x_2\right) \\ &= \frac{s_1^2}{4} - \left(2\left(\frac{s_1}{4}\right)^2 + x_2^2\right) \\ &= \frac{s_1^2}{4} - \left(\frac{s_1^2}{8} + x_2^2\right) \\ &= \frac{s_1^2}{8} + x_2^2. \end{aligned}$$

Por tanto, la transformada de Legendre que surge al cambiar la variable x_1 por s_1 es $g(s_1, x_2) = \frac{s_1^2}{8} + x_2^2$.

Supongamos ahora que queremos cambiar las variables x_1 y x_2 de la función f por las variables $s_1 = \frac{df}{dx_1} = 4x_1$ y $s_2 = \frac{df}{dx_2} = 2x_2$ respectivamente, la transformada de Legendre será la función

$$g(s_1, s_2) = \sum_{i=1}^2 x_i s_i - f(x_1, x_2). \quad (2.17)$$

Dado que $s_1 = 4x_1$ y $s_2 = 2x_2$, se sigue que $x_1 = \frac{s_1}{4}$ y $x_2 = \frac{s_2}{2}$, sustituyendo x_1 y x_2 en (2.17) se tiene que

$$\begin{aligned}
g(s_1, s_2) &= \sum_{i=1}^2 x_i s_i - f\left(\frac{s_1}{4}, \frac{s_2}{2}\right) \\
&= \frac{s_1^2}{4} + \frac{s_2^2}{2} - \left(2\left(\frac{s_1}{4}\right)^2 + \left(\frac{s_2}{2}\right)^2\right) \\
&= \frac{s_1^2}{4} + \frac{s_2^2}{2} - \left(\frac{s_1^2}{8} + \frac{s_2^2}{4}\right) \\
&= \frac{s_1^2}{8} + \frac{s_2^2}{4}.
\end{aligned}$$

Concluyendo, al cambiar ambas variables se obtiene la transformada $g(s_1, s_2) = \frac{s_1^2}{8} + \frac{s_2^2}{4}$.

Una propiedad importante de las transformadas de Legendre es que podemos recuperar la función f a partir de su transformada de Legendre g al aplicarle a ésta nuevamente una transformada de Legendre que cambie las variables s_i anexadas a g . Esto ocurre porque dada la función f y su transformada g

$$g(\mathbf{s}, \mathbf{x}) = \sum_{i \in S} s_i x_i - f(x_1, x_2, \dots, x_n),$$

se obtiene que

$$f(x_1, x_2, \dots, x_n) = \sum_{i \in S} s_i x_i - g(\mathbf{s}, \mathbf{x}),$$

y dado que $\frac{\partial g}{\partial s_i} = x_i$ con $i \in S$, se cumplirá por definición que f es la transformada de Legendre de g , esta propiedad se ilustra en el ejemplo 3.

Ejemplo 3. Retomemos el segundo caso del ejemplo 2, en donde se encontró la transformada $g(s_1, s_2) = \frac{s_1^2}{8} + \frac{s_2^2}{4}$ de la función f , mostraremos que si aplicamos una transformada de Legendre que cambie ambas variables a la función g , obtendremos de nuevo, a la función f .

Denotemos por g^* a la transformada de Legendre de la función g , dada por

$$f^*(x_1^*, x_2^*) = \sum_{i=1}^2 x_i^* s_i - g(s_1, s_2), \quad (2.18)$$

donde $x_1^* = \frac{\partial g}{\partial s_1} = \frac{s_1}{4}$ y $x_2^* = \frac{\partial g}{\partial s_2} = \frac{s_2}{2}$, de estas igualdades se sigue que $s_1 = 4x_1^*$ y $s_2 = 2x_2^*$, sustituyendo s_1 y s_2 en (2.18), obtenemos que

$$\begin{aligned}
f^*(x_1^*, x_2^*) &= 4x_1^{*2} + 2x_2^{*2} - g(4x_1^* + 2x_2^*) \\
&= 4x_1^{*2} + 2x_2^{*2} - \left(\frac{(4x_1^*)^2}{8} + \frac{(2x_2^*)^2}{4} \right) \\
&= 4x_1^{*2} + 2x_2^{*2} - (2x_1^{*2} + x_2^{*2}) \\
&= 2x_1^{*2} + x_2^{*2},
\end{aligned}$$

llegando a que $g^* = f$.

Si aplicamos una transformada de Legendre al lagrangiano, en donde se cambien las n velocidades generalizadas, obtendremos el hamiltoniano H , cuya regla de correspondencia está dada por

$$H(q(t), p(t), t) = \sum_{i=1}^n p_i(t) \frac{dq_i(t)}{dt} - L\left(q(t), \frac{dq(t)}{dt}, t\right), \quad (2.19)$$

donde

$$p_i(t) = \frac{\partial L\left(q(t), \frac{dq_i(t)}{dt}, t\right)}{\partial \left(\frac{dq_i(t)}{dt}\right)} \quad \text{con } i \in \{1, 2, \dots, n\}. \quad (2.20)$$

Las funciones $p_i(t)$ presentadas en (2.20), son conocidas como los momentos generalizados. Denotaremos por $p(t)$ al vector $(p_1(t), p_2(t), \dots, p_n(t))$ y definiremos como espacio fase, representado por Q con $Q \subseteq \mathbb{R}^{2n}$, al espacio al que pertenecen las coordenadas y momentos generalizados.

Lo que se consigue al aplicar la transformada de Legendre al lagrangiano es un mapeo del espacio fase de velocidades al espacio fase, en donde puede describirse el movimiento de las coordenadas y momentos generalizados a través de ecuaciones diferenciales de primer orden, con la particularidad de que podemos volver a aplicar una transformada de Legendre para obtener las velocidades generalizadas y así las trayectorias en el espacio fase de velocidades.

Con el objetivo de hacer menos enrevesada la notación del hamiltoniano (2.19), omitiremos escribir la dependencia del tiempo de las coordenadas, velocidades y momentos generalizados, como se muestra a continuación:

$$H(q, p, t) = \sum_{i=1}^n p_i \frac{dq_i}{dt} - L\left(q, \frac{dq}{dt}, t\right). \quad (2.21)$$

Las ecuaciones de Hamilton describen las trayectorias que siguen las coordenadas y momentos generalizados asociados a los sistemas de partículas en el espacio fase. Las primeras n ecuaciones de Hamilton surgen al calcular el cambio que sufre el hamiltoniano al desplazar infinitesimalmente la j -ésima coordenada generalizada

$$\frac{\partial H(q, p, t)}{\partial q_j} = \frac{\partial}{\partial q_j} \left(\sum_{i=1}^n p_i \frac{dq_i}{dt} - L \left(q, \frac{dq}{dt}, t \right) \right). \quad (2.22)$$

Aplicando la regla de la cadena para el tiempo t , las n coordenadas y n velocidades generalizadas en el lado derecho de la igualdad (2.22), se obtiene que

$$\begin{aligned} \frac{\partial H(q, p, t)}{\partial q_j} &= \sum_{i=1}^n p_i \frac{\partial \left(\frac{dq_i}{dt} \right)}{\partial q_j} - \sum_{i=1}^n \frac{\partial L}{\partial q_i} \frac{\partial q_i}{\partial q_j} \\ &\quad - \sum_{i=1}^n \frac{\partial L}{\partial \left(\frac{dq_i}{dt} \right)} \frac{\partial \left(\frac{dq_i}{dt} \right)}{\partial q_j} - \frac{\partial L}{\partial t} \frac{\partial t}{\partial q_j}, \end{aligned}$$

observemos que $\frac{\partial t}{\partial q_j} = 0$ y $\frac{\partial q_i}{\partial q_j} = 0$, ya que el tiempo no depende de las coordenadas generalizadas, ni q_i de q_j cuando $i \neq j$, por tanto,

$$\begin{aligned} \frac{\partial H(q, p, t)}{\partial q_j} &= \sum_{i=1}^n p_i \frac{\partial \left(\frac{dq_i}{dt} \right)}{\partial q_j} - \frac{\partial L}{\partial q_j} - \sum_{i=1}^n \frac{\partial L}{\partial \left(\frac{dq_i}{dt} \right)} \frac{\partial \left(\frac{dq_i}{dt} \right)}{\partial q_j} \\ &= \sum_{i=1}^n \left[p_i \frac{\partial \left(\frac{dq_i}{dt} \right)}{\partial q_j} - \frac{\partial L}{\partial \left(\frac{dq_i}{dt} \right)} \frac{\partial \left(\frac{dq_i}{dt} \right)}{\partial q_j} \right] - \frac{\partial L}{\partial q_j}, \end{aligned}$$

dado que $p_i = \frac{\partial L}{\partial \left(\frac{dq_i}{dt} \right)}$, se sigue que

$$\begin{aligned} &= \sum_{i=1}^n \left[p_i \frac{\partial \left(\frac{dq_i}{dt} \right)}{\partial q_j} - p_i \frac{\partial \left(\frac{dq_i}{dt} \right)}{\partial q_j} \right] - \frac{\partial L}{\partial q_j} \\ &= -\frac{\partial L}{\partial q_j}. \end{aligned} \quad (2.23)$$

Teniendo en cuenta las ecuaciones de Euler-Lagrange $\frac{\partial L}{\partial q_i(t)} = \frac{d}{dt} \left(\frac{\partial L}{\partial \left(\frac{dq_i}{dt} \right)} \right)$, presentadas en (2.10), podemos reescribir (2.23) como

$$\begin{aligned}\frac{\partial H(q, p, t)}{\partial q_j} &= -\frac{d}{dt} \left(\frac{\partial L}{\partial \left(\frac{dq_j}{dt} \right)} \right) \\ &= -\frac{dp_j}{dt},\end{aligned}$$

y por tanto,

$$\frac{\partial H(q, p, t)}{\partial q_j} = -\frac{dp_j}{dt}, \quad \text{para todo } j \in \{1, 2, \dots, n\}. \quad (2.24)$$

Las ecuaciones (2.24) son n de las $2n$ ecuaciones de Hamilton, para obtener las n restantes, debemos derivar respecto al j -ésimo momento generalizado,

$$\begin{aligned}\frac{\partial H(q, p, t)}{\partial p_j} &= \frac{\partial}{\partial p_j} \left(\sum_{i=1}^n p_i \frac{dq_i}{dt} - L \left(q, \frac{dq}{dt}, t \right) \right) \\ &= \sum_{i=1}^n \frac{dq_i}{dt} \frac{\partial p_i}{\partial p_j} - \frac{\partial L}{\partial p_j},\end{aligned}$$

nuevamente, observemos que $\frac{\partial L}{\partial p_j} = 0$ y $\frac{\partial p_i}{\partial p_j} = 0$, porque el lagrangiano no depende de los momentos generalizados ni p_i de p_j cuando $i \neq j$, así se sigue que las n ecuaciones de Hamilton restantes, son

$$\frac{\partial H(q, p, t)}{\partial p_j} = \frac{dq_j}{dt}, \quad \text{para todo } j \in \{1, 2, \dots, n\}. \quad (2.25)$$

Concluyendo, las $2n$ ecuaciones diferenciales de Hamilton dadas en (2.24) y (2.25) para $j \in \{1, 2, \dots, n\}$ están dadas por

$$\frac{\partial H(q, p, t)}{\partial p_j} = \frac{dq_j}{dt} \quad \text{y} \quad \frac{\partial H(q, p, t)}{\partial q_j} = -\frac{dp_j}{dt}. \quad (2.26)$$

La solución de las ecuaciones (2.26), determinan la trayectoria de las coordenadas y momentos generalizados asociados a un sistema de partículas.

En el caso concreto en el que se utiliza el lagrangiano estándar definido en (2.11), los momentos generalizados son

$$\begin{aligned}
p_i &= \frac{\partial L(q_i, \frac{dq_i}{dt})}{\partial (\frac{dq_i}{dt})} \\
&= \frac{\partial}{\partial (\frac{dq_i}{dt})} \left(T \left(\frac{dq_i}{dt} \right) - U(q_i) \right) \\
&= \frac{\partial}{\partial (\frac{dq_i}{dt})} \left(\frac{m}{2} \left(\frac{dq_i}{dt} \right)^2 - U(q_i) \right) \\
&= m \frac{dq_i}{dt}.
\end{aligned} \tag{2.27}$$

Con el resultado (2.27) y considerando al lagrangiano estándar en la función del hamiltoniano (2.21), se obtiene que

$$\begin{aligned}
H(q, p) &= \sum_{i=1}^n p_i \frac{dq_i}{dt} - L \left(q, \frac{dq}{dt} \right) \\
&= \sum_{i=1}^n m \frac{dq_i}{dt} \frac{dq_i}{dt} - \left(T \left(\frac{dq}{dt} \right) - U(q) \right) \\
&= \sum_{i=1}^n m \left(\frac{dq_i}{dt} \right)^2 - \left(T \left(\frac{dq}{dt} \right) - U(q) \right) \\
&= 2T \left(\frac{dq}{dt} \right) - \left(T \left(\frac{dq}{dt} \right) - U(q) \right) \\
&= T \left(\frac{dq}{dt} \right) + U(q).
\end{aligned} \tag{2.28}$$

Así, el hamiltoniano inducido por el lagrangiano estándar está dado por

$$H(q, p) = T \left(\frac{dq}{dt} \right) + U(q). \tag{2.29}$$

A (2.29) se le conoce como hamiltoniano estándar. Es interesante notar que el hamiltoniano estándar es la suma de la energía cinética y la energía potencial, por lo que es igual a la energía mecánica definida en la sección A.0.4.

Escribir la energía cinética del hamiltoniano estándar en función de los momentos generalizados dados en (2.27), será de utilidad en la sección 4.1 donde se presenta la construcción del método HMC. Si definimos a K como la energía cinética en función de los momentos generalizados, K tendrá la regla de correspondencia $K(p) = \frac{p^2}{2m}$, donde p^2 es el producto

punto del vector p consigo mismo. De esta forma será equivalente a la función T , pero su argumento ya no será la velocidad generalizada sino el momento generalizado, como se muestra a continuación:

$$\begin{aligned} K(p) &= \frac{p^2}{2m} \\ &= \frac{\left(m \frac{dq}{dt}\right)^2}{2m} \\ &= \frac{m \frac{dq^2}{dt}}{2} \\ &= T\left(\frac{dq}{dt}\right), \end{aligned}$$

por lo que podemos sustituir a $K(p)$ en (2.29) y obtenemos que

$$\begin{aligned} H(q, p) &= K(p) + U(q) \\ &= \frac{p^2}{2m} + U(q). \end{aligned} \tag{2.30}$$

Para obtener las ecuaciones de Hamilton presentadas en (2.26), en el caso concreto del hamiltoniano estándar (2.30), observemos que

$$\begin{aligned} \frac{\partial H(q, p)}{\partial p_j} &= \frac{\partial (K(p) + U(q))}{\partial p_j} \\ &= \frac{\partial (K(p))}{\partial p_j} \\ &= \frac{\partial \left(\frac{p^2}{2m}\right)}{\partial p_j} \\ &= \frac{\partial}{\partial p_j} \left(\frac{\sum_{i=1}^n p_i^2}{2m}\right) \\ &= \frac{p_j}{m}. \end{aligned} \tag{2.31}$$

Además, notemos que

$$\begin{aligned} \frac{\partial H(q, p)}{\partial q_j} &= \frac{\partial (K(p) + U(q))}{\partial q_j} \\ &= \frac{\partial U(q)}{\partial q_j}. \end{aligned} \tag{2.32}$$

En conclusión, las ecuaciones presentadas en (2.26) para el caso del hamiltoniano estándar, sustentados en (2.31) y (2.32), están dadas por

$$\frac{p_j}{m} = \frac{dq_j}{dt} \quad \text{y} \quad \frac{\partial U(q)}{\partial q_j} = \frac{-dp_j}{dt} \quad \text{para } j \text{ en } \{1, 2, \dots, n\}. \quad (2.33)$$

El hamiltoniano (2.30) y las ecuaciones de Hamilton (2.33) son fundamentales para la construcción del algoritmo HMC presentado en la sección 4.1.

Ejemplo 4. Resolveremos el mismo problema que en el ejemplo 1, pero ahora bajo el enfoque de la mecánica hamiltoniana, con el objetivo de mostrar cómo aplicar las ecuaciones de Hamilton y resaltar que el enfoque newtoniano, lagrangiano y hamiltoniano conducen a los mismos resultados.

Consideremos a $q(t)$ como la coordenada generalizada, que al tiempo t devuelve la posición de la partícula del sistema masa-resorte en el eje x . Recordemos además que en el ejemplo 7 se dedujo que la energía potencial para este sistema está dada por $U(q) = \frac{kq^2}{2}$, de tal manera que el hamiltoniano estándar presentado en (2.30) para este ejemplo está dado por

$$\begin{aligned} H(q, p) &= K(p) + U(q) \\ &= \frac{p^2}{2m} + \frac{kq^2}{2}. \end{aligned} \quad (2.34)$$

Para aplicar las ecuaciones de Hamilton asociadas al hamiltoniano estándar presentadas en (2.33), notemos que

$$\begin{aligned} \frac{\partial U(q)}{\partial q} &= \frac{\partial \left(\frac{kq^2}{2} \right)}{\partial q} \\ &= kq. \end{aligned} \quad (2.35)$$

Por el resultado (2.35) y (2.33), se sigue que las ecuaciones de Hamilton están dadas por

$$\frac{p}{m} = \frac{dq}{dt} \quad \text{y} \quad kq = -\frac{dp}{dt}. \quad (2.36)$$

Las incógnitas del sistema de ecuaciones diferenciales (2.36) son la coordenada generalizada q y el momento generalizado p . Notemos que si despejamos a p en la primer ecuación de (2.36) y derivamos respecto al tiempo obtenemos que

$$\frac{dp}{dt} = m \frac{d^2q}{dt^2}. \quad (2.37)$$

Si sustituimos (2.37) en la segunda ecuación de (2.36), se obtiene que

$$kq = -m \frac{d^2q}{dt^2}$$

o bien, $m \frac{d^2q}{dt^2} + kq = 0,$ (2.38)

que es la misma ecuación diferencial a la que se llegó en el ejemplo 1, cuya solución es $q(t) = A \operatorname{sen}(wt)$.

2.1.5. Propiedades del Hamiltoniano

En esta sección se presentan propiedades de las dinámicas hamiltonianas que resultan imprescindibles para la construcción del método HMC.

2.1.5.1. Reversibilidad

Con el objetivo de explicar el concepto de reversibilidad en las dinámicas hamiltonianas, definamos a $\delta : \mathbb{R} \rightarrow \mathbb{R}^n$ como la función que a un tiempo t devuelve el punto del espacio fase en el que se encuentra un sistema de partículas, es decir, $\delta(t) = (q_1(t), \dots, q_n(t), p_1(t), \dots, p_n(t))$. De este modo, se puede definir la reversibilidad como la capacidad que tiene el sistema de partículas de volver al estado $\delta(t)$ partiendo de $\delta(t + \varepsilon)$ con $\varepsilon > 0$.

La manera en que se consigue la reversibilidad en las dinámicas hamiltonianas, como se afirma en [8] y se enuncia en el teorema 1, es siguiendo en un periodo de tiempo ε la trayectoria que surge de las ecuaciones de Hamilton cambiando el signo del momento generalizado p .

Teorema 1. *Las dinámicas hamiltonianas asociadas al hamiltoniano estándar son reversibles bajo la transformación $H(q, p) \rightarrow H(q, -p)$.*

Demostración. Al cambiar el signo al momento p , las ecuaciones de Hamilton quedan determinadas por

$$\frac{p_j}{m} = -\frac{dq_j}{dt} \quad \text{y} \quad \frac{\partial U(q)}{\partial q_j} = \frac{dp_j}{dt} \quad \text{para } j \text{ en } \{1, 2, \dots, n\} \quad (2.39)$$

dado que $-\frac{dq_j}{dt} = \frac{dq_j}{d(-t)}$ y $\frac{dp_j}{dt} = -\frac{dp_j}{d(-t)}$, se sigue que

$$\frac{p_j}{m} = \frac{dq_j}{d(-t)} \quad \text{y} \quad \frac{\partial U(q)}{\partial q_j} = -\frac{dp_j}{d(-t)} \quad (2.40)$$

lo que prueba que al considerar $H(q, -p)$ en lugar de $H(q, p)$, se obtienen las ecuaciones de Hamilton reversibles en el tiempo, es decir, con la transformación $t \rightarrow -t$.

□

Poder obtener las dinámicas inversas al cambiar el signo de los momentos en las ecuaciones de Hamilton resulta de utilidad en el método HMC, porque facilita obtener las trayectorias inversas en el algoritmo. Conseguir la reversibilidad es fundamental ya que permite que la cadena de Markov generada por el algoritmo HMC, satisfaga las ecuaciones de balance detallado, presentadas en 4.2.2, y además logra que se puedan proponer nuevos estados para la cadena de Markov que no anulen la razón de aceptación y rechazo, como se explica en la sección 4.1.

2.1.5.2. Conservación

Otra propiedad que es importante para la construcción del método HMC, es que las dinámicas hamiltonianas, mantienen constante al hamiltoniano estándar $H(q, p)$, presentado en (2.30).

Teorema 2. *El valor del hamiltoniano estándar $H(q, p)$, permanece invariante en las trayectorias de las dinámicas hamiltonianas.*

Demostración. Para probarlo, notaremos que se cumple que $\frac{\partial H(q, p)}{\partial t} = 0$.

Obteniendo la derivada parcial respecto al tiempo de $H(q, p)$, tenemos que

$$\frac{\partial H(q, p)}{\partial t} = \sum_{i=1}^n \frac{\partial H}{\partial q_i} \frac{dq_i}{dt} + \sum_{i=1}^n \frac{\partial H}{\partial p_i} \frac{dp_i}{dt}, \quad (2.41)$$

por las ecuaciones de Hamilton dadas en (2.26), sabemos que $\frac{\partial H}{\partial p_i} = \frac{dq_i}{dt}$ y $\frac{\partial H}{\partial q_i} = -\frac{dp_i}{dt}$, sustituyendo estas igualdades en (2.41), se sigue que

$$\begin{aligned}
\frac{\partial H}{\partial t} &= \sum_{i=1}^n \frac{dq_i}{dt} \frac{dp_i}{dt} - \sum_{i=1}^n \frac{dp_i}{dt} \frac{dq_i}{dt} \\
&= \sum_{i=1}^n \left(\frac{dq_i}{dt} \frac{dp_i}{dt} - \frac{dq_i}{dt} \frac{dp_i}{dt} \right) \\
&= \sum_{i=1}^n 0 \\
&= 0.
\end{aligned}$$

□

Este resultado implica que la probabilidad de aceptación en el método Monte Carlo Hamiltoniano sea uno, y que los puntos en las trayectorias de las dinámicas hamiltonianas tengan la misma densidad asociada de acuerdo a la distribución de Boltzman, como se menciona en la sección 4.1.

2.1.5.3. Preservacion del volumen

Las dinámicas hamiltonianas preservan su volumen en el espacio fase de momentos, a este resultado se le conoce como el teorema de Liouville.

Teorema 3. *Teorema de Liouville.* El volumen de una región ω en el espacio fase de momentos se conserva si los puntos de su frontera $d(\omega)$ se mueven de acuerdo a las ecuaciones de Hamilton.

Demostración. Las ecuaciones de Hamilton describen la evolución en el tiempo de la función $\frac{d\delta(t)}{dt} = \left(\frac{dq_1}{dt}, \frac{dq_2}{dt}, \dots, \frac{dq_n}{dt}, \frac{dp_1}{dt}, \frac{dp_2}{dt}, \dots, \frac{dp_n}{dt} \right)$, para probar el teorema de Liouville basta con demostrar que el campo vectorial $\frac{d\delta}{dt}$ tiene divergencia cero, porque como se demuestra en [1, pp.69-70], los campos vectoriales con divergencia nula preservan su volumen.

Notemos que en efecto, la divergencia de $\frac{d\delta}{dt}$ es cero,

$$\begin{aligned}
div \left(\frac{d\delta}{dt} \right) &= \left(\frac{\partial}{\partial q_1}, \dots, \frac{\partial}{\partial q_n}, \frac{\partial}{\partial p_1}, \dots, \frac{\partial}{\partial p_n} \right) \left(\frac{dq_1}{dt}, \dots, \frac{dq_n}{dt}, \frac{dp_1}{dt}, \dots, \frac{dp_n}{dt} \right)^t \\
&= \sum_{i=1}^n \frac{\partial}{\partial q_i} \left(\frac{dq_i}{dt} \right) + \sum_{i=1}^n \frac{\partial}{\partial p_i} \left(\frac{dp_i}{dt} \right),
\end{aligned} \tag{2.42}$$

sutituyendo las ecuaciones de Hamilton (2.26) en (2.42), obtenemos que

$$\begin{aligned}
\operatorname{div} \left(\frac{d\delta}{dt} \right) &= \sum_{i=1}^n \frac{\partial}{\partial q_i} \left(\frac{\partial H}{\partial p_i} \right) - \sum_{i=1}^n \frac{\partial}{\partial p_i} \left(\frac{\partial H}{\partial q_i} \right) \\
&= \sum_{i=1}^n \frac{\partial^2 H}{\partial q_i \partial p_i} - \sum_{i=1}^n \frac{\partial^2 H}{\partial q_i \partial p_i} \\
&= 0,
\end{aligned}$$

probando así el teorema de Liouville. □

La importancia de la preservación del volumen para el método HMC como se explica en [8, p.117], radica en que si no se preserva, se tendría que calcular el determinante de la matriz jacobiana de la transformación cada vez que se propusiera un nuevo estado en la cadena, para modificar de manera oportuna la razón de aceptación y rechazo, lo que resultaría costoso computacionalmente.

2.1.5.4. Dinámicas simplécticas

Para introducir la propiedad simpléctica de las dinámicas hamiltonianas, es conveniente tener presente las definiciones (2) y (3).

Definición 2 (Matriz simpléctica). *Decimos que una matriz M es simpléctica si $M^t J M = J$, con $J = \begin{pmatrix} 0_{n \times n} & I_{n \times n} \\ -I_{n \times n} & 0_{n \times n} \end{pmatrix}$.*

Definición 3 (Transformación simpléctica). *Se dice que una transformación φ es simpléctica si su matriz jacobiana M es simpléctica.*

Notemos que si una transformación φ es simpléctica, entonces el valor absoluto de su matriz jacobiana M , es uno, ya que

$$\begin{aligned}
\det(M^t) \det(J) \det(M) &= \det(J) \\
\det(M)^2 &= 1 \\
|\det(M)| &= 1,
\end{aligned} \tag{2.43}$$

este resultado se utiliza en la sección 3.1, donde se prueba que el método de Euler es un método numérico no simpléctico. Además, es importante tener presente que la transformación φ , preserva el volumen, debido a que su matriz jacobiana M satisface la igualdad

(2.43), que como se afirma en [8], implica la preservación del volumen.

El teorema 4, probado en 1899 por Henri Poincaré³, bajo la hipótesis de un hamiltoniano $H(q, p)$ con segundas derivadas parciales continuas, sostiene que las dinámicas hamiltonianas son simplécticas, su demostración se puede consultar en [6].

Teorema 4. *Si $H(q, p)$ es un hamiltoniano con segundas derivadas parciales continuas, entonces la transformación implicada por las ecuaciones de Hamilton asociadas a $H(q, p)$ al tiempo t , es una transformación simpléctica.*

La propiedad de preservación de volumen en las dinámicas hamiltonianas, es también una consecuencia de que las dinámicas sean simplécticas, ya que como sostiene el teorema 4, la transformación implicada por las ecuaciones de Hamilton en el tiempo es simpléctica y este tipo de transformaciones preserva el volumen, como se mencionó anteriormente.

En la sección 3.1, se debe tener presente la propiedad simpléctica de las dinámicas hamiltonianas, porque como se exhibe en ésta, al utilizar métodos numéricos que posean la propiedad simpléctica se obtienen aproximaciones más exactas de las soluciones de las ecuaciones de Hamilton, las cuales son indispensables para la aplicación del algoritmo HMC.

