

Cálculo del Volumen en Dimensión Alta

Santosh Vempala

Apunte basado en el curso *Computing the Volumen in High Dimension* de la XII Escuela de Matemáticas Discretas de Valparaíso, 6 al 9 de Enero 2017.

Escrito por Patricio Foncea y revisado por Santosh Vempala.

Índice

1. Introducción	2
1.1. Problemas de Dimensión Alta	2
1.2. La frontera de complejidad. Convexidad	3
1.3. Aprendizaje por Muestreo	4
1.4. Distribución del Volumen. Log-concavidad	5
1.5. Elipsoides	7
2. Algoritmos Aleatorizados	9
2.1. Redondeo por muestreo	10
2.2. Complejidad del Muestreo	11
2.3. Recocido Simulado	11
2.4. Optimización por Recocido	13
3. Muestreo y Probabilidad	13
3.1. Paseos Aleatorios Geométricos	13
3.2. Conductancia y Convergencia	14
4. Geometría	17
4.1. Isoperimetría	17
4.2. Localización	20
4.3. Enfriamiento Gaussiano	21
5. Problemas Abiertos	23

1. Introducción

El cálculo de volumen es un problema muy antiguo. Formalmente, tenemos un conjunto K en un espacio de dimensión n y queremos estimar su volumen. Ejemplos varían desde calcular el volumen de pirámides y barriles de vino, hasta politopos o la intersección de un politopo con una elipsoide. Por otro lado, analizar problemas en dimensión alta no solo permite modelar distintos aspectos de un problema (cantidad de variables), sino que además otorga nuevas técnicas y algoritmos. En este curso, miraremos esta pregunta desde un punto de vista computacional, es decir, nos interesa el *cuánto*, o que tan costoso es el cálculo de volumen.

El modelo computacional con el que trabajaremos consta de un conjunto K compacto para el cual supondremos tenemos un oráculo de membresía bien garantizado. En otras palabras, tenemos un oráculo de membresía del cual podemos obtener una respuesta certera de si $x \in K$. Además, supondremos que conocemos un punto $x_0 \in K$ y dos números r, R tales que

$$x_0 + rB^n \subseteq K \subseteq RB^n.$$

1.1. Problemas de Dimensión Alta

Tradicionalmente, el input de nuestro problema será un conjunto S de puntos en \mathbf{R}^n o una distribución de probabilidad \mathbf{R}^n . Más en general, podría ser una función que mapea puntos a reales. La pregunta que intentaremos responder será sobre cuál es la complejidad del problema a medida que la dimensión n crece. Recordemos que la dimensión nos importa ya que esta representa el número de variables.

Aunque el problema del cálculo de volumen tiene un interés teórico Muestre al azar e integración: aplicación en diversas áreas (stadística, redes, biología, vision computacional, privacidad, investigación de operaciones, etc.)

En este curso estudiaremos el problema de muestreo e integración. Aunque el enfoque será en los fundamentos matemáticos y algorítmicos de estos problemas, cabe destacar que estos tienen aplicaciones en diversas áreas como stadística, redes, biología, vision computacional, privacidad, investigación de operaciones, etc. Existen rutinas en MATLAB para desarrollar estas aplicaciones en caso de interés.

Los tipos de problemas que estudiaremos son los siguientes:

1. Optimización

Input: Una función $f : \mathbf{R}^n \rightarrow \mathbf{R}$ (dada por un oráculo), un punto x y un error ϵ .

Output: Un punto y tal que

$$f(y) > \max f - \epsilon.$$

Algoritmo Clásico: Método de la elipsoide es polinomial cuando la función y el conjunto sobre el que se optimiza son convexos.

2. Integración.

Input: Una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$ (dada por un oráculo), un punto x y un error ϵ .

Output: Valor A tal que

$$(1 - \epsilon) \int f < A < (1 + \epsilon) \int f.$$

Algoritmo Clásico: Algoritmo DFK es polinomial cuando el conjunto es convexo.

3. Muestreo al azar

Input: Una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$ (dados por un oráculo), un punto x y un error ϵ .

Output: Un punto y de una distribución a distancia ϵ de una distribución proporcional a f .

