Pregrado / Ingeniería de Sistemas y Computación
Ingeniería Industrial

Protección de la Privacidad en Datos de Ubicación mediante Privacidad Diferencial Local: Evaluación de Mecanismos de Perturbación.

---

por

Carol Sofia Florido Castro

Asesores:

Valérie Gauthier Umaña & Juan Fernando Perez Bernal

2025

Índice general

Índice de figuras

Índice de tablas

Capítulo 1

Introducción

La proliferación de aplicaciones móviles y servicios basados en ubicación ha transformado radicalmente la manera en que los datos personales son recolectados, procesados y utilizados. Desde que los dispositivos móviles se convirtieron en extensiones inseparables de la vida cotidiana, la recopilación continua de información de ubicación se ha vuelto no solo común, sino esperada. Navegadores GPS, plataformas de movilidad, redes sociales, aplicaciones de citas, servicios de entrega y sistemas de publicidad basada en proximidad dependen del acceso constante a los datos espaciales de los usuarios. Esta dependencia, sin embargo, ha convertido la ubicación en uno de los vectores más vulnerables para la exposición involuntaria de información personal. A diferencia de otros tipos de datos, los registros de ubicación permiten inferir aspectos profundamente sensibles y estructurales de la vida de un individuo: sus hábitos diarios, patrones de desplazamiento, domicilio, lugar de trabajo, prácticas religiosas, estado de salud e incluso sus vínculos sociales o políticos. Estudios recientes han demostrado que basta una pequeña cantidad de puntos geográficos para reidentificar a una persona con alta probabilidad, incluso si los datos han sido anonimizados. Esto ha despertado crecientes preocupaciones en la comunidad académica, en organismos reguladores y entre los propios usuarios, quienes, si bien desean servicios personalizados y contextualmente relevantes, también exigen mayor control sobre su privacidad. Este escenario se enmarca en un contexto más amplio de tensiones entre innovación tecnológica y derechos fundamentales. A medida que se afianzan regulaciones como el Reglamento General de Protección de Datos (GDPR) en Europa, la Ley de Privacidad del Consumidor de California (CCPA) y otros marcos legales emergentes en América Latina y Asia, las organizaciones enfrentan una presión creciente para implementar técnicas que permitan ex-

plotar valor en los datos sin comprometer la privacidad individual. En este punto, la privacidad diferencial ha emergido como uno de los paradigmas más robustos y formalmente garantizados para proteger los datos personales. La privacidad diferencial, al introducir ruido estadísticamente controlado sobre los datos o las respuestas a consultas, permite realizar análisis útiles sin revelar información específica sobre ningún individuo. Su adopción por parte de empresas como Apple, Google y Microsoft ha reforzado su legitimidad técnica y operativa. Sin embargo, la mayoría de los desarrollos iniciales se centraron en bases de datos tabulares o conteos agregados, y solo más recientemente se han explorado con profundidad sus aplicaciones en el dominio espacial-temporal. En particular, proteger trayectorias de usuarios bajo el modelo de privacidad diferencial plantea desafíos técnicos únicos. Las trayectorias presentan una fuerte dependencia espacio-temporal, alta dimensionalidad y complejidad estructural. Además, los usuarios tienen distintos niveles de tolerancia a la perturbación, dependiendo del contexto de uso, el tipo de aplicación y sus propias percepciones de riesgo. Por ello, se requiere evaluar cuidadosamente los mecanismos disponibles, entendiendo cómo afectan no solo la privacidad garantizada, sino también la utilidad práctica de los datos para distintos tipos de servicios. En este proyecto, se propone comparar distintos métodos de privacidad diferencial aplicados a datos de ubicación, con el fin de evaluar cómo afectan las trayectorias y qué tan bien se adaptan a diferentes perfiles de usuario. Al integrar criterios cuantitativos como la distancia esperada, la pérdida de precisión espacial y la preservación de patrones de movilidad, junto con parámetros personalizables de privacidad, esta tesis busca contribuir al diseño de sistemas que protejan los datos de ubicación de forma flexible, robusta y centrada en el usuario. Asimismo, se pretende ofrecer lineamientos que orienten la implementación práctica de mecanismos de privacidad diferencial en sistemas de movilidad, geolocalización y análisis urbano, asegurando un equilibrio entre innovación, utilidad y protección de los derechos individuales.

Capítulo 2

Privacidad Diferencial

La Privacidad Diferencial (DP, por sus siglas en inglés) es un marco matemático diseñado para proteger la información individual dentro de un conjunto de datos, garantizando que el resultado de cualquier análisis estadístico no se vea significativamente afectado por la inclusión o exclusión de un solo individuo. En palabras de Dwork [2008], una función satisface privacidad diferencial si la probabilidad de obtener un determinado resultado es casi la misma independientemente de si un registro individual está o no presente en la base de datos. Formalmente, un mecanismo aleatorizado K satisface ϵ-privacidad diferencial si, para toda pareja de bases de datos D y D' que difieren en una sola fila, y para cualquier subconjunto de posibles salidas S, se cumple que:

$$\Pr[K(D) \in S] \leq e^{\epsilon} \cdot \Pr[K(D') \in S] \tag{2.1}$$

En donde el parámetro ϵ (épsilon) cuantifica el nivel de privacidad, controlando el máximo factor por el cual pueden diferir las probabilidades de obtener un resultado al modificar un único dato en el conjunto. Valores pequeños ($\epsilon < 1$) garantizan mayor privacidad al hacer las salidas casi indistinguibles, mientras que valores mayores ($\epsilon > 5$) permiten mayor precisión pero reducen la protección. Es decir, ϵ determina cuánto puede revelar un dato perturbado sobre la posición real del usuario.

En el contexto de datos de ubicación, este principio puede aplicarse para proteger la privacidad de usuarios que comparten sus coordenadas geográficas con servicios que dependen de información basada en localización, como aplicaciones de navegación, recomendaciones

cercanas o análisis de movilidad urbana.

A partir de esto suponemos que los usuarios exponen sus datos a través de un sistema para obtener algún tipo de servicio (utilidad) basado en su información compartida. Sin embargo, desean proteger ciertos aspectos sensibles de su comportamiento, como los lugares exactos que visitan, las etiquetas semánticas asociadas (por ejemplo, "hospital", ïglesia.º "protesta") o las actividades inferidas a partir de esas ubicaciones.

Para preservar su privacidad, se asume que el usuario ofusca sus coordenadas antes de enviarlas o publicarlas. La Figura 2.1 ilustra el flujo de información que se asume en este trabajo.

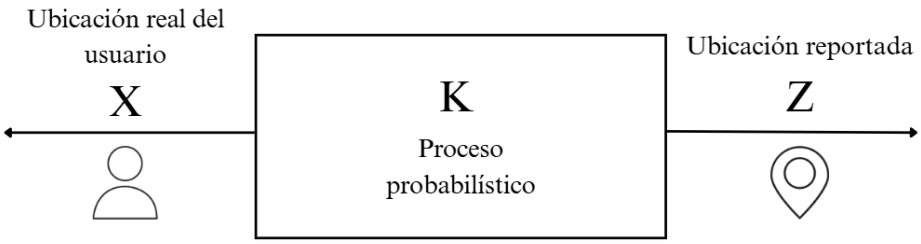

Figura 2.1: Flujo de información del mecanismo de ofuscación aplicado a coordenadas de ubicación.

Sea K una función aleatorizada que transforma coordenadas reales en coordenadas perturbadas por medio de algún mecanismo que satisface ϵ-privacidad diferencial. Más formalmente, definimos el mecanismo de ofuscación como una función probabilística

$$K : X \to \mathcal{P}(Z)$$

donde X es el conjunto de ubicaciones reales (por ejemplo, coordenadas (lat, lon)), y $\mathcal{P}(Z)$ representa un conjunto de distribuciones de probabilidad sobre las ubicaciones reportadas Z.

El mecanismo toma como entrada una ubicación secreta $x \in X$, y devuelve una coordenada reportada $z \in Z$ que es compartida con el proveedor del servicio. Es decir, $K(x) = z$.

Desde la perspectiva del flujo de información cuantitativa, el mecanismo puede modelarse como un canal de comunicación con entrada x y salida z. En general, la salida z puede pertenecer al conjunto potencia de X, incluyendo casos donde el mecanismo generaliza la ubicación, por ejemplo, reduciendo su precisión espacial (pasar de coordenadas exactas a una zona, barrio o ciudad).

Esta estrategia permite al usuario mantener un balance entre privacidad y utilidad del dato compartido, dependiendo del nivel de perturbación introducido por el mecanismo K.

2.1. Mecanismo de Laplace para Perturbación de Coordenadas

El mecanismo de Laplace para protección de privacidad en coordenadas se fundamenta en la adición controlada de ruido siguiendo una distribución de Laplace. Dado un punto original $x \in \mathbb{R}^2$ y un parámetro de privacidad $\epsilon > 0$, el proceso genera un punto perturbado z mediante una transformación probabilística. El mecanismo de perturbación implementado sigue el enfoque propuesto por [3], que aplica ruido Laplace para preservar privacidad en trayectorias.

La densidad de probabilidad para el punto perturbado z viene dada por:

$$D_\epsilon(z|x) = \underbrace{\frac{\epsilon^2}{2\pi}}_{\text{Factor de normalización}} e^{-\epsilon d(x,z)}$$

donde $d(x,z) = \sqrt{(x_1 - z_1)^2 + (x_2 - z_2)^2}$ es la distancia euclidiana entre los puntos x y z. Esta formulación garantiza que puntos más cercanos al original tengan mayor probabilidad de ser seleccionados (debido al término $e^{-\epsilon d(x,z)}$), mientras que los más distantes tienen probabilidad que decae exponencialmente con la distancia.