³Jules Henri Poincaré (1854-1912), matemático francés, científico teórico y filósofo de la ciencia, considerado a menudo como el último “universalista”.

Capítulo 3

Algoritmos y temas precedentes

Para aplicar el método HMC, es necesario conocer o tener una aproximación de la solución de las ecuaciones de Hamilton presentadas en el capítulo 2. Generalmente estas ecuaciones no tienen una solución analítica por lo que se debe recurrir a métodos numéricos. En este capítulo se introduce el método de Euler, Euler simpléctico y el algoritmo Leapfrog, se ejemplifica cómo utilizarlos para aproximar las dinámicas hamiltonianas, y se comparan con el objetivo de resaltar qué método numérico es más adecuado utilizar para el algoritmo HMC.

Por otro lado, se introducen los métodos de cadenas de Markov Monte Carlo que son la base del algoritmo HMC, en particular, se profundiza en el algoritmo *Random Walk Metropolis-Hastings* que es uno de los más conocidos y utilizados por su sencillez.

3.1. Métodos para solución de ecuaciones diferenciales ordinarias

Las trayectorias que siguen las coordenadas y los momentos generalizados están determinadas por las ecuaciones de Hamilton (2.26), cuya solución frecuentemente no tiene forma analítica, debido a esto se deben aproximar las trayectorias con métodos numéricos.

Aunque pueden utilizarse diversos algoritmos para aproximar las soluciones de las trayectorias, es conveniente utilizar métodos numéricos simplécticos (su concepto se da en la definición 4), porque producen aproximaciones más exactas, debido a que comparten la propiedad simpléctica de las dinámicas hamiltonianas y en consecuencia poseen propiedades como la preservación del volumen, además de que conservan la estructura simpléctica de las dinámicas.

Definición 4 (Método numérico simpléctico). *Un método numérico se denomina simpléctico si su transformación a un paso de tamaño ϵ dada por*

$$y_1 = \Phi_\epsilon(y_0),$$

es una transformación simpléctica cuando se aplica a un sistema hamiltoniano suave.

En esta sección presentaremos dos métodos numéricos simplécticos y el método de Euler que no lo es, con el objetivo de comparar sus aproximaciones.

3.1.1. Método de Euler

El método de Euler¹, como su nombre indica, fue propuesto por Leonhard Euler, en el año 1768. Es un método numérico útil para aproximar soluciones de ecuaciones diferenciales ordinarias con condiciones iniciales.

Dado un sistema de ecuaciones diferenciales de la forma

$$\begin{cases} \frac{dx_1}{dt} = f_1(x_1, x_2, \dots, x_m) \\ \frac{dx_2}{dt} = f_2(x_1, x_2, \dots, x_m) \\ \dots \\ \frac{dx_m}{dt} = f_m(x_1, x_2, \dots, x_m), \end{cases} \quad (3.1)$$

con condiciones iniciales $x_i(t_0) = x_i^\alpha$ y $y_i(t_0) = y_i^\alpha$ con $i \in \{1, 2, \dots, m\}$. El método de Euler permite aproximar las soluciones del sistema (3.1) a través de la sucesión iterativa

$$x_i(t + \epsilon) = x_i(t) + \epsilon f(x_1, x_2, \dots, x_m), \quad (3.2)$$

en donde ϵ representa el tamaño de paso.

Para generar un aproximación de la solución de las ecuaciones de Hamilton inducidas por el hamiltoniano estándar con el método de Euler recordemos que, como se dedujo en (2.33), este sistema está dado por

$$\begin{cases} \frac{p_j}{m} = \frac{dq_j}{dt} \\ \frac{\partial U(q)}{\partial q_j} = -\frac{dp_j}{dt} \end{cases} \quad \text{con } j \in \{1, 2, \dots, n\}. \quad (3.3)$$

Teniendo presente el sistema (3.3) y la sucesión iterativa (3.2), se puede deducir que la aproximación para este sistema está dada por

¹Leonhard Paul Euler (1707-1783), matemático y físico suizo, de los más destacados de la historia.

$$q_i(t + \epsilon) = q_i(t) + \epsilon \frac{p(t)}{m_i}, \quad (3.4)$$

$$p_i(t + \epsilon) = p_i(t) - \epsilon \frac{\partial U(q_i(t))}{\partial q_i}. \quad (3.5)$$

Es fácil notar que este algoritmo no es un método numérico simpléctico mostrando que el valor absoluto del determinante de la matriz jacobiana de la transformación dada por (3.4) y (3.5) en general es distinto de uno, esto se debe a que como se explicó en (2.43) todas las matrices jacobianas de transformaciones simplécticas tienen determinante uno.

$$\begin{aligned} |\det(J(q_i(t + \epsilon), p_i(t + \epsilon)))| &= \left| \det \begin{pmatrix} \frac{\partial q_i(t+\epsilon)}{\partial q_i(t)} & \frac{\partial q_i(t+\epsilon)}{\partial p_i(t)} \\ \frac{\partial p_i(t+\epsilon)}{\partial q_i(t)} & \frac{\partial p_i(t+\epsilon)}{\partial p_i(t)} \end{pmatrix} \right| \\ &= \left| \det \begin{pmatrix} 1 & \frac{\epsilon}{m_i} \\ -\epsilon \frac{\partial^2 U(q_i(t))}{\partial^2 q_i} & 1 \end{pmatrix} \right| \\ &= \left| 1 + \frac{\epsilon^2}{m_i} \frac{\partial^2 U(q_i(t))}{\partial^2 q_i} \right| \\ &\neq 1. \end{aligned}$$

3.1.2. Método de Euler Simpléctico

El método de Euler simpléctico es una modificación del método de Euler, que puede ser consultada en [18]. Este método permite aproximar soluciones de sistemas de ecuaciones diferenciales de la forma

$$\begin{cases} \frac{dx}{dt} = \psi(x, y) \\ \frac{dy}{dt} = \phi(x, y) \end{cases},$$

a partir de las sucesiones

$$x(t + \epsilon) = x(t) + \epsilon \psi(x(t), y(t + \epsilon)), \quad (3.6)$$

$$y(t + \epsilon) = y(t) + \epsilon \phi(x(t), y(t + \epsilon)). \quad (3.7)$$

El método de Euler simpléctico es un método implícito, ya que la variable $y(t + \epsilon)$ no está dada de manera explícita en (3.6) ni (3.7). Para aproximar su valor debe solucionarse el sistema de ecuaciones.

Al aplicar el método de Euler simpléctico a las ecuaciones de Hamilton (2.26), se obtienen las sucesiones

$$\begin{aligned} p_i(t + \varepsilon) &= p_i(t) - \varepsilon \frac{\partial H(q(t), p(t + \varepsilon), t)}{\partial q_i}, \\ q_i(t + \varepsilon) &= q_i(t) + \varepsilon \frac{\partial H(q(t), p(t + \varepsilon), t)}{\partial p_i}. \end{aligned}$$

Notemos que cuando se aplica este método a las ecuaciones de Hamilton inducidas por el hamiltoniano estándar, presentadas en (2.33), no es necesario resolver el sistema de ecuaciones para aproximar las variables implícitas, ya que las iteraciones, en este caso, quedan determinadas de forma explícita, como se muestra a continuación

$$\begin{aligned} p_i(t + \varepsilon) &= p_i(t) - \varepsilon \frac{\partial H(q(t), p(t + \varepsilon))}{\partial q_i} \\ &= p_i(t) - \varepsilon \frac{\partial (K(p(t + \varepsilon)) + U(q(t)))}{\partial q_i} \\ &= p_i(t) - \varepsilon \frac{\partial U(q(t))}{\partial q_i}, \end{aligned}$$

y,

$$\begin{aligned} q_i(t + \varepsilon) &= q_i(t) + \varepsilon \frac{\partial H(q(t), p(t + \varepsilon))}{\partial p_i} \\ &= q_i(t) + \varepsilon \frac{\partial (K(p(t + \varepsilon)) + U(q(t)))}{\partial p_i} \\ &= q_i(t) + \varepsilon \frac{\partial K(p(t + \varepsilon))}{\partial p_i} \\ &= q_i(t) + \varepsilon \frac{p_i(t + \varepsilon)}{m_i}. \end{aligned}$$

Concluyendo, las sucesiones para aproximar las ecuaciones de Hamilton (2.33), quedan determinadas de forma explícita y son

$$\begin{aligned} p_i(t + \varepsilon) &= p_i(t) - \varepsilon \frac{\partial U(q(t))}{\partial q_i}, \text{ y} \\ q_i(t + \varepsilon) &= q_i(t) + \varepsilon \frac{p_i(t + \varepsilon)}{m_i}. \end{aligned}$$

En [18, p.189] se demuestra que este algoritmo es un método numérico simpléctico cuando es empleado para aproximar la solución de las ecuaciones de Hamilton inducidas por un hamiltoniano de la forma $H = K(p) + U(q)$, como es el caso del hamiltoniano estándar. Para demostrarlo se prueba que la matriz jacobiana de la transformación inducida por el método numérico es una matriz simpléctica.

3.1.3. Método Leapfrog

El método Leapfrog también llamado Störmer-Verlet *Scheme* fue propuesto por Loup Verlet² en el año 1967 con el objetivo de estudiar las dinámicas moleculares como se menciona en [18, p.7].

La aproximación de la solución de las ecuaciones de Hamilton (2.26), generada con el método Leapfrog está dado por los siguientes pasos iterativos

$$\begin{aligned} p_i\left(t + \frac{\varepsilon}{2}\right) &= p_i(t) - \frac{\varepsilon}{2} \frac{\partial H(q(t), p(t + \frac{\varepsilon}{2}), t)}{\partial q_i}, \\ q_i(t + \varepsilon) &= q_i(t) + \frac{\varepsilon}{2} \left(\frac{\partial H(q(t), p(t + \frac{\varepsilon}{2}))}{\partial p_i} + \frac{\partial H(q(t + \varepsilon), p(t + \frac{\varepsilon}{2}), t)}{\partial p_i} \right), \\ p_i(t + \varepsilon) &= p_i\left(t + \frac{\varepsilon}{2}\right) - \frac{\varepsilon}{2} \left(\frac{\partial H(q(t + \varepsilon), p(t + \frac{\varepsilon}{2}), t)}{\partial q_i} \right). \end{aligned}$$

Del mismo modo que el algoritmo de Euler simpléctico, el algoritmo Leapfrog es un método numérico implícito que cuando se aplica a las ecuaciones de Hamilton (2.33) inducidas por el hamiltoniano estándar, queda determinado de manera explícita, como se muestra a continuación

$$\begin{aligned} p_i\left(t + \frac{\varepsilon}{2}\right) &= p_i(t) - \frac{\varepsilon}{2} \frac{\partial U(q(t))}{\partial q_i} \\ q_i(t + \varepsilon) &= q_i(t) + \frac{\varepsilon}{2} \left(\frac{p_i\left(t + \frac{\varepsilon}{2}\right)}{m_i} + \frac{p_i\left(t + \frac{\varepsilon}{2}\right)}{m_i} \right) \\ &= q_i(t) + \varepsilon \left(\frac{p_i\left(t + \frac{\varepsilon}{2}\right)}{m_i} \right) \\ p_i(t + \varepsilon) &= p_i\left(t + \frac{\varepsilon}{2}\right) - \frac{\varepsilon}{2} \left(\frac{\partial U(q(t + \varepsilon))}{\partial q_i} \right). \end{aligned}$$

²(1931-2019), físico francés pionero en simulaciones computacionales de dinámicas moleculares.

La demostración de que este método numérico es simpléctico cuando se aplica a las ecuaciones de Hamilton implicadas por un hamiltoniano de la forma $H = K(p) + U(q)$ puede encontrarse en [18, p.190].

3.1.4. Comparación

Retomemos el ejemplo 4, en donde se obtuvo el hamiltoniano $H(q, p) = \frac{p^2}{2m} + \frac{kq^2}{2}$ y las ecuaciones de Hamilton

$$\frac{p}{m} = \frac{dq}{dt} \quad \text{y} \quad kq = -\frac{dp}{dt}, \quad (3.8)$$

consideremos además, los parámetros propuestos en el ejemplo 8 para este sistema de partículas, es decir una masa $m = 2Kg$ y la proporcionalidad del resorte, $k = 30N/m$, de tal forma que el sistema (3.9) pueda reescribirse como

$$\frac{p}{2} = \frac{dq}{dt} \quad \text{y} \quad 30q = -\frac{dp}{dt}. \quad (3.9)$$

Supongamos ahora, las condiciones iniciales $p(0) = 0$ y $q(0) = 1$. Si despejamos a p de la primer ecuación de (3.9) obtenemos que $p = 2\frac{dq}{dt}$, derivando respecto al tiempo en ambos lados de la igualdad se sigue que $\frac{dp}{dt} = 2\frac{d^2q}{dt^2}$ y al sustituirlo en la segunda ecuación de (3.9) obtenemos la ecuación diferencial

$$2\frac{d^2q}{dt^2} + 30q = 0,$$

equivalente a

$$\frac{d^2q}{dt^2} + 15q = 0. \quad (3.10)$$

La ecuación (3.10), es una ecuación de segundo orden con coeficientes constantes, por lo que su solución será proporcional a $e^{\lambda t}$ para alguna constante λ , y por tanto se cumplirá que

$$\begin{aligned} 0 &= \frac{d^2}{dt^2} (e^{\lambda t}) + 15e^{\lambda t} \\ &= \lambda^2 e^{\lambda t} + 15e^{\lambda t} \\ &= e^{\lambda t} (\lambda^2 + 15). \end{aligned} \quad (3.11)$$

Dado que $e^{\lambda t} > 0$, (3.11) implica que

$$\begin{aligned} 0 &= \lambda^2 + 15 \\ \sqrt{-15} &= \lambda, \end{aligned}$$

por lo que las posibles raíces de λ están dadas por

$$\lambda_1 = -i\sqrt{15} \quad \text{o} \quad \lambda_2 = i\sqrt{15}.$$

Así, se sigue que la solución de (3.10) es de la forma

$$q(t) = c_1 e^{i\sqrt{15}t} + c_2 e^{-i\sqrt{15}t}. \quad (3.12)$$

Aplicando la ecuación de Euler $e^{\alpha+i\beta} = e^\alpha \cos(\beta) + ie^\alpha \sin(\beta)$ a (3.12), obtenemos que

$$\begin{aligned} q(t) &= c_1 \left(\cos(\sqrt{15}t) + i \operatorname{sen}(\sqrt{15}t) \right) + c_2 \left(\cos(\sqrt{15}t) - i \operatorname{sen}(\sqrt{15}t) \right) \\ &= (c_1 + c_2) \cos(\sqrt{15}t) + i(c_1 - c_2) \sin(\sqrt{15}t). \end{aligned}$$

Dado que c_1 y c_2 son constantes arbitrarias, es conveniente definir $\alpha_1 = i(c_1 - c_2)$ y $\alpha_2 = c_1 + c_2$, de modo que la solución de la ecuación diferencial (3.10) se reduzca a

$$q(t) = \alpha_1 \operatorname{sen}(\sqrt{15}t) + \alpha_2 \cos(\sqrt{15}t). \quad (3.13)$$

Además, por la condición inicial $q(0) = 1$ y (3.13),

$$\begin{aligned} 1 &= \alpha_1 \operatorname{sen}(0) + \alpha_2 \cos(0) \\ 1 &= \alpha_2. \end{aligned} \quad (3.14)$$

Sustituyendo el valor de (3.14) en (3.13), se sigue que

$$q(t) = \alpha_1 \operatorname{sen}(\sqrt{15}t) + \cos(\sqrt{15}t). \quad (3.15)$$

Al derivar (3.15), se llega a que

$$\frac{dq(t)}{dt} = \alpha_1 \sqrt{15} \cos(\sqrt{15}t) - \sqrt{15} \operatorname{sen}(\sqrt{15}t), \quad (3.16)$$

por lo que podemos obtener la función p al sustituir el valor de (3.16) en la igualdad $p = 2 \frac{dq}{dt}$ que surge al despejar a p de la primer ecuación del sistema (3.9), así obtenemos que

$$p(t) = 2 \left(\alpha_1 \sqrt{15} \cos(\sqrt{15}t) - \sqrt{15} \operatorname{sen}(\sqrt{15}t) \right). \quad (3.17)$$

Por la condición inicial $p(0) = 0$,

$$\begin{aligned}
0 &= 2 \left(\alpha_1 \sqrt{15} \cos(0) - \sqrt{15} \operatorname{sen}(0) \right) \\
&= 2\alpha_1 \sqrt{15} \\
&= \alpha_1,
\end{aligned} \tag{3.18}$$

y finalmente al sustituir (3.18) en (3.17), se obtiene que la función p está dada por

$$p(t) = -2\sqrt{15} \operatorname{sen}(\sqrt{15}t). \tag{3.19}$$

Para obtener la función q sustituimos (3.18) en (3.15), resultando que

$$q(t) = \cos(\sqrt{15}t). \tag{3.20}$$

Concluyendo, las trayectorias en el tiempo que siguen las coordenadas generalizadas y los momentos generalizados en este sistema, están dadas por la solución (3.19) y (3.20) del sistema de ecuaciones diferenciales (3.9).

La imagen 3.1 muestra las trayectorias que siguen los momentos generalizados y las coordenadas generalizadas en el espacio fase en un intervalo de tiempo $[0, \frac{2\pi}{w}]$ donde $w = \sqrt{\frac{k}{m}}$, como se mencionó en el ejemplo 7, $\frac{2\pi}{w}$ es el tiempo que tarda la partícula del sistema masa-resorte en dar un ciclo completo.

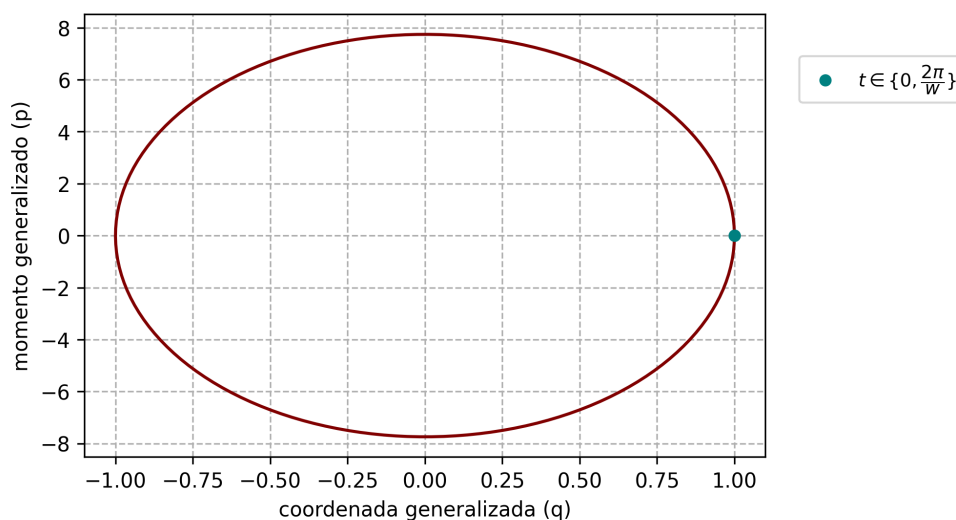


Figura 3.1: Trayectoria en el espacio fase del sistema masa-resorte.

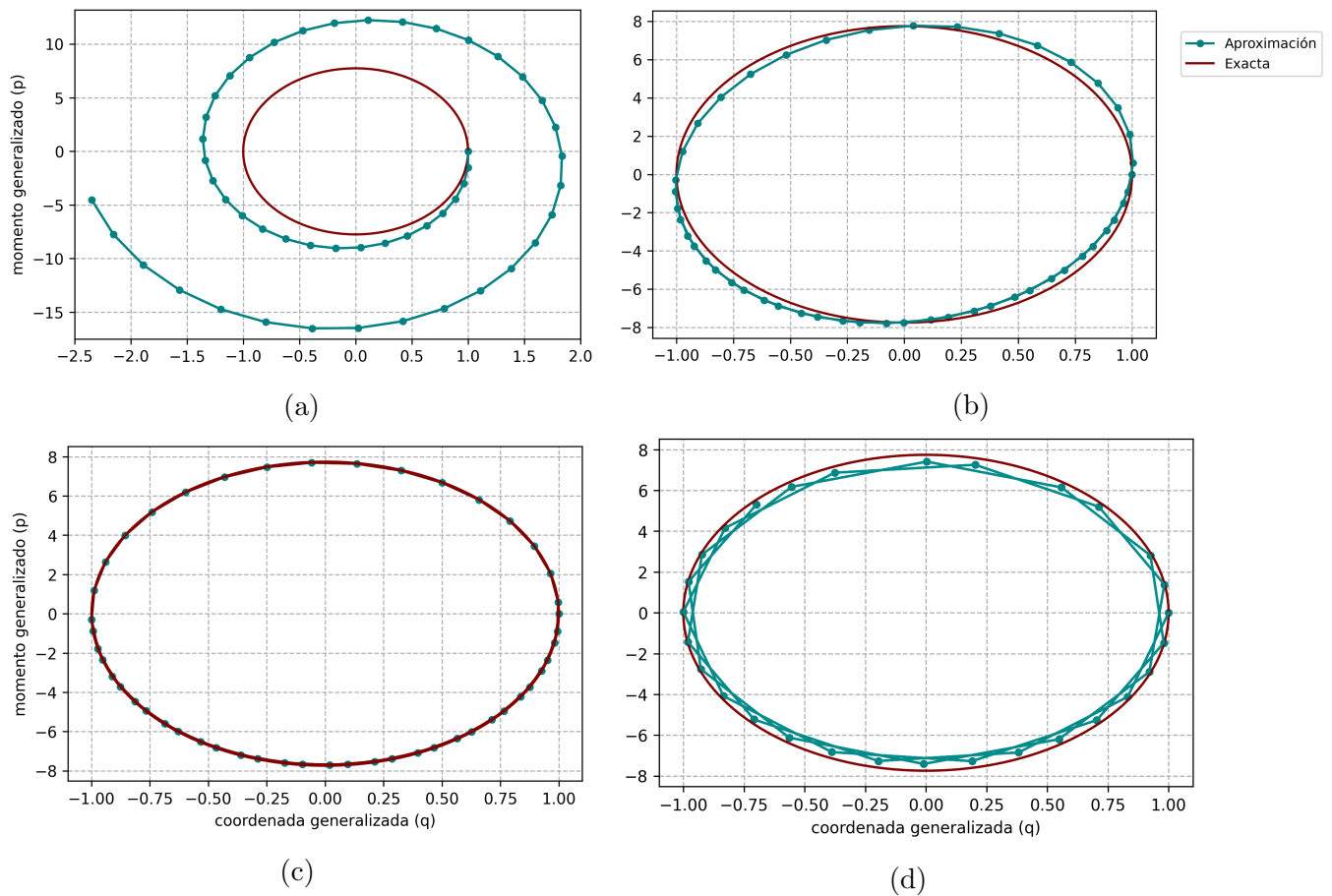


Figura 3.2: Las aproximaciones por el método de Euler, Euler simpléctico y Leapfrog presentadas en (a), (b) y (c) respectivamente, muestran 50 iteraciones con un tamaño de paso $\epsilon = 0.05$. La imagen (d) muestra una aproximación de 50 iteraciones con el algoritmo Leapfrog utilizando $\epsilon = 0.15$.

Las aproximaciones de la solución del sistema de ecuaciones (3.9), obtenidas al utilizar los métodos numéricos descritos en esta sección, gráficamente se presentan en la imagen 3.2.

Como se observa en la figura 3.2 (a), la aproximación con el método de Euler no es buena, produce resultados alejados de los valores exactos de la trayectoria y con el tiempo diverge.

El resultado obtenido con el método de Euler simpléctico graficado en 3.2 (b), produce mejores resultados: se elimina la divergencia y la aproximación es más cercana a la solución analítica.

Por último, la figura 3.2 (c), muestra la aproximación con el algoritmo Leapfrog, notemos que en esta imagen el error no es perceptible, y que incluso tomando un tamaño de paso tres veces más grande, $\epsilon = 0.15$, la aproximación que se consigue es muy cercana a la trayectoria exacta como se muestra en 3.2 (d).

En este [repositorio](#) se puede consultar el código en Python del método de Euler, Euler simplético, Leapfrog y los ejemplos presentados en las imágenes 3.1 y 3.2.

Consideremos ahora el hamiltoniano estándar $H(q, p) = K(p) + U(q)$ asociado a una partícula de masa $m = 1$, es decir, donde la energía cinética $K(p) = \frac{p^2}{2}$ y definamos a la energía potencial U como $U(q) = -\log(f(q, 5, 1))$, donde $f(x, \alpha, \beta)$ es la función de densidad de la distribución Gamma dada en (3.21). Al definir a $H(q, p)$ de esta manera se puede simular de la distribución $Gamma(5, 1)$ empleando el método HMC, como se muestra en la sección 4.3.1.

$$f(x, \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}. \quad (3.21)$$

Notemos que la función $U(q) = -\log(f(q, 5, 1))$, de forma explícita está dada por

$$\begin{aligned} U(q) &= -\log\left(\frac{1}{\Gamma(5)} q^4 e^{-q}\right) \\ &= -\log\left(\frac{q^4}{\Gamma(5)}\right) - \log(e^{-q}) \\ &= -4 \log(q) + \log(\Gamma(5)) + q \\ &= -4 \log(q) + \log(4!) + q, \end{aligned} \quad (3.22)$$

y por tanto,

$$\begin{aligned} H(q, p) &= K(p) + U(q) \\ &= \frac{p^2}{2} - 4 \log(q) + \log(4!) + q. \end{aligned} \quad (3.23)$$

Al derivar (3.22) respecto a q , se obtiene que

$$\frac{dU}{dq} = \frac{-4}{q} + 1. \quad (3.24)$$

Además, recordemos que las ecuaciones de Hamilton deducidas en (2.33) están dadas por

$$\frac{p}{m} = \frac{dq}{dt} \quad \text{y} \quad \frac{\partial U(q)}{\partial q} = \frac{-dp}{dt}. \quad (3.25)$$

Teniendo presente (3.24) y (3.25), se sigue que las ecuaciones de Hamilton para el hamiltoniano definido en (3.23), son

$$\frac{p}{2} = \frac{dq}{dt} \quad \text{y} \quad \frac{-4}{q} + 1 = \frac{-dp}{dt}. \quad (3.26)$$

El sistema de ecuaciones diferenciales (3.26), no tiene una solución analítica, por lo que debe aproximarse con métodos numéricos. La gráfica 3.3 muestra la aproximación de la solución con el método de Euler, Euler simpléctico y Leapfrog, considerando los puntos iniciales $q(0) = 1$ y $p(0) = 1$. El código en Python de este ejercicio se puede consultar en este [repositorio](#).

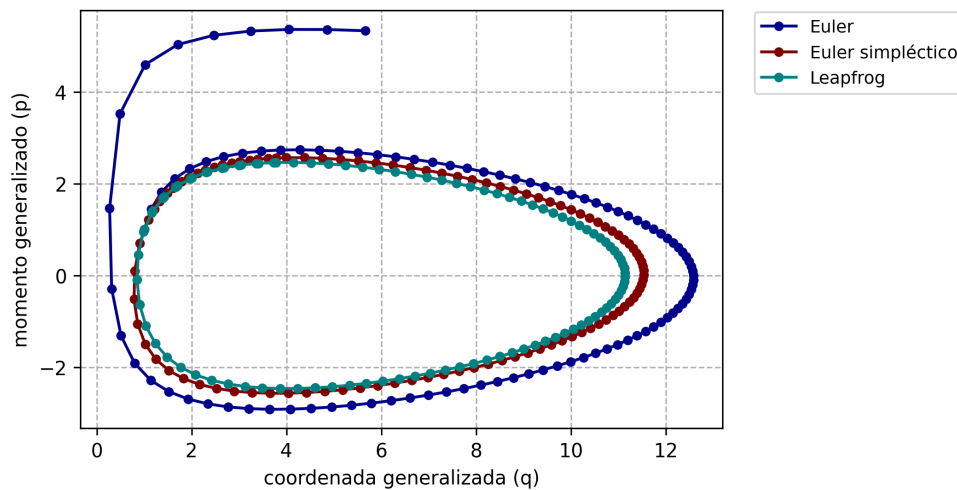


Figura 3.3: Aproximación de la solución de las ecuaciones de Hamilton, generando 100 iteraciones en los métodos numéricos con $\epsilon = 0.15$.