4. Redondear

Input: Una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$ o un conjunto (dada por un oráculo), un punto x y un error ϵ .

Output: Una transformación afón que aproximadamente redondea f entre dos bolas.

5. Aprendizaje.

Input: Puntos i.i.d. etiquetados según una distribución desconocida y un error ϵ .

Output: Una regla para etiquetar correctamente una proporción $1 - \epsilon$ de los puntos del input.

Todos estos problemas son difíciles de resolver y de aproximar en el caso general.

1.2. La frontera de complejidad. Convexidad

Determinar si una matriz A es de la forma $A = BB^t$ se puede resolver en tiempo polinomial, pero si se impone la condición extra de que las entradas de B sean positivas, entonces se vuelve difícil. Esto, sumado a los ejemplos anteriores donde existían algoritmos cuando habían estructuras convexas, lleva a pensar que lo que se necesita para poder resolver estos problemas es convexidad (o alguna extensión de esta).

Se dirá que f es *log-cóncava* si f es no negativa y $\log(f)$ es una función cóncava, y se dirá que f es *quasi-cóncava* si para $\lambda \in [0, 1]$, $f(\lambda x + (1 - \lambda)y) \geq \min\{f(x), f(y)\}$. Por último, un conjunto S se dice *conjunto estrella* si existe un punto $x \in S$ tal que toda línea que pasa por este punto y otro punto cualquiera en S , está contenida en el conjunto.

Para especificar un conjunto convexo se puede dar mediante una serie de restricciones como $Ax \leq b$, por ejemplo. Sin embargo, hay otros que no pueden ser descritos de esta forma. Un ejemplo de esto es el conjunto de matrices semidefinidas positivas, o el conjunto de vectores sobre los arcos de un grafo que tienen peso al menos uno en cada corte.

De ahora en adelante pensaremos que cada conjunto convexo K tiene un oráculo de separación. Es decir, para cada punto x podemos determinar si x pertenece a K , y en caso contrario, el oráculo entrega un hiperplano separador entre el punto y el convexo.

En los casos anteriores, un oráculo verificaría si todos los valores propios de la matriz son semidefinidos positivos, o entregaría el vector propio respectivo; o buscaría el corte mínimo y chequearía si este es mayor a uno.

1.3. Aprendizaje por Muestreo

Sea una secuencia de puntos X_1, X_2, \dots y una función f desconocida que toma valores 1 o -1 . Para cada uno de los puntos X_i , buscamos adivinar su etiqueta correcta $f(X_i)$. El objetivo es minizar el número de errores cometidos al adivinar las etiquetas.

Si el problema es planteado de esta forma, no podemos tener ninguna garantía. Necesitamos estructura en la función f . Para ello supongamos que f es tal que existe $w \in \mathbb{R}^n$ desconocido, donde cada entrada w_i es un entero de b bits, para el cual

$$f(X) = 1 \text{ ssi } w^T X > 0.$$

El problema ahora es que no conocemos el semiespacio determinado por w . Notemos que este es un problema de dimensión alta, ya que en dimensión baja podemos adivinar con un tamaño pequeño de errores, sin embargo, a medida que crece la dimensión el tamaño de la envoltura convexa crece exponencialmente.

Nos gustaría tener un método en donde la cantidad de preguntas que se deben realizar antes de aprender el etiquetado (es decir, la cantidad mínima de errores que debemos arriesgar a cometer) fuese polinomial en la dimensión. Proponemos el siguiente algoritmo (llamado *Algoritmo de Mayoría*) para atacar el problema anterior.

Realizar k preguntas y construir el conjunto

$$S_k = \{w : w^T(\text{sgn}(X_i)X_i) > 0, i = 1, \dots, k\},$$

donde $\text{sgn}(X_i) = f(X_i)$ corresponde al etiquetado del punto X_i .

El algoritmo adivina $f(X_i)$ como la mayoría de las predicciones de cada w en S_k . Notemos que cada vez que nos equivocamos, podemos desechar la mitad de los elementos y reducir el tamaño del conjunto. La pregunta ahora es cómo calculamos la mayoría, ya que en principio el conjunto S_k podría ser de tamaño $2^{(b+1)n}$. Para esto introducimos un algoritmo aleatorio, el cual escoge un punto w de S_k al azar y lo utilizamos para adivinar $w^T X_{k+1}$.