A partir de lo anterior, para implementar eficientemente este mecanismo en el contexto de datos de ubicación, se debe utilizar una representación en coordenadas polares centradas en x. Un punto z se expresa como (r, θ), donde r es la distancia radial desde x y θ el ángulo polar. La densidad conjunta en estas coordenadas se factoriza como:

$$D_\epsilon(r,\theta) = \underbrace{\frac{1}{2\pi}}_{\text{uniforme en ángulo}} \cdot \underbrace{\epsilon^2 r e^{-\epsilon r}}_{\text{distribución radial}}$$

La generación del punto perturbado aprovecha esta factorización. Primero se selecciona un ángulo θ uniformemente en $[0, 2\pi)$. Para la componente radial, se utiliza la función de distribución acumulativa:

$$C_\epsilon(r) = 1 - (1 + \epsilon r)e^{-\epsilon r}$$

Mediante el método de transformada inversa, se genera un valor uniforme $p \in [0, 1)$ y se

resuelve $C_\epsilon(r) = p$ para obtener el radio:

$$r = -\frac{1}{\epsilon}\left(W_{-1}\left(\frac{p-1}{e}\right) + 1\right)$$

donde W_{-1} es la rama negativa de la función W de Lambert. Finalmente, las coordenadas cartesianas del punto perturbado se calculan como:

$$z = x + (r\cos\theta, r\sin\theta)$$

Este mecanismo preserva la privacidad diferencial geométrica, garantizando que la probabilidad de generar cualquier punto z decae exponencialmente con su distancia al punto original x, controlado por el parámetro ϵ que determina el nivel de ruido añadido.

Ejemplo ilustrativo. Consideremos un punto original $x = (0, 0)$ y un presupuesto de privacidad $\epsilon = 1$. Para generar un punto perturbado, primero se obtiene un ángulo θ muestreado uniformemente en el intervalo $(0, 2\pi)$; en este caso, supongamos que $\theta = \pi/4$ (45°). Luego se genera un valor $p \sim$ Uniforme$(0, 1)$, obteniendo $p = 0{,}5$. A partir de este valor, se resuelve la ecuación $1 - (1 + r)e^{-r} = 0{,}5$, cuyo resultado es $r \approx 1{,}67835$. Finalmente, se calculan las coordenadas cartesianas perturbadas como

$$z = (0 + 1{,}67835\cos(\pi/4),\ 0 + 1{,}67835\sin(\pi/4)) \approx (1{,}1868,\ 1{,}1868).$$

Este resultado muestra cómo el punto original $(0, 0)$ se desplazó a $(1{,}1868,\ 1{,}1868)$ mediante el mecanismo de perturbación, manteniendo la privacidad diferencial.

2.1.1. Implementación del Mecanismo de Laplace

El siguiente pseudocodigo describe la implementación del mecanismo de Laplace, mostrando sus entradas, proceso y salida para perturbar coordenadas preservando privacidad:

Algorithm 1 Mecanismo de Laplace para Perturbación Bidimensional

Require: Punto original $x = (x_1, x_2) \in \mathbb{R}^2$, parámetro de privacidad $\epsilon > 0$
Ensure: Punto perturbado $z = (z_1, z_2)$ que garantiza ϵ-privacidad diferencial
1: Muestrear $\theta \sim \text{Uniforme}(0, 2\pi)$
2: Muestrear $p \sim \text{Uniforme}(0, 1)$
3: Calcular $r \leftarrow -\frac{1}{\epsilon}\left(W_{-1}\left(\frac{p-1}{e}\right)+1\right)$
4: Calcular $z_1 \leftarrow x_1 + r \cdot \cos(\theta)$
5: Calcular $z_2 \leftarrow x_2 + r \cdot \sin(\theta)$
6: **return** $z \leftarrow (z_1, z_2)$

La Figura 2.2 muestra los resultados obtenidos al ejecutar el código de perturbación de Laplace. El proceso completo consta de tres etapas fundamentales:

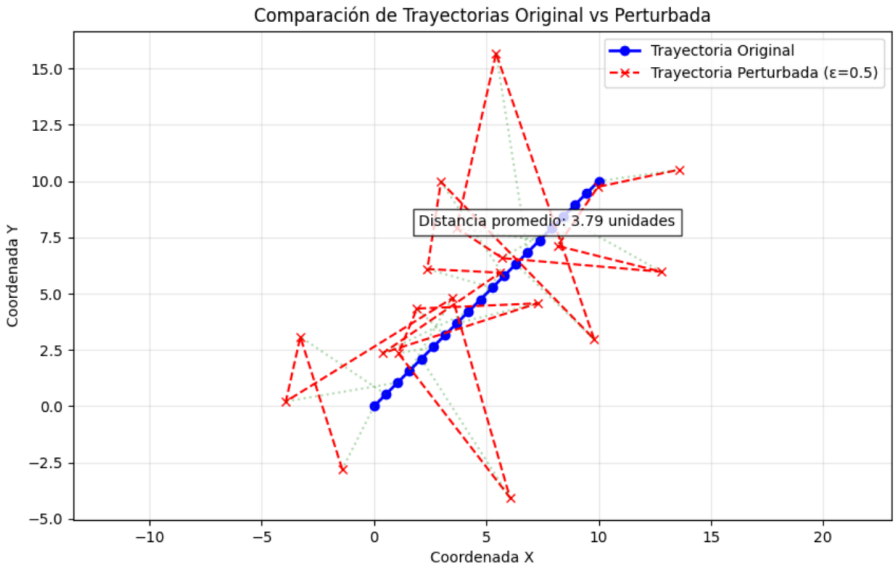

Figura 2.2: Comparación entre trayectoria original y perturbada Laplace

Primero, se generó una trayectoria original compuesta por 20 puntos equidistantes a lo largo de la línea recta que conecta $(0, 0)$ con $(10, 10)$. Segundo, cada uno de estos puntos fue perturbado aplicando el mecanismo de Laplace con parámetro $\epsilon = 0{,}5$, donde para cada punto (x_i, y_i) se calculó un punto perturbado (z_i, w_i) mediante el algoritmo descrito en la Tabla **??**.

La distancia promedio de $3{,}79$ unidades se calculó específicamente mediante:

$$\bar{d} = \frac{1}{20}\sum_{i=1}^{20}\sqrt{(x_i - z_i)^2 + (y_i - w_i)^2}$$

Este valor concreto de $3{,}79$ unidades representa la magnitud típica de perturbación observada en este experimento particular. Cabe destacar que:

- Cada ejecución del código produce resultados ligeramente diferentes debido a la naturaleza aleatoria del mecanismo
- El valor exacto puede variar, pero se mantiene en el mismo orden de magnitud para el mismo ϵ

La gráfica resultante demuestra empíricamente cómo el mecanismo preserva la forma general de la trayectoria (útil para análisis agregados) mientras protege la ubicación exacta de cada punto individual.

2.2. Algoritmo Pivot Sampling para Privacidad en Trayectorias (TP)

El algoritmo *Pivot Sampling* es un mecanismo de *privacidad diferencial local* (LDP) que protege trayectorias geográficas mediante perturbación en dos etapas: perturbación independiente de puntos pivote y perturbación direccional de puntos no pivote. Esta explicación se fundamenta en lo expuesto por Zhang et al. [2023].

2.2.1. Conceptos Fundamentales

Dada una trayectoria original $\tau = p_i \in \mathcal{P} \mid 1 \leq i \leq |\tau|$, se define como una secuencia ordenada de puntos p_i que pertenecen a un dominio espacial $\mathcal{P}$, el cual representa el conjunto de todas las ubicaciones posibles. Cada punto p_i está compuesto por un par de coordenadas (x_i, y_i) que indican su posición en el espacio bidimensional. La longitud de la trayectoria, denotada por $|\tau|$, corresponde al número total de puntos que la conforman.

Primero, el algoritmo crea dos copias idénticas τ' y $\tau*$ de τ. En cada copia se selecciona un subconjunto de puntos seleccionados alternadamente conocidos como **puntos pivote**, (ej. puntos con índice par o impar) que se perturban independientemente usando el Mecanismo Exponencial (EM):

$$Pr[\hat{p} = r] = \frac{\exp\left(\frac{\epsilon_{ind} \cdot u(p,r)}{2\Delta u}\right)}{\sum_{r' \in \mathcal{P}} \exp\left(\frac{\epsilon_{ind} \cdot u(p,r')}{2\Delta u}\right)} \tag{2.2}$$

En el contexto de perturbación de trayectorias, se considera un punto real original p perteneciente a la trayectoria y su correspondiente versión perturbada $\hat{p}$. Ambos puntos forman parte del dominio espacial $\mathcal{P}$, que contiene todas las ubicaciones posibles. Dentro de este dominio, r y r' representan dos puntos candidatos que podrían ser seleccionados como versiones perturbadas del punto original. Para decidir entre estos candidatos, se utiliza una función de utilidad definida como $u(p, r) = -\text{dist}(p, r)$, donde $\text{dist}(p, r)$ representa la distancia Haversine entre los puntos $p = (lat_1, lon_1)$ y $r = (lat_2, lon_2)$. Esta distancia se calcula mediante la fórmula:

$$dist(p,r) = 2R \arcsin\left(\sqrt{\sin^2\left(\frac{\Delta lat}{2}\right) + \cos(lat_1)\cos(lat_2)\sin^2\left(\frac{\Delta lon}{2}\right)}\right) \tag{2.3}$$

donde R es el radio de la Tierra (aproximadamente 6,371 km), $\Delta\text{lat} = \text{lat}_2 - \text{lat}_1$ y $\Delta\text{lon} = \text{lon}_2 - \text{lon}_1$, expresadas en radianes. El signo negativo en la función de utilidad permite que el mecanismo exponencial favorezca la selección de puntos cercanos, ya que estos presentan mayor utilidad (menor distancia).

La sensibilidad de esta función 2.2, denotada como Δu, se define como el máximo cambio posible en la utilidad al comparar dos candidatos diferentes respecto a un mismo punto, es decir, $\Delta u = \max_{p,r,r'} |u(p,r) - u(p,r')|$. Finalmente, el parámetro ϵ_{ind} representa el presupuesto de privacidad asociado a la perturbación de cada punto de manera independiente, lo que determina el grado de protección garantizado frente a la inferencia adversaria.

Para seleccionar los puntos de pivote que servirán de base para la perturbación, el algoritmo implementa un mecanismo de intercambio estratégico que reparte los puntos de la trayectoria original entre las dos copias: τ' y $\tau*$. De manera inicial, los puntos pares de la trayectoria se asignan como pivotes en τ' (es decir, $\mathcal{P}ind^{\tau'} = p_2, p_4, \ldots$), mientras que los impares se utilizan como pivotes en $\tau*$ (es decir, $\mathcal{P}ind^{\tau*} = p_1, p_3, \ldots$). Esta asignación se complementa con la relación $\mathcal{P}ind^{\tau*} = \mathcal{P} - \mathcal{P}ind^{\tau'}$, que asegura una cobertura total y no redundante del conjunto de puntos.