Como se observa en la imagen 3.3, de nuevo, la aproximación por el método de Euler diverge en el tiempo, mientras que las aproximaciones por el método de Euler simpléctico y Leapfrog permanecen estables y similares entre ellas.

Las aproximaciones con el método de Euler simpléctico y Leapfrog son más acertadas porque comparten las propiedad simpléctica de las dinámicas hamiltonianas, a diferencia del método de Euler que como se mostró en la sección 3.1.1 no es un método numérico simpléctico.

De los métodos numéricos simplécticos presentados, Leapfrog es el que produce resultados más cercanos a la solución exacta, ya que a diferencia del método de Euler simpléctico, realiza un paso intermedio que, en efecto, lo hace computacionalmente más costoso, pero a su vez más preciso.

El método numérico utilizado en HMC para aproximar la solución de las ecuaciones de Hamilton, juega un papel importante porque la tasa de aceptación del algoritmo HMC depende de la exactitud de las aproximaciones de las dinámicas hamiltonianas, entre más exacta es, mayor es la tasa de aceptación, como se explica en la sección 4.2.1.

El cuadro 3.1, obtenido en [8] muestra el orden de error de los métodos numéricos presentados en esta sección. Recordemos, que el error local se refiere al error que se genera del tiempo t al tiempo $t + \epsilon$, mientras que el error global se refiere al error acumulado en un intervalo de tiempo R que se recorre en $\frac{R}{\epsilon}$ pasos.

	Euler	Euler simpléctico	Leapfrog
Error Local	ϵ^2	ϵ^2	ϵ^3
Error Global	ϵ	ϵ	ϵ^2

Cuadro 3.1: Orden de error de métodos numéricos.

Generalmente, Leapfrog es el método numérico utilizado en HMC para aproximar la solución de las ecuaciones de Hamilton, debido a su propiedad simpléctica y que como se muestra en el cuadro 3.1, el orden de sus errores es preferible. Por esta razón, y por los resultados obtenidos en los ejemplos presentados en esta sección, en este trabajo también se empleará el algoritmo Leapfrog para aproximar las dinámicas hamiltonianas requeridas en HMC.

3.2. Estadística bayesiana

La estadística bayesiana, considera la probabilidad desde una perspectiva subjetiva. En este enfoque, se adhiere un grado de creencia o expectativa a las probabilidades de la ocurrencia de fenómenos inciertos. En este sentido, es una herramienta eficiente que permite aprovechar la información disponible sobre algún problema para solucionarlo.

En la actualidad, la estadística bayesiana ha tenido un desarrollo notable, derivado en parte del gran avance que ha ocurrido en la parte computacional. Tiempo atrás, aunque las

soluciones que brindaba la estadística bayesiana eran atractivas y adecuadas, en la práctica no podían realizarse por limitaciones computacionales. Sin embargo, hoy en día, estas limitaciones se han reducido, y ha sido posible aplicar la estadística bayesiana a diversas áreas como la econometría, bioestadística, ciencias sociales, ingenierías, entre otras.

La estadística bayesiana se sustenta en el teorema de Bayes, enunciado en 5. Su demostración se puede consultar en [16].

Teorema 5. *Teorema de Bayes.* Dado un conjunto de observaciones D con verosimilitud asociada $L(q | D)$, y una medida de probabilidad a priori G que cuantifica la incertidumbre sobre q , se sigue que la densidad de $q | D$, está dada por

$$S'(q | D) = \frac{\mathcal{L}(D | q)G(q)}{\int_{q \in \mathbb{R}^d} \mathcal{L}(D | q)G(q) dq}, \quad (3.27)$$

y es conocida como la densidad posterior.

En general, para hacer inferencia bayesiana es necesario simular de la densidad posterior (3.27), cuyo divisor habitualmente no se conoce de manera analítica y muchas veces resulta inviable aproximarlos numéricamente. De aquí radica la utilidad de los métodos MCMC en estadística bayesiana, porque permiten simular de la distribución objetivo conociendo la densidad (3.27) salvo constantes, es decir, no es indispensable conocer analíticamente la integral de la densidad para simular de ella.

3.3. Métodos de cadenas de Markov Monte Carlo

Los métodos de cadenas de Markov Monte Carlo (*Markov Chain Monte Carlo*, por sus siglas en inglés MCMC), son una familia de algoritmos utilizados para simular muestras de una variable aleatoria X con función de densidad S' , donde $X \in \Theta$ y $\Theta \subseteq \mathbb{R}^d$. Para obtener las simulaciones, estos algoritmos generan una cadena de Markov $\{X_n : n = 0, 1, \dots, N - 1\}$ ergódica con distribución estacionaria igual que la distribución objetivo. La definición de distribución estacionaria y cadena ergódica se presenta en las secciones 4.2.3 y 4.2.4 respectivamente.

El objetivo de los algoritmos MCMC es poder simular de variables aleatorias y son útiles para aproximar integrales de la forma

$$\mathbb{E}_{S'}(g(X)) = \int_{\Theta} g(x) S'(x) dx,$$

a través del promedio

$$\hat{\mu}_g = \frac{1}{N} \sum_{i=0}^{N-1} g(X_i), \quad (3.28)$$

en donde $g : \Theta \rightarrow \mathbb{R}$ es una función real. En la sección 4.2.4 se enuncia el teorema ergódico que bajo ciertos supuestos prueba que $\hat{\mu}_g \xrightarrow{c.s.} \mathbb{E}_{S'}(g(X))$.

Debemos tener en cuenta que las cadenas generadas por los algoritmos MCMC, no son una muestra aleatoria e independiente de la variable objetivo X . En principio, puede suceder que los primeros estados no sean representativos de la distribución objetivo porque el punto inicial de la cadena puede estar lejos de zonas en donde la densidad S' sea alta. Además, dado que las simulaciones provienen de una cadena de Markov no son independientes, ya que el estado X_{i+1} depende del estado anterior X_i . Sin embargo, se pueden aminorar estas limitaciones al aplicar un *burn-in* y un lag a la cadena, tales conceptos son presentadas en la sección 3.3.2.2 y 3.3.2.3, respectivamente.

3.3.1. Metropolis-Hastings

El algoritmo Metropolis-Hastings permite simular de una distribución objetivo con función de densidad S' cuando se puede evaluar de una función S proporcional a S' , lo cual denotaremos como $S' \propto S$. Poder simular utilizando sólo S resulta de gran utilidad en estadística bayesiana ya que en esta área no es habitual conocer las constantes de normalización de las densidades posteriores. En general, es común definir a S como el kernel de S' , es decir la función de densidad S' salvo constantes.

Para simular una cadena de Markov $\{X_n : n = 0, 1, \dots, N - 1\}$, el algoritmo comienza simulando el primer estado de la cadena X_0 de una distribución inicial μ cuyo soporte esté contenido en el de la distribución que se desea simular; denotaremos la simulación del primer estado como $X_0 \leftarrow \mu$.

Posteriormente, se define a Y_i con $i \in \{0, 1, \dots, N - 2\}$ como una simulación de una variable aleatoria con densidad q condicionada a X_i que puede ser conocida salvo constantes o simétrica (es decir, que satisfaga que $q(x | y) = q(y | x)$), el soporte de la densidad objetivo debe estar contenido en el soporte de q y debe cumplirse que $q(x | y) = 0$ si y sólo si $q(y | x) = 0$. Además, es importante tener presente que en la práctica es conveniente que sea fácil simular de q . Denotaremos esta simulación por

$$Y_i \leftarrow q(\cdot | X_i),$$

de este modo, Y_i es la propuesta para el estado X_{i+1} de la cadena de Markov.

Posteriormente, se define al estado X_{i+1} de la cadena de Markov como la realización

$$X_{i+1} = \begin{cases} Y_i & \text{con probabilidad } \rho(X_i, Y_i) \\ X_i & \text{con probabilidad } 1 - \rho(X_i, Y_i), \end{cases}$$

en donde $\rho(X_i, Y_i)$ está dada por

$$\begin{aligned} \rho(X_i, Y_i) &= \min\left(1, \frac{S'(Y_i)q(X_i|Y_i)}{S'(X_i)q(Y_i|X_i)}\right) \\ &= \min\left(1, \frac{S(Y_i)q(X_i|Y_i)}{S(X_i)q(Y_i|X_i)}\right). \end{aligned} \quad (3.29)$$

La igualdad (3.29) se sigue de que $S' \propto S$.

De forma resumida se presenta este procedimiento en el algoritmo 1.

Algoritmo 1: Metropolis-Hastings

Valores de entrada: función de densidad S proporcional a la distribución objetivo,
función de densidad condicional q , tamaño de la cadena N

Simulamos punto inicial $X_0 \leftarrow \mu$

para $i = 0, 1, \dots, N - 2$ **hacer**

 Generamos un valor $Y_i \leftarrow q(\cdot | X_i)$.

 Calculamos $\rho(X_i, Y_i) = \min\left(1, \frac{f(Y_i)q(X_i|Y_i)}{f(X_i)q(Y_i|X_i)}\right)$.

 Simulamos $u \leftarrow U(0, 1)$

si $u \leq \rho(X_i, Y_i)$ **entonces**

 | aceptamos el estado Y_i , es decir $X_{i+1} = Y_i$

en otro caso

 | rechazamos Y_i , es decir $X_{i+1} = X_i$

fin

 Devolvemos la cadena $\{X_0, X_1, \dots, X_{N-1}\}$

fin

Si el lector desea profundizar en el algoritmo Metropolis-Hastings puede consultar [32], [12], [16].

3.3.2. Random Walk Metropolis-Hastings

Un caso particular del algoritmo Metropolis-Hastings, conocido como *Random Walk Metropolis-Hastings* (RWMH por sus siglas en inglés), surge al generar las propuestas Y_i

para los estados de la cadena simulando de una variable aleatoria con densidad q condicionada al punto X_i y simétrica en ese mismo punto, de tal forma que la simulación $Y_i \leftarrow q(\cdot | X_i)$, puede conseguirse simulando una variable ε de una distribución W simétrica en cero, a través de la suma

$$Y_i = X_i + \varepsilon. \quad (3.30)$$

Del mismo modo que en el algoritmo Metropolis-Hastings, el estado X_{i+1} de la cadena es una variable aleatoria definida como

$$X_{i+1} = \begin{cases} Y_i & \text{con probabilidad } \rho(X_i, Y_i) \\ X_i & \text{con probabilidad } 1 - \rho(X_i, Y_i), \end{cases}$$

pero, en este caso la función $\rho(X_i, Y_i)$ definida en (3.29) puede simplificarse, pues notemos que si consideramos a w como la función de densidad de ε , se cumple que

$$q(Y_i | X_i) = w(\varepsilon), \quad (3.31)$$

y además,

$$\begin{aligned} q(X_i | Y_i) &= w(-\varepsilon) \\ &= w(\varepsilon). \end{aligned} \quad (3.32)$$

De (3.31) y (3.32) se sigue que $q(Y_i | X_i) = q(X_i | Y_i)$, por lo que

$$\begin{aligned} \rho(X_i, Y_i) &= \min\left(1, \frac{S(Y_i) q(X_i | Y_i)}{S(X_i) q(Y_i | X_i)}\right) \\ &= \min\left(1, \frac{S(Y_i)}{S(X_i)}\right). \end{aligned} \quad (3.33)$$

Los pasos para aplicar el algoritmo RWMH de manera estructurada se presentan en el algoritmo 2.

3.3.2.1. Escalamiento óptimo en RWMH

Un caso concreto del algoritmo RWMH surge al considerar que la variable aleatoria ε , de la propuesta $Y_i = X_i + \varepsilon$ presentada en (3.30), tiene distribución normal multivariada $N(\mu, I_d \sigma^2)$, en donde $\mu \in \mathbb{R}^d$ es el vector de ceros $\mu = (0, 0, \dots, 0)^t$, σ^2 la varianza y I_d es la matriz identidad

Algoritmo 2: *Random Walk Metropolis-Hastings*

Valores de entrada: función de densidad S proporcional a la distribución objetivo, distribución W simétrica en cero, tamaño de la cadena N

Simulamos punto inicial $X_0 \leftarrow \mu$

para $i = 0, 1, \dots, N - 2$ **hacer**

 Generamos un valor $Y_i = X_i + \varepsilon$ donde $\varepsilon \leftarrow W$

 Calculamos $\rho(X_i, Y_i) = \min\left(1, \frac{S(Y_i)}{S(X_i)}\right)$.

 Simulamos $u \leftarrow U(0, 1)$

si $u \leq \rho(X_i, Y_i)$ **entonces**

 | aceptamos el estado Y_i , es decir $X_{i+1} = Y_i$;

en otro caso

 | rechazamos Y_i , es decir $X_{i+1} = X_i$;

fin

fin

Devolvemos la cadena $\{X_0, X_1, \dots, X_{N-1}\}$

$$I_d = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

de tal forma que

$$I_d \sigma^2 = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix}.$$

La elección de la varianza σ^2 , es un tema de especial interés, ya que si la varianza σ^2 es pequeña la propuesta Y_i será similar al último valor de la cadena X_i , lo que ocasionará que la probabilidad de aceptación definida en (3.33) sea cercana a uno, sin embargo esto no es necesariamente óptimo porque los estados simulados en la cadena tendrán una dependencia alta y se tardará en explorar completamente el soporte de la distribución objetivo. En contraste, cuando σ^2 es grande, se logra reducir la dependencia en la cadena y se explora de manera más eficiente, pero la probabilidad de aceptar un nuevo estado en la cadena disminuye.

Una forma óptima de elegir el vector de varianzas en el caso concreto en que $\varepsilon \leftarrow N(\mu, I_d \sigma_d^2)$, $d \rightarrow \infty$ y la densidad objetivo tiene forma

$$S'(x_1, x_2, \dots, x_d) = \prod_{i=1}^d f(x_i),$$

donde f es una función de densidad, es consiguiendo que la tasa de aceptación sea de 0.234, este resultado se demuestra en [34] al maximizar la eficiencia del algoritmo MCMC. En este mismo artículo se menciona que aunque la tasa de aceptación 0.234 es óptima para una dimensión d infinita, existen resultados numéricos que muestran que aún en una dimensión $d = 5$, la tasa de aceptación óptima es muy cercana a 0.234, así mismo se afirma que en el caso concreto donde la dimensión $d = 1$ la tasa de aceptación óptima es 0.44.

Para profundizar en temas de optimización del algoritmo RWMH puede consultarse [15].

Ejemplo 5. En este ejemplo se muestra una aplicación de RWMH para simular una muestra de la distribución Gamma con función de densidad

$$S'(x, \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x},$$

tomando los parámetros $\alpha = 5$ y $\beta = 1$, es decir

$$S'(x, 5, 1) = \frac{1}{\Gamma(5)} x^4 e^{-x}. \quad (3.34)$$

Como se mencionó anteriormente, se puede aplicar RWMH conociendo una función proporcional a la función de densidad objetivo (3.34). En este caso, se utiliza como dicha función al kernel de S' que denotaremos por S y que está dado por

$$S(x) = x^4 e^{-x}.$$

Además, consideraremos que las propuestas surgen a través de

$$Y_i = X_i + \varepsilon,$$

donde $\varepsilon \sim N(0, \sigma^2)$.

Debido a que la distribución $N(0, \sigma^2)$ es simétrica en cero la razón de aceptación para este ejemplo es de la forma (3.33), que de manera explícita se puede calcular como

$$\begin{aligned}
\rho(X_i, Y_i) &= \min\left(1, \frac{S(Y_i)}{S(X_i)}\right) \\
&= \min\left(1, \frac{Y_i^4 e^{-Y_i}}{X_i^4 e^{-X_i}}\right) \\
&= \min\left(1, \frac{Y_i^4}{X_i^4} e^{(X_i - Y_i)}\right) \\
&= \min\left(1, \left(\frac{Y_i}{X_i}\right)^4 e^{(X_i - Y_i)}\right). \tag{3.35}
\end{aligned}$$

Al programar este algoritmo es recomendable utilizar la razón simplificada (3.35), para reducir el costo computacional y evitar desbordamientos de punto flotante.

En este ejemplo, se simuló una cadena de 100,000 estados y se eligió un punto inicial $X_0 = 500$, alejado de zonas de alta probabilidad con la intención de resaltar el número de transiciones necesarias para llegar a zonas más representativas de la distribución objetivo. Se escogió $\sigma^2 = 25$ porque de esta manera se consiguió una tasa de aceptación de 0.4403, que es cercana a la tasa óptima de 0.44 para el caso de una dimensión, mencionada previamente en esta sección.

El algoritmo RWMH para este ejemplo, tardó 10.53 segundos en simular los 100,000 estados de la cadena de Markov. El código en Python utilizado, se puede consultar en este [repositorio](#).

En las secciones 3.3.2.2 y 3.3.2.3 se muestra de qué manera tratar esta cadena para obtener una muestra pseudoaleatoria representativa de la distribución objetivo.

3.3.2.2. Burn-in

El *burn-in* en los métodos MCMC, es un término coloquial que se refiere a desechar las primeras b muestras de la cadena que toman valores en los que la densidad objetivo es baja, con el objetivo de tener una cadena cuyos estados sean representativos de la distribución objetivo. Al conjunto de simulaciones $\{X_n : n = 0, 1, \dots, b - 1\}$ con $b - 1 < N - 1$ se le conoce como el periodo de *burn-in* de la cadena. El valor b depende del punto inicial X_0 y de cómo se proponen nuevos estados.

Una manera en la que se suele determinar el periodo de *burn-in*, es graficando los puntos

$$(X_i, \ln(S'(X_i))) \text{ con } i \in \{0, 1, \dots, N - 1\}, \tag{3.36}$$

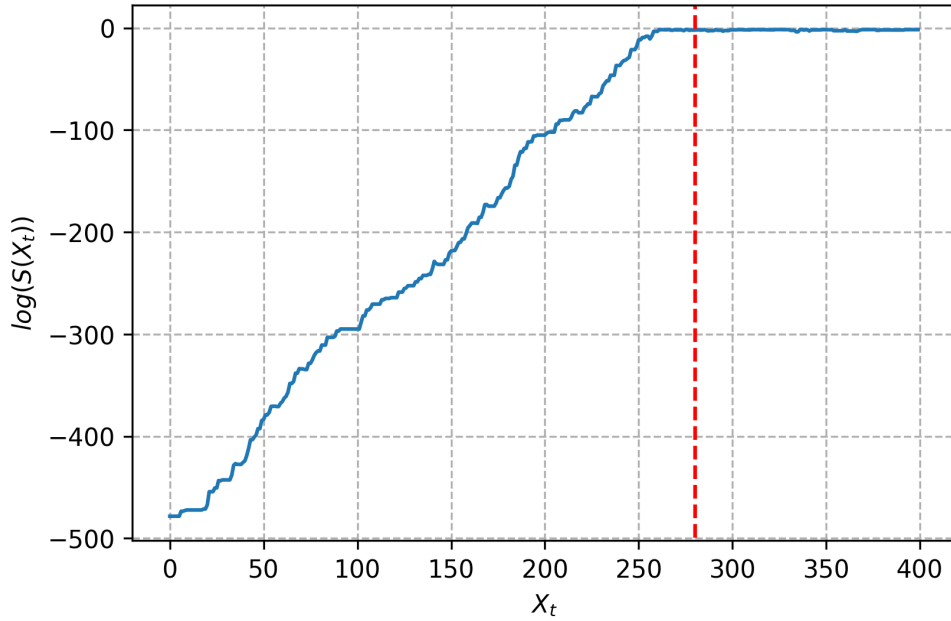


Figura 3.4: Selección del *burn-in* a través de las log-densidades de los estados de la cadena de Markov, la línea roja punteada representa la longitud de la cadena en que las log-densidades dejan de tener un comportamiento creciente.

y notando a partir de qué estado de la cadena los logaritmos de las densidades empiezan a estabilizarse. Si a partir del estado X_α los valores $\ln(S'(X_i))$ con $i \in \{\alpha, \alpha + 1, \dots, N - 1\}$ dejan de tener un comportamiento creciente, se elige a b como el valor α , y se desechan los estados X_i con $i \in \{0, 1, 2, \dots, \alpha - 1\}$ de la cadena.

Para ejemplificar este método, en la gráfica 3.4, se muestran los puntos (3.36) de la cadena generada en el ejemplo 5 con el objetivo de simular de la distribución $Gamma(5, 1)$. Como se muestra en la imagen, aproximadamente después del estado 280 de la cadena los logaritmos de las densidades parecen estabilizarse y dejar de crecer, por lo que de acuerdo a esta perspectiva se sugiere que se eliminen los primeros 280 estados de la cadena.

Por otro lado, una manera analítica en la que se puede determinar la longitud del periodo de *burn-in*, es analizando disimilaridades entre conjuntos consecutivos de la cadena de Markov a través de la distancia de Hellinger. La definición 5, presenta el concepto de distancia de Hellinger consultado en [5].

Definición 5 (Distancia de Hellinger). Sean f y g funciones de densidad. Definimos la distancia de Hellinger como

$$H(f, g) = \sqrt{\frac{1}{2} \int (\sqrt{f(x)} - \sqrt{g(x)})^2 dx}. \quad (3.37)$$

La distancia de Hellinger está acotada entre cero y uno, $0 \leq H(f, g) \leq 1$, entre mayor sea la distancia más grande es la disimilaridad.

El método para determinar el *burn-in* con las distancias de Hellinger, de manera concisa, consiste en dividir en r subconjuntos de la misma longitud la cadena $\{X_n : n = 0, 1, \dots, N - 1\}$, donde r es un divisor de N , de tal forma que se formen las subcadenas:

$$\begin{aligned} Z_1 &= \left\{ X_n : n = 0, 1, \dots, \frac{N}{r} - 1 \right\} \\ Z_2 &= \left\{ X_n : n = \frac{N}{r}, \frac{N}{r} + 1, \dots, 2 \left(\frac{N}{r} \right) - 1 \right\} \\ &\vdots \\ Z_r &= \left\{ X_n : n = (r - 1) \left(\frac{N}{r} \right), (r - 1) \left(\frac{N}{r} \right) + 1, \dots, N - 1 \right\}, \end{aligned}$$

Posteriormente, se aproxima la distancia de Hellinger entre subconjuntos consecutivos, denotaremos por $d_{i,i+1}$ con i en $\{1, 2, \dots, r - 1\}$ a la distancia de Hellinger entre el subconjunto Z_i y Z_{i+1} . Si las distancias $d_{\alpha,\alpha+1}$ con α en $\{\alpha, \alpha + 1, \dots, r - 1\}$ son menores que un umbral η en el que se supone que la distancia ya no es significativa para considerar que el conjunto Z_α y $Z_{\alpha+1}$ provienen de distribuciones distintas, se elige como periodo de *burn-in* al subconjunto Z_i con i en $\{0, 1, \dots, \alpha - 1\}$. Esta idea es presentada y detallada en el artículo [5].

En R la función *BMK.Diagnostic* de la paquetería “LaplacesDemon” permite determinar el *burn-in* a través del método descrito anteriormente, pide como parámetros de entrada la cadena de Markov y el número de subconjuntos en que será dividida y devuelve las distancias de Hellinger entre los subconjuntos.

Para ejemplificar este método se aplicó la función *BMK.Diagnostic* a la cadena generada en el ejemplo 5. Con el objetivo de que fuera ilustrativo se dividió en 200 subconjuntos, de tal forma que cada uno tenía 500 muestras y se eligió como umbral el valor de $\eta = 0.5$, que viene por default en la función *burn-in* de la paquetería “LaplacesDemon” que del mismo modo aplica el método de las distancias de Hellinger para determinar el periodo de *burn-in* de la cadena. En este ejemplo, las distancias entre los subconjuntos consecutivos fueron

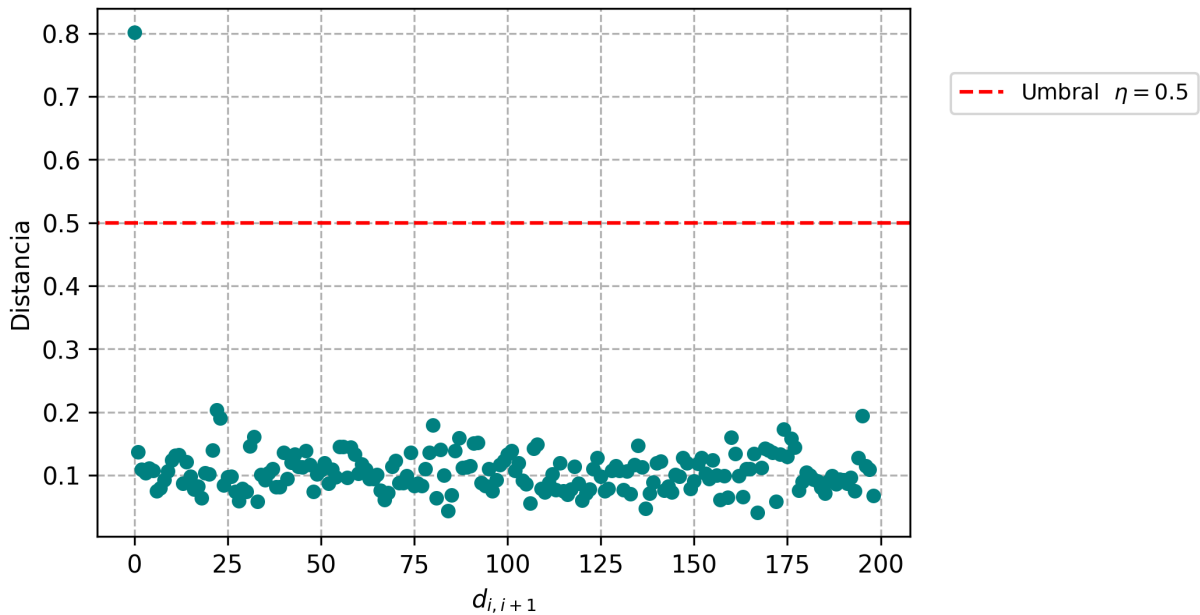


Figura 3.5: Distancias de Hellinger entre los 200 subconjuntos consecutivos de la cadena de Markov.

menores al umbral η a excepción de la distancia entre el primero y segundo, donde se obtuvo una distancia de 0.8018, por lo que de acuerdo con el método se tomó como periodo de *burn-in* a las 500 muestras del primer subconjunto. La imagen 3.5 muestra las distancias de Hellinger entre los subconjuntos.

En este [repositorio](#) se puede consultar el código en Python, en el que se obtiene el periodo de *burn-in* para la cadena del ejemplo 5, bajo las metodologías presentadas en esta sección.

Eliminar el periodo de *burn-in* de la cadena, como se afirma en [8], no es imprescindible para obtener simulaciones de la distribución S' , pues teóricamente si la cadena satisface las hipótesis del teorema ergódico presentado en 7, al simular N estados de la cadena, con N tendiendo a infinito se cumplirá que $\hat{\mu}_g \xrightarrow{c.s.} \mathbb{E}_{S'}(g(X))$ que es el objetivo de los métodos MCMC, sin embargo, dado que es imposible generar infinitas muestras, tomar un periodo de *burn-in* es crucial para tener simulaciones que sean representativas de las distribución objetivo y de este modo tener un buen estimador.

3.3.2.3. Lag

Como se mencionó previamente, las simulaciones que surgen de los métodos MCMC provienen de una cadena de Markov y por tanto no son independientes. Para reducir esta dependencia se conserva solo un estado cada l transiciones, de tal forma que de la cadena $\{X_n : n = 0, 1, \dots, N - 1\}$, sólo se utilizan como muestras de la distribución objetivo el subconjunto

$$\{X_n : n = b, b + l, b + 2l, \dots, b + al\}$$

donde $a = \left\lfloor \frac{(N-1)-b}{l} \right\rfloor$, $\lfloor * \rfloor$ es la función piso y b es el *burn-in* introducido en la sección 3.3.2.2.