Lema 1.1 (Ejercicio 1). $\mathbb{E}(\text{número de errores}) \leq 2(b+1)n$.

(Idea: cada vez que una adivinanza está equivocada, el algoritmo de la mayoría tiene probabilidad al menos $1/2$ de estar equivocado.)

Solo queda determinar como escoger eficientemente un punto w de S_k . Esto se verá más adelante, pero se debe a que S_k es convexo.

1.4. Distribución del Volumen. Log-concavidad

Por ahora asumiremos que tenemos K determinado por un oráculo de membresía (no de separación) y un punto dentro de K . Recordemos que nuestro objetivo es determinar el volumen de K .

Nuestro primer intento para determinar este volumen será la clásica estrategia de dividir y conquistar: tomamos una partición en cubos del espacio y para cada uno, si su centro se encuentra dentro del conjunto, lo contamos. La estimación del volumen será la proporción de cubos dentro del conjunto con respecto al total, ponderado por el volumen total. El problema es que la cantidad de cubos que se necesita para este procedimiento crece exponencialmente en la dimensión.

Es importante también estudiar la distribución de volumen. El volumen de la bola de radio uno decae exponencialmente en la dimensión n . De hecho, el volumen es aproximadamente $(c/n)^{n/2}$. Esto significa que cuando intentemos muestrear la bola inscrita en un cubo, necesitaremos muchos puntos ya que el tamaño de la bola con respecto al cubo se vuelve muy pequeña.

Aún más, se puede demostrar que para cada hiperplano que pasa por el centro de la bola, la mayoría de la masa de la bola está a distancia $1/\sqrt{n}$ de este hiperplano.

Sin embargo, también es cierto que casi todo el volumen de la bola está cerca de la frontera! Como $(1 - 1/n)^n$ converge a $1/e$, que es un valor constante, tenemos que

$$\text{vol}((1 - \varepsilon)K) = (1 - \varepsilon)^n \text{vol}(K) \approx \text{vol}(K).$$

Definimos la suma Minkowski como $A + B = \{x + y : x \in A, y \in B\}$. Dado un conjunto K , para un punto x , podemos tomar su sección transversal K_x a través de K y por convexidad tenemos que

$$\frac{A_x + A_y}{2} \subseteq A_{\frac{x+y}{2}}.$$

El siguiente teorema nos llevará a encontrar como distribuye el volumen, al relacionar el volumen (en $n - 1$ dimensiones) de las proyecciones con respecto al volumen de la proyección de su promedio (el cual siempre está en el conjunto).

Teorema 1.2 (Desigualdad de Brunn-Minkowski). *La función sobre conjuntos $\text{vol}(\cdot)^{1/n}$ es una función cóncava. Es decir, para $\lambda \in [0, 1]$,*

$$\text{vol}(\lambda A + (1 - \lambda)B)^{\frac{1}{n}} \geq \lambda \text{vol}(A)^{\frac{1}{n}} + (1 - \lambda) \text{vol}(B)^{\frac{1}{n}}.$$

Demostración. Basta probar que

$$\text{vol}(A + B)^{1/n} \geq \text{vol}(A)^{1/n} + \text{vol}(B)^{1/n}.$$

Comencemos asumiendo que $A = [0, a_1] \times \dots \times [0, a_n]$ y $B = [0, b_1] \times \dots \times [0, b_n]$ son

cuboides. Esto implica que $A + B = [0, a_1 + b_1] \times \dots \times [0, a_n + b_n]$ y luego

$$\begin{aligned} \frac{vol(A)^{\frac{1}{n}} + vol(B)^{\frac{1}{n}}}{vol(A + B)^{\frac{1}{n}}} &\leq \left(\prod_i \frac{a_i}{a_i + b_i} \right)^{\frac{1}{n}} \left(\prod_i \frac{b_i}{a_i + b_i} \right)^{\frac{1}{n}} \\ &\leq \frac{1}{n} \sum_i \frac{a_i}{a_i + b_i} + \frac{1}{n} \sum_i \frac{b_i}{a_i + b_i} \\ &= 1. \end{aligned}$$

Donde la segunda desigualdad se debe a la desigualdad de promedio geométrico y promedio aritmético.