Durante el proceso de perturbación, en la trayectoria τ' los puntos pares son perturbados de forma independiente mediante el Mecanismo Exponencial, mientras que los puntos impares aprovechan la información direccional proveniente de sus vecinos ya perturbados. En la trayectoria $\tau*$ ocurre el proceso inverso: los puntos impares se perturban de forma independiente y los pares utilizan la información de sus vecinos impares para guiar su perturbación. Este intercambio de roles promueve un equilibrio entre privacidad y utilidad, ya que evita que toda la trayectoria dependa únicamente de perturbaciones independientes o correlacionadas.

Ejemplo ilustrativo.

Para una trayectoria $\tau = \{p_1, p_2, p_3, p_4, p_5\}$:

Tabla 2.1: Asignación de puntos en el proceso de intercambio

Copia	Puntos Pivote	Puntos No Pivote
τ'	p_2, p_4	p_1, p_3, p_5
τ''	p_1, p_3, p_5	p_2, p_4

2.2.2. Perturbación Direccional

Para los puntos no pivote, el algoritmo introduce ruido en la dirección relativa respecto a los puntos pivote vecinos, con el fin de preservar la privacidad manteniendo cierta coherencia espacial. Para ello, el espacio angular se divide en un número finito de sectores iguales, cada uno representando una dirección posible. La dirección real de un punto se asigna inicialmente a uno de estos sectores discretos según su ángulo geométrico. Esto por medio de la siguiente formula:

$$\mathcal{D} = \{0, 1, \ldots, |\mathcal{D}| - 1\}, \text{ donde cada } d \in \mathcal{D} \text{ cubre } \left[(2d-1)\frac{\pi}{|\mathcal{D}|}, (2d+1)\frac{\pi}{|\mathcal{D}|}\right] \tag{2.4}$$

Luego, se aplica el mecanismo de respuesta aleatorizada k-RR, que introduce incertidumbre en la dirección reportada: con alta probabilidad (controlada por el parámetro de privacidad ϵ), se selecciona el sector verdadero, pero con una pequeña probabilidad se elige otro sector al azar. Esto asegura que el valor final comunicado no revele exactamente la dirección original, dificultando la identificación precisa del punto, pero sin perder completamente la información direccional, lo que ayuda a preservar la utilidad de la trayectoria perturbada.

$$Pr[\mathcal{K}(d) = d'] = \begin{cases} \frac{e^\epsilon}{|\mathcal{D}|-1+e^\epsilon} & \text{si } d = d' \text{ (dirección correcta)} \\ \frac{1}{|\mathcal{D}|-1+e^\epsilon} & \text{si } d \neq d' \text{ (dirección incorrecta)} \end{cases} \tag{2.5}$$

donde:

- $\mathcal{K}$: Mecanismo de perturbación
- ϵ: Presupuesto de privacidad para direcciones
- d: Dirección original (índice del sector)
- d': Dirección reportada

Ejemplo ilustrativo. Supongamos que el espacio angular se divide en $|\mathcal{D}| = 8$ sectores. Cada sector cubre un ángulo de $\frac{2\pi}{8} = \frac{\pi}{4}$ radianes (45 grados). Si un punto tiene una dirección relativa de $\theta = \frac{5\pi}{8}$, entonces este ángulo cae dentro del sector:

$$\left[(2d-1)\frac{\pi}{|\mathcal{D}|},(2d+1)\frac{\pi}{|\mathcal{D}|}\right]=\left[(2d-1)\frac{\pi}{8},(2d+1)\frac{\pi}{8}\right]$$

Buscando el valor de d que satisface esta desigualdad para $\theta=\frac{5\pi}{8}$, obtenemos $d=3$, pues:

$$\left[(2\cdot 3-1)\frac{\pi}{8},(2\cdot 3+1)\frac{\pi}{8}\right]=\left[\frac{5\pi}{8},\frac{7\pi}{8}\right]$$

Por tanto, la dirección original se asigna al sector $d=3$. Luego, usando el mecanismo k-RR con $\epsilon=1{,}0$, se calcula la probabilidad de reportar la dirección verdadera como:

$$Pr[\mathcal{K}(3)=3]=\frac{e^1}{7+e^1}\approx\frac{2{,}718}{7+2{,}718}\approx 0{,}28$$

y la probabilidad de reportar cualquiera de los sectores incorrectos ($d'\neq 3$) como:

$$Pr[\mathcal{K}(3)=d']=\frac{1}{7+e^1}\approx\frac{1}{9{,}718}\approx 0{,}103$$

Este ejemplo muestra cómo se introduce aleatoriedad en la dirección sin eliminar completamente la información espacial.

2.2.3. Determinación del Dominio de Perturbación

Para cada punto no pivote p_j, el dominio de perturbación $\mathcal{P}^j$ se define como el conjunto de posibles ubicaciones en el espacio que están restringidas a un rango angular específico determinado por las direcciones perturbadas provenientes de sus puntos pivote vecinos. Este dominio se calcula utilizando un procedimiento que, dado un pivote perturbado $\hat{p}_k$ y una dirección angular perturbada $\hat{d}_{k,j}$, obtiene todos los puntos $p\in\mathcal{P}$ cuya dirección relativa con respecto a $\hat{p}_k$ cae dentro del sector angular delimitado por los ángulos $\theta_{\text{mín}}$ y $\theta_{\text{máx}}$. Estos límites se determinan según el índice del sector angular $\hat{d}_{k,j}$ y la cantidad total de sectores $|\mathcal{D}|$, de modo que

$$\theta_{\text{mín}}=(2\hat{d}_{k,j}-1)\frac{\pi}{|\mathcal{D}|},\quad \theta_{\text{máx}}=(2\hat{d}_{k,j}+1)\frac{\pi}{|\mathcal{D}|}.$$

El dominio de perturbación final para el punto p_j resulta de la intersección de estos conjuntos angulares obtenidos desde los pivotes vecinos $\hat{p}_{j-1}$ y $\hat{p}_{j+1}$, es decir,

$$\mathcal{P}^j = \text{GetPointSet}(\hat{p}_{j-1}, \hat{d}_{j-1,j}) \cap \text{GetPointSet}(\hat{p}_{j+1}, \hat{d}_{j+1,j}).$$

Esta intersección restringe el espacio posible del punto no pivote, preservando la coherencia direccional y limitando la perturbación para mantener la utilidad de la trayectoria.

Ejemplo ilustrativo. Supongamos que el espacio $\mathcal{P}$ está discretizado en una cuadrícula y que existen 8 sectores direccionales (es decir, $|\mathcal{D}| = 8$), cada uno de $\frac{\pi}{8}$ radianes. Para el punto no pivote p_j, su vecino anterior $\hat{p}_{j-1}$ le asigna una dirección perturbada $\hat{d}_{j-1,j} = 2$, mientras que su vecino posterior $\hat{p}_{j+1}$ le asigna $\hat{d}_{j+1,j} = 3$. Entonces:

$$\text{Desde } \hat{p}_{j-1} : \theta^{(1)}_{\text{mín}} = \frac{3\pi}{8}, \quad \theta^{(1)}_{\text{máx}} = \frac{5\pi}{8}$$

$$\text{Desde } \hat{p}_{j+1} : \theta^{(2)}_{\text{mín}} = \frac{5\pi}{8}, \quad \theta^{(2)}_{\text{máx}} = \frac{7\pi}{8}$$

Por tanto, el dominio angular permitido para p_j está contenido entre $\frac{5\pi}{8}$ y $\frac{5\pi}{8}$, es decir, una intersección mínima (un solo borde compartido), lo que restringe significativamente su posible ubicación. Este tipo de intersección garantiza que la dirección entre p_j y sus pivotes vecinos no varíe de forma abrupta, manteniendo la consistencia espacial de la trayectoria original.

2.2.4. Selección de Granularidad Óptima

La selección de la granularidad óptima, denotada como $|\mathcal{D}|$, resulta fundamental para lograr un balance adecuado entre la precisión y la privacidad en la perturbación direccional. Este proceso se formaliza como la búsqueda de la granularidad g dentro de un conjunto de candidatos $\mathcal{D}_c$ que maximice la función objetivo definida como el promedio ponderado de la cobertura geométrica y la probabilidad de acierto sobre distintos rangos angulares de consulta. Matemáticamente, esto se expresa como:

$$|\mathcal{D}|_{opt} = \arg\max_{g \in \mathcal{D}_c} \frac{1}{|\Theta|} \sum_{\theta_j \in \Theta} \sum_{d_i \in \mathcal{R}(g)} \underbrace{\varphi(d_i; \theta_j)}_{\text{Cobertura geométrica}} \cdot \underbrace{\lambda(d_i; \epsilon; g)}_{\text{Probabilidad de acierto}} .$$

Los conjuntos de referencia involucrados en este cálculo son: el conjunto de granularidades candidatas $\mathcal{D}_c = \{2, 4, 6, 8, 12\}$, que representan los posibles números de sectores angulares; el conjunto de índices de dirección para una granularidad dada, $\mathcal{R}(g) = \{0, 1, \ldots, g-1\}$; y el

conjunto de rangos angulares de consulta $\Theta = \{\pi/2, \pi/4, \pi/6, \pi/12\}$, que definen diferentes amplitudes angulares de interés.

La función de cobertura geométrica $\varphi(d_i; \theta_j)$ cuantifica qué fracción del sector angular d_i está contenida dentro del rango de consulta θ_j, calculándose como la razón entre el área de intersección del sector con el rango de consulta y el área total del sector:

$$\varphi(d_i; \theta_j) = \frac{|[(2d_i - 1)\frac{\pi}{g}, (2d_i + 1)\frac{\pi}{g}] \cap [-\theta_j, \theta_j]|}{2\pi/g}.$$

Por ejemplo, para una granularidad $g = 4$ y un rango de consulta $\theta_j = \pi/4$, la cobertura geométrica para el sector $d_i = 0$ es $\varphi(0; \pi/4) = 0{,}5$, indicando que solo la mitad del sector está dentro del rango de interés.