Al número natural l se le conoce como lag y debe ser un valor suficientemente grande para eliminar las autocorrelaciones significativas en los estados de la cadena.

Una forma común de seleccionar al lag l es observando la gráfica de autocorrelaciones de los estados de la cadena y notando a partir de cuántos retardos las autocorrelaciones ya no son significativamente distintas de cero. Para ejemplificar este procedimiento, la imagen 3.6 muestra las autocorrelaciones de la cadena generada en el ejemplo 5, desechando los primeros 280 estados que se identificaron como periodo de *burn-in* a través del método gráfico en la sección 3.3.2.2. Como se puede observar después de ocho retardos las autocorrelaciones ya son cercanas a cero, por lo que éste sería un valor razonable para el lag de la cadena.

Otra manera en la que se puede determinar el lag, es tomando la función techo del *Integrated Autocorrelation Time* (IAT, por sus siglas en inglés). Para introducir el IAT recordemos que como se mencionó en la sección 3.3, el objetivo de los métodos MCMC es estimar $\mathbb{E}_{S'}(g(X))$ a través del promedio

$$\hat{\mu}_g = \frac{1}{N} \sum_{i=0}^{N-1} g(X_i),$$

y definamos además, a las autocovarianzas de orden k como

$$\gamma_k = Cov_{S'}(g(X_n), g(X_{n+k})).$$

Notemos que

$$\begin{aligned} \gamma_0 &= Cov_{S'}(g(X_n), g(X_n)) \\ &= Var(g(X_n)). \end{aligned}$$

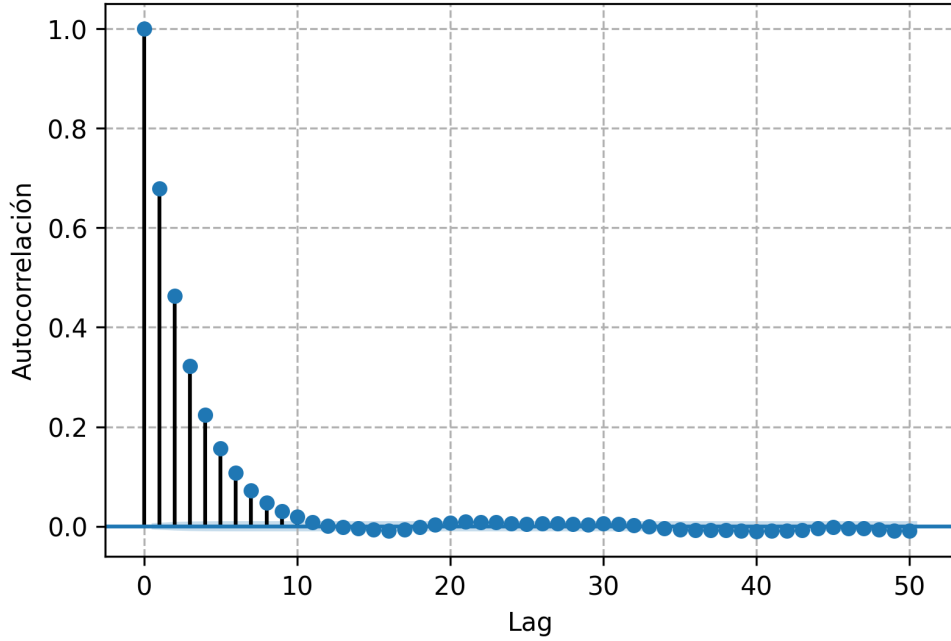


Figura 3.6: Autocorrelaciones en la cadena de Markov del ejemplo
5

Denotemos a $Var(g(X_n))$ por σ^2 , de tal forma que $\gamma_0 = \sigma^2$, y definamos a las autocorrelaciones de orden h como

$$\rho_h = \frac{\gamma_h}{\sigma^2}.$$

En [36] se prueba que la varianza del estimador $\hat{\mu}_g$ está dada por

$$\begin{aligned} Var(\hat{\mu}_g) &= Var\left(\frac{1}{N} \sum_{i=0}^{N-1} g(X_i)\right) \\ &= \frac{\sigma^2}{N} \tau_g(N), \end{aligned} \quad (3.38)$$

en donde $\tau_g(N) = \left(1 + 2 \sum_{h=0}^{N-1} \frac{N-(h+1)}{N} \rho_{h+1}\right)$. Notemos que en un caso hipotético donde los estados de la cadena $\{X_n : n = 0, 1, 2, \dots, N-1\}$ son independientes las autocorrelaciones ρ_{h+1} se anulan y en consecuencia la varianza del estimador $\hat{\mu}_g$ se reduce a

$$Var(\hat{\mu}_g) = \frac{\sigma^2}{N}. \quad (3.39)$$

Observemos que en (3.39) la $Var(\hat{\mu}_g)$ decrece de acuerdo a la razón $1/N$. En los métodos MCMC las muestras no son independientes por lo que la $Var(\hat{\mu}_g)$ es de la forma (3.38) y en este caso decrece de acuerdo a la razón $\frac{\tau_g(N)}{N}$, así es necesario generar $N\tau_g(N)$ simulaciones con los métodos MCMC para tener asintóticamente la misma varianza que una muestra aleatoria independiente de tamaño N .

Si definimos además a τ_g como

$$\tau_g = \lim_{N \rightarrow \infty} \tau_g(N), \quad (3.40)$$

y suponemos que en el límite (3.40) converge, se cumplirá que

$$\tau_g = 1 + 2 \sum_{h=1}^{\infty} \rho_h. \quad (3.41)$$

El valor τ_g es el IAT de la cadena y en la práctica la función techo del IAT indica el valor l que debe tomarse para tener una muestra pseudo-independiente.

En R, la función *IAT* de la paquetería *Rtwalk* permite aproximar el valor del *Integrated Autocorrelation time* tomando como parámetro la cadena generada por el método MCMC. Para ejemplificarlo, se aplicó esta función a la cadena generada en el ejemplo 5, descartando el periodo de *burn-in* y el valor obtenido fue de 5.2574.

El lag l que utilizaremos para este ejemplo es la función techo del IAT obtenido, es decir 6.

Concluyendo, la cadena generada en el ejemplo 5 no es la muestra aleatoria de la distribución $\text{Gamma}(5,1)$; para llegar a ésta se descartaron los primeros 280 estados referentes al periodo de *burn-in*, y se tomó además un lag $l = 6$, de tal forma que de los 100,000 estados, la muestra obtenida de la distribución objetivo fue de tamaño 16,620.

La imagen 3.7 muestra la densidad de la cadena después de considerar el *burn-in* y *lag*.

El código en Python, en el que se implementa el lag a la cadena de Markov del ejemplo 5, de acuerdo a los métodos descritos en esta sección, se puede consultar en este [repositorio](#).

3.3.3. T-walk

En esta sección se introduce el algoritmo T-walk, que es empleado en el capítulo 5 para hacer inferencia bayesiana de un modelo normal jerárquico con el objetivo de comparar sus

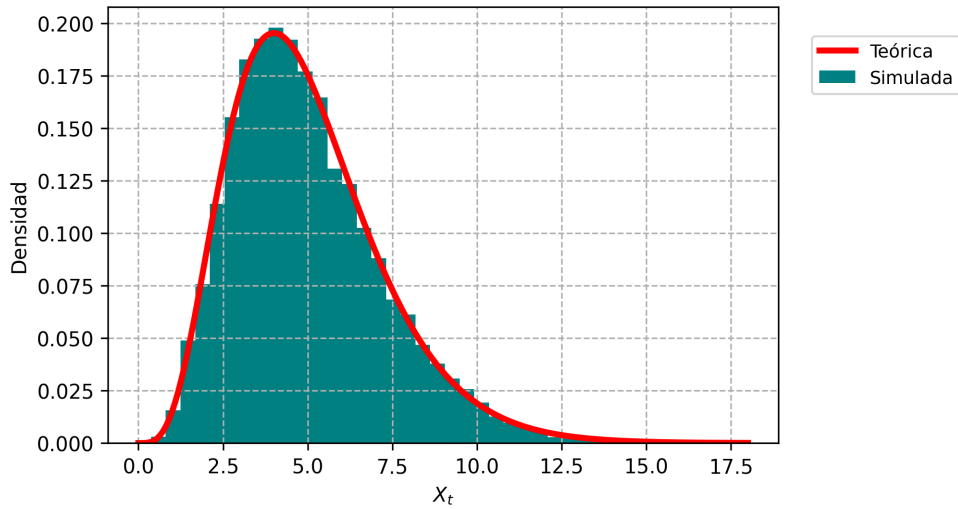


Figura 3.7: Densidad de la muestra de la distribución $\text{Gamma}(5,1)$ obtenida con RWMH.

resultados con los obtenidos al utilizar HMC. Si el lector está interesado puede profundizar en el tema consultando el artículo [9] donde fue presentado el algoritmo.

La principal característica de T-walk es que a diferencia de otros métodos MCMC no requiere parámetros de calibración ni de adaptación. Además, está diseñado para permanecer invariante a la escala y a la estructura de correlación de la distribución objetivo.

Para aplicar el algoritmo T-walk se requiere conocer el logaritmo de la densidad objetivo S' y dos puntos en su soporte que denotaremos por x y x' , con $(x, x') \in \mathbb{R}^{2d}$. Las propuestas para los siguientes estados de la cadena se determinan a través de la regla

$$(y, y') = \begin{cases} (x, h(x, x')), & \text{con probabilidad } 0.5 \\ (h(x, x'), x'), & \text{con probabilidad } 0.5, \end{cases} \quad (3.42)$$

donde $h(x, x')$ es una variable aleatoria.

Una vez que se tiene la propuesta, se utiliza la misma razón empleada en Metropolis Hastings para aceptar o rechazar el estado propuesto. Si consideramos el primer caso de (3.42) y denotamos a la función de densidad de $h(x, x')$ por $g(\cdot | x, x')$, la razón de aceptación y rechazo se puede escribir como

$$\frac{S(y')g(x' | y', x)}{S(x')g(y' | x', x)},$$

mientras que para el segundo caso está dada por

$$\frac{S(y)g(x | y, x')}{S(x)g(y | x, x')}.$$

El algoritmo T-walk genera las propuestas a través de cuatro movimientos, para cada uno de ellos se define la variable aleatoria $h(x, x')$ de una manera específica. Antes de obtener la propuesta se simulan d variables, I_1, I_2, \dots, I_d , de la distribución Bernoulli con parámetro p . Si $I_j = 0$, en la propuesta se mantiene invariante la coordenada j -ésima del estado anterior de la cadena. Es decir, el valor esperado de las coordenadas a cambiar, que denotaremos por n_1 , está dada por el producto dp .

El primer movimiento, llamado *Walk Move*, resulta útil cuando existe poca correlación entre las variables de la distribución objetivo. La variable aleatoria $h(x, x')$, en este caso se define como

$$h_w(x, x')_j = \begin{cases} x_j + (x_j - x'_j) \alpha_j & I_j = 1 \\ x_j & I_j = 0 \end{cases}$$

con $j \in \{1, 2, \dots, n\}$, donde $\alpha_j \in \mathbb{R}$ es una variable aleatoria con función de densidad

$$\psi_w(\alpha) = \begin{cases} \frac{1}{k\sqrt{1+\alpha}}, & \alpha \in \left[\frac{-w}{1+w}, w\right] \\ 0, & \text{en otro caso,} \end{cases} \quad (3.43)$$

donde $k = 2 \left(\sqrt{1+a_w} - \frac{1}{\sqrt{1+a_w}} \right)$ y $a_w > 0$.

El segundo movimiento se conoce como *traverse move* y resulta útil para simular de distribuciones que presentan fuerte correlación entre sus variables. Este movimiento resulta eficaz porque rota y da escala al soporte de la densidad objetivo a través de la variable aleatoria

$$h_t(x, x')_j = \begin{cases} x'_j + \beta (x'_j - x_j) & I_j = 1 \\ x_j & I_j = 0, \end{cases}$$

donde $\beta \in \mathbb{R}^+$ es una variable aleatoria con función de densidad

$$\psi_t(\beta) = \frac{a_t - 1}{2a_t} \{ (a_t + 1) \beta^{a_t} I_{(0,1]}(\beta) \} + \frac{a_t + 1}{2a_t} \{ (a_t - 1) \beta^{-a_t} I_{(1,\infty]}(\beta) \}. \quad (3.44)$$

Los movimientos *Walk Move* y *traverse move* no son suficientes para garantizar que la cadena de Markov simulada sea irreducible. Por lo tanto, para garantizar la irreducibilidad de la cadena y además mejorar el desempeño del algoritmo al simular de distribuciones con

alta correlación entre sus variables, se utilizan también los movimientos denominados *Hop* y *Blow*.

El movimiento *Hop* se define por

$$h_h(x, x')_j = \begin{cases} x_j + \frac{\sigma(x, x')}{3} z_j & I_j = 1 \\ x_j & I_j = 0, \end{cases} \quad (3.45)$$

donde $z_j \sim N(0, 1)$ y $\sigma(x, x') = \max_{I_j=1} |x_j - x'_j|$, mientras que el movimiento *Blow* corresponde a la variable aleatoria

$$h_b(x, x')_j = \begin{cases} x'_j + \sigma(x, x') z_j & I_j = 1 \\ x_j & I_j = 0, \end{cases} \quad (3.46)$$

con $z_j \sim N(0, 1)$.

Es importante mencionar que en el artículo [9], se utilizan probabilidades de 0.4918, 0.4918, 0.0082, 0.0082, para los movimientos *Walk*, *Traverse*, *hop* y *Blow*, respectivamente, ya que se menciona que de esta manera se minimizó el IAT para diversos ejemplos realizados en el estudio. Así mismo, para n_1 , a_t y a_w , que corresponden al valor esperado de coordenadas a modificar en las propuestas, al parámetro de la densidad (3.43) y al de la densidad (3.44), respectivamente, se fijan los valores $n_1 = 4$, $a_w = 1.5$, $a_w = 6$ porque nuevamente son los valores redondeados que minimizaban el IAT en los ejemplos estudiados en el artículo.

El algoritmo T-walk, como se menciona en el artículo ya mencionado, resulta eficiente para simular de distribuciones con diferentes escalas, distintos niveles de correlación y grandes dimensiones. Sin embargo, se menciona que no resulta tan eficaz cuando se busca simular una distribución con grandes dimensiones y fuerte coorelación.

Los resultados que sustentan que la densidad de la distribución estacionaria en efecto es S' y que la cadena es ergódica bajo ciertos supuestos, puede consultarse en [9, pp.8-9].

Capítulo 4

Monte Carlo Hamiltoniano

El método Monte Carlo Hamiltoniano, originalmente llamado *Hybrid Montecarlo*, es un método de cadenas de Markov Monte Carlo. Este fue propuesto en el año 1987 en el artículo [11], con el objetivo de simular de sistemas relacionados al área de la cromodinámica cuántica, un campo de la física que estudia la interacción fuerte, es decir, la fuerza que mantiene juntos a los neutrones y protones que subsisten en un núcleo atómico.

El término *Hamiltonian Monte Carlo* fue utilizado en lugar de *Hybrid Montecarlo* por primera vez en [23], aunque ambos términos son correctos, en la actualidad se suele emplear más *Hamiltonian Monte Carlo* ya que resulta más explícito.

La particularidad del método Monte Carlo Hamiltoniano (HMC, por sus siglas en inglés), es que a diferencia de otros algoritmos MCMC, los estados propuestos para la cadena de Markov surgen de manera determinista a partir de dinámicas hamiltonianas, lo que permite explorar la distribución objetivo de una manera más eficiente.

4.1. Construcción del algoritmo HMC

El objetivo del método HMC es simular de una distribución objetivo continua en \mathbb{R}^d , con función de densidad S' . Para aplicar este método es necesario poder evaluar de una función S proporcional a S' , y además deben de existir las derivadas parciales de la función $-\log(S(q))$ respecto a q , donde $q = (q_1, q_2, \dots, q_d)$ es el argumento de S .

El método HMC logra simular de la distribución objetivo construyendo una cadena de Markov $\{(q_i, p_i) : i = 0, 1, \dots, N - 1\}$, con $q_i, p_i \in \mathbb{R}^d$, cuya distribución estacionaria, tiene por marginal respecto a q a la distribución objetivo.

La distribución estacionaria de la cadena de Markov generada en HMC, es la distribución de Boltzmann¹ también conocida como distribución de Gibbs, considerando ciertas particularidades que mencionamos más adelante en esta sección.

La densidad que asocia la distribución de *Boltzmann* a que un sistema de partículas se encuentre en algún estado $x \in \mathbb{R}^{2d}$, es

$$B(x) = \frac{1}{z} \exp\left(\frac{-E(x)}{c}\right), \quad (4.1)$$

en donde z es una constante de proporcionalidad, E es una función que devuelve la energía del sistema en el estado x , y c es el producto $k\psi$, donde k es la constante de Boltzmann² y ψ es la temperatura del sistema.

Para adaptar la distribución de *Boltzmann* al método HMC, debemos considerar un sistema de partículas con d coordenadas generalizadas y d momentos generalizados denotados por $q = (q_1, q_2, \dots, q_d)$ y $p = (p_1, p_2, \dots, p_d)$, respectivamente, de tal modo que la densidad asociada al punto $(q, p) \in \mathbb{R}^{2d}$, está dada por

$$B(q, p) = \frac{1}{z} \exp\left(\frac{-E(q, p)}{c}\right). \quad (4.2)$$

Además, debemos utilizar al hamiltoniano estándar como la función de energía E , en (4.2), lo que tiene sentido, porque como se explicó en la sección 2.1.4 el hamiltoniano estándar es la energía mecánica de un sistema de partículas. Así mismo, es necesario considerar a la temperatura ψ como $\frac{1}{k}$ con el objetivo de evitar la constante c . Teniendo esto en cuenta, podemos reescribir (4.2) como

$$B(q, p) = \frac{1}{z} \exp(-H(q, p)). \quad (4.3)$$

Es importante resaltar que las dinámicas hamiltonianas, de acuerdo con (4.3), se moverán en superficies con la misma densidad asociada, ya que como se probó en 2.1.5.2, estas dinámicas mantienen invariante al hamiltoniano $H(q, p)$.

Tengamos presente que la función (4.3) puede escribirse de la siguiente manera

¹(1844-1906), físico austríaco, pionero de la mecánica estadística.

²Constante de proporcionalidad entre la temperatura y la energía cinética media de un sistema termodinámico, es aproximadamente $1.380649 \times 10^{-23} \text{ Julio/Kelvin}$.

$$\begin{aligned}
B(q, p) &= \frac{1}{z} \exp(- (K(p) + U(q))) \\
&= \frac{1}{z} \exp(-K(p)) \exp(-U(q)) \\
&= \frac{1}{z} \exp\left(-\sum_{i=1}^d \frac{p_i^2}{2m_i}\right) \exp(-U(q)). \tag{4.4}
\end{aligned}$$

Notemos que el paréntesis $\exp\left(-\sum_{i=1}^d \frac{p_i^2}{2m_i}\right)$ en (4.4), es el kernel de la distribución normal multivariada dada por

$$f(x | \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} \det(\Sigma)^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (x - \mu)^t \Sigma^{-1} (x - \mu)\right) \tag{4.5}$$

donde $x \in \mathbb{R}^d$, $\mu = (0, 0, \dots, 0)^t$ y $\Sigma = \begin{bmatrix} m_1 & 0 & \cdots & 0 \\ 0 & m_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m_d \end{bmatrix}$.

Si definimos a $\frac{1}{w}$ como la constante de normalización de f , se seguirá que $w = (2\pi)^{\frac{n}{2}} \det(\Sigma)^{\frac{1}{2}}$ y además

$$\exp\left(-\sum_{i=1}^d \frac{p_i^2}{2m_i}\right) = wf(p). \tag{4.6}$$

De (4.6) y (4.4) se sigue que

$$\begin{aligned}
B(q, p) &= \frac{1}{z} \exp\left(-\sum_{i=1}^d \frac{p_i^2}{2m_i}\right) \exp(-U(q)) \\
&= \frac{w}{z} f(p) \exp(-U(q)). \tag{4.7}
\end{aligned}$$

Adicionalmente, en (4.7) se debe considerar a la energía potencial U del hamiltoniano estándar como

$$U(q) = -\log(S(q)), \tag{4.8}$$

de tal forma que pueda reescribirse como

$$\begin{aligned}
B(q, p) &= \frac{w}{z} f(p) \exp(\log(S(q))) \\
&= \frac{w}{z} f(p) S(q).
\end{aligned} \tag{4.9}$$

Finalmente, notemos que al construir de este modo a $B(q, p)$, su densidad marginal respecto a q será la densidad objetivo S'

$$\begin{aligned}
\int_{\mathbb{R}^d} B(q, p) dp &= \int_{\mathbb{R}^d} \frac{w}{z} f(p) S(q) dp \\
&= \frac{w}{z} S(q) \int_{\mathbb{R}^d} f(p) dp \\
&= \frac{w}{z} S(q) \\
&= S'(q).
\end{aligned} \tag{4.10}$$

La igualdad (4.10), se sigue de que $\frac{w}{z}$ es la constante de normalización de la función S , y debido a que S es proporcional a la densidad objetivo S' , debe cumplirse que $S' = \frac{w}{z} S$.

Además, notemos que la marginal de $B(q, p)$ respecto a p es la función normal multivariada definida en (4.5), como se muestra a continuación.

$$\begin{aligned}
\int_{\mathbb{R}^d} B(q, p) dq &= \int_{\mathbb{R}^d} \frac{w}{z} f(p) S(q) dq \\
&= f(p) \int_{\mathbb{R}^d} \frac{w}{z} S(q) dq \\
&= f(p).
\end{aligned} \tag{4.11}$$

El método HMC, a través del algoritmo presentado en 3, construye una cadena de Markov $\{q_i, p_i : i = 1, \dots, N - 1\}$, cuya distribución estacionaria tiene por densidad a $B(q, p)$, logrando así que el conjunto $\{q_i : i = 1, \dots, N - 1\}$ tenga la distribución de interés.

El primer estado de la cadena generada con el algoritmo HMC es el punto $(q^{(0)}, p^{(0)})$, donde $q^{(0)}$ es un valor de inicio que debe pertenecer al soporte de la distribución objetivo con densidad S' , y $p^{(0)}$ es una simulación de la distribución $N(\mu, \Sigma)$, cuya densidad es presentada en (4.5).

Posteriormente, se debe resolver o aproximar la solución de las ecuaciones de Hamilton

$$\frac{p_j}{m} = \frac{dq_j}{dt} \quad \text{y} \quad \frac{\partial U(q)}{\partial q_j} = \frac{-dp_j}{dt} \quad \text{para } j \text{ en } \{1, 2, \dots, d\}, \quad (4.12)$$

a un tiempo T , considerando como punto inicial $(q^{(i-1)}, p^{(i-1)})$. El valor T es un parámetro de calibración (*tunning parameter*). Su elección depende de la distribución que se quiera simular, debe ser lo suficientemente grande para que las propuestas estén a una distancia aceptable de los estados actuales y se pueda explorar de manera eficiente el soporte de la distribución objetivo.

La propuesta para el estado $(q^{(i)}, p^{(i)})$ de la cadena, es el punto $(q^*, -p^*)$ donde (q^*, p^*) es la aproximación o la solución del sistema (4.12). El signo negativo en la propuesta es de vital importancia, ya que si no se utiliza, la razón de aceptación utilizada en el algoritmo Metropolis-Hastings presentado en la sección 3.3.1 siempre se hace cero. Para probarlo notemos que de acuerdo a esta razón, las propuestas deben de aceptarse con una probabilidad $\rho((q_{i-1}, p_{i-1}), (q_i, p_i)) = \min(1, r)$, donde

$$\begin{aligned} r &= \frac{f(p_i) S'(q_i) \mathbb{P}(q_{i-1}, p_{i-1} | q_i, p_i)}{f(p_{i-1}) S'(q_{i-1}) \mathbb{P}(q_i, p_i | q_{i-1}, p_{i-1})} \\ &= \frac{f(p_i) S(q_i) \mathbb{P}(q_{i-1}, p_{i-1} | q_i, p_i)}{f(p_{i-1}) S(q_{i-1}) \mathbb{P}(q_i, p_i | q_{i-1}, p_{i-1})}. \end{aligned} \quad (4.13)$$

La igualdad (4.13) se sigue teniendo de que la densidad S es proporcional a S' . La razón (4.13) se anula porque la probabilidad de llegar al punto (q_{i-1}, p_{i-1}) a un tiempo T partiendo del punto (q_i, p_i) al tiempo cero, denotada por $\mathbb{P}(q_{i-1}, p_{i-1} | q_i, p_i)$ es cero, debido a que para regresar al mismo estado del que se partió debe obtenerse un valor p concreto si es que existe de las distribución normal $N(\mu, \Sigma)$, y dado que ese valor de p es una cantidad numerable su valor se obtiene con probabilidad cero.

Al proponer $(q^*, -p^*)$ en lugar de (q^*, p^*) , se logra que la razón (4.13) no se anule, pues como se explicó en el apartado 2.1.5.1 la reversibilidad de las dinámicas hamiltonianas implicadas por el hamiltoniano estándar se consigue al cambiar el valor del momento generalizado p por $-p$, así, de manera determinista la cadena llegará al estado (q_{i-1}, p_{i-1}) partiendo de $(q_i, -p_i)$, es decir $\mathbb{P}(q_{i-1}, p_{i-1} | q_i, -p_i) = 1$. Además, como $\mathbb{P}(q_i, -p_i | q_{i-1}, p_{i-1}) = 1$, debido a que la transición del punto (q_{i-1}, p_{i-1}) a (q_i, p_i) surge de manera determinista de las ecuaciones de Hamilton, la razón (4.13) se reduce a

$$\begin{aligned}
r &= \frac{f(-p_i) S(q_i) \mathbb{P}(q_{i-1}, p_{i-1} \mid q_i, -p_i)}{f(p_{i-1}) S(q_{i-1}) \mathbb{P}(q_i, -p_i \mid q_{i-1}, p_{i-1})} \\
&= \frac{f(-p_i) S(q_i)}{f(p_{i-1}) S(q_{i-1})} \\
&= \frac{f(p_i) S(q_i)}{f(p_{i-1}) S(q_{i-1})}.
\end{aligned} \tag{4.14}$$

La igualdad (4.14), se sigue de que f es una densidad simétrica en cero, y por tanto $f(-p_i) = f(p_i)$.

Simplificaremos la razón (4.14), para que las operaciones realizadas computacionalmente al calcular la razón de aceptación y rechazo en cada iteración sean menos costosas. Por la ecuación (4.3) sabemos que $B(q, p) = \frac{1}{z} \exp(-H(q, p))$ y por (4.9) tenemos que $B(q, p) = \frac{w}{z} f(p) S(q)$. De esta manera se sigue que

$$\begin{aligned}
r &= \frac{f(p_i) S(q_i)}{f(p_{i-1}) S(q_{i-1})} \\
&= \frac{\frac{w}{z} f(p_i) S(q_i)}{\frac{w}{z} f(p_{i-1}) S(q_{i-1})} \\
&= \frac{\frac{1}{z} \exp(-H(q_i, p_i))}{\frac{1}{z} \exp(-H(q_{i-1}, p_{i-1}))} \\
&= \exp(H(q_{i-1}, p_{i-1}) - H(q_i, p_i)).
\end{aligned} \tag{4.15}$$

Por último, se simula un valor u con distribución uniforme $U(0, 1)$; si $u < \min(1, r)$ aceptamos el punto (q_i, p_i) como el siguiente estado de la cadena, de lo contrario, el siguiente estado que se asigna a la cadena es el estado actual. De forma resumida se anexa este procedimiento en el algoritmo 3.

En el caso particular donde se utiliza el algoritmo Leapfrog para aproximar la solución del sistema de ecuaciones (4.12), se define a T como $L\epsilon$ donde L denota el número de pasos de tamaño ϵ con el algoritmo Leapfrog.

El parámetro ϵ controla la exactitud con la que el algoritmo leapfrog aproxima la trayectoria de las dinámicas hamiltonianas. En la sección (4.2.1), se explica que una adecuada aproximación de la solución de las ecuaciones de Hamilton implica una alta tasa de aceptación en el algoritmo HMC.