En general, basta que probemos la desigualdad para uniones finitas de cuboides disjuntos, ya que el volumen de cualquier conjunto puede ser aproximada por una unión finita de cuboides disjuntos. Es decir, $A = \cup_i A_i$ y $B = \cup_j B_j$.

Primero, notemos que se puede probar que existe un hiperplano paralelo a algún eje tal que existe al menos un cuboide completo de los que forman A en cada lado. Luego, particionamos los cuboides que forman A y B según en que lado del hiperplano se encuentran en dos conjuntos A^+ y A^- , y B^+ y B^- . Ya que a un lado del hiperplano siempre hay al menos un cuboide A_i completamente contenido en él, sabemos que al otro lado hay al menos un cuboide menos de los que forman originalmente al conjunto A . Esto nos permite usar inducción en el número de cuboides que se necesitan para formar cada conjunto, con el caso base probado anteriormente.

Antes de concluir, necesitamos trasladar B para que se cumpla que

$$\frac{vol(A)}{vol(B)} = \frac{vol(A^+)}{vol(B^+)}.$$

Sigue que

$$\begin{aligned} vol(A + B) &\geq vol(A^+ + B^+) + vol(A^- + B^-) \\ &\geq (vol(A^+)^{\frac{1}{n}} + vol(B^+)^{\frac{1}{n}})^n + (vol(A^-)^{\frac{1}{n}} + vol(B^-)^{\frac{1}{n}})^n \\ &\geq vol(A^+) \left(1 + \frac{vol(B^+)^{\frac{1}{n}}}{vol(A^+)^{\frac{1}{n}}} \right)^n + vol(A^-) \left(1 + \frac{vol(B^-)^{\frac{1}{n}}}{vol(A^-)^{\frac{1}{n}}} \right)^n \\ &\geq (vol(A^+) + vol(A^-)) \left(1 + \frac{vol(B)^{\frac{1}{n}}}{vol(A)^{\frac{1}{n}}} \right)^n \\ &\geq (vol(A)^{\frac{1}{n}} + vol(B)^{\frac{1}{n}})^n. \end{aligned}$$

□

Recordemos que f es log-cóncava si para $\lambda \in [0, 1]$

$$f(\lambda x + (1 - \lambda)y) \geq f(x)^\lambda f(y)^{1-\lambda}.$$

Entre las funciones log-cóncavas encontramos las indicatrices de conjuntos convexos y algunas densidades de probabilidad, como la gaussiana y la exponencial.

Antes de enunciar algunas de propiedades de estas funciones, presentamos el siguiente teorema:

Teorema 1.3 (Desigualdad de Prekopa-Leindler). Sean $f, g, h : \mathbf{R}^n \rightarrow \mathbf{R}$ tales que

$$h(\lambda x + (1 - \lambda)y) \geq f(x)^\lambda g(y)^{1-\lambda},$$

entonces,

$$\int h \geq \left(\int f \right)^\lambda \left(\int g \right)^{1-\lambda}.$$

Propiedades de las funciones log-cóncavas:

1. El producto y el mínimo de dos funciones log-cóncavas es una función log-cóncava. La suma en general no lo es.
2. La convolución de dos funciones log-cóncavas es log-concava.
3. Las densidades marginales de una función log-concava es log-cóncava.

1.5. Elipsoides

Intentemos ahora calcular el volumen usando elipsoides. Para un conjunto K convexo, se define la elipsoide de John E como la elipsoide de máximo volumen contenida en K . Luego, para todo convexo K la elipsoide de John satisface que

$$E \subseteq K \subseteq nE.$$

Si K además es simétrica con respecto a su centro (i.e., $x \in K \Rightarrow -x \in K$) entonces $E \subseteq K \subseteq \sqrt{n}E$.

Ejercicio 2. Encuentre un ejemplo ajustado de elipsoide de John.