La función $\lambda(d_i; \epsilon; g)$ modela la probabilidad de preservación o acierto del mecanismo de perturbación k-RR, que depende críticamente del presupuesto de privacidad ϵ. Esta función asigna mayor probabilidad a la dirección correcta y menor a las incorrectas según:

$$\lambda(d_i; \epsilon; g) = \begin{cases} \frac{e^\epsilon}{g-1+e^\epsilon} & \text{si } d_i = d \text{ (dirección correcta)} \\ \frac{1}{g-1+e^\epsilon} & \text{si } d_i \neq d \text{ (dirección incorrecta).} \end{cases}$$

Así, un mayor valor de ϵ incrementa la probabilidad de seleccionar la dirección correcta, mejorando la precisión del mecanismo.

En cuanto a los criterios de diseño, una granularidad gruesa, con un valor pequeño de g, ofrece una mayor probabilidad de acierto, pero genera dominios de perturbación más amplios, lo que puede afectar la precisión espacial. Por otro lado, una granularidad fina, con un valor grande de g, permite dominios de perturbación más precisos, aunque reduce la probabilidad de preservar la dirección exacta debido a la mayor dispersión en la selección de sectores.

2.2.5. Composición Final y Análisis de Privacidad

El proceso de composición final busca optimizar la trayectoria perturbada combinando las dos versiones obtenidas mediante el mecanismo de pivotes pares e impares. Formalmente, se define una trayectoria $\hat{\tau}$ que minimiza la suma de las distancias Haversine entre cada punto candidato $\hat{p}_i$ y sus correspondientes puntos perturbados en ambas trayectorias $\hat{\tau}'$ y $\hat{\tau}*$:

$$\hat{\tau} = \arg\min_{\hat{\tau}} \sum_{i=1}^{|\tau|} \left(d_{\text{Hav}}(\hat{p}_i, \hat{p}'_i) + d_{\text{Hav}}(\hat{p}_i, \hat{p}_i*) \right).$$

Aquí, $\hat{\tau}'$ representa la trayectoria perturbada donde los puntos pares funcionan como pivotes, mientras que $\hat{\tau}*$ corresponde a la trayectoria con pivotes en posiciones impares. El objetivo es balancear la influencia de ambas perturbaciones para obtener una representación final que reduzca el error acumulado en comparación con las trayectorias individuales.

El proceso de optimización se realiza para cada punto i de la trayectoria original, generando un conjunto de candidatos $\mathcal{P}^i$ dentro del dominio restringido definido por la intersección de dominios vecinos. Para cada candidato se calcula la suma de las distancias Haversine respecto a los puntos perturbados correspondientes en $\hat{\tau}'$ y $\hat{\tau}*$, y se selecciona aquel que minimiza esta suma, garantizando así una mejor precisión espacial en la trayectoria final.

A continuación, se presenta un pseudocódigo que ilustra este procedimiento de combinación óptima:

Algorithm 2 Combinación Óptima de Trayectorias

Require: Trayectorias perturbadas $\hat{\tau}'$, $\hat{\tau}*$, presupuesto de privacidad ϵ, dominio espacial $\mathcal{P}$
Ensure: Trayectoria final optimizada $\hat{\tau}$

```
Inicializar τ̂ ← ∅
for cada punto p_i ∈ τ do
    Obtener dominio restringido P^i ← GetPerturbationDomain(p_i, τ̂', τ̂'')
    Seleccionar p̂_i = arg min_{r∈P^i} (d_Hav(r, p̂'_i) + d_Hav(r, p̂''_i))
    Actualizar trayectoria final τ̂ ← τ̂ ∪ {p̂_i}
end for
return τ̂
```

2.2.6. Implementación y Visualización del Algoritmo

El algoritmo de **Pivot Sampling** fue implementado en Python para evaluar su desempeño en un entorno controlado. La trayectoria original corresponde a una línea recta que va desde el punto $(0, 0)$ hasta $(10, 10)$, compuesta por 20 puntos equidistantes. El dominio espacial se definió como una rejilla discreta de 50×50 puntos, distribuidos uniformemente dentro del rango bidimensional $[0, 10] \times [0, 10]$. En cuanto al presupuesto de privacidad, se consideró un valor total de $\epsilon = 1{,}0$. La Figura 2.3 muestra una comparación visual entre la trayectoria original y la trayectoria resultante tras la aplicación del mecanismo de privacidad. Como puede observarse, el algoritmo logra preservar la estructura global de la trayectoria mientras introduce ruido de

manera controlada. Las líneas verdes conectan cada punto original con su versión perturbada, ilustrando el desplazamiento generado. Para cuantificar la magnitud del ruido introducido, se calcula la **distancia promedio de perturbación**, la cual alcanza un valor de 5.07 unidades y se define mediante la expresión:

$$\bar{d} = \frac{1}{|\tau|} \sum_{i=1}^{|\tau|} \sqrt{(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2}$$

donde (x_i, y_i) representa cada punto de la trayectoria original y $(\hat{x}_i, \hat{y}_i)$ su correspondiente versión perturbada.

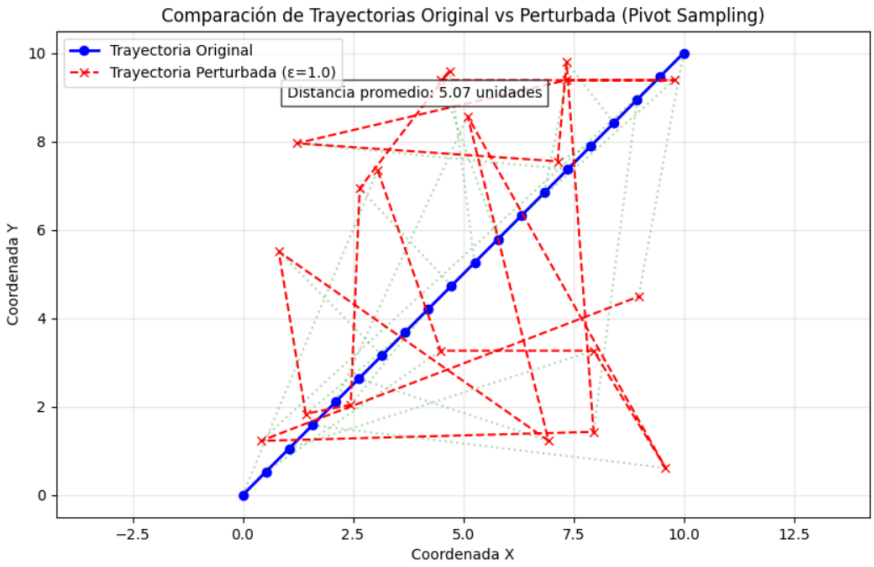

Figura 2.3: Comparación entre la trayectoria original y la perturbada Pivot Sampling.

2.3. Anchor-Based Pivot Sampling (ATP)

El mecanismo **Anchor-Based Pivot Sampling** (ATP) mejora las técnicas tradicionales de muestreo con pivotes al incorporar anclajes espaciales y dominios de perturbación adaptativos. Esto permite reducir significativamente el ruido introducido en trayectorias geográficas al aplicar mecanismos de privacidad diferencial, especialmente cuando el presupuesto de privacidad ϵ es bajo. Esta explicación se fundamenta en lo expuesto por Zhang et al. [2023].

Según la *Primera Ley de la Geografía de Tobler*, "todo está relacionado con todo lo demás, pero las cosas más cercanas están más relacionadas entre sí". En el contexto de trayectorias humanas, esto implica que los puntos de una trayectoria tienden a agruparse espacialmente en regiones compactas. Sin embargo, los mecanismos tradicionales, como el *Pivot Sampling* (TP), generan perturbaciones sobre dominios muy amplios $\mathcal{P}$, lo cual conduce a una alta probabilidad de seleccionar puntos ruidos alejados del recorrido original, particularmente cuando el presupuesto de privacidad ϵ es pequeño.

Para mitigar este problema, ATP restringe adaptativamente el dominio de perturbación a una región circular centrada en un *pivote ancla*, calculado para cada trayectoria individual. Este enfoque permite preservar la utilidad geográfica de los datos al tiempo que se mantiene el cumplimiento del modelo de privacidad diferencial.

2.3.1. Cálculo del Anclaje

Sea $\tau = \{p_1, p_2, \ldots, p_n\}$ una trayectoria compuesta por coordenadas geográficas. Se define un *anclaje* α como el promedio de todas las posiciones:

$$\alpha = \left(\frac{\sum_{i=1}^{n} \text{lat}(p_i)}{n}, \ \frac{\sum_{i=1}^{n} \text{lon}(p_i)}{n} \right) \tag{2.6}$$

Este punto α no necesariamente pertenece al conjunto discreto de ubicaciones posibles $\mathcal{P}$. Por ello, se selecciona como *pivote ancla* p_α el punto más cercano a α en $\mathcal{P}$, empleando como métrica la distancia geodésica (por ejemplo, la distancia de Haversine):

$$p_\alpha = \arg\min_{p \in \mathcal{P}} \text{dist}(\alpha, p) \tag{2.7}$$

El pivote ancla p_α se perturba para garantizar privacidad diferencial mediante el **Mecanis-**

mo Exponencial (EM). Dicho mecanismo selecciona una ubicación $\hat{p}_\alpha \in \mathcal{P}$ con probabilidad:

$$\Pr[\hat{p}_\alpha = p] = \frac{\exp\left(\frac{\epsilon \cdot u(p_\alpha, p)}{2\Delta u}\right)}{\sum_{p' \in \mathcal{P}} \exp\left(\frac{\epsilon \cdot u(p_\alpha, p')}{2\Delta u}\right)} \tag{2.8}$$

donde la función de utilidad favorece puntos cercanos:

$$u(p_\alpha, p) = -\mathsf{dist}(p_\alpha, p) \tag{2.9}$$

y la sensibilidad Δu se calcula como:

$$\Delta u = \max_{p, p' \in \mathcal{P}} \left|\mathsf{dist}(p_\alpha, p') - \mathsf{dist}(p_\alpha, p)\right| \tag{2.10}$$

Un ϵ alto concentra la probabilidad en puntos próximos a p_α, mientras que un ϵ bajo produce una distribución más uniforme.