El parámetro L que debe elegirse depende del ϵ seleccionado. Si se elige un ϵ pequeño, L debe ser lo suficientemente grande para que el valor $L\epsilon$, que representa la longitud de la trayectoria que hace la cadena en una iteración, sea adecuada para explorar la distribución objetivo de una manera eficiente, además de evitar la autocorrelación en la cadena.

La dificultad de la elección de los parámetros de calibración L y ϵ radica en que por lo general los parámetros óptimos son distintos antes y después de converger a la distribución objetivo, más aún puede suceder que en distintas partes de la distribución objetivo los parámetros óptimos no sean los mismos. Derivado de las dificultades para la elección, en [8, pp.135-137] se propone que L sea elegido a prueba y error, se sugiere como punto de inicio $L = 100$, si después se simulan valores de la distribución objetivo y esta logra tener autocorrelaciones no significativas puede intentarse con un L menor, de lo contrario se deberá elegir uno más grande, y para el valor de ϵ se sugiere utilizar un valor aleatorio en un intervalo pequeño.

Algoritmo 3: Método Monte Carlo Hamiltoniano

Valores de entrada: Energía potencial $U(q) = -\log(S(q))$, donde $S(q)$ es proporcional a la densidad objetivo S' , valor inicial $q^{(0)}$, longitud de trayectoria de la dinámica T , tamaño de la cadena N .

$q[0]=q^{(0)}$

para $i = 1, \dots, N - 1$ **hacer**

Generamos un valor $p^{(i-1)} \sim N(0, \Sigma)$.

Resolvemos o aproximamos las ecuaciones de Hamilton

$$\frac{\partial H(q, p)}{\partial p_j} = \frac{-dq_j}{dt} \quad \frac{\partial H(q, p)}{\partial q_j} = \frac{dp_j}{dt} \quad j \in \{1, 2, \dots, longitud(q)\}$$

con punto inicial $(q^{(i-1)}, p^{(i-1)})$ a un tiempo T , denotamos su solución por (q^*, p^*) .

Calculamos la razón $r = \exp(H(q^{(i-1)}, p^{(i-1)}) - H(q^*, -p^*))$.

Generamos $u \sim U(0, 1)$

si $u \leq \text{mín}(1, r)$ **entonces**

| aceptamos el estado q^* , es decir $q[i] = q^*$;

en otro caso

| rechazamos q^* , es decir $q[i] = q^{(i-1)}$;

fin

Devolvemos la cadena $\{q[0], q[1], \dots, q[N - 1]\}$

fin

4.2. Resultados teóricos

En esta sección se presentan resultados teóricos sobre la tasa de aceptación y la teoría de cadenas de Markov, que sustenta el funcionamiento del método HMC.

4.2.1. Tasa de aceptación

Las propuestas en el método HMC, teóricamente se aceptan con probabilidad uno. Para probar este hecho, tengamos presente que en la sección 4.1 se exhibió que la probabilidad de aceptación está dada por el mín $(1, r)$, donde

$$r = \exp(H(q_i, p_i) - H(q_{i+1}, p_{i+1})). \quad (4.16)$$

Además, recordemos que en la sección 2.1.5.2, se probó que las dinámicas hamiltonianas mantienen invariante al hamiltoniano, por lo que $H(q_i, p_i) = H(q_{i+1}, p_{i+1})$, y de esta manera

$$\begin{aligned} r &= \exp(H(q_{i-1}, p_{i-1}) - H(q_i, p_i)) \\ &= \exp(0) \\ &= 1. \end{aligned} \quad (4.17)$$

De (4.17) se sigue que

$$\begin{aligned} \text{mín}(1, r) &= \text{mín}(1, 1) \\ &= 1, \end{aligned} \quad (4.18)$$

por lo que la probabilidad de aceptar un nuevo estado en la cadena es uno.

Es importante resaltar que en la práctica no se aceptan las propuestas con probabilidad uno, debido a que el Hamiltoniano no permanece constante a causa de los errores inherentes a los métodos numéricos utilizados para aproximar las dinámicas hamiltonianas. Notemos que entre más exacto sea nuestro método de aproximación, la diferencia $H(q_{i-1}, p_{i-1}) - H(q_i, p_i)$ será más cercano a cero y el valor r en (4.17) será más cercano a uno, por lo que entre mayor precisión tenga el método numérico utilizado la tasa de aceptación será más cercana a uno.

Resulta imprescindible destacar que, aunque teóricamente la tasa de aceptación en HMC es del cien por ciento, lo óptimo no es aceptar todas las propuestas. En el artículo [3] se prueba que cuando la dimensión de la distribución objetivo tiende a infinito y bajo

otras hipótesis hechas al algoritmo Leapfrog, se cumple que la tasa de aceptación óptima en HMC converge a 0.651.

4.2.2. Ecuaciones de balance detallado

Para introducir las ecuaciones de balance detallado se presentan las siguientes definiciones consultadas en [21]:

Definición 6 (Kernel de transición). *Sea $\{X_n : n = 0, 1, 2, \dots, N - 1\}$ una cadena de Markov con espacio de estados \mathbb{X} y σ -álgebra \mathcal{B} . El kernel de transición de la cadena es una función $P(A | x)$ tal que para todo $n \geq 0$*

$$\mathbb{P}(X_{n+1} \in A | X_n = x) = P(A | x)$$

con $x \in \mathbb{X}$ y $A \in \mathcal{B}$.

Definición 7 (Ecuaciones de balance detallado). *Sea $\{X_n : n = 0, 1, 2, \dots, N - 1\}$ una cadena de Markov con espacio de estados \mathbb{X} y $\pi(dy)$ una medida de probabilidad en $(\mathbb{X}, \mathcal{B})$. Se dice que la distribución $\pi(dy)$ satisface las ecuaciones de balance detallado respecto a la la cadena $\{X_n : n = 0, 1, 2, \dots, N - 1\}$ con kernel de transición $P(dy | x)$ si*

$$\int_B \int_A \pi(dx) P(dy | x) = \int_A \int_B \pi(dy) P(dx | y),$$

donde $A, B \in \mathcal{B}$. Si denotamos por $f(x)$ y $Q(y | x)$ a las derivadas de Radon-Nikodym respecto a la medida de Lebesgue de las medidas $\pi(dy)$ y $P(x, dy)$ respectivamente, tendremos que

$$\int_B \int_A f(x) Q(y | x) dx dy = \int_A \int_B f(y) Q(x | y) dx dy,$$

y por tanto

$$f(x) Q(y | x) = f(y) Q(x | y). \quad (4.19)$$

Notemos que en (4.19), f es la densidad de la distribución $\pi(dy)$ y $Q(y | x)$ es la densidad condicionada al punto x del kernel $P(dy | x)$.

En esta sección probaremos que la función $B(q, p)$ presentada en (4.3) satisface las ecuaciones de balance detallado (4.19) respecto a la cadena de Markov $\{(q_i, p_i) : i = 0, 1, \dots, N - 1\}$ generada por el algoritmo HMC. Este resultado será utilizado en la sección 4.2.3 donde probaremos que la distribución estacionaria de la cadena $\{(q_i, p_i) : i = 0, 1, \dots, N - 1\}$ tiene

función de densidad $B(q, p)$.

Para probar que $B(q, p)$ en efecto satisface las ecuaciones de balance detallado respecto a $\{(q_i, p_i) : i = 0, 1, \dots, N - 1\}$, recordemos que como se mencionó en la sección 4.1, el punto (q', p') que se propone para el siguiente estado de la cadena en el método HMC es la solución de las ecuaciones de Hamilton (4.12) a un tiempo T considerando como punto inicial al último estado de la cadena (q, p) . Si la solución de las ecuaciones está dada por (q_T, p_T) el punto propuesto será $(q_T, -p_T)$ y de esta forma la probabilidad de transición de un estado a otro en el espacio fase estará dada por

$$\mathbb{P}((q', p') | (q, p)) = \delta((q', p') - (q_T, -p_T)),$$

donde

$$\delta(x) = \begin{cases} 1 & \text{si } x = 0 \\ 0 & \text{si } x \neq 0. \end{cases}$$

Tengamos presente además, que como se mencionó en la sección 2.1.5.1, al cambiar el momento p por $-p$ en las dinámicas hamiltonianas se consigue que la trayectoria descrita por las ecuaciones de Hamilton en el espacio fase sea inversa, implicando, en este caso, que la probabilidad de ir del punto $(q_T, -p_T)$ partiendo del punto (q', p') sea uno y por tanto se cumpla que

$$\mathbb{P}((q', p') | (q, p)) = \mathbb{P}((q, p) | (q', p')). \quad (4.20)$$

El lado derecho de la igualdad (4.20) de forma explícita está dado por

$$\mathbb{P}((q, p) | (q', p')) = \delta((q, p) - (q'_T, -p'_T)), \quad (4.21)$$

donde (q'_T, p'_T) es la solución de las ecuaciones de Hamilton con punto inicial (q', p') a un tiempo T .

Notemos además, que la probabilidad de aceptación $\rho((q, p), (q', p'))$ dada en la sección 4.1, cumple que

$$\begin{aligned} & \exp(-H(q, p)) \rho((q, p), (q', p')) \\ &= \exp(-H(q, p)) \min(1, \exp(H(q, p) - H(q', p'))) \\ &= \exp(-H(q, p)) \min(1, \exp(H(q, p) - H(q', p'))) \\ &= \min(\exp(-H(q, p)), \exp(-H(q', p'))) \\ &= \exp(-H(q', p')) \min(\exp(-H(q, p) + H(q', p')), 1) \\ &= \exp(-H(q', p')) \rho((q', p'), (q, p)), \end{aligned}$$

concluyendo, satisface que

$$\exp(-H(q, p)) \rho((q, p), (q', p')) = \exp(-H(q', p')) \rho((q', p'), (q, p)). \quad (4.22)$$

Multiplicando las probabilidades (4.20) en (4.22), obtenemos que

$$\begin{aligned} & \exp(-H(q, p)) \mathbb{P}((q', p') | (q, p)) \rho((q, p), (q', p')) \\ &= \exp(-H(q', p')) \mathbb{P}((q, p) | (q', p')) \rho((q', p'), (q, p)). \end{aligned} \quad (4.23)$$

Para hacer la notación más clara, denotemos por $Q((q', p') | (q, p))$ a la probabilidad de moverse y aceptar como siguiente estado de la cadena al punto (q', p') partiendo de (q, p) , es decir

$$Q((q', p') | (q, p)) = \mathbb{P}((q', p') | (q, p)) \rho((q, p), (q', p')),$$

de tal forma que podemos reescribir a (4.23) como

$$\exp(-H(q, p)) Q((q', p') | (q, p)) = \exp(-H(q', p')) Q((q, p) | (q', p')). \quad (4.24)$$

De la ecuación (4.24) y debido a que en (4.3) mostramos que $B(q, p) = \frac{1}{z} \exp(-H(q, p))$, se sigue que

$$B(q, p) Q((q', p') | (q, p)) = B(q', p') Q((q, p) | (q', p')). \quad (4.25)$$

El resultado (4.25) prueba que la función de densidad $B(q, p)$ satisface la ecuación de balance detallado (4.19), respecto a la cadena de Markov generada por el algoritmo HMC.

4.2.3. Distribución estacionaria de la cadena

Los métodos MCMC, como se explicó en la sección 3.3 permiten simular de una distribución objetivo con función de densidad S' , mediante la construcción de una cadena de Markov ergódica con distribución estacionaria igual a la distribución objetivo.

En esta sección mostraremos que la distribución estacionaria de la cadena de Markov generada por el algoritmo HMC tiene función de densidad $B(q, p)$, cuya densidad marginal respecto a q , como se mostró en la sección 4.1, es la densidad objetivo S' . Para mostrarlo es conveniente tener presente la definición de distribución estacionaria presentada en 8 y el teorema 6.

Definición 8 (Distribución estacionaria). Sea $\{X_n : n = 0, 1, 2, \dots, N - 1\}$ una cadena de Markov en el espacio de estados \mathbb{X} y $\pi(dy)$ una medida de probabilidad en $(\mathbb{X}, \mathcal{B})$. Se dice que la cadena con kernel de transición $P(dy | x)$ tiene una distribución estacionaria $\pi(dy)$ si

$$\pi(dy) = \int_{\mathbb{X}} \pi(dx) P(dy | x),$$

si denotamos por f a la función de densidad de la distribución π y a $P(dy | x)$ por la función de densidad condicionada $Q(y | x)$, podemos reescribir la condición (8) como

$$f(y) = \int_{\mathbb{X}} f(x)Q(y | x) dx. \quad (4.26)$$

Teorema 6. Sea $\{X_n : n = 0, 1, 2, \dots, N - 1\}$ una cadena de Markov en el espacio de estados \mathbb{X} con kernel de transición $P(dy | x)$ y $\pi(dy)$ una medida de probabilidad en $(\mathbb{X}, \mathcal{B})$. Denotemos por f a la densidad de la distribución $\pi(dy)$ y por $Q(y | x)$ a la densidad condicionada del kernel $P(dy | x)$, si f satisface las ecuaciones de balance detallado, cumplirá también la condición (4.26) y por tanto $\pi(dy)$ será la distribución estacionaria de $\{X_n : n = 0, 1, 2, \dots, N - 1\}$

Demostración.

$$\begin{aligned} \int_{\mathbb{X}} f(x) Q(y | x) dx &= \int_{\mathbb{X}} f(y) Q(x | y) dx \\ &= f(y) \int_{\mathbb{X}} Q(x | y) dx \\ &= f(y). \end{aligned} \quad (4.27)$$

La igualdad (4.27) se debe a que f y $Q(y | x)$ satisfacen las ecuaciones de balance detallado presentadas en 4.19

□

Dado que en la sección 4.2.2, probamos que la densidad $B(q, p)$ satisface las ecuaciones de balance detallado (4.19), respecto a la cadena de Markov generada por el algoritmo HMC, aunado al teorema 6, se sigue que la densidad de la distribución estacionaria de la cadena $\{(q_i, p_i) : i = 0, 1, \dots, N - 1\}$ en efecto es $B(q, p)$.

4.2.4. Ergodicidad de la cadena

En la sección 4.2.3 probamos que la distribución estacionaria de la cadena generada por el algoritmo HMC tiene por densidad a $B(q, p)$, sin embargo este resultado no es suficiente para afirmar que la cadena simulada convergerá a la distribución objetivo, para asegurar

que esto suceda la cadena debe satisfacer las hipótesis del teorema ergódico enunciado en 7, las definiciones 9, 10 y 11 consultadas en [2], [33], [21] respectivamente, son fundamentales para enunciar este teorema.

Definición 9 (Cadena Harris recurrente positiva).

Sea $\{X_n : n = 0, 1, 2, \dots, N - 1\}$ una cadena de Markov con medida estacionaria $\pi(dy)$ en $(\mathbb{X}, \mathcal{B})$. Se dice que la cadena de Markov es Harris recurrente positiva si $\pi(dy)$ puede ser normalizada a una medida de probabilidad en el espacio medible $(\mathbb{X}, \mathcal{B})$.

Definición 10 (Cadena de Markov aperiódica).

Sea $\{X_n : n = 0, 1, 2, \dots, N - 1\}$ una cadena de Markov en el espacio de estados \mathbb{X} con distribución estacionaria $\pi(dy)$ en $(\mathbb{X}, \mathcal{B})$ y kernel de transición $P(A | x)$. Se dice que la cadena es aperiódica si no existe $d \geq 2$ y conjuntos ajenos no vacíos $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_d \subseteq \mathcal{B}$, tal que $P(\mathcal{X}_{i+1} | x) = 1$ para toda x en \mathcal{X}_i con $1 \leq i \leq d - 1$ y $P(\mathcal{X}_1 | x) = 1$ para todo x en \mathcal{X}_d con $\pi(\mathcal{X}_1) \geq 0$.

Definición 11 (Cadena de Markov ergódica).

Una cadena de Markov $\{X_n : n = 0, 1, 2, \dots, N - 1\}$ es ergódica si es Harris recurrente positiva y aperiódica.

El teorema ergódico consultado en [16] y enunciado en 7, sustenta que las cadenas ergódicas convergen a su distribución estacionaria. El lector puede profundizar en la teoría ergódica de las cadenas de Markov en la sección 6.6 de [32].

Teorema 7. *Teorema ergódico.* Sea $\{X_n : n = 0, 1, 2, \dots, N - 1\}$ una cadena de Markov ergódica con espacio de estados \mathbb{X} , distribución estacionaria $\pi(dx)$ en $(\mathbb{X}, \mathcal{B})$ y g una función real. Si $\mathbb{E}_\pi(|g(x)|) < \infty$, entonces

$$\hat{\mu}_g \xrightarrow{c.s.} \mathbb{E}_\pi(g(X)), \quad (4.28)$$

donde $\hat{\mu}_g = \frac{1}{N} \sum_{i=0}^{N-1} g(X_i)$ y $\mathbb{E}_\pi(g(X)) = \int_{\mathbb{X}} g(x) \pi(dx)$.

Una propiedad más restrictiva que se busca en los métodos MCMC es que la cadena generada sea geoméricamente ergódica. Este concepto recuperado de [37], es presentado en la definición 12, la importancia de esta propiedad como se afirma en [20], es que implica que el teorema de límite central, enunciado en 8 y consultado en [16], se satisfaga para los estados simulados de la cadena y además logra que la convergencia a la distribución estacionaria ocurra a una velocidad exponencial.

Definición 12 (Cadena de Markov geoméricamente ergódica).

Sea $\{X_n : n = 0, 1, 2, \dots, N - 1\}$ una cadena de Markov con espacio de estados \mathbb{X} , distribución estacionaria $\pi(dy)$ y kernel de transición en n pasos $P^n(dy | x)$, se dice que la cadena $\{X_n : n = 0, 1, 2, \dots, N - 1\}$ es geoméricamente ergódica si existe una constante λ con $0 \leq \lambda < 1$ tal que

$$\|P^n(x, \cdot) - \pi(\cdot)\| \leq M(x) \lambda^n, \quad (4.29)$$

donde $M(x)$ es una función real e integrable.

Teorema 8. *Teorema de límite central.* Sea $\{X_n : n = 0, 1, 2, \dots, N - 1\}$ una cadena de Markov geoméricamente ergódica con espacio de estados \mathbb{X} , distribución estacionaria $\pi(dy)$ en $(\mathbb{X}, \mathcal{B})$ y g una función real.

Si $\int |g|^{2+\varepsilon} d\pi < \infty$ para alguna $\varepsilon > 0$, entonces

$$\sigma_g^2 = \text{Var}_\pi(g(X_0)) + 2 \sum_{i=1}^{\infty} \text{Cov}_\pi(g(X_0), g(X_i)),$$

está bien definida, es no negativa y finita. Además $\sqrt{N}(\hat{\mu}_g - \mathbb{E}_{\mathcal{S}'}(g(X)))$ converge en distribución a $N(0, \sigma_g^2)$.

En general, no se puede garantizar que la cadena generada por el algoritmo HMC sea geoméricamente ergódica, sin embargo en el artículo [22] se estudia bajo qué supuestos la cadena generada por el algoritmo HMC es o no geoméricamente ergódica. Los resultados no son triviales y no son presentados en este trabajo porque exceden los objetivos del mismo.

4.3. Ejemplos y comparación entre HMC y RWMH

Con el objetivo de comparar el método HMC y el algoritmo RWMH, en las siguientes secciones se presentan ejemplos en los que se resaltan algunas de sus diferencias respecto al IAT de la cadena simulada, el tiempo de ejecución, la convergencia y las tasas de aceptación.

4.3.1. Simulación de la distribución Gamma(5,1)

En esta sección, simularemos una muestra de la distribución $\text{Gamma}(5, 1)$, empleando el algoritmo HMC, y lo compararemos con el ejemplo 5, en donde se se simuló de esta misma distribución utilizando RWMH. El código en Python de este ejercicio se puede consultar en este [repositorio](#).

Una manera de hacer comparable las propuestas en los algoritmos HMC y RWMH mencionada en [8], es seleccionando el parámetro del tamaño de paso ϵ del algoritmo HMC igual a la desviación estándar de la distribución normal con la que se proponen nuevos estados en RWMH y tomando al número de iteraciones L del algoritmo Leapfrog utilizado en HMC igual al lag utilizado en RWMH, ya que de esta manera la longitud de las trayectorias de ambos algoritmos es similar.

Para generar las simulaciones de la distribución $Gamma(5, 1)$ con el algoritmo HMC, debemos tener en cuenta que como se mencionó en el ejemplo 5, el kernel de esta distribución está dado por

$$S(q) = q^4 e^{-q}.$$

Posteriormente, debemos definir a la energía potencial del hamiltoniano estándar $H(q, p)$, como se especificó en (4.8), es decir,

$$\begin{aligned} U(q) &= -\log(S(q)) \\ &= -\log(q^4 e^{-q}) \\ &= -4 \log(q) + q. \end{aligned} \tag{4.30}$$

Notemos que la derivada de (4.30) respecto a la coordenada generalizada q está dada por

$$\frac{\partial U}{\partial q} = \frac{-4}{q} + 1. \tag{4.31}$$

La función (4.30) y su derivada (4.31), fueron utilizadas en el algoritmo HMC para simular de la distribución $Gamma(5, 1)$. Además, para hacer el ejercicio comparable con el ejemplo 5, se tomó como punto inicial $q_0 = 500$ y como parámetros de calibración, $\epsilon = 5$ y $L = 6$. Sin embargo al utilizar estos parámetros la tasa de rechazo del algoritmo HMC fue de 99.99%, debido a que el tamaño de paso ϵ utilizado en la aproximación numérica de la solución de las ecuaciones de Hamilton es grande, provocando que la aproximación esté lejos de la solución real. Esto como se explicó en la sección 4.2.1, conlleva a una tasa de aceptación pequeña.

Teniendo presente lo anterior, se simuló de este mismo ejemplo pero ahora tomando los parámetros $\epsilon = 0.09$ y $L = 47$. De esta manera, la longitud de las trayectorias de las cadenas en ambos algoritmos no es similar, sin embargo, se busca resaltar que aún teniendo una tasa de aceptación cercana a la óptima en RWMH, el algoritmo HMC es capaz de superarlo, incluso cuando sus parámetros no fueron elegidos bajo un criterio de optimalidad.

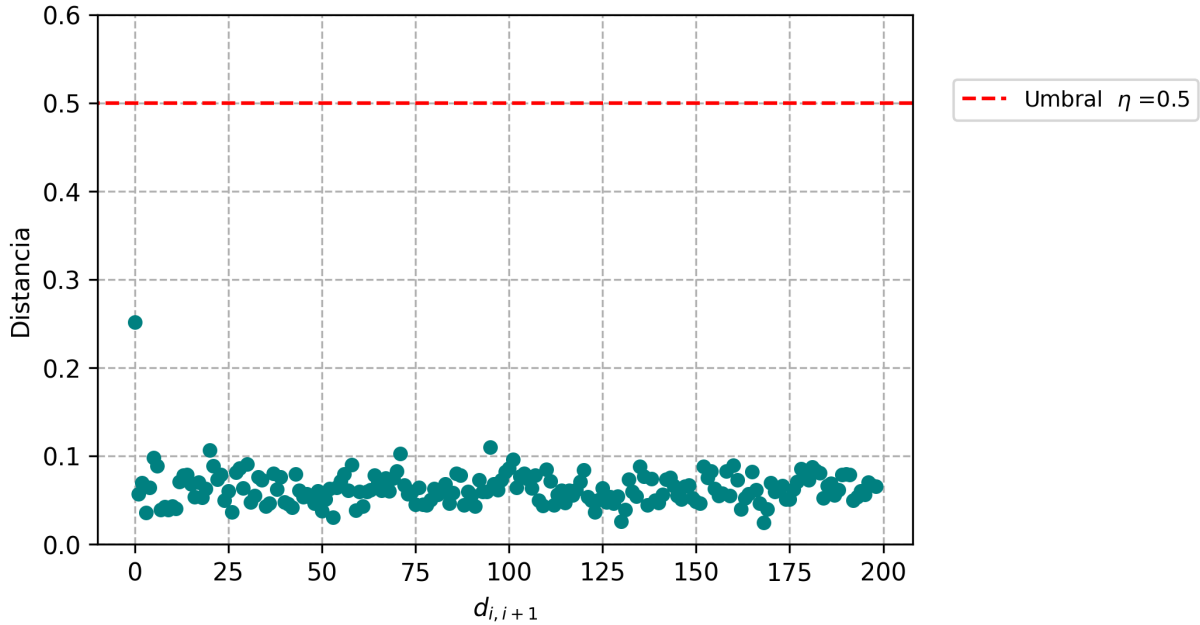


Figura 4.1: Distancias de Hellinger entre los 200 subconjuntos consecutivos de la cadena de Markov.

La imagen 4.2 ilustra el periodo de *burn-in* de la cadena, introducido en la sección 3.3.2.2. Como se puede observar a partir del estado 65, la gráfica del logaritmo de las densidades parece dejar de crecer, así, bajo esta perspectiva se sugiere desechar los primeros 65 estados de la cadena.

Por otro lado, el método descrito en la sección 3.3.2.2, para determinar el periodo de *burn-in* a través de las distancias de Hellinger, sostiene que no debe desecharse ninguna muestra, debido a que la cadena logra llegar pronto a zonas con una alta densidad asociada, de tal modo que los subconjuntos formados por este método no superen el umbral utilizado de $\eta = 0.5$, a partir del cual consideramos que los subconjuntos provienen de distintas distribuciones. Es importante destacar que al utilizar este método resulta conveniente desechar al menos el primer subconjunto, para evitar, como en este caso, que la cadena tenga estados atípicos de la distribución objetivo. En la imagen 4.1 se muestran las distancias de Hellinger entre los subconjuntos de estados obtenidos en este ejemplo.

Para determinar el lag de la cadena, utilizaremos el método descrito en la sección 3.3.2.3, en el cual se elegía a l como la función techo del IAT de la cadena. En este ejemplo, el IAT obtenido, descartando los primeros 65 estados que se determinaron como periodo de *burn-in* a través del método gráfico, fue de 0.9818, por lo que tomaremos $l = 1$. La imagen

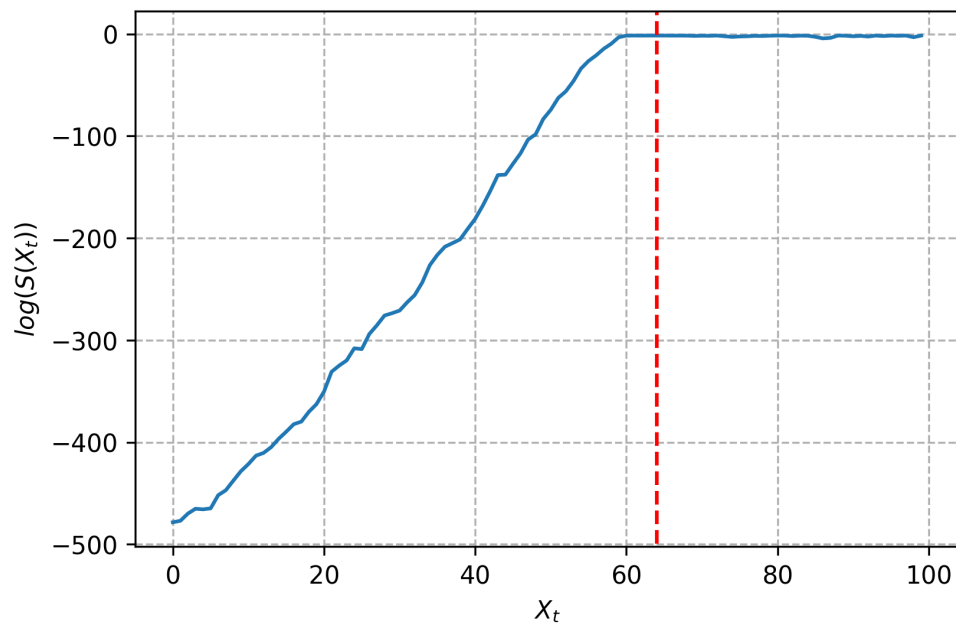


Figura 4.2: Selección del *burn-in* a través de las log-densidades de la cadena de Markov, la línea roja punteada representa la longitud de la cadena en que las log-densidades dejan de tener un comportamiento creciente.