Podemos utilizar una pequeña modificación en el algoritmo de la elipsoide para construir una aproximación de la elipsoide de John. Supongamos que el centro actual es z y los semiejes son a_1, \dots, a_n . Verificamos si $z + -a_i/n \in K$. Si se cumple, entregamos esta elipsoide. Si no, tomamos otra elipsoide de volumen mínimo que contenga la intersección de la elipsoide anterior y el hiperplano separador.

$\text{vol}(E_{i+1})/\text{vol}(E_i) \leq \exp(-c/n) \approx 1 - c/n$. Luego, número de iteraciones $O(n^2 \log(R/r))$.

Teorema 1.4. El algoritmo anterior entrega un elipsoide E con $E \subseteq K \subseteq n^{1.5}E$, en tiempo polinomial.

Lo anterior implica que $\text{vol}(E) \leq \text{vol}(K) \leq n^{1.5n} \text{vol}(E)$. Es decir, tenemos una aproximación exponencial en tiempo polinomial.

Veamos ahora un segundo tipo de elipsoide: la elipsoide de inercia. Para un cuerpo K convexo, se define su matriz de inercia por

$$M = \mathbb{E}_K((x - \bar{x})(x - \bar{x})^t).$$

La matriz de inercia corresponde a $E = \{y \in \mathbb{R}^n : y^t M^{-1} y \leq 1\}$.

Teorema 1.5 (Kannan, Lovász, y Simonovits (1995)). *La matriz de inercia cumple que*

$$\sqrt{\frac{n+1}{n}} E \subseteq K \subseteq \sqrt{n(n+1)} E.$$

Podemos realizar una transformación isotrópica a cualquier elipsoide que tenga segundos momentos acotados. Es decir, volver a la elipsoide una bola centrada en el origen. Más en general, podemos aplicar esta transformación a un conjunto convexo K , tal que su centro sea el origen y su matriz de inercia sea la identidad. Este proceso se llama redondeo y deja a cualquier cuerpo convexo lo más redondo posible.

Aún más, la elipsoide inercial puede ser aproximada a cualquier valor constante. Por lo tanto,

$$E \subseteq K \subseteq 2nK \Rightarrow \text{vol}(E) \leq \text{vol}(K) \leq (2n)^n \text{vol}(E).$$

¿Podemos hacer algo mejor que una $O(n^n)$ aproximación en tiempo polinomial?

Consideremos una tercera elipsoide. Para ello, si A y B son dos conjuntos compactos, definamos por $N(A, B)$ = número de traslaciones de A necesarias para cubrir a B . Dado un conjunto convexo K , la elipsoide de Milman E es aquella que cumple que

$$N(K, E)N(E, K) \leq 2^{O(n)}.$$

Su existencia está garantizada por el Teorema de Milman, y puede ser encontrada algorítmicamente. En particular,

Teorema 1.6 (Dadush y Vempala (2012)). *La complejidad determinista de calcular una elipsoide de Milman es $2^{O(n)}$.*

¿Podemos mejorar la $2^{O(n)}$ aproximación en $2^{O(n)}$ tiempo? Los siguientes resultados sobre complejidad del volumen nos dicen que la respuesta es negativa.

Teorema 1.7 (Elekes (1986); Furedi y Barany (1986)). *Todo algoritmo determinista que usa a lo mas n^α llamados a un oráculo de membresía de un conjunto convexo K y calcula dos números A y B tales que $A \leq \text{vol}(K) \leq B$, existe un convexo para el cual el ratio B/A es al menos*

$$\left(\frac{cn}{a \log(n)} \right)^{\frac{n}{2}}.$$

Teorema 1.8 (Furedi y Barany (1986)). *Se puede obtener una $(1 - \epsilon)$ aproximación con $(1/\epsilon)$ llamadas al oráculo.*

Teorema 1.9 (Dyer y Frieze (1988)). *Calcular el volumen de un polítopo de la forma $Ax \leq b$ es un problema $\#P$ -completo (incluso para una matriz unimodular y un vector racional).*

Teorema 1.10 (Dadush y Vempala (2012)). *Se puede lograr un factor de aproximación de $(1+\epsilon)^n$ en tiempo $(1/\epsilon)^{O(n)} \text{poly}(n)$.*

Nos gustaría ahora obtener un límite inferior para todos los algoritmos deterministas. Probaremos un límite exponencial. Consideraremos un oráculo de membresía que responde "SÍ" para todos los puntos en la bola unitaria, y "NO" para los puntos afuera. Después de m preguntas, el volumen de K está entre la envoltura convexa de estos m puntos y el volumen de la bola. Sin embargo, la proporción entre estos dos volúmenes es exponencial.