Ejemplo ilustrativo. Considérese una trayectoria $\tau = \{p_1 = (0,0),\ p_2 = (2,1),\ p_3 = (1,3)\}$. El anclaje es:

$$\alpha = \left(\frac{0+2+1}{3},\ \frac{0+1+3}{3}\right) = (1,\ 1{,}33)$$

Si $\mathcal{P} = \{q_1 = (0,0),\ q_2 = (1,1),\ q_3 = (2,2)\}$, las distancias euclidianas a α son:

$$\mathsf{dist}(\alpha, q_1) \approx 1{,}67, \quad \mathsf{dist}(\alpha, q_2) \approx 0{,}33, \quad \mathsf{dist}(\alpha, q_3) \approx 0{,}75$$

El punto más cercano es $q_2 = (1,1)$, elegido como p_α. Aplicando el Mecanismo Exponencial, la probabilidad de cada q_i dependerá de su cercanía a p_α y del valor de ϵ.

2.3.2. Determinación del Radio Adaptativo

Una vez obtenido el anclaje perturbado $\hat{p}_\alpha$, el siguiente paso es definir una región circular que restrinja el dominio de perturbación de los puntos de la trayectoria. Esta región está caracterizada por un *radio adaptativo* $\hat{R}$, que debe cubrir de forma razonable los puntos originales sin ser excesivamente amplio. El mecanismo ATP sigue tres etapas: 1) cálculo del radio base, 2) perturbación del radio mediante un mecanismo para datos acotados, y 3) calibración final

para evitar valores extremos.

1) Cálculo del radio base. Sea τ la trayectoria original. Se define el radio máximo necesario para cubrir todos los puntos respecto al anclaje perturbado:

$$R^{\tau}_{\mathsf{max}} = \max_{p \in \tau} \mathsf{dist}(\hat{p}_{\alpha}, p) \tag{2.11}$$

2) Perturbación con SW. Para evitar filtraciones de información, se perturba R^{τ}_{max} usando el **Staircase Mechanism for Bounded Data** (SW), adecuado para valores en el rango $[0, \Delta R]$, donde:

$$\Delta R = \max_{p \in \mathcal{P}} \mathsf{dist}(\hat{p}_{\alpha}, p) \tag{2.12}$$

Se normaliza el valor original:

$$r = \frac{R^{\tau}_{\mathsf{max}}}{\Delta R}$$

Se aplica el mecanismo SW para obtener un valor perturbado $\hat{r}$, y luego se reescala:

$$\hat{R}^{\tau}_{\mathsf{max}} = \frac{(\hat{r} + b)\,\Delta R}{2b + 1}. \tag{2.13}$$

3) Calibración final. El valor perturbado puede ser demasiado pequeño (limitando en exceso la región) o demasiado grande (introduciendo demasiado ruido). Para suavizar estas variaciones, se introduce una calibración que ajusta $\hat{R}^{\tau}_{\mathsf{max}}$ hacia un centro de calibración η, con intensidad decreciente al aumentar ϵ:

$$\hat{R} = \hat{R}^{\tau}_{\mathsf{max}} + \xi\, e^{-\epsilon} \tag{2.14}$$

donde el factor de ajuste no lineal es:

$$\xi = (\eta - \hat{R}^{\tau}_{\mathsf{max}}) \cdot \frac{1}{1 - e^{-\beta/2}} \tag{2.15}$$

y

$$\beta = \begin{cases} \frac{\eta - \hat{R}^{\tau}_{\mathsf{max}}}{\eta}, & \text{si } \hat{R}^{\tau}_{\mathsf{max}} \leq \eta, \\ \frac{\hat{R}^{\tau}_{\mathsf{max}} - \eta}{\Delta R - \eta}, & \text{si } \hat{R}^{\tau}_{\mathsf{max}} > \eta \end{cases} \tag{2.16}$$

Así, para valores bajos de ϵ (mayor aleatoriedad), el radio se aproxima a η, mientras que para valores altos de ϵ el efecto de calibración se atenúa, preservando mayor fidelidad a los datos reales. La forma de obtener η se detalla en la siguiente sección.

Ejemplo ilustrativo. Supóngase que la trayectoria original es:

$$\tau = \{(0,\,0),\ (2,\,1),\ (1,\,3)\}$$

y que el anclaje perturbado obtenido es $\hat{p}_\alpha = (1,\,1)$.

Las distancias euclidianas desde $\hat{p}_\alpha$ a cada punto de τ son:

$$\mathsf{dist}((1,1),(0,0)) \approx 1{,}41, \quad \mathsf{dist}((1,1),(2,1)) = 1{,}00, \quad \mathsf{dist}((1,1),(1,3)) = 2{,}00$$

Por lo tanto:

$$R^{\tau}_{\mathsf{max}} = 2{,}00$$

Si el conjunto de ubicaciones posibles es:

$$\mathcal{P} = \{(0,0),\ (1,1),\ (2,2),\ (3,3)\}$$

entonces:

$$\Delta R = \max_{p \in \mathcal{P}} \mathsf{dist}((1,1),p) \approx 2{,}83$$

y el valor normalizado es:

$$r = \frac{2{,}00}{2{,}83} \approx 0{,}707$$

Aplicando el mecanismo SW (con parámetros $b = 0{,}2$ y un ϵ dado) obtenemos, por ejemplo, $\hat{r} = 0{,}75$. El mapeo al dominio real usando la fórmula definida es:

$$\hat{R}^{\tau}_{\mathsf{max}} = \frac{(\hat{r} + 0{,}2) \cdot 2{,}83}{2 \cdot 0{,}2 + 1} \approx \frac{0{,}95 \cdot 2{,}83}{1{,}4} \approx 1{,}92$$

Si el centro de calibración es $\eta = 2{,}00$ y $\epsilon = 0{,}5$, el factor de ajuste no lineal ξ se calcula

según:

$$\beta = \frac{\eta - 1{,}92}{\eta} = 0{,}04, \quad \xi = (2{,}00 - 1{,}92) \cdot \frac{1}{1 - e^{-0{,}04/2}} \approx 0{,}08 \cdot \frac{1}{1 - e^{-0{,}02}} \approx 4{,}00$$

El radio calibrado final es:

$$\hat{R} \approx 1{,}92 + 4{,}00 \cdot e^{-0{,}5} \approx 1{,}92 + 2{,}43 \approx 4{,}35$$

Este valor final $\hat{R}$ se usará para restringir el dominio de perturbación de todos los puntos de la trayectoria.

2.3.3. Dominio de Perturbación Restringido

Con el radio adaptativo calibrado $\hat{R}$ obtenido en la etapa anterior, se define el **dominio de perturbación restringido** como el subconjunto de ubicaciones en $\mathcal{P}$ cuya distancia geodésica al anclaje perturbado $\hat{p}_\alpha$ no supera dicho radio. Este dominio actúa como espacio de búsqueda para aplicar perturbaciones diferenciales a cada punto de la trayectoria, evitando que el ruido se disperse en zonas geográficas irrelevantes.

Formalmente, el dominio se expresa como:

$$\hat{\mathcal{P}}_\alpha = \{p \in \mathcal{P} \mid \mathsf{dist}(\hat{p}_\alpha, p) \leq R\} \tag{2.17}$$

La restricción espacial que impone $\hat{\mathcal{P}}_\alpha$ ofrece dos ventajas clave. En primer lugar, mejora la utilidad de los datos perturbados al mantenerlos próximos a la trayectoria original, lo que reduce el error espacial introducido. En segundo lugar, al limitar el espacio de probabilidad sobre el que opera el mecanismo de perturbación, se reduce la cantidad de ruido necesario para alcanzar un mismo nivel de privacidad diferencial. De este modo, el diseño adaptativo y localizado optimiza el equilibrio entre privacidad y precisión. Con esto se obtiene el conjunto de puntos perturbados $\hat{\mathcal{P}}_\alpha$ de la trayectoria τ basado en $\hat{p}_\alpha$ y $\hat{R}$.

Ejemplo ilustrativo. Supóngase que las coordenadas en $\mathcal{P}$ se expresan en kilómetros dentro de un plano 2D y que el anclaje perturbado obtenido es:

$$\hat{p}_\alpha = (5{,}0,\ 5{,}0)$$

Si el radio adaptativo calibrado es $\hat{R} = 3{,}0$ km, el dominio restringido incluirá únicamente los puntos cuya distancia euclidiana a $(5{,}0, 5{,}0)$ sea menor o igual a $3{,}0$ km. Sea:

$$\mathcal{P} = \{(4,4),\ (7,7),\ (8,5),\ (5,2)\}$$

Calculando las distancias:

$$\text{dist}((5,5),(4,4)) \approx 1{,}41 \text{ km}$$

$$\text{dist}((5,5),(7,7)) \approx 2{,}83 \text{ km}$$

$$\text{dist}((5,5),(8,5)) = 3{,}00 \text{ km}$$

$$\text{dist}((5,5),(5,2)) = 3{,}00 \text{ km}$$

Dado que todas estas distancias son menores o iguales a $3{,}0$ km, se cumple que:

$$\hat{\mathcal{P}}_\alpha = \mathcal{P}$$

Si en cambio el radio hubiese sido $\hat{R} = 2{,}5$ km, únicamente los puntos $(4,4)$ y $(7,7)$ habrían permanecido dentro del dominio restringido.

2.3.4. Calibración del Radio

La calibración del radio busca evitar que el valor perturbado $\hat{R}^{\tau}_{\text{max}}$ resulte excesivamente pequeño o grande, ya que cualquiera de estos extremos puede degradar la utilidad del dominio de perturbación. Para lograrlo, se introduce un *centro de calibración* η que actúa como valor de referencia hacia el cual se desplaza suavemente el radio perturbado, especialmente cuando el presupuesto de privacidad ϵ es bajo y la aleatoriedad del mecanismo es más pronunciada.

El cálculo de η se basa en una estimación empírica del comportamiento del mecanismo SW. En primer lugar, se define un conjunto de valores de prueba uniformemente espaciados en el intervalo $[0, 1]$:

$$\mathcal{V} = \{0{,}0, 1{,}0{,}2, \ldots, 1{,}0\} \tag{2.18}$$

De este conjunto, se identifica un subconjunto $\mathcal{V}^R \subseteq \mathcal{V}$ compuesto por los valores que se consideran *válidos*, es decir, aquellos cuya salida bajo el mecanismo SW tiene una alta probabilidad de haber producido el valor normalizado:

$$r = \frac{\hat{R}^{\tau}_{\mathsf{max}}}{\Delta R}$$

El criterio de validez se define usando un umbral de probabilidad q:

$$q = \frac{e^{\epsilon}}{2be^{\epsilon} + 1} \tag{2.19}$$

donde b y c son parámetros del mecanismo SW que controlan la forma de su distribución y su concentración.