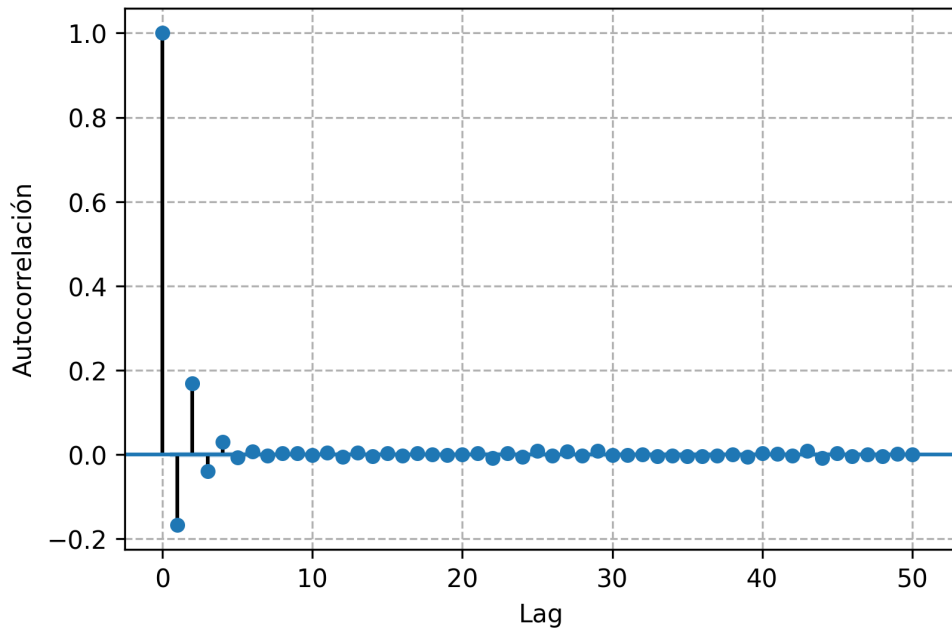


Figura 4.3: Autocorrelaciones en la cadena de Markov.

4.3 muestra las autocorrelaciones en los estados de la cadena sin considerar el periodo de *burn-in*.

La densidad de la cadena generada por el algoritmo HMC en este ejemplo (considerando el *burn-in* y lag), así como su aproximación a la densidad teórica de la distribución $\text{Gamma}(5, 1)$, se muestran en la imagen 4.4.

En el cuadro 4.1, se muestran las principales diferencias al simular de la distribución $\text{Gamma}(5, 1)$ con el algoritmo HMC y RWMH. Como se puede observar, el algoritmo HMC presenta mejores resultados, pues de la cadena de 100,000 estados se utilizaron como muestra de la distribución objetivo 99,934 mientras que en RWMH sólo 16,620. Notemos además que aunque el algoritmo HMC tardó más en ejecutarse, es más rápido en el sentido de que generó por segundo aproximadamente 2,272 muestras mientras que RWMH sólo 1,578.

4.3.2. Normal bivariada

Con el objetivo de comparar el comportamiento que sigue la cadena del método HMC y el algoritmo RWMH, se simuló una cadena de Markov, teniendo como distribución objetivo una distribución normal bivariada con vector de medias $\mu = (0, 0)$ y matriz de varianzas y covarianzas

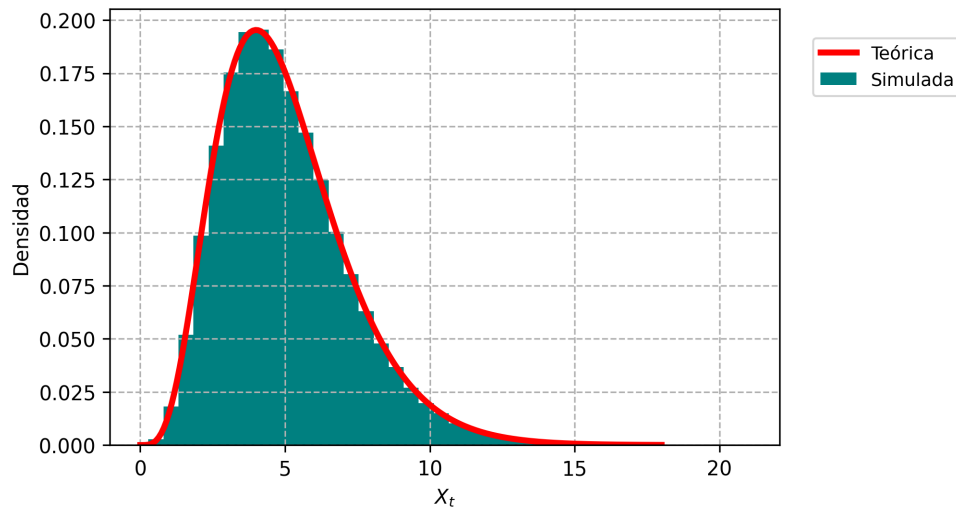


Figura 4.4: Densidad de la cadena de Markov generada con HMC, considerando *burn-in* y lag.

	RWMH	HMC
<i>Burn-in</i>	280	65
IAT	5.2574	0.9818
Tasa de aceptación	44.03 %	99.96 %
Muestra efectiva	16,620	99,935
Tiempo de ejecución	10.53 s	43.99 s
Segundos por muestra	0.0006 s	0.0004 s

Cuadro 4.1: Comparación entre HMC y RWMH al simular de la distribución Gamma(5,1).

$$\Sigma = \begin{pmatrix} 1 & -0.85 \\ -0.85 & 1 \end{pmatrix},$$

cuyo kernel está dado por

$$S(q) = \exp\left(-\frac{q^T \Sigma^{-1} q}{2}\right).$$

Para simular de esta distribución con el método HMC, debemos definir a la energía potencial del hamiltoniano estándar como

$$\begin{aligned}
U(q) &= -\log(S(q)) \\
&= -\log\left(\exp\left(-\frac{q^T \Sigma^{-1} q}{2}\right)\right) \\
&= \frac{q^T \Sigma^{-1} q}{2}.
\end{aligned} \tag{4.32}$$

Además, notemos que

$$\frac{\partial U}{\partial q} = \Sigma^{-1} q, \tag{4.33}$$

esta igualdad se prueba en [29].

En este ejemplo, para aplicar el algoritmo HMC se utilizaron las funciones (4.32) y (4.33), además se tomaron como parámetros de calibración $\epsilon = 0.15$ y $L = 35$. En RWMH, se eligió para la propuesta del nuevo estado en la caminata aleatoria una distribución normal multivariada $N(\mu, I_2 \sigma^2)$, con $\mu = (0, 0)$ y $\sigma^2 = 0.15^2$, y se aplicó un lag de 35, con el objetivo de hacer la longitud de la trayectoria de la cadena comparable con la de HMC. Para ambos métodos se tomó como punto inicial $q = (-7, -7)$ y se simularon 30 estados.

La gráfica 4.5, muestra las trayectorias de las cadenas generadas con HMC y RWMH. Como se puede observar, las primeras transiciones de la cadena generada con HMC van directo a la zona con mayor densidad asociada, mientras que en RWMH debido a su aleatoriedad, no se logra percibir un comportamiento similar que lleve la cadena a la zona de interés. Por otro lado, con HMC se logra explorar gran parte de la zona con alta densidad, mientras que la cadena generada con RWMH no parece haberla recorrido.

La tasa de aceptación de ambos métodos, para esta simulación, se muestran en el cuadro 4.2.

RWMH	HMC
83 %	100 %

Cuadro 4.2: Tasa de aceptación en la simulación de la cadena

El código en Python de este ejemplo se puede consultar en este [repositorio](#).

4.3.3. Mezcla gaussiana

Es conocido que el método RWMH presenta dificultades al simular de distribuciones multimodales. Con el objetivo de exhibir si HMC comparte esta misma limitación, en este

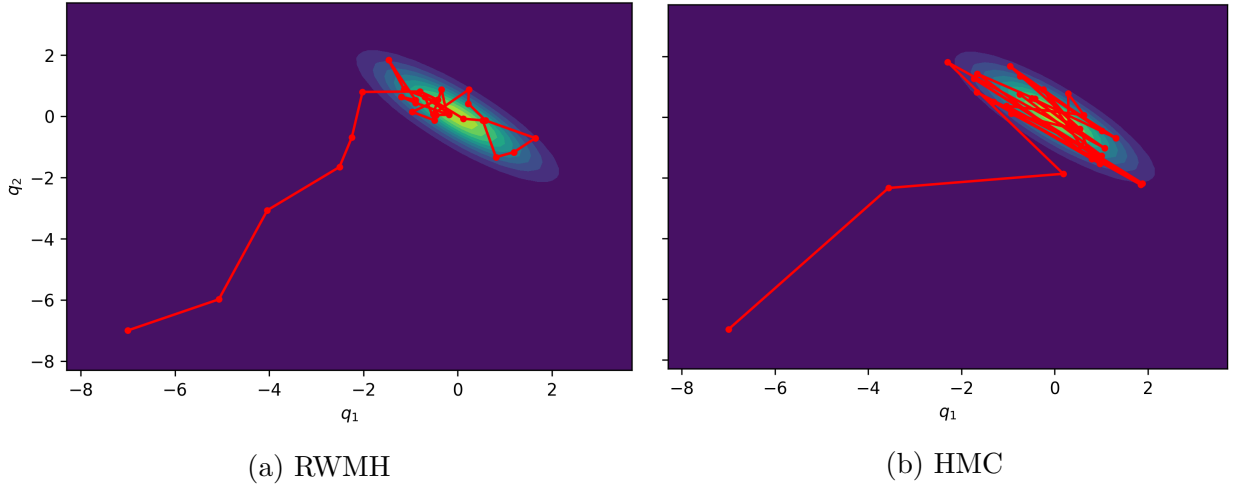


Figura 4.5: Trayectorias de las cadenas de Markov simuladas.

apartado simularemos una cadena de Markov, considerando como distribución objetivo a la mezcla gaussiana con densidad

$$S(q) = 0.4f(x | \mu_1, \Sigma_1) + 0.6f(x | \mu_2, \Sigma_2) \quad (4.34)$$

donde f es la función de densidad de la distribución normal multivariada dada en (4.35) y

$$\begin{aligned} \mu_1 &= (-4, -4), & \Sigma_1 &= \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \\ \mu_2 &= (5, 5), & \Sigma_2 &= \begin{pmatrix} 1 & -0.3 \\ -0.3 & 1 \end{pmatrix}. \end{aligned}$$

$$f(x | \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} \det(\Sigma)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x - \mu)^t \Sigma^{-1}(x - \mu)\right). \quad (4.35)$$

La gráfica 4.6 representa la densidad de la mezcla gaussiana dada en (4.34). Para simular de ésta utilizando el método HMC, debemos definir a la energía potencial U como

$$U(q) = -\log(0.4f(q | \mu_1, \Sigma_1) + 0.6f(q | \mu_2, \Sigma_2)). \quad (4.36)$$

Notemos además que

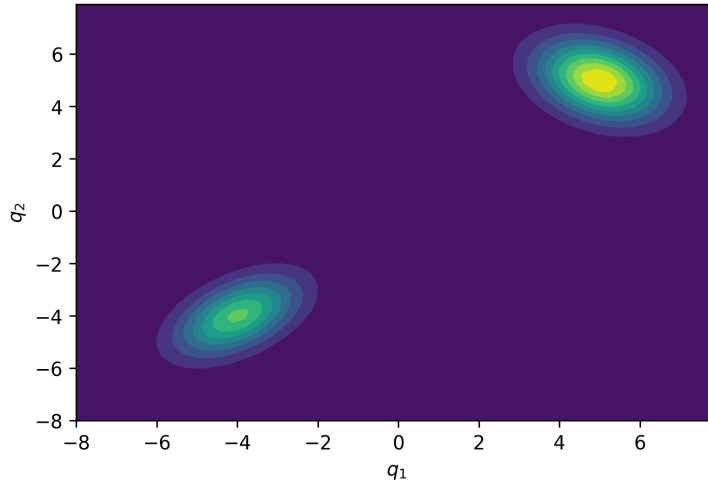


Figura 4.6: Densidad de la mezcla gaussiana.

$$\begin{aligned}
 \frac{\partial U}{\partial q} &= \frac{\partial}{\partial q} (-\log(0.4f(q | \mu_1, \Sigma_1) + 0.6f(q | \mu_2, \Sigma_2))) \\
 &= -\frac{0.4 \frac{\partial}{\partial q} f(q | \mu_1, \Sigma_1) + 0.6 \frac{\partial}{\partial q} f(q | \mu_2, \Sigma_2)}{0.4f(q | \mu_1, \Sigma_1) + 0.6f(q | \mu_2, \Sigma_2)}.
 \end{aligned} \tag{4.37}$$

En [29] se prueba que

$$\frac{\partial}{\partial q} f(q | \mu, \Sigma) = f(q | \mu, \Sigma) \Sigma^{-1} (q - \mu),$$

por lo que (4.37) puede reescribirse como

$$\frac{\partial U}{\partial q} = -\frac{0.4f(q | \mu_1, \Sigma_1) \Sigma_1^{-1} (q - \mu_1) + 0.6f(q | \mu_2, \Sigma_2) \Sigma_2^{-1} (q - \mu_2)}{0.4f(q | \mu_1, \Sigma_1) + 0.6f(q | \mu_2, \Sigma_2)}. \tag{4.38}$$

En este ejemplo, para generar las simulaciones con el algoritmo HMC se requirieron las funciones (4.36) y (4.38), además se tomaron como parámetros de calibración $\epsilon = 0.2$ y $L = 30$. En RWMH, para la propuesta de la caminata aleatoria se empleó una distribución normal multivariada $N(\mu, I_2\sigma^2)$, con $\mu = (0, 0)$ y $\sigma^2 = 0.2^2$, y se aplicó un lag de 35, con el objetivo de hacer la longitud de la trayectoria de la cadena comparable con la de HMC. Con ambos algoritmos se generaron dos cadenas de 5,000 estados, considerando como puntos iniciales $(-9, -9)$ y $(2.5, 2, 5)$, respectivamente.

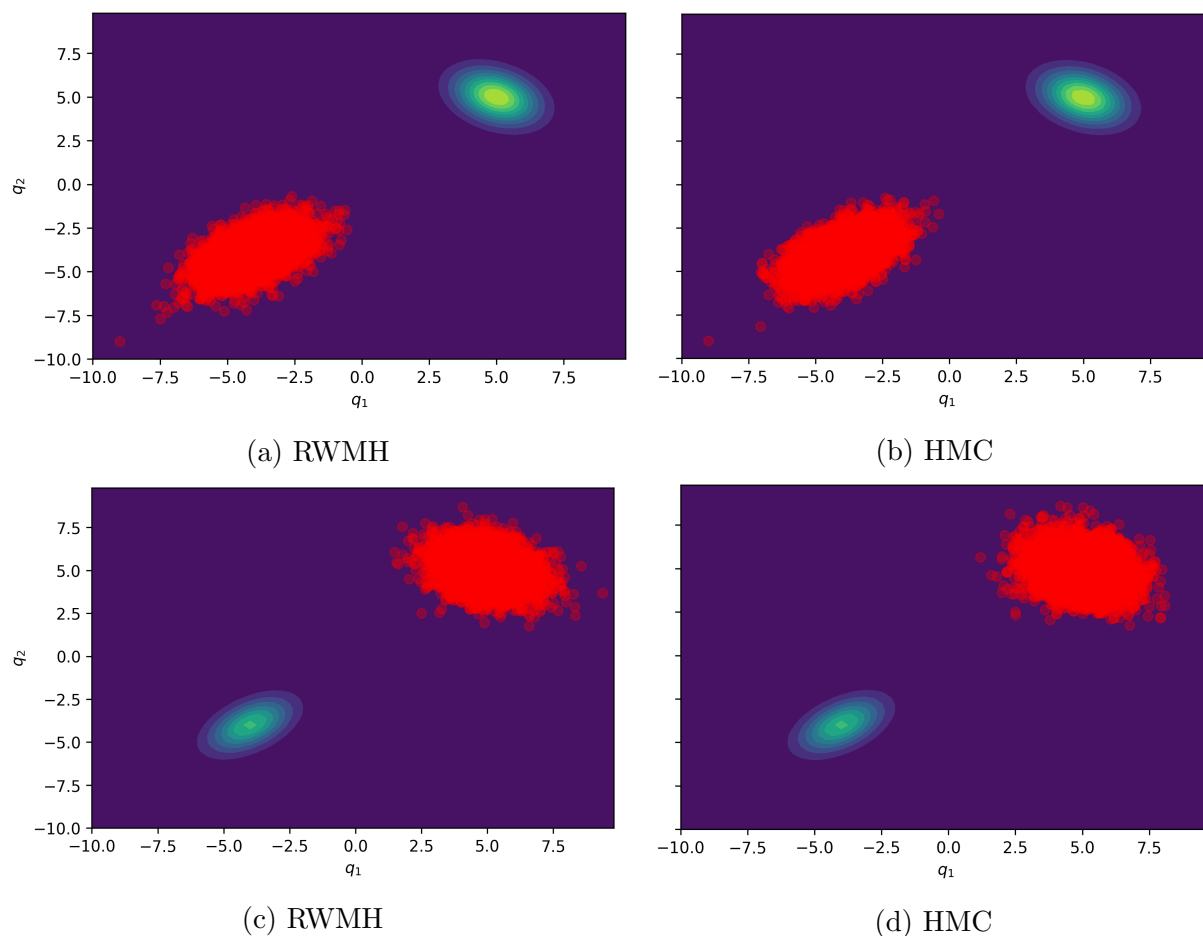


Figura 4.7: Simulación de mezcla gaussiana, para (a) y (b) se tomó como punto iniciales $(-9,-9)$, mientras que para (c) y (d) se usó $(2.5,2.5)$.

La imagen 4.7 muestra las simulaciones obtenidas en este ejemplo, como se puede observar con ninguno de los dos métodos se logra simular adecuadamente de la mezcla gaussiana, debido a que las trayectorias de las cadenas quedan atrapadas en una moda de la distribución, dependiendo del punto inicial que se utilice.

Los métodos MCMC presentan problemas al simular distribuciones multimodales. Sin embargo, sí se puede simular de éstas, en el artículo [26], se exponen posibles soluciones para lograr una simulación efectiva de una distribución multimodal utilizando métodos MCMC.

En el cuadro 4.3, se muestra el tiempo de ejecución de los algoritmos RWMH y HMC para simular las cadenas de este ejemplo. Notemos que los tiempos requeridos en HMC son

preferibles, debido a que en RWMH para generar los 5,000 estados, por el lag utilizado, se tuvo que simular una cadena de tamaño 150,000, mientras que en HMC sólo se simuló la cadena de 5,000 estados.

Punto inicial	RWMH	HMC
(-9, -9)	171.4 s	115.8 s
(2.5, 2.5)	170.5 s	120.7 s

Cuadro 4.3: Tiempos de ejecución de la simulación de la cadena de Markov.

Las tasas de aceptación de las cadenas generadas en esta sección se muestran en 4.4.

Punto inicial	RWMH	HMC
(-9, -9)	88.9 %	93.9 %
(2.5, 2.5)	90 %	91.5 %

Cuadro 4.4: Tasas de aceptación en la simulación de la cadena de Markov, con la mezcla gaussiana como distribución objetivo.

El código en Python de este ejemplo se puede consultar en este [repositorio](#).

4.3.4. Distribución normal 150-dimensional

Con el objetivo de comparar el funcionamiento del método RWMH y HMC al simular de distribuciones en grandes dimensiones, en esta sección simularemos con ambos algoritmos una distribución normal multivariada con vector de medias $\mu = \{1, 2, \dots, 150\}$ y matriz de varianzas y covarianzas $\Sigma = \text{diag}(\frac{1}{2}, 1, \dots, \frac{150}{2})$, cuyo kernel está dado por

$$S(q) = \exp\left(-\frac{(q - \mu)^T \Sigma^{-1} (q - \mu)}{2}\right).$$

Para simular de esta distribución con el método HMC, debemos definir a la energía potencial del hamiltoniano estándar como

$$\begin{aligned} U(q) &= -\log(S(q)) \\ &= -\log\left(\exp\left(-\frac{(q - \mu)^T \Sigma^{-1} (q - \mu)}{2}\right)\right) \\ &= \frac{(q - \mu)^T \Sigma^{-1} (q - \mu)}{2}. \end{aligned} \tag{4.39}$$

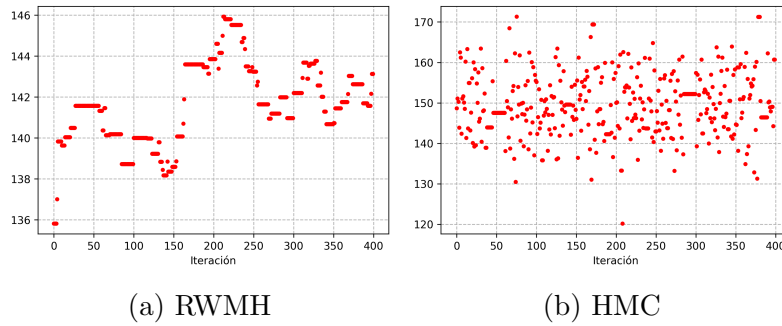


Figura 4.8: Últimos mil estados de la distribución marginal $N\left(150, \frac{150}{2}\right)$ de las cadenas generadas.

Además, notemos que

$$\frac{\partial U}{\partial q} = \Sigma^{-1} (q - \mu). \quad (4.40)$$

Con el objetivo de simular la cadena de este ejemplo con RWMH para la propuesta del nuevo estado de la caminata aleatoria se empleó una distribución normal multivariada $N(\mu, I_{150}\sigma^2)$ con $\mu = (0, 0, \dots, 0)$ y $\sigma^2 = 0.9^2$, ya que de esta forma se consiguió una tasa de aceptación de 0.237 que es cercana a la tasa óptima de 0.234 introducida en la sección 3.3.2.1. En el caso de HMC se utilizaron las funciones (4.39) y (4.40), además se tomaron como parámetros de calibración $\epsilon = 0.63$ y $L = 32$, porque de este modo se consiguió una tasa de aceptación de 0.628 que es cercana a la tasa óptima de 0.651 mencionada en la sección 4.2.1.

Si bien los parámetros utilizados para HMC y RWMH en este ejemplo no los hace comparables respecto al costo computacional, resulta interesante compararlos teniendo en cuenta que en ambos algoritmos se obtuvieron tasas de aceptación cercanas a las óptimas. La imagen 4.8, muestra los últimos 400 estados en la cadena de la distribución marginal $N(150, \frac{150}{2})$. Como se puede observar, las simulaciones en RWMH presentan un patrón que se traduce en una fuerte correlación en los estados simulados, mientras que en HMC no se percibe tal dependencia.

El cuadro 4.5 muestra los tiempos de ejecución, las tasas de aceptación y el IAT en ambos algoritmos al simular cadenas de 10,000 estados con la distribución objetivo de este ejemplo.

Concluyendo, aunque generar la cadena con RWMH es más rápido, en este ejemplo resulta más eficiente utilizar HMC porque tiene un IAT mucho menor, lo que se traduce en que sus estados no tienen una correlación tan grande como la de la cadena simulada

	RWMH	HMC
Tasa de aceptación	23.7 %	62.8 %
Tiempo de ejecución	16.81 s	24.3 s
IAT	3,083	8

Cuadro 4.5: Comparación del tiempo de ejecución y tasas de aceptación en las simulaciones de la cadena con distribución objetivo $N(\mu, I_{150}\sigma^2)$.

con RWMH, y por tanto no se tienen que desechar tantos estados para tener una muestra pseudo-independiente.

El código en Python de este ejemplo se puede consultar en este [repositorio](#).

Capítulo 5

Aplicación de HMC en un modelo normal jerárquico

En este capítulo se muestra una aplicación del método HMC para hacer inferencia bayesiana de un modelo gaussiano jerárquico, y se replica el procedimiento utilizando el algoritmo RWMH y T-walk con el objetivo de comparar sus resultados.

El algoritmo T-walk es un método MCMC que no requiere parámetros de calibración, propuesto en el año 2010 por Christen, J. y Fox, C. presentado en la sección 3.3.3. Se puede emplear en Python utilizando la función *pytwalk* de la paquetería *pytwalk-1.6.0.tar.gz* que puede ser descargada en esta [liga](#).

5.1. Modelo de las 8 escuelas

El ejemplo de las ocho escuelas, es un modelo bayesiano presentado en [14], donde se estudia el efecto de los programas de entrenamiento de ocho escuelas para el examen SAT (*Scholastic Aptitude Test*). Este examen mide las aptitudes de los estudiantes de preparatoria en USA con el objetivo de ayudar a la universidades a tomar decisiones sobre la admisión de sus postulantes.

El SAT está diseñado para reflejar los conocimientos adquiridos en varios años de educación, por lo que se pretende que los esfuerzos realizados en los últimos momentos para acrecentar la calificación de la prueba no sean útiles.

El objetivo del modelo de las ocho escuelas es identificar el impacto que tuvieron los programas de entrenamiento en los resultados del examen. El cuadro 5.1 muestra una estimación de la diferencia en el puntaje del examen entre los alumnos que tomaron y no

el curso de entrenamiento, con su respectiva desviación muestral.

Escuela	Efecto y_i	Desviación estándar muestral κ_i^2
1	2.8	0.8
2	0.8	0.5
3	-0.3	0.8
4	0.7	0.6
5	-0.1	0.5
6	0.1	0.6
7	1.8	0.5
8	1.2	0.4

Cuadro 5.1: Efectos del entrenamiento para el examen SAT y su desviación muestral. Los datos fueron obtenidos de [27].

El modelo propuesto supone que el impacto de entrenamiento de las escuelas, $y_i \mid \theta_i, \kappa_i$, es condicionalmente independiente y satisface que

$$y_i \mid \theta_i, \kappa_i \sim N(\theta_i, \kappa_i^2) \text{ para } i \in \{1, 2, \dots, 8\},$$

donde κ_i es conocida y está dada por la desviación muestral de la i -ésima escuela presentada en el cuadro 5.1, y

$$\theta_i = \mu + \tau\eta_i.$$

En este ejercicio haremos inferencia sobre los parámetros μ, τ y η_i con $i \in \{1, 2, \dots, 8\}$, siguiendo el enfoque presentado en el artículo [27], donde se asignan distribuciones independientes $N(0, 1)$ como las distribuciones *a priori* para las variables μ, τ y η_i , de tal modo que la densidad *a priori* conjunta, p , satisfaga que

$$p(\eta, \mu, \tau) \propto \exp\left(-\frac{\mu^2}{2}\right) \exp\left(-\frac{\tau^2}{2}\right) \prod_{i=1}^8 \exp\left(-\frac{\eta_i^2}{2}\right), \quad (5.1)$$

donde $\eta = (\eta_1, \eta_2, \dots, \eta_8)$.