Lema 1.11 (Elekes (1986)). $\text{vol}(\text{conv}\{x_1, \dots, x_m\}) \leq \frac{m}{2^n} \text{vol}(B^n)$.

Demostración. Sean B_i las bolas de centro $x_i/2$ y radio $\|x_i\|/2$. Es claro que $\sum_i \text{vol}(B_i) \leq m \frac{\text{vol}(B^n)}{2^n}$. Además, si $y \notin \cup_i B_i$, entonces para todo i , el triángulo $0yx_i$ tiene un ángulo agudo en el vértice donde está y , y por lo tanto, el hiperplano ortogonal que pasa por este punto deja en el mismo lado a 0 y a x_i . Como esto es para todo i , se concluye que y no puede estar en la envoltura convexa de los x_i , y por lo tanto $\text{conv}\{x_1, \dots, x_m\} \subseteq \cup_i B_i$. \square

2. Algoritmos Aleatorizados

Ya vimos que los algoritmos deterministas no podían superar la barrera exponencial en términos de aproximación de volumen. Comenzaremos esta sección dando un par de resultados que mejoran las cotas anteriores cuando el algoritmo es aleatorizado.

Teorema 2.1 (Dyer, Frieze, y Kannan (1989)). *Existe un algoritmo aleatorizado que en tiempo $\text{poly}(n, 1/\epsilon, \log(1/\delta))$ entrega una estimación con un error relativo de $1 + \epsilon$ con probabilidad al menos $1 - \delta$.*

Teorema 2.2 (Applegate y Kannan (1991)). *Existe un algoritmo aleatorizado que en tiempo polynomial entrega una aproximación $(1 + \epsilon)$ de la integral de cualquier función log-cóncava y Lipschitz.*

Realizaremos un tercer intento para estimar volúmenes, esta vez usando muestreo. Tomaremos muestra aleatorias dentro de una bola o un cubo que contenga al conjunto K y contaremos la fracción de puntos que están dentro de K y la usaremos para calcular el volumen del convexo. Sin embargo, tal como está descrito este algoritmo, se requerirían demasiadas muestras para obtener garantías decentes en la aproximación (la razón entre el volumen del convexo y el cuerpo que lo contiene puede ser exponencial).

Para superar esta dificultad, no estimaremos directamente el volumen del conjunto K , sino que iremos acercándonos a través de una sucesión decreciente de bolas. Si $K_i = K \cap 2^{i/n} B$ para $i = 0, \dots, m = n \log(R)$, tenemos que

$$\text{vol}(K) = \text{vol}(B) \frac{\text{vol}(K_1)}{\text{vol}(K_0)} \cdots \frac{\text{vol}(K_m)}{\text{vol}(K_{m-1})}.$$

Sigue que si podemos obtener un muestreo de los puntos de K_i con respecto a las bolas K_{i+1} , obtenemos un estimado para los volúmenes sucesivos. La pregunta de como muestrear la dejaremos para más adelante; por ahora analizaremos la complejidad del algoritmo.

Nos preguntamos primero cuántos puntos necesitamos. Comencemos notando que $\text{vol}(K_{i+1}) \leq \text{vol}(K_i)$. Para encontrar la cantidad de puntos que necesitamos, podemos estimar la varianza de la muestra y luego estimar la cantidad de puntos con la expresión

$$\frac{\mathbb{V}\text{ar}(Y_1 \cdot \dots \cdot Y_m)}{\mathbb{E}(Y_1 \cdot \dots \cdot Y_m)^2} = \prod_i \frac{\mathbb{E}(Y_i^2)}{\mathbb{E}(Y_i)^2} - 1 = \prod_i \left(1 + \frac{\mathbb{V}\text{ar}(Y_i)}{\mathbb{E}(Y_i)^2} \right) - 1 \leq \frac{c}{k},$$

donde k es el número de muestras que se toman en cada fase. Esto implica que hacen falta m/ϵ^2 muestras para garantizar la aproximación. Es decir, tenemos una complejidad de $m^2/\epsilon^2 = O^*(n^2)$, con $O^*(\cdot)$ es la notación para la complejidad de un algoritmo dejando implícitos cualquier término logarítmico en n o que dependa de otros parámetros como ϵ .