Una vez identificados los valores válidos en $\mathcal{V}^R$, se selecciona el subconjunto de puntos $\mathcal{P}^R \subset \mathcal{P}$ cuya distancia a $\hat{p}_\alpha$ se encuentra en el rango asociado a dichos valores. El resto de los puntos, fuera de este rango, se denota como $\mathcal{P} \setminus \mathcal{P}^R$.

El centro de calibración η se calcula como el promedio ponderado de las distancias, asignando peso q a los puntos de $\mathcal{P}^R$ y peso $(1-q)$ a los demás:

$$\eta = \frac{\sum\limits_{p \in \mathcal{P}^R} q \cdot \mathsf{dist}(\hat{p}_\alpha, p) + \sum\limits_{p' \in \mathcal{P} - \mathcal{P}^R} (1-q) \cdot \mathsf{dist}(\hat{p}_\alpha, p')}{q|\mathcal{P}^R| + (1-q)|\mathcal{P} - \mathcal{P}^R|} \tag{2.20}$$

Este valor η sirve como referencia contextual basada en la distribución empírica de distancias en el espacio geográfico, permitiendo suavizar el efecto del ruido cuando se trabaja con presupuestos de privacidad bajos.

Ejemplo ilustrativo. Supóngase que el anclaje perturbado es:

$$\hat{p}_\alpha = (5{,}0, 5{,}0)$$

y que el conjunto de ubicaciones es:

$$\mathcal{P} = \{(4,4),\ (7,7),\ (8,5),\ (5,2),\ (9,9)\}$$

El radio máximo perturbado es $\hat{R}^{\tau}_{\text{max}} = 3{,}0$ km y el máximo posible $\Delta R = 6{,}0$ km, por lo que:

$$r = \frac{3{,}0}{6{,}0} = 0{,}5$$

Supongamos que, para un $\epsilon = 1{,}0$ y un parámetro $b = 0{,}5$, el umbral de probabilidad según la ecuación (2.19) resulta:

$$q \approx 0{,}42$$

Si la salida del mecanismo SW para $r = 0{,}5$ indica que los valores válidos $\mathcal{V}^R$ son $\{0{,}4, 0{,}5, 0{,}6\}$, entonces $\mathcal{P}^R$ contendrá los puntos cuya distancia a $(5{,}0, 5{,}0)$ esté entre $0{,}4\Delta R = 2{,}4$ km y $0{,}6\Delta R = 3{,}6$ km. En este caso, dichos puntos podrían ser $(7, 7)$ y $(8, 5)$. Los demás puntos pertenecerán a $\mathcal{P} \setminus \mathcal{P}^R$.

Finalmente, aplicando la ecuación de η, se obtiene un valor intermedio entre las distancias de ambos subconjuntos, ponderado por los pesos q y $(1 - q)$. Este valor servirá para ajustar $\hat{R}^{\tau}_{\text{max}}$ en la etapa de calibración final, evitando radios extremos y mejorando la estabilidad del mecanismo.

2.3.5. Representación Gráfica de la Perturbación

Para ilustrar visualmente el funcionamiento del algoritmo de perturbación, se implementó el mecanismo ATP (Anchor-based Trajectory Perturbation) sobre una trayectoria sintética en un espacio discretizado.

Se definió una grilla de tamaño 21×21, excluyendo tres ubicaciones públicas en $(3, 3)$, $(8, 8)$ y $(15, 15)$. La trayectoria original consistió en 20 puntos alineados sobre la diagonal principal, es decir, desde $(1, 1)$ hasta $(20, 20)$. Utilizando un presupuesto de privacidad $\epsilon = 1{,}0$, el mecanismo ATP perturbó cada punto de la trayectoria original, seleccionando ubicaciones alternativas cercanas entre las ubicaciones públicas restantes.

La Figura 2.4 muestra tanto la trayectoria original como la trayectoria perturbada resultante del proceso. Los puntos públicos utilizados por el mecanismo también se representan como referencia espacial.

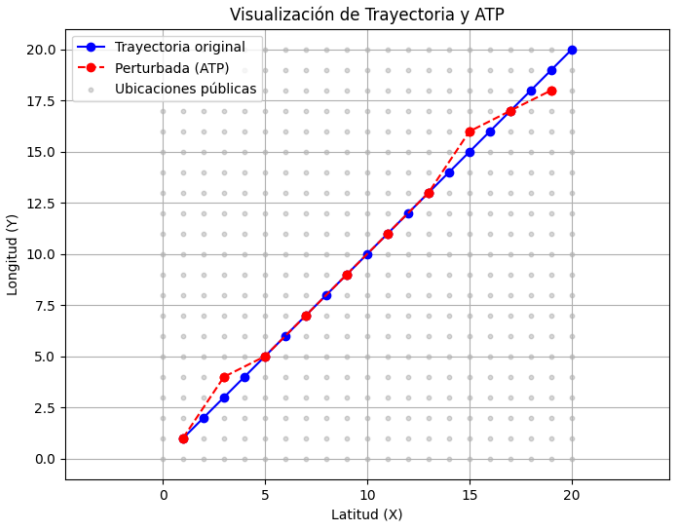

Figura 2.4: Visualización de una trayectoria original y su versión perturbada usando el mecanismo ATP con $\epsilon = 1{,}0$.

2.4. LDPTrace

El algoritmo *LDPTrace* opera bajo un marco teórico riguroso que combina los principios de privacidad diferencial local (LDP) con un modelo probabilístico para la síntesis de trayectorias geoespaciales. Su objetivo principal es generar un conjunto de trayectorias sintéticas que preserven patrones de movilidad representativos de la población, sin comprometer la privacidad de los individuos. Esta descripción del algoritmo *LDPTrace* se fundamenta en el trabajo de Du et al. [2023].

En esencia, LDPTrace transforma trayectorias continuas en secuencias discretas utilizando una cuadrícula espacial, y extrae tres características clave que capturan la movilidad del usuario: la longitud de la trayectoria, las transiciones entre celdas adyacentes y los puntos de inicio/terminación. Estos elementos constituyen un resumen estadístico privado que se utiliza para sintetizar nuevas trayectorias.

El presupuesto de privacidad ϵ se asigna estratégicamente mediante:

$$\epsilon = \epsilon_1 + \epsilon_2 + \epsilon_3 \tag{2.21}$$

correspondiendo a:

- ϵ_1: Preservación de longitudes.
- ϵ_2: Protección de transiciones.
- ϵ_3: Resguardo de puntos terminales.

El modelo probabilístico subyacente se formaliza como:

$$\mathcal{P} = (\mathcal{L}, \mathcal{S}, \mathcal{A}) \tag{2.22}$$

donde $\mathcal{L}$ representa la distribución de longitudes, $\mathcal{S}$ el modelo de transiciones intra-trayectoria basado en cadenas de Markov de primer orden, y $\mathcal{A}$ el conjunto de transiciones desde y hacia los puntos virtuales de inicio y fin.

2.4.1. Discretización Geoespacial

La representación de trayectorias como secuencias de puntos en un dominio bidimensional continuo (coordenadas latitud-longitud) presenta desafíos para su modelado. Para abordar este problema, *LDPTrace* emplea un enfoque de discretización del espacio geográfico en celdas de cuadrícula. Este proceso se realiza mediante la partición del espacio en celdas de igual tamaño utilizando una granularidad de cuadrícula N.

Cada trayectoria T se transforma en una secuencia enumerable de celdas:

$$T = \{C_1, C_2, \ldots, C_{|T|}\} \tag{2.23}$$

donde $T[i]$ denota la i-ésima celda visitada por la trayectoria y $|T|$ representa la longitud de la trayectoria en unidades de celdas. La elección de N determina el tamaño de cada celda y, en consecuencia, la granularidad de la trayectoria discretizada.

La elección del parámetro N, que define la granularidad de la grilla utilizada para discretizar el espacio geográfico, tiene un impacto crucial en la representación de las trayectorias. Cuando N es pequeño, el espacio se divide en un número reducido de celdas, lo que implica que cada celda abarca una región espacial extensa. En consecuencia, varios puntos distintos de la trayectoria original pueden proyectarse sobre una misma celda. Esta agrupación conlleva una pérdida de detalle, y los patrones de movilidad resultantes tienden a ser más generales y, por tanto, menos informativos.

Por el contrario, si se selecciona un valor grande para N, cada celda cubre una región muy pequeña del espacio. Esto permite capturar con mayor precisión los movimientos de los usuarios. Sin embargo, esta mayor resolución también introduce un riesgo: muchas de las celdas pueden quedar vacías, es decir, no ser visitadas por ninguna trayectoria. En tales casos, al aplicar mecanismos de perturbación sobre celdas vacías, se introduce ruido innecesario y se reduce la eficiencia del modelo, afectando negativamente la utilidad de los datos sintetizados.

Ejemplo ilustrativo. Supongamos que el área de estudio corresponde a una ciudad de $10\,\mathrm{km} \times 10\,\mathrm{km}$.

- Si $N = 5$, cada celda mide $2\,\mathrm{km} \times 2\,\mathrm{km}$, por lo que un trayecto que cruza el centro podría proyectarse en tan solo 3–4 celdas, perdiendo detalles de calles intermedias.

- Si $N = 50$, cada celda mide $200\,\mathrm{m} \times 200\,\mathrm{m}$, permitiendo representar con mayor fidelidad cada giro o desplazamiento. Sin embargo, muchas celdas periféricas podrían no ser visitadas, introduciendo ruido extra cuando se aplica el mecanismo de privacidad diferencial local.

2.4.2. Distribución de Longitud de Trayectorias

LDPTrace estima la distribución de longitudes de trayectorias utilizando privacidad diferencial local (LDP), mediante el mecanismo *Optimized Unary Encoding* (OUE). Para proteger la longitud individual de cada trayectoria, el proceso comienza con la codificación del valor. Si una trayectoria tiene longitud m, esta se representa como un vector binario V de tamaño $|C|$, donde $|C|$ corresponde a la longitud máxima posible. El vector contiene un único valor 1 en la posición m, y ceros en las demás:

$$V[i] = \begin{cases} 1 & \text{si } i = m, \\ 0 & \text{en otro caso.} \end{cases}$$

Este vector se perturba localmente para preservar la privacidad del usuario. La perturbación se realiza siguiendo la distribución del mecanismo OUE, que define la probabilidad de que el valor perturbado $\hat{V}[i]$ sea uno como:

$$\Pr[\hat{V}[i] = 1] = \begin{cases} \frac{1}{2}, & \text{si } V[i] = 1, \\ \frac{1}{e^{\epsilon}+1}, & \text{si } V[i] = 0, \end{cases}$$

donde ϵ es el presupuesto de privacidad asignado a esta etapa.