Notemos además que la verosimilitud, $\mathcal{L}(y \mid \eta, \mu, \tau)$, cumple que

$$\mathcal{L}(y \mid \eta, \mu, \tau) \propto \prod_{i=1}^8 \exp\left(-\frac{(y_i - (\mu + \tau\eta_i))^2}{2\kappa_i^2}\right), \quad (5.2)$$

y que la densidad posterior está dada por

$$S'(y | \eta, \mu, \tau) = \frac{p(\eta, \mu, \tau) \mathcal{L}(y | \eta, \mu, \tau)}{\int_{(\eta, \mu, \tau) \in \mathbb{R}^{10}} \mathcal{L}(y | \eta, \mu, \tau) p(\eta, \mu, \tau) d\eta d\mu d\tau}. \quad (5.3)$$

Sustituyendo (5.1) y (5.2) en (5.3), podemos observar que la densidad posterior es proporcional a la función

$$S(y | \eta, \mu, \tau) = \exp\left(-\frac{\mu^2}{2}\right) \exp\left(-\frac{\tau^2}{2}\right) \prod_{i=1}^8 \exp\left(-\frac{\eta_i^2}{2}\right) \prod_{i=1}^8 \exp\left(-\frac{(y_i - (\mu + \tau\eta_i))^2}{2\kappa_i^2}\right).$$

Para simular de la distribución posterior, utilizando HMC, debemos definir la energía del hamiltoniano estandar como

$$U(-\log(S(y | \eta, \mu, \tau))) = \frac{\mu^2}{2} + \frac{\tau^2}{2} + \sum_{i=1}^8 \frac{\eta_i^2}{2} + \sum_{i=1}^8 \left(\frac{(y_i - (\mu + \tau\eta_i))^2}{2\kappa_i^2}\right). \quad (5.4)$$

Además, es necesario obtener el vector gradiente $\left(\frac{\partial U}{\partial \eta_1}, \dots, \frac{\partial U}{\partial \eta_8}, \frac{\partial U}{\partial \mu}, \frac{\partial U}{\partial \tau}\right)$ de (5.4), cuyas entradas están dadas por

$$\begin{aligned} \frac{\partial U}{\partial \eta_i} &= \eta_i - \left(\frac{(y_i - (\mu + \tau\eta_i))}{\kappa_i^2}\right) \tau \text{ para } i \in \{1, 2, \dots, 8\}, \\ \frac{\partial U}{\partial \mu} &= \mu - \sum_{i=1}^8 \left(\frac{(y_i - (\mu + \tau\eta_i))}{\kappa_i^2}\right), \\ \frac{\partial U}{\partial \tau} &= \tau - \sum_{i=1}^8 \left(\frac{(y_i - (\mu + \tau\eta_i))}{\kappa_i^2}\right) \eta_i. \end{aligned}$$

En este ejemplo, simulamos una cadena de 500,000 estados teniendo como densidad objetivo a (5.3) con los métodos HMC, RWMH y T-walk.

Para simular de la distribución objetivo con RWMH, se utilizó como punto inicial $(2, 2, \dots, 2)$ de manera arbitraria y para la propuesta del nuevo estado de la caminata aleatoria se eligió una distribución normal multivariada $N(\mu, I_{10}\sigma^2)$, con $\mu = (0, 0, \dots, 0)$ y $\sigma^2 = 0.32^2$, consiguiendo un porcentaje de aceptación en la ejecución del algoritmo de 24.6%; la elección de σ^2 se hizo buscando que la tasa de aceptación fuera cercana a la tasa óptima de 23.4% introducida en la sección 3.3.2.1.

Para generar las simulaciones con el método HMC, se utilizó el gradiente $\left(\frac{\partial U}{\partial \eta_1}, \dots, \frac{\partial U}{\partial \eta_8}, \frac{\partial U}{\partial \mu}, \frac{\partial U}{\partial \tau}\right)$ y la función (5.4), como punto inicial se tomó $(2, 2, \dots, 2)$ y como

parámetros de calibración $\epsilon = 0.08$ y $L = 60$, que si bien no se seleccionaron bajo un criterio de óptimalidad, se eligieron considerando un ϵ lo suficientemente fino y un L vasto para que las transiciones en la cadena fueran adecuadas y produjeran buenos resultados.

El algoritmo T-walk no requiere parametros de calibración, pero su ejecución demanda dos puntos iniciales, para este caso, de manera arbitraria, se utilizaron $(2, 2, \dots, 2)$ y $(3, 3, \dots, 3)$.

Para visualizar los resultados, presentaremos las densidades conjuntas de (μ, τ) y (τ, η_1) , para ello, es necesario determinar el periodo de *burn-in* y lag de las cadenas simuladas.

En la figura 5.1 se muestra el periodo de *burn-in* de las cadenas simuladas, como se puede observar la cadena generada con el método HMC es la que parece llegar con menos iteraciones a las zonas con mayor densidad asociada.

Para seleccionar el lag, l , y conseguir muestras pseudo-independientes de cada cadena, se analizaron las autocorrelaciones de las variables marginales, y con los tres métodos la que presentaba una mayor autocorrelación era τ , por lo que l se eligió como el lag en el que las autocorrelaciones de τ fueran cercanas a cero. En la figura 5.2 se muestra las autocorrelaciones de la variable τ ; como se puede observar la cadena generada con HMC tiene una menor dependencia en los estados que los de la cadena generada con RWMH y T-walk.

En la figura 5.3 se presenta las densidades conjuntas de la muestra efectiva obtenida de las cadenas simuladas con HMC, RWMH, y T-walk.

En el cuadro 5.2 se muestran las principales diferencias en los algoritmos HMC, RWMH y T-walk al simular la cadena de Markov con densidad objetivo S' . Como se puede observar, el método HMC resulta más eficiente respecto a tiempo que RWMH y T-walk, ya que tarda aproximadamente 23 veces menos en generar una muestra que T-walk, y cerca de 28 veces menos de lo que tarda RWMH, esto se debe principalmente a que los estados de las cadenas simuladas con RWMH y T-walk tienen una alta correlación, por lo que se tiene que desechar una gran parte de las simulaciones con el objetivo de obtener la muestra pseudo-independiente buscada.

Como último ejercicio, se generó una muestra efectiva de tamaño 500 con los métodos HMC, RWMH, y T-walk, considerando el *burn-in* y lag determinado para cada método en el cuadro 5.2. En la gráfica 5.4 se muestra las densidades de las distribuciones marginales (μ, τ) y (τ, η_1) , obtenidas en estas simulaciones.

El cuadro 5.3 presenta las diferencias entre HMC, RWMH y T-walk, respecto al tiempo

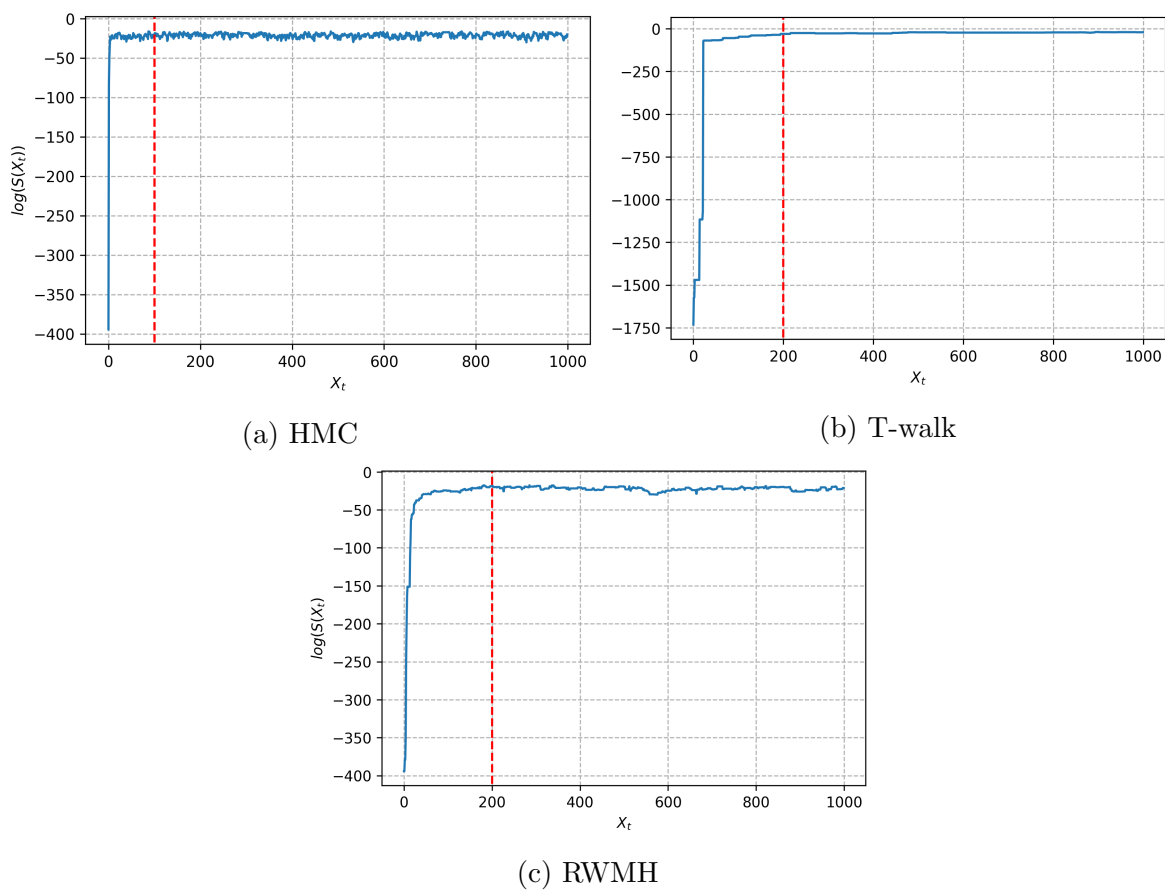


Figura 5.1: Periodo de *burn-in* en las cadenas simuladas con densidad objetivo S' . La línea roja punteada representa la longitud de la cadena en que las log-densidades dejan de tener un comportamiento creciente.

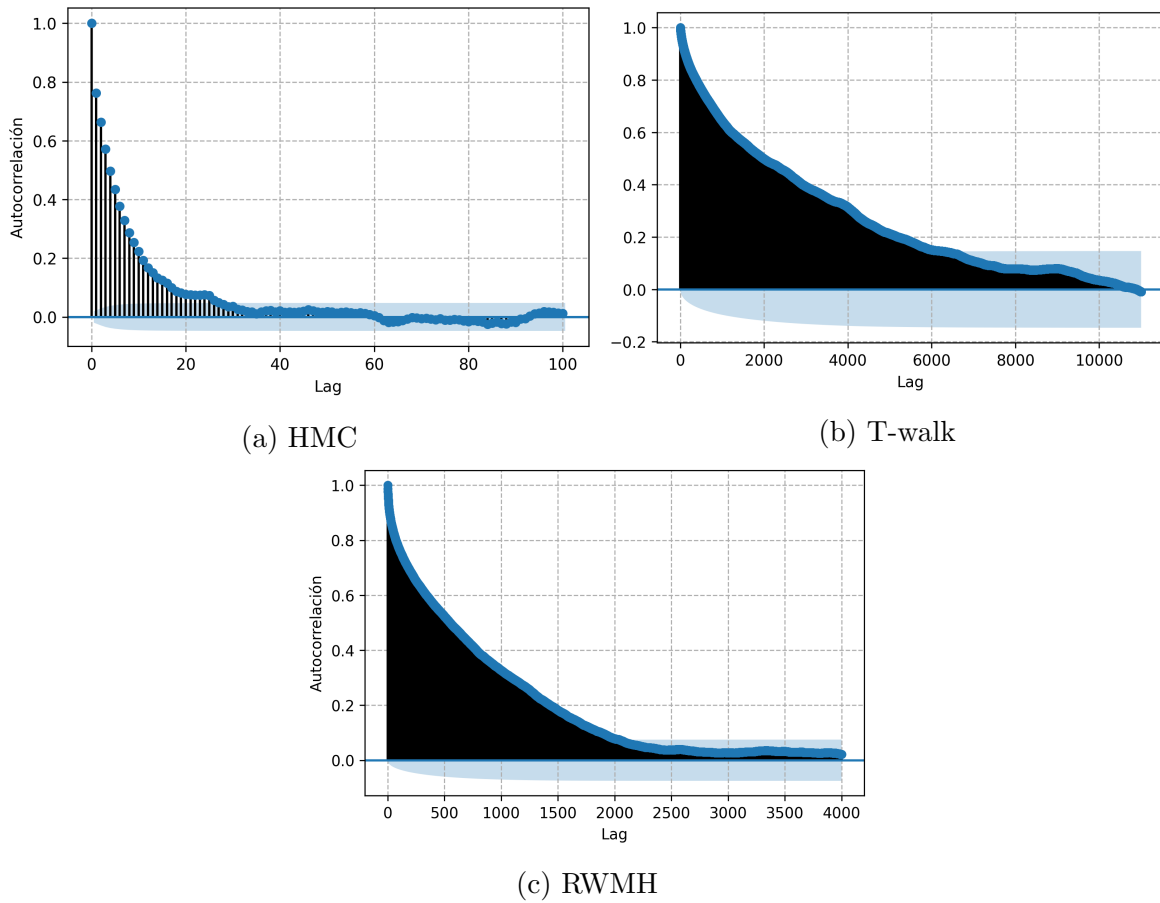


Figura 5.2: Autocorrelaciones de la distribución marginal de τ en las cadenas simuladas con densidad objetivo S' .

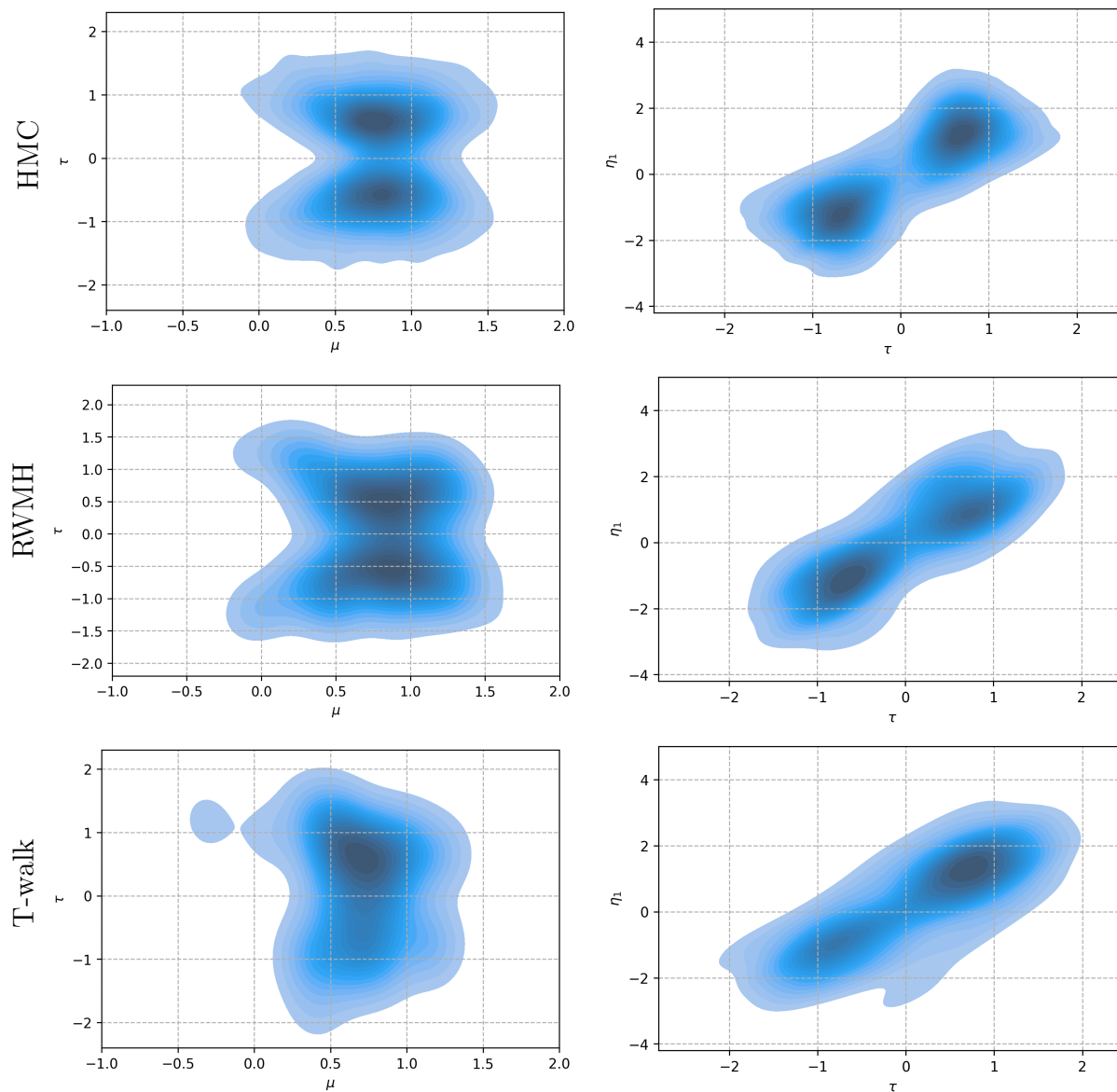


Figura 5.3: Densidades de las distribuciones marginales (μ, τ) y (τ, η_1) simuladas en las cadenas de Markov con densidad objetivo S' , para una cadena cruda de 500,000 estados con muestra efectiva de 14238, 218, y 56 generadas con HMC, RWMH, Y T-walk, respectivamente.

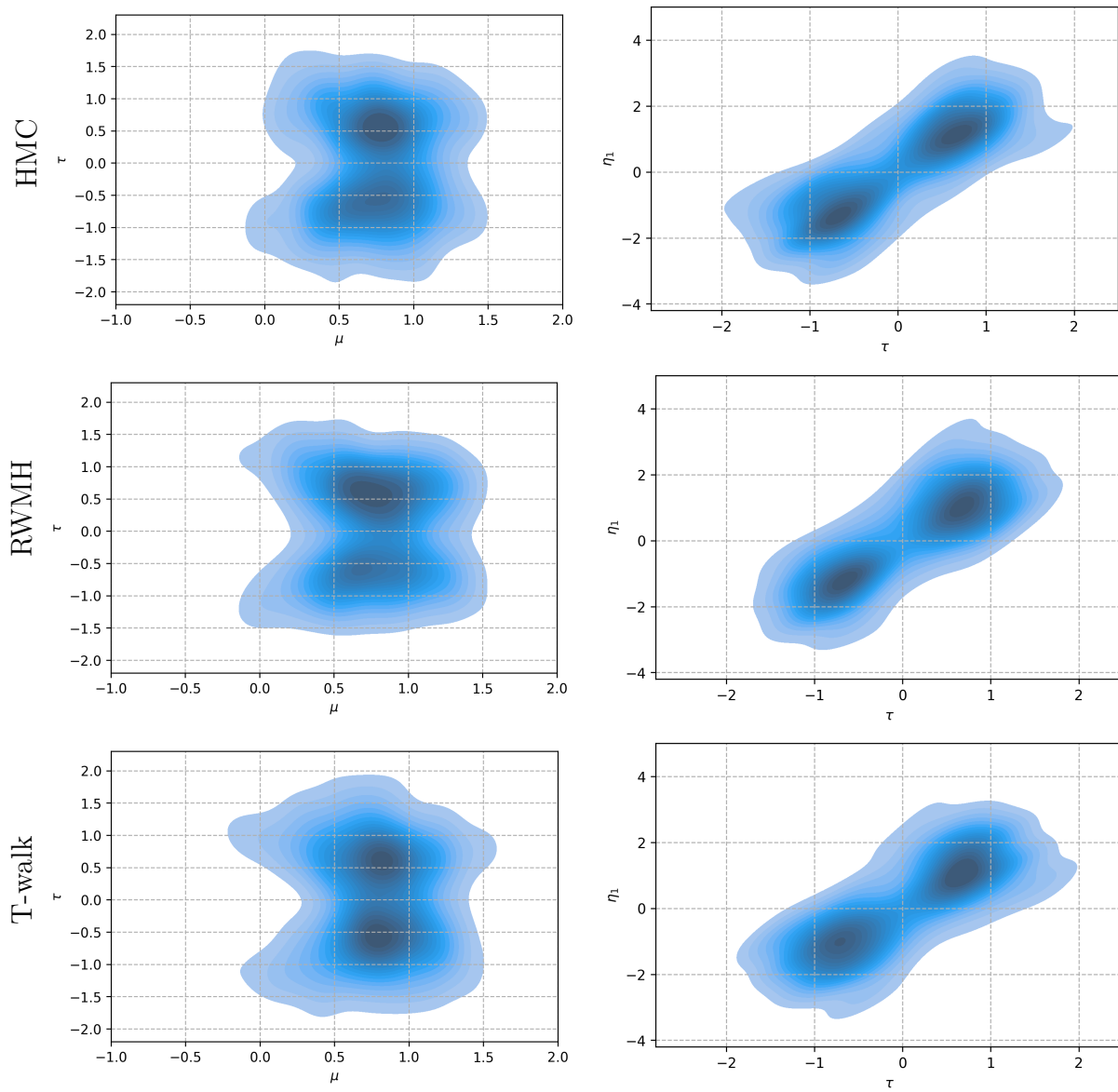


Figura 5.4: Densidades de las 500 muestras simuladas de las distribuciones marginales (μ, τ) y (τ, η_1) con densidad objetivo S' .

	HMC	RWMH	T-walk
<i>Burn-in</i>	100	200	200
<i>lag</i>	35	2300	9000
Tasa de aceptación	93.05 %	24.68 %	18.22 %
Muestra efectiva	14283	218	56
Tiempo de ejecución	9968 s	4345 s	907 s
Segundos por muestra	0.698	19.93	16.2

Cuadro 5.2: Comparativo entre HMC, RWMH, y T-walk, al simular la distribución objetivo con densidad S' .

de ejecución, tasas de aceptación, y la cantidad de iteraciones necesarias para generar la muestra de tamaño 500 de la distribución de interés. Es importante notar que el tiempo que tarda HMC en generar las simulaciones es considerablemente menor que lo que tarda RWMH y T-walk.

	HMC	RWMH	T-walk
Tasa de aceptación	92.49 %	24.77 %	18.52 %
Tiempo de ejecución	166 s	16861 s	7888 s
Simulaciones necesarias	17,600	1,150,200	4,500,200

Cuadro 5.3: Comparativo al simular de 500 muestras de la distribución objetivo con densidad S' .

El código en Python utilizado para las simulaciones de este capítulo, se puede consultar en este [repositorio](#).

Capítulo 6

Conclusiones

En las comparaciones del método HMC y RWMH presentadas en la sección 4.3, se puede notar que el método HMC resulta más eficiente en tiempo que RWMH. Su ventaja se debe a la información que utiliza del gradiente de la densidad objetivo y a la aplicación de las dinámicas hamiltonianas, que logran que las transiciones de la cadena sean adecuadas para explorar la distribución objetivo.

Cada iteración del método HMC es más costosa que las del algoritmo RWMH, porque se necesita resolver o aproximar las ecuaciones de Hamilton. Sin embargo, resulta más eficiente ya que HMC usualmente necesita menos iteraciones para obtener una muestra efectiva de la distribución de interés, debido a que el periodo de *burn-in* generalmente, es menor en HMC que RWMH, ya que cuando la cadena está en zonas con baja densidad asociada, HMC logra desplazar la cadena de una manera rápida a las zonas de interés, aunado a que las autocorrelaciones en los estados de la cadena también suelen ser considerablemente menores en HMC que en RWMH.

Otra diferencia importante entre el algoritmo HMC y RWMH es su tasa de aceptación óptima. Como se mencionó en la sección 3.3.2.1, bajo ciertos supuestos, el algoritmo RWMH tiene una tasa de aceptación óptima de 0.234. En el caso de HMC, un resultado equivalente se presentó en la sección 4.2.1, donde se menciona que bajo ciertas hipótesis la tasa óptima en este algoritmo es de 0.651. Esto también resulta ser una causa de que el algoritmo HMC pueda conservar más simulaciones como una muestra efectiva de la distribución objetivo que RWMH.

Es crucial resaltar que HMC no resulta una panacea para las limitaciones de los métodos MCMC, pues como se mostró en el ejemplo de la sección 4.3.3, también presenta dificultades al simular de distribuciones con más de una moda.

Además, resulta plausible la eficiencia en tiempo del método HMC, al hacer inferencia

del modelo presentado en la sección 5.1, pues como se muestra en éste, genera las simulaciones de una manera mucho más rápida que RWMH y T-walk.

Por último, resulta fundamental resaltar que el método HMC presenta un limitación importante, ya que requiere información del gradiente de la densidad objetivo, y ésta en muchas ocasiones no es conocida, lo que impide su aplicación.

Apéndice A

Breve introducción a la mecánica clásica

En [13] se afirma de manera lacónica que: “*La Mecánica es el estudio de cómo la interacción entre los cuerpos causa el cambio de su estado de movimiento.*”.

La mecánica clásica, generalmente se considera iniciada por Isaac Newton¹ en 1687, con la publicación de “*Philosophiae Naturalis Principia Mathematica*” donde se establecen las bases de la mecánica clásica por medio de cuatro leyes: tres sobre el movimiento y una llamada ley de la gravitación universal que permite unificar la mecánica terrestre y celeste.

El objeto de estudio de la mecánica newtoniana es el movimiento de partículas en espacios vectoriales reales, tridimensionales y completos con producto interior (euclidianos).

Las partículas son una abstracción matemática que permite considerar a cuerpos u objetos como puntos en el espacio, a estos se les puede atribuir propiedades físicas como la masa que es la cantidad de materia que contiene un cuerpo y cuyas unidades de medida en el Sistema Internacional de Unidades (S.I., por sus siglas en francés), ² son el kg.

El movimiento de una partícula en el espacio euclidiano puede ser descrito por una función vectorial $r(t) = (x(t), y(t), z(t))$ que indica la posición de la partícula a un tiempo t , en función de cada eje de coordenadas.

El cambio infinitesimal de la posición respecto al tiempo se denomina velocidad, se denota por $v(t)$, y se calcula como la derivada del vector de posición respecto al tiempo,

¹(1642-1727), físico, teólogo, alquimista y matemático inglés, hizo importantes contribuciones a la física, mecánica y matemáticas.

²Creado por el Comité Internacional de Pesos y Medidas en Francia en el año 1960, establece medidas a utilizar internacionalmente.

en notación matemática

$$v(t) = \frac{dr(t)}{dt} = \left(\frac{dx(t)}{dt}, \frac{dy(t)}{dt}, \frac{dz(t)}{dt} \right).$$

Así mismo, la aceleración al tiempo t , se define como el cambio infinitesimal de velocidad respecto al tiempo. Se calcula como la derivada del vector de velocidad, y se denota por $a(t)$, donde

$$a(t) = \frac{dv(t)}{dt} = \left(\frac{d^2x(t)}{dt^2}, \frac{d^2y(t)}{dt^2}, \frac{d^2z(t)}{dt^2} \right).$$

Ejemplo 6. Suponga que la posición de una partícula está dada por la función $r(t) = (t^2, t + 5, \text{sen}(t))$ para todo $t \geq 0$, entonces su velocidad será $v(t) = (2t, 1, \text{cos}(t))$ y su aceleración $a(t) = (2, 0, -\text{sen}(t))$. En la figura A.1 se muestra la dinámica de esta partícula, generada con el código en Python presentado en este [repositorio](#).

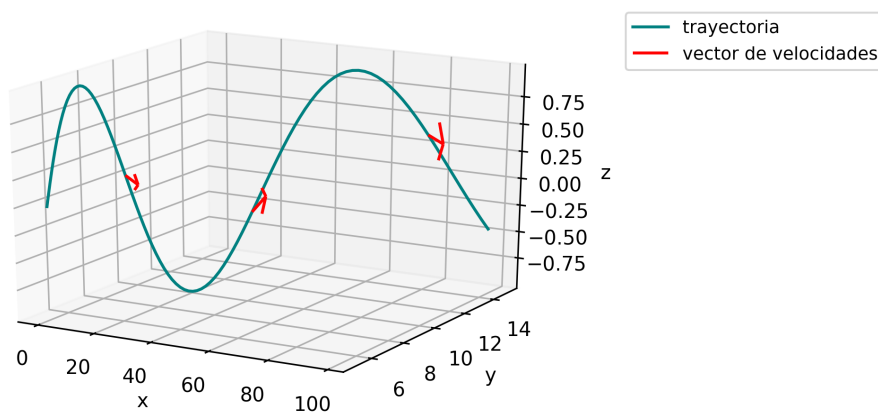


Figura A.1: Trayectoria de la partícula en el tiempo, con $t \in [0, 10]$, y $v(t)$ para t en $\{3, 6, 9\}$.

A.0.1. Leyes de Newton

A.0.1.1. Primera Ley de Newton:

La primera ley de Newton, también llamada ley de la inercia, consultada en [30, p.11], afirma que “*Un cuerpo no sometido a acciones externas conserva el estado de movimiento en que se encuentra cuando se le observa desde un sistema inercial, es decir, no acelerado*”.