Una vez recolectados los vectores perturbados $\hat{V}$ de todos los usuarios, el curador de datos estima la frecuencia real de cada longitud m a partir del conteo de unos observados en la posición correspondiente, denotado por $\hat{g}(m)$. Para corregir el sesgo introducido por el ruido, se aplica un ajuste que produce un estimador insesgado $\tilde{g}(m)$:

$$\tilde{g}(m) = \frac{\hat{g}(m) - nq}{\frac{1}{2} - q}, \quad \text{donde} \quad q = \frac{1}{e^{\epsilon} + 1},$$

y n es el número total de usuarios.

Finalmente, se obtiene la distribución estimada de longitudes normalizando los valores ajustados:

$$\Pr(m) = \frac{\tilde{g}(m)}{\sum_{i=1}^{|C|} \tilde{g}(i)}.$$

Esta distribución proporciona una visión global de las longitudes de trayectorias, permitiendo generar trayectorias sintéticas coherentes, mientras se mantiene la privacidad de los usuarios individuales gracias a los mecanismos de LDP.

Ejemplo ilustrativo. Supongamos que $|C| = 6$ y que la trayectoria de un usuario tiene longitud $m = 4$.

- El vector binario inicial es $V = [0, 0, 0, 1, 0, 0]$.
- Aplicando OUE con $\epsilon = 1$, se calcula $q = \frac{1}{e^1+1} \approx 0{,}2689$.
- Para la posición $i = 4$ (donde $V[4] = 1$), la probabilidad de que $\hat{V}[4] = 1$ es $0{,}5$.
- Para cualquier otra posición $i \neq 4$ (donde $V[i] = 0$), la probabilidad de que $\hat{V}[i] = 1$ es $\approx 0{,}2689$.

Si $n = 1000$ usuarios envían sus vectores perturbados, el curador podrá aplicar la fórmula de $\tilde{g}(m)$ para corregir el sesgo y estimar la verdadera distribución de longitudes.

2.4.3. Modelo de Movilidad Intra-Trayectoria

Para generar trayectorias sintéticas con mayor realismo, *LDPTrace* modela los patrones de movimiento entre celdas adyacentes mediante una **cadena de Markov de primer orden**. En este enfoque, cada trayectoria discretizada $\mathcal{T} = [C_1, C_2, \ldots, C_m]$ se representa como una secuencia de transiciones $(C_i \rightarrow C_{i+1})$, donde cada salto es modelado como un estado de transición s_{ij}. Con el fin de garantizar la continuidad espacial, las trayectorias se interpolan para asegurar que cada celda C_i sea adyacente a su sucesora C_{i+1}.

La probabilidad de transición entre celdas queda definida como:

$$\Pr(s_{ij}) = \begin{cases} \Pr(T[l+1] = C_j \mid T[l] = C_i), & \text{si } C_j \in \mathcal{N}_{C_i}, \\ 0, & \text{en otro caso,} \end{cases}$$

donde $\mathcal{N}_{C_i}$ es el conjunto de celdas vecinas a C_i, y el conjunto completo de transiciones $\mathcal{S}$ se construye a partir de todas las observadas.

Cada usuario transforma su trayectoria en una secuencia de transiciones $S_{\mathcal{T}} = [s_1, s_2, \ldots, s_{m-1}]$, y representa cada transición $s \in \mathcal{S}$ mediante un vector binario de longitud $|\mathcal{S}| \approx 8|C|$, ya que cada celda puede tener hasta ocho vecinas. Este vector es perturbado localmente usando el mecanismo *Optimized Unary Encoding* (OUE), utilizando un presupuesto de privacidad ϵ_2, y la versión perturbada $\hat{V}$ es enviada al curador de datos.

Para mantener un presupuesto constante entre usuarios, se limita el número de transiciones reportadas utilizando el cuantil-k de la distribución de longitudes, denotado como L_k. Así, se reportan como máximo L_k transiciones por trayectoria, lo cual permite distribuir equitativamente el presupuesto de privacidad ϵ_2 entre ellas.

El curador estima la frecuencia de cada transición utilizando un estimador insesgado derivado del mecanismo OUE, y evalúa su varianza como:

$$N(s, \epsilon_2, L_k) = \mathsf{Var}^* \left[\Pr(s), \frac{\epsilon_2}{L_k} \right],$$

donde $\Pr(s) = \frac{f_s}{f_{L_k}}$ es la frecuencia relativa de la transición s con respecto al total de transiciones permitidas.

La varianza ajustada se obtiene mediante:

$$\mathsf{Var}^* \left[\Pr(s), \frac{\epsilon_2}{L_k} \right] = \frac{f_s^2}{f_{L_k}^2} \left(\frac{\sigma_s^2}{f_s^2} - 2\frac{\mathsf{Cov}(s, L_k)}{f_s f_{L_k}} + \frac{\sigma_{L_k}^2}{f_{L_k}^2} \right),$$

donde $\sigma_s^2 = n \cdot \frac{e^{\epsilon_2/L_k}}{(e^{\epsilon_2/L_k}+1)^2}$ corresponde a la varianza del estimador OUE para la frecuencia de s, y $\mathsf{Cov}(s, L_k)$ representa la covarianza entre f_s y el total f_{L_k}.

El error total asociado a la estimación de los estados de transición incorpora dos componentes: la varianza del estimador (ruido) y el sesgo debido a la omisión de transiciones más allá de L_k. Esta combinación se expresa como:

$$\mathsf{Error}(\mathcal{S}, \epsilon_2, L_k) = \sum_{s \in \mathcal{S}} \left[N(s, \epsilon_2, L_k) + (1-k)^2 \cdot f_s^2 \right].$$

El hiperparámetro $k \in (0, 1]$ regula el compromiso entre ruido y sesgo: un valor de k bajo reduce el ruido al concentrar presupuesto por transición, pero introduce sesgo al omitir parte

de la trayectoria; mientras que un valor alto reduce el sesgo al capturar más transiciones, a costa de una mayor varianza. En la práctica, se selecciona el valor óptimo de k que minimiza el error total sobre el conjunto de transiciones $\mathcal{S}$.

Ejemplo ilustrativo. Supongamos un espacio discretizado en una cuadrícula de $N = 3 \times 3$ celdas, numeradas del C_1 al C_9. Un usuario reporta la trayectoria $[C_5, C_6, C_9]$.

- Las transiciones reales son: $s_{5,6}$ y $s_{6,9}$.
- El conjunto $\mathcal{S}$ contiene todas las transiciones posibles entre celdas adyacentes (máximo 8 por celda), lo que da $|\mathcal{S}| \approx 8|C| = 72$.
- Cada transición se codifica como un vector binario de tamaño 72, con un único 1 en la posición correspondiente a la transición observada.
- Con $\epsilon_2 = 1{,}5$ y $L_k = 2$, el presupuesto por transición es $\epsilon_2/L_k = 0{,}75$.
- El mecanismo OUE asigna probabilidad $0{,}5$ de mantener un 1 en su posición real, y $q = \frac{1}{e^{0{,}75}+1} \approx 0{,}3208$ para asignar un 1 a las posiciones que originalmente eran 0.
- Tras recibir los vectores perturbados de múltiples usuarios, el curador aplica la corrección de sesgo, estima $\Pr(s)$ para cada transición y calcula el error total para ajustar k.

Este proceso permite modelar patrones de movilidad realistas protegiendo las transiciones individuales de cada usuario.

2.4.4. Modelado de Transiciones de Inicio y Fin

Además de capturar los patrones de movimiento intra-trayectoria representados por el conjunto $\mathcal{S}$, es esencial modelar de manera explícita las transiciones que definen el inicio y el final de cada trayectoria. En aplicaciones reales, como trayectos en taxi o viajes casa-trabajo, los puntos de partida y llegada no se distribuyen de forma uniforme, sino que tienden a concentrarse en regiones semánticamente relevantes, como áreas residenciales, estaciones de transporte o zonas comerciales.

Para reflejar esta distribución no uniforme, se introducen dos celdas virtuales que actúan como nodos auxiliares: C_a, que representa un punto de inicio virtual conectado a todas las celdas reales; y C_b, un punto de finalización virtual al que todas las celdas reales pueden

conectarse. Con estas celdas se definen dos clases de transiciones especiales: las transiciones de inicio, representadas como $A_j = (C_a \to C_j)$, y las de fin, como $B_i = (C_i \to C_b)$. Cada una de estas transiciones ocurre exactamente una vez por trayectoria y se modela por separado, permitiendo un tratamiento más preciso del ruido bajo privacidad diferencial local (LDP).

Cada usuario codifica estas transiciones mediante el protocolo OUE, asignándoles un presupuesto de privacidad independiente ϵ_3. Para ello, las transiciones se representan como vectores binarios sobre un dominio extendido $\mathcal{M}$, el cual abarca todas las transiciones intra-trayectoria y las nuevas transiciones de inicio y fin:

$$\mathcal{M} = \mathcal{S} \cup \{A_j\}_{j \in C} \cup \{B_i\}_{i \in C}.$$

Una vez recolectados los vectores perturbados, el curador estima la frecuencia de cada transición del dominio $\mathcal{M}$, y construye el modelo de movilidad completo. La probabilidad de transición entre dos celdas C_i y C_j se define como:

$$\Pr(M_{ij}) = \frac{\tilde{g}(M_{ij})}{\sum\limits_{r \in \mathcal{N}^*_{C_i}} \tilde{g}(M_{ir})},$$

donde $\tilde{g}(M_{ij})$ representa la frecuencia estimada de la transición M_{ij}. Esta puede corresponder a una transición de inicio, fin o intra-trayectoria, según:

$$M_{ij} = \begin{cases} A_j, & \text{si } C_i = C_a, \\ B_i, & \text{si } C_j = C_b, \\ s_{ij}, & \text{en otro caso.} \end{cases}$$

El denominador considera un conjunto extendido de vecinos $\mathcal{N}^*_{C_i}$, que permite la integración de transiciones desde y hacia las celdas virtuales:

$$\mathcal{N}^*_{C_i} = \begin{cases} \mathcal{N}_{C_i} \cup \{C_b\}, & \text{si } C_i \in C, \\ C, & \text{si } C_i \in \{C_a, C_b\}. \end{cases}$$

Este modelo extendido $\mathcal{M}$ proporciona una representación más completa del comportamiento de los usuarios, al capturar no solo las dinámicas locales de movimiento, sino también

las distribuciones espaciales de origen y destino. Al integrarse con el modelo intra-trayectoria, permite sintetizar trayectorias que reflejan de manera más realista los patrones observados en la base de datos original.

Ejemplo ilustrativo. Consideremos nuevamente la cuadrícula 3×3 con celdas C_1 a C_9. Supongamos que un usuario realiza la trayectoria $[C_2, C_5, C_8]$.

- La transición de inicio es $A_2 = (C_a \to C_2)$ y la de fin es $B_8 = (C_8 \to C_b)$.
- Las transiciones intra-trayectoria son $s_{2,5}$ y $s_{5,8}$.
- El dominio extendido $\mathcal{M}$ contiene:
  - Todas las transiciones intra-trayectoria $\mathcal{S}$.
  - 9 transiciones de inicio A_1 a A_9.
  - 9 transiciones de fin B_1 a B_9.
- Si $|\mathcal{S}| = 72$, entonces $|\mathcal{M}| = 72 + 9 + 9 = 90$.
- Con $\epsilon_3 = 1$, el OUE para cada transición especial asigna $p = \frac{e^1}{e^1+1} \approx 0{,}7311$ a la posición real y $q = \frac{1}{e^1+1} \approx 0{,}2689$ a las demás.
- El curador, tras recibir múltiples reportes perturbados, estima las frecuencias $\tilde{g}(M_{ij})$ y normaliza con $\mathcal{N}^*_{C_i}$ para obtener probabilidades de inicio y fin realistas.

Este procedimiento asegura que los patrones espaciales de origen y destino queden preservados en el modelo sintético bajo garantías de privacidad.

2.4.5. Distribución de Longitud de Trayectorias

Para estimar la distribución de longitudes de trayectorias sin comprometer la privacidad individual, LDPTrace emplea el mecanismo de codificación *Optimized Unary Encoding* (OUE), el cual se ajusta al marco de privacidad diferencial local.

Cada usuario codifica la longitud m de su trayectoria como un vector binario $V \in \{0, 1\}^{|C|}$, donde $|C|$ representa la longitud máxima considerada. En este vector, únicamente la posición correspondiente a m contiene un uno, mientras que el resto de las posiciones se establecen en

cero:

$$V[i] = \begin{cases} 1, & \text{si } i = m, \\ 0, & \text{en otro caso.} \end{cases}$$

Luego, para preservar la privacidad, cada componente del vector se perturba de forma independiente utilizando el mecanismo OUE. Específicamente, si una posición i tiene un valor original de uno, se reporta como uno con probabilidad $\frac{1}{2}$; en cambio, si originalmente es cero, se reporta como uno con una probabilidad más baja $q = \frac{1}{e^{\epsilon}+1}$, donde ϵ representa el presupuesto de privacidad asignado:

$$\Pr[\hat{V}[i] = 1] = \begin{cases} \frac{1}{2}, & \text{si } V[i] = 1, \\ q = \frac{1}{e^{\epsilon}+1}, & \text{si } V[i] = 0. \end{cases}$$

Una vez recolectados los vectores perturbados $\hat{V}$ de todos los usuarios, el curador estima la frecuencia de cada posible longitud i sumando los unos observados en la posición correspondiente, lo que se denota como $\hat{g}(i)$. Estas observaciones se corrigen aplicando un estimador insesgado, el cual compensa el sesgo introducido por el ruido aleatorio:

$$\tilde{g}(i) = \frac{\hat{g}(i) - nq}{\frac{1}{2} - q},$$

donde n es el número total de usuarios participantes.

Finalmente, la distribución de longitudes de trayectorias se construye como una distribución categórica, normalizando las frecuencias estimadas:

$$\Pr(m) = \frac{\tilde{g}(m)}{\sum_{i=1}^{|C|} \tilde{g}(i)}.$$

Este enfoque permite capturar patrones agregados de comportamiento en cuanto a la longitud de las trayectorias, garantizando al mismo tiempo la privacidad local de cada usuario mediante un proceso de codificación y perturbación controlada.

Ejemplo ilustrativo. Supongamos que la longitud máxima considerada es $|C| = 5$ y que un usuario tiene una trayectoria de longitud $m = 3$.

- El vector original es $V = [0, 0, 1, 0, 0]$.

- Para un presupuesto de privacidad $\epsilon = 1$, se tiene $q = \frac{1}{e^1+1} \approx 0{,}2689$.

- La posición $i = 3$ (valor real 1) se reporta como 1 con probabilidad $0{,}5$; las demás (valor real 0) se reportan como 1 con probabilidad $0{,}2689$.

- Tras recibir múltiples reportes de n usuarios, supongamos que para $i = 3$ se observan $\hat{g}(3) = 120$ unos, siendo $n = 300$.

- El estimador corrige el sesgo:

$$\tilde{g}(3) = \frac{120 - 300 \times 0{,}2689}{0{,}5 - 0{,}2689} \approx \frac{39{,}33}{0{,}2311} \approx 170{,}3.$$

- La distribución final $\Pr(m)$ se obtiene normalizando todas las $\tilde{g}(i)$ para $i = 1, \ldots, 5$.

De esta forma, el curador obtiene una estimación precisa de la frecuencia de longitudes de trayectorias, preservando la privacidad individual de los usuarios.

2.4.6. Representación Gráfica de LDPTrace

La implementación considera un área cuadrada de 20×20 unidades, discretizada en una grilla de 10×10 celdas. La trayectoria original está compuesta por 11 puntos equidistantes que conectan los extremos del espacio, es decir, desde $(0, 0)$ hasta $(20, 20)$.

El algoritmo emplea el mecanismo *Optimized Unary Encoding* (OUE) con un presupuesto total de privacidad $\epsilon = 3{,}0$, distribuido equitativamente entre las tres fases del modelo: $\epsilon_1 = 1{,}0$ para la longitud de la trayectoria, $\epsilon_2 = 1{,}0$ para las transiciones entre celdas adyacentes, y $\epsilon_3 = 1{,}0$ para los puntos de inicio y final.

La Figura 2.5 presenta la trayectoria original discretizada y la trayectoria sintética generada por LDPTrace. Se observa que la trayectoria sintética respeta la estructura de la cuadrícula y reproduce de forma realista patrones de movimiento similares a los originales, pese a las perturbaciones introducidas para garantizar la privacidad.

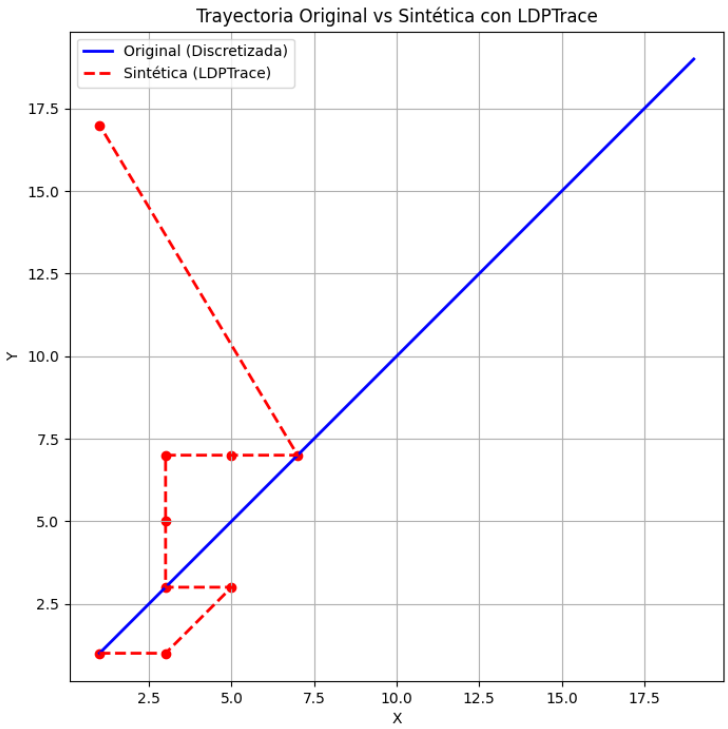

Figura 2.5: Visualización de una trayectoria original (azul) y una trayectoria sintética (roja) generada mediante el algoritmo LDPTrace con $\epsilon = 3{,}0$.

Bibliografía

[1] Yuntao Du, Yujia Hu, Zhikun Zhang, Ziquan Fang, Lu Chen, Baihua Zheng, and Yunjun Gao. Ldptrace: Locally differentially private trajectory synthesis. In *Proceedings of the VLDB Endowment*, volume 16, pages 1854–1866. VLDB Endowment, 2023. doi: 10.14778/3594512.3594520. URL https://doi.org/10.14778/3594512.3594520. Licensed under the Creative Commons BY-NC-ND 4.0 International License.

[2] Cynthia Dwork. Differential privacy: A survey of results. In Manindra Agrawal, Dingzhu Du, Zhenhua Duan, and A. Li, editors, *Theory and Applications of Models of Computation. TAMC 2008*, volume 4978 of *Lecture Notes in Computer Science*, pages 1–19. Springer, 2008. doi: 10.1007/978-3-540-79228-4_1.

[3] Daiyong Quan, Lihua Yin, and Yunchuan Guo. Enhancing the trajectory privacy with laplace mechanism. In *2015 IEEE Trustcom/BigDataSE/ISPA*, pages 758–765. Institute of Information Engineering, Chinese Academy of Sciences, IEEE, 2015. doi: 10.1109/Trustcom-BigDataSE.2015.408.

[4] Yuemin Zhang, Qingqing Ye, Rui Chen, Haibo Hu, and Qilong Han. Trajectory data collection with local differential privacy. *Proceedings of the VLDB Endowment*, 16(10): 2591–2604, 2023. doi: 10.14778/3603581.3603597. URL https://doi.org/10.14778/3603581.3603597.