Se define a un sistema de referencia inercial como aquel en que se satisface la primera ley de Newton. Estos sistemas son indispensables por la relatividad del movimiento, pues éste

depende de la posición de donde se observe. Por ejemplo: en un tren en movimiento, un pasajero puede ver por la ventana a una persona pasando de manera muy veloz, cuando en realidad estaba detenida o en reposo.

La inercia es la resistencia que tienen los cuerpos a cambiar su estado de reposo o movimiento, mientras que a la acción capaz de modificar ese estado se le conoce como fuerza. Existen distintos tipos de fuerzas, como la fuerza de gravedad, de tensión, de fricción, entre otras, se representan por vectores, y a la suma vectorial de todas las fuerzas que actúan sobre un cuerpo se le conoce como fuerza resultante.

La primera ley de Newton, por tanto, afirma que para que un cuerpo altere su movimiento debe existir alguna fuerza que lo provoque, o en términos matemáticos, donde F_i denota la i -ésima fuerza aplicada a un cuerpo,

$$\sum_{i=1}^n F_i(x(t), y(t), z(t)) = (0, 0, 0) \text{ si y sólo si } \frac{dv}{dt} = (0, 0, 0),$$

es decir, la fuerza resultante aplicada a una partícula es cero si y sólo si no se modifica su velocidad.

A.0.1.2. Segunda ley de Newton

La segunda ley de Newton utiliza el término de masa inercial, que debe entenderse como una medida de resistencia de la masa a cambiar su estado de movimiento en relación con un sistema de referencia inercial.

Esta ley consultada en [30, p.11], menciona que “*Visto desde un marco de referencia inercial, el cambio en la velocidad de un cuerpo es proporcional a la fuerza ejercida sobre él, siendo la constante de proporcionalidad un escalar característico del cuerpo llamado masa inercial*”.

De este modo, la segunda ley de Newton afirma que en un marco de referencia inercial se cumple que para cualquier tiempo t ,

$$F(x(t), y(t), z(t)) = \frac{dv(t)}{dt} m(t), \quad (\text{A.1})$$

donde $m(t)$ y $F(t)$ representan la masa inercial y la fuerza a un tiempo t , respectivamente. De (A.1) se sigue la ecuación que relaciona la fuerza con la masa y aceleración

$$F(x(t), y(t), z(t)) = m(t) a(t).$$

A.0.1.3. Tercera ley de Newton

La tercera ley de Newton, consultada en [30, p.12], afirma que: “*Las acciones mutuas que dos cuerpos ejercen entre sí están representadas por fuerzas de igual magnitud, dirección y de sentidos opuestos*”.

Esta ley se conoce también como el principio de acción y reacción. Afirma que si un cuerpo aplica una fuerza sobre otro, éste último ejercerá una fuerza de igual magnitud y en la dirección opuesta sobre el primero, es decir

$$f_{ij} = -f_{ji},$$

donde f_{ij} es la fuerza producida por la acción del cuerpo i sobre un cuerpo j .

El ejemplo 7 muestra cómo obtener la función $r(t)$ para un objeto adherido a un resorte, utilizando las leyes de Newton.

Ejemplo 7. Consideremos el sistema presentado en el ejemplo 1. Podemos abstraerlo, considerando al cuerpo como una partícula que se mueve solo sobre el eje x , es decir, donde $y(t) = z(t) = 0$ para cualquier tiempo t . Denotaremos por $r : \mathbb{R} \rightarrow \mathbb{R}$ a la función que al tiempo t devuelve la posición de la partícula en el eje x .

El movimiento que sigue esta partícula inducido por el resorte se conoce como movimiento armónico simple, éste es un movimiento periódico derivado de una fuerza proporcional a la distancia entre la partícula y su posición de equilibrio (aquella donde la fuerza del resorte es cero). Si consideramos que el punto de equilibrio ocurre en $x = 0$, entonces debe cumplirse que para algún tiempo t , existe una $k \in \mathbb{R}^+$ tal que

$$F(t) = -kr(t). \tag{A.2}$$

El signo negativo en (A.2) es utilizado para modelar el sistema, de esta manera cuando la partícula se encuentre en una posición x , con $x < 0$ (es decir, del lado izquierdo del punto de equilibrio) la fuerza del resorte será positiva logrando mover a la partícula al lado derecho del punto de equilibrio, de manera análoga, cuando $x > 0$ la fuerza será negativa logrando desplazar a la partícula al lado izquierdo del punto de equilibrio.

Para encontrar la función $r(t)$, igualemos (A.2) con el resultado de la segunda ley de Newton que afirma que $F(t) = ma(t)$, suponiendo la masa constante

$$-kr(t) = m \frac{d^2r(t)}{dt^2},$$

esto implica que

$$0 = m \frac{d^2 r(t)}{dt^2} + kr(t). \quad (\text{A.3})$$

La expresión (A.3) es una ecuación diferencial homogénea con coeficientes constantes y de segundo orden, cuyas soluciones son de la forma $C_1 e^{\alpha t}$ como se afirma en [7, p.132], por tanto, de (A.3) se sigue que

$$\begin{aligned} 0 &= m \frac{d^2 (e^{\alpha t})}{dt^2} + k e^{\alpha t} \\ &= m \alpha^2 e^{\alpha t} + k e^{\alpha t} \\ &= m \alpha^2 + k, \end{aligned}$$

despejando a α se obtiene que

$$\alpha = \pm \sqrt{-\frac{k}{m}}. \quad (\text{A.4})$$

Dado que $-\frac{k}{m}$ siempre es negativo, los valores α que satisfacen la ecuación (A.4) están dados por

$$\alpha_1 = i \sqrt{\frac{k}{m}} \quad \text{y} \quad \alpha_2 = -i \sqrt{\frac{k}{m}},$$

así se sigue que la solución de la ecuación diferencial es de la forma

$$r(t) = C_1 e^{iwt} + C_2 e^{-iwt}, \quad (\text{A.5})$$

donde $w = \sqrt{\frac{k}{m}}$ y $C_1, C_2 \in \mathbb{R}$.

Si aplicamos las ecuaciones de Euler

$$\begin{aligned} e^{i\theta} &= \cos(\theta) + i \operatorname{sen}(\theta) \\ e^{-i\theta} &= \cos(\theta) - i \operatorname{sen}(\theta), \end{aligned}$$

a la función dada en (A.5), obtenemos que

$$\begin{aligned} r(t) &= C_1 (\cos(wt) + i \operatorname{sen}(wt)) + C_2 (\cos(wt) - i \operatorname{sen}(wt)) \\ &= (C_1 i - C_2 i) (\operatorname{sen}(wt)) + (C_1 + C_2) \cos(wt) \\ &= A \operatorname{sen}(wt) + B \cos(wt), \end{aligned} \quad (\text{A.6})$$

donde $B = C_1 + C_2$ y $A = C_1i - C_2i$.

Las funciones de la forma de (A.6), son soluciones de la ecuación diferencial (A.3). Si suponemos como condición inicial del sistema, $r(0) = 0$, es decir, al tiempo cero la posición de la partícula es cero, se sigue que

$$\begin{aligned} 0 &= B \cos(w0) + A(\sin(w0)) \\ &= B, \end{aligned} \tag{A.7}$$

finalmente, sustituyendo (A.7) en (A.6), obtenemos la trayectoria $r(t)$ que seguirá la partícula en el tiempo,

$$r(t) = A \sin(wt). \tag{A.8}$$

Al valor A en (A.8) se le llama amplitud del movimiento y se refiere al punto más alejado que toma la partícula del punto de equilibrio, al valor $\frac{2\pi}{w}$ se le conoce como periodo y es el tiempo que tarda la partícula en recorrer un ciclo completo.

A.0.2. Trabajo y energía cinética

El trabajo es un escalar, cuya unidad de medida en el S.I. es el Julio³ denotado por J . En mecánica, se dice que una fuerza aplicada a una partícula realiza trabajo cuando es capaz de desplazarla de un punto a otro. La definición de trabajo se basa en una integral de línea, cuyo concepto se presenta en la definición 13, si el lector no está familiarizado con integrales de línea puede consultar [25, pp. 419-436].

Definición 13 (Integral de línea). *Dado $F : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$ un campo vectorial, continuo sobre una trayectoria C de clase C^1 (es decir, con derivadas parciales continuas), parametrizada por una función $r(t)$ con t en el intervalo cerrado $[a, b]$. La integral de línea a lo largo de C , se define como*

$$\int_C F(r(t)) \cdot dr = \int_a^b F(r(t)) \cdot \frac{dr(t)}{dt} dt.$$

donde \cdot denota el producto punto.

Considere una partícula de masa m , que se mueve de acuerdo a una trayectoria C bajo la acción de una fuerza resultante F dependiente de la posición de la partícula, de tal forma que el trabajo, denotado por W_{ab} , depende de la trayectoria y se calcula como

³es equivalente a $kg \left(\frac{m}{s}\right)^2$.

$$W_{ab} = \int_C F(r(t)) \cdot dr = \int_a^b F(r(t)) \cdot \frac{dr(t)}{dt} dt,$$

donde $r(t)$ es la parametrización de la curva C por la que se mueve la partícula y $\frac{dr(t)}{dt}$ como se definió anteriormente, es el vector de velocidad. Utilizando la segunda ley de Newton y suponiendo la masa constante, $F = ma(t)$, se sigue que

$$\begin{aligned} W_{ab} &= \int_a^b m \frac{dv(t)}{dt} v(t) dt \\ &= m \int_a^b \frac{dv(t)}{dt} v(t) dt \\ &= m \int_a^b \frac{d}{dt} \left(\frac{1}{2} v(t)^2 \right) dt \end{aligned} \tag{A.9}$$

$$= \frac{m}{2} (v(b)^2 - v(a)^2), \tag{A.10}$$

en la igualdad (A.9) se utilizó la regla de la cadena, el cuadrado en el vector $v(t)$ hace referencia al producto punto de $v(t)$ consigo mismo, y en (A.10) se aplicó el teorema fundamental del cálculo.

Si se define una función T dependiente de la velocidad, con la regla de correspondencia $T(v(t)) = \frac{1}{2}mv(t)^2$, podemos reescribir el trabajo dado en (A.10) como

$$W_{ab} = T(v(b)) - T(v(a)). \tag{A.11}$$

A $T(v(t))$ se le conoce como la energía cinética de la partícula al tiempo t . En mecánica, la energía se refiere a la capacidad que tienen los cuerpos de realizar un trabajo, y la energía cinética es aquella que posee un cuerpo por el hecho de estar en movimiento, su unidad de medida en el S.I. es el Julio (J).

Concluyendo, el resultado (A.11) afirma que el trabajo realizado por una fuerza es igual al cambio en energía cinética en el tiempo $b - a$.

Es importante mencionar además, que en un sistema de n partículas, la energía cinética T a un tiempo t , que denotaremos por T_t se calcula como la suma de las energías cinéticas de cada partícula, es decir

$$T_t = \sum_{i=1}^n T(v_i(t)).$$

A.0.3. Campos de fuerzas conservativas y energía potencial

Un campo de fuerzas conservativas es aquel en que el trabajo realizado por una fuerza para desplazar a una partícula de un punto A a un punto B, es independiente de la trayectoria.

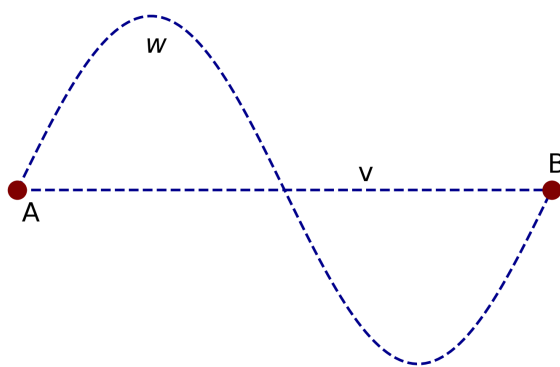


Figura A.2: Campos de fuerzas conservativas

Por definición, en los campos de fuerzas conservativas, para cualesquiera trayectorias v y w de un punto A a un punto B (como se ilustra en la gráfica A.2, generada con el código presentado es este [repositorio](#)), se cumplirá que

$$\int_w F(r(t)) dr = \int_v F(r(t)) dr,$$

o bien,

$$\int_w F(r(t)) dr - \int_v F(r(t)) dr = 0. \quad (\text{A.12})$$

El lado izquierdo de la igualdad (A.12) puede reescribirse como una integral de línea sobre una curva Z, definida como una trayectoria cerrada que comienza y termina en A y pasa por el punto B,

$$\int_Z F(r(t)) dr = 0. \quad (\text{A.13})$$

Para encontrar las fuerzas F que cumplan la igualdad (A.13), se hace uso del teorema de Stokes, que relaciona las integrales de superficie con integrales de línea. Su demostración puede ser encontrada en [25, pp.510-516] y es enunciado en el teorema 9.

Teorema 9. *Teorema de Stokes.* Sea S una superficie orientada, definida por una parametrización uno a uno $\Phi : D \subset \mathbb{R}^2 \rightarrow S$. Denotemos por r a la frontera orientada de S , y sea F un campo vectorial C^1 en S . Entonces se cumple que

$$\int_S (\nabla \times F) dS = \int_r F dr,$$

donde ∇ denota al vector gradiente y \times al producto cruz.

Así, sustentados por el teorema 9 y la ecuación (A.13), se sigue que las fuerzas conservativas son aquellas que para alguna superficie S que satisfaga las hipótesis del teorema 9, cumplen que

$$\int_S (\nabla \times F) dS = 0. \quad (\text{A.14})$$

Notemos que $F = -\nabla U(r(t))$, donde U es una función escalar que depende de la posición de la partícula, satisface la ecuación A.14, debido a que $\nabla \times (-\nabla U) = 0$, como se muestra a continuación $\times(-U)$:

$$\begin{aligned} \nabla \times (-\nabla U) &= \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) \times \left(-\frac{\partial U}{\partial x}, -\frac{\partial U}{\partial y}, -\frac{\partial U}{\partial z} \right) \\ &= \det \begin{pmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ -\frac{\partial U}{\partial x} & -\frac{\partial U}{\partial y} & -\frac{\partial U}{\partial z} \end{pmatrix} \\ &= \left(\det \begin{pmatrix} \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ -\frac{\partial U}{\partial y} & -\frac{\partial U}{\partial z} \end{pmatrix}, -\det \begin{pmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial z} \\ -\frac{\partial U}{\partial x} & -\frac{\partial U}{\partial z} \end{pmatrix}, \det \begin{pmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} \\ -\frac{\partial U}{\partial x} & -\frac{\partial U}{\partial y} \end{pmatrix} \right) \\ &= \left(\left(\frac{\partial}{\partial z} \frac{\partial U}{\partial y} - \frac{\partial}{\partial y} \frac{\partial U}{\partial z} \right), -\left(\frac{\partial}{\partial z} \frac{\partial U}{\partial x} - \frac{\partial}{\partial x} \frac{\partial U}{\partial z} \right), \left(\frac{\partial}{\partial y} \frac{\partial U}{\partial x} - \frac{\partial}{\partial x} \frac{\partial U}{\partial y} \right) \right) \\ &= \left(\left(\frac{\partial^2 U}{\partial y \partial z} - \frac{\partial^2 U}{\partial y \partial z} \right), -\left(\frac{\partial^2 U}{\partial x \partial z} - \frac{\partial^2 U}{\partial x \partial z} \right), \left(\frac{\partial^2 U}{\partial x \partial y} - \frac{\partial^2 U}{\partial x \partial y} \right) \right) \\ &= (0, 0, 0). \end{aligned}$$

Por tanto, las fuerzas conservativas son aquellas que pueden escribirse de la forma

$$F = \left(-\frac{\partial U}{\partial x}, -\frac{\partial U}{\partial y}, -\frac{\partial U}{\partial z} \right) = -\nabla U(r(t)). \quad (\text{A.15})$$

El valor $U(r(t))$ en (A.15) se conoce como la energía potencial y se refiere a la energía que posee una partícula debido a su posición.

A.0.4. Conservación de la energía mecánica

La energía mecánica de un sistema de partículas a un tiempo t , denotada por $E(t)$, se define como la suma de la energía cinética y la energía potencial del sistema, lo que se formula como

$$E(t) = T(v(t)) + U(r(t)).$$

La energía mecánica se conserva, es decir no cambia su valor en el tiempo, tal resultado puede probarse mostrando que la derivada de $E(t)$ respecto al tiempo t , es cero, como se muestra a continuación:

$$\begin{aligned} \frac{dE(t)}{dt} &= \frac{d}{dt} (T(v(t)) + U(r(t))) \\ &= \frac{d}{dt} \left(\frac{1}{2}mv(t)^2 + U(r(t)) \right), \end{aligned}$$

por la regla de la cadena en funciones multivariadas,

$$\frac{dE(t)}{dt} = \frac{1}{2}m \frac{d}{dt}v(t)^2 + \nabla U(r(t)) \frac{dr(t)}{dt},$$

aplicando la derivada del producto de funciones,

$$\begin{aligned} \frac{dE(t)}{dt} &= \frac{1}{2}m (v(t)a(t) + a(t)v(t)) + \nabla U(r(t))v(t) \\ &= mv(t)a(t) + \nabla U(r(t))v(t) \\ &= v(t)(ma(t) - F), \end{aligned}$$

por la segunda ley de Newton,

$$\begin{aligned} \frac{dE(t)}{dt} &= v(t)(0) \\ &= 0. \end{aligned}$$

Dado que $\frac{dE(t)}{dt} = 0$, se sigue que la energía mecánica es constante en el tiempo, a este resultado se le conoce como la ley de la conservación de la energía.

En el ejemplo 8 se muestra cómo calcular la energía cinética, potencial y mecánica del sistema masa-resorte.

Ejemplo 8. Retomemos el ejemplo 7. Calcularemos la energía mecánica del sistema de la partícula sujeta al resorte horizontal.

La energía mecánica del sistema está dada por la suma de la energía cinética y la energía potencial. En el ejemplo 7 se mostró que la posición de la partícula a un tiempo t , está dada por la función $r(t) = A \sin(\omega t)$, y por tanto, la velocidad con que se mueve está dada por

$$\begin{aligned} v(t) &= \frac{dr(t)}{dt} \\ &= \frac{d(A \sin(\omega t))}{dt} \\ &= A \omega \cos(\omega t), \end{aligned}$$

con este resultado, podemos calcular la energía cinética como

$$\begin{aligned} T(v(t)) &= \frac{1}{2} m v(t)^2 \\ &= \frac{1}{2} m (A \omega \cos(\omega t))^2. \end{aligned} \tag{A.16}$$

Por otro lado, la energía potencial U , como se mostró en la sección A.0.3 debe satisfacer que $F = -\nabla U(x)$, donde x representa la posición de la partícula, al trabajar con una dimensión el gradiente no es más que la derivada respecto al eje en que se da el movimiento, es decir $F = -\frac{d}{dx} U(x)$. Tengamos presente también, que en el ejemplo 7 se mostró que la fuerza del sistema en una posición x es $F = -kx$, igualando estos resultados obtenemos que

$$-kx = -\frac{d}{dx} U(x),$$

integrando en ambos lados tenemos

$$\begin{aligned} -\int kx dx &= -U(x) + c_1 \\ \frac{kx^2}{2} + c &= U(x), \end{aligned}$$

donde c_1 y c son las constantes de integración.

Dado que en el punto de equilibrio $x = 0$ la energía potencial es cero, la constante de integración c de (A.17) debe ser cero, y de este modo la energía potencial se reduce a

$$U(x) = \frac{kx^2}{2}.$$

Considerando a la posición dependiente del tiempo, puede reescribirse a la energía potencial como

$$U(r(t)) = \frac{k(A\text{sen}(wt))^2}{2}. \quad (\text{A.17})$$

Finalmente, calculamos la energía total del sistema como la suma de la energía cinética (A.16) y la energía potencial (A.17),

$$\begin{aligned} E(t) &= T(v(t)) + U(q(t)) \\ &= \frac{m(Aw\cos(wt))^2}{2} + \frac{k(A\text{sen}(wt))^2}{2} \\ &= \frac{1}{2}(mA^2w^2\cos^2(wt) + kA^2\text{sen}^2(wt)), \end{aligned}$$

sustituyendo el valor $w = \sqrt{\frac{k}{m}}$,

$$\begin{aligned} &= \frac{1}{2}kA^2(\cos^2(wt) + \text{sen}^2(wt)) \\ &= \frac{1}{2}kA^2. \end{aligned}$$

Concluyendo, $E(t) = \frac{1}{2}kA^2$, como se puede observar, no depende del tiempo ni de la posición de la partícula por lo que se mantendrá constante para cualquier tiempo t .

Para observar el resultado de la conservación de energía gráficamente, consideremos que la masa m de la partícula son 2 Kg, la amplitud $A=10\text{m}$, y la constante de proporcionalidad del resorte $k=30\text{ N/m}$ (Newton sobre metro). En la imagen A.3 se muestra el valor de la energía cinética, potencial y mecánica en un periodo de tiempo $[0, \frac{2\pi}{w}]$.

El código en Python de este ejercicio se puede consultar en este [repositorio](#).

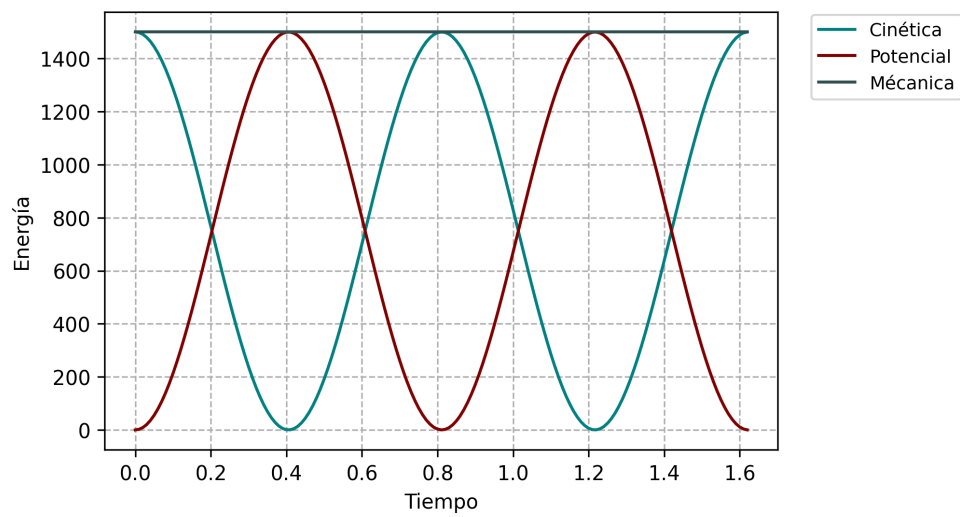


Figura A.3: Energía en al tiempo t , con $t \in [0, \frac{2\pi}{\omega}]$

Bibliografía

- [1] Arnold, V. (1989). *Graduate texts in mathematics* (Vol. 60). Springer New York.
- [2] Baxendale, P. (2011). TE Harris's contributions to recurrent Markov processes and stochastic flows. *The Annals of Probability*, 417-428.
- [3] Beskos, A., Pillai, N., Roberts, G., Sanz-Serna, J.-M., Stuart, A. y col. (2013). Optimal tuning of the hybrid Monte Carlo algorithm. *Bernoulli*, 19(5A), 1501-1534.
- [4] Betancourt, M. (2017). A conceptual introduction to Hamiltonian Monte Carlo. *arXiv preprint arXiv:1701.02434*.
- [5] Boone, E. L., Merrick, J. R. & Krachey, M. J. (2014). A Hellinger distance approach to MCMC diagnostics. *Journal of Statistical Computation and Simulation*, 84(4), 833-849.
- [6] Borzì, A. (2020). *Modelling with Ordinary Differential Equations: A Comprehensive Approach*. CRC Press.
- [7] Braun, M. & Golubitsky, M. (1983). *Differential equations and their applications* (Vol. 1). Springer.
- [8] Brooks, S., Gelman, A., Jones, G. & Meng, X.-L. (2011). *Handbook of markov chain monte carlo*. CRC press.
- [9] Christen, J. A. & Fox, C. (2010). A general purpose sampling algorithm for continuous distributions (the t-walk). *Bayesian Analysis*, 5(2), 263-281.
- [10] del Castillo, G. F. T. (2018). *An Introduction to Hamiltonian Mechanics*. Springer.
- [11] Duane, S., Kennedy, A. D., Pendleton, B. J. & Roweth, D. Hybrid monte carlo. *Physics letters B*, 195(2), 216-222.
- [12] Gamerman, D. & Lopes, H. F. (2006). *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. CRC Press.
- [13] Garrigós, J. E., MAESTRO, M. d. M. S. & López, I. Z. (2012). *Mecánica clásica*. Editorial UNED.
- [14] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.
- [15] Gelman, A., Roberts, G. O., Gilks, W. R. y col. (1996). Efficient Metropolis jumping rules. *Bayesian statistics*, 5(599-608), 42.
- [16] Gilks, R. & Richardson, S. (1996). Spiegelhalter. Markov Chain Monte Carlo in practice.

- [17] Greiner, W. (2009). *Classical mechanics: systems of particles and Hamiltonian dynamics*. Springer Science & Business Media.
- [18] Hairer, E., Lubich, C. & Wanner, G. (2006). *Geometric numerical integration: structure-preserving algorithms for ordinary differential equations* (Vol. 31). Springer Science & Business Media.
- [19] Hamill, P. (2014). *A student's guide to Lagrangians and Hamiltonians*. Cambridge University Press.
- [20] Jarner, S. F. & Hansen, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic processes and their applications*, 85(2), 341-361.
- [21] Liang, F., Liu, C. & Carroll, R. (2011). *Advanced Markov chain Monte Carlo methods: learning from past samples* (Vol. 714). John Wiley & Sons.
- [22] Livingstone, S., Betancourt, M., Byrne, S., Girolami, M. y col. (2019). On the geometric ergodicity of Hamiltonian Monte Carlo. *Bernoulli*, 25(4A), 3109-3138.
- [23] MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge university press.
- [24] Mann, P. (2018). *Lagrangian and Hamiltonian dynamics*. Oxford University Press.
- [25] Marsden, J. E., Tromba, A. J. & Muñiz, P. C. (1991). *Cálculo vectorial* (Vol. 69). Addison-Wesley Iberoamericana.
- [26] Medina-Aguayo, F. J. & Christen, J. A. (2020). Penalised t-walk MCMC. *arXiv preprint arXiv:2012.02293*.
- [27] Mescheder, L., Nowozin, S. & Geiger, A. Adversarial variational bayes: Unifying variational autoencoders and generative adversarial networks. En: *International Conference on Machine Learning*. PMLR. 2017, 2391-2400.
- [28] Neal, R. M. (1996). *Bayesian Learning for Neural Networks*. Springer-Verlag.
- [29] Petersen, K. B. & Pedersen, M. S. (2012). The matrix cookbook, nov 2012. URL <http://www2.imm.dtu.dk/pubdb/p.php,3274>, 14.
- [30] Ponce, V. H. (2010). Mecánica clásica. *Mendoza: EDIUNC*.
- [31] Reyes Flores, M. (2008). *El método de Monte Carlo via cadenas de Markov y su uso en problemas de epidemiología* (Tesis de licenciatura). Universidad Nacional Autónoma de México. CDMX, Mexico.
- [32] Robert, C. & Casella, G. (2013). *Monte Carlo statistical methods*. Springer Science & Business Media.
- [33] Roberts, G. O. & Rosenthal, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probability surveys*, 1, 20-71.
- [34] Roberts, G. O., Rosenthal, J. S. y col. (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical science*, 16(4), 351-367.
- [35] Rosa, A. (2005). *Integradores symplecticos basados en funciones generatrices*. Tesis de licenciatura, UNAM.
- [36] Santana Cibrian, M. (2011). *Gibbs Direccional Optimo: Aproximación Normal*. Tesina de maestría, CIMAT.

-
- [37] Tierney, L. (1994). Markov chains for exploring posterior distributions. *the Annals of Statistics*, 1701-1728.