Lo que queda por determinar es cómo realizaremos el muestreo. En general, el problema de muestreo toma como input una función $f : \mathbf{R}^n \rightarrow \mathbf{R}$ dada por un oráculo, un punto x con $f(x) > 0$ y un parámetro $\epsilon > 0$ que representa el error, y entrega un punto y que proviene de una distribución que se encuentra a una distancia ϵ de la distribución dada por f . El problema de muestreo es importante ya que si tenemos un algoritmo para muestrear distribuciones dadas por densidades log-cóncavas, podemos utilizar este algoritmo para redondear, optimizar, calcular volumen o integrar y para algunos problemas de aprendizaje.

2.1. Redondeo por muestreo

En el caso de redondeo por muestreo, utilizando una caja negra la cual muestrea puntos de un conjunto eficientemente, podemos redondear un conjunto convexo K de la siguiente manera:

1. Muestrear m puntos de K .
2. Calcular el promedio empírico $z = \mathbb{E}(X)$ y la matriz de covarianza empírica $A = \mathbb{E}((x - z)(x - z)^t)$.
3. Entregar $B = A^{-1/2}$.

Luego, $B(K - z)$ será aproximadamente isotrópico y el algoritmo será eficiente, en el siguiente sentido:

Teorema 2.3 (Adamczak, Litvak, Pajor, y Tomczak-Jaegermann (2010)). *Muestrear $C(\epsilon)n$ puntos aleatorios de una distribución isotrópica son suficientes para obtener un redondeo donde la matriz de covarianza \tilde{A} del conjunto redondeado cumple que $\mathbb{E}(\|\tilde{A} - I\|_2) \leq \epsilon$. Equivalentemente, para cualquier vector unitario v ,*

$$1 - \epsilon \leq \mathbb{E}((v^t x)^2) \leq 1 + \epsilon.$$

2.2. Complejidad del Muestreo

Hasta ahora habíamos dejado pendiente el problema de como realizar el muestreo. Aunque las técnicas específicas se explicarán con detalle más adelante, por ahora se pueden enunciar los siguientes teoremas.

Teorema 2.4 (Kannan, Lovász, y Simonovits (1997)). *En un conjunto convexo, el paseo aleatorio en la bola con una partida M -warm alcanza un punto (independiente) casi aleatorio en una cantidad polinomial en n , R y m de pasos. M debe ser igual a $\sup Q_0(S)/Q(S)$ o $\mathbb{E}_{Q_0}(Q_0(x)/Q(x))$, con Q_0 la ley inicial del paseo aleatorio.*

Teorema 2.5 (Lovász y Vempala (2007)). *El mismo resultado se cumple para densidades log-cóncavas, con complejidad $O^*(M^2 n^2 R^2)$.*

Observamos que bajo una transformación isotrópica, $R = O(\sqrt{n})$ y que M puede mantenerse constante. En total, el la complejidad del algoritmo de volumen puede ser expresada como $O^*(n^5)$ ya que son n fases, n muestras y n^3 llamados al oráculo por cada muestra (Kannan y cols. (1997)).

Cuadro 1: Avances en el cálculo de la complejidad del problema del cálculo de volumen.

Referencia	Complejidad	Nuevo Ingrediente
Dyer y cols. (1989)	n^{23}	Todo
Lovász y Simonovits (1990)	n^{16}	Lema de Localización
Applegate y Kannan (1991)	n^{10}	Muestreo log-cóncavo
Lovász (1990)	n^{10}	Paseo en Bola
Dyer y Frieze (1991)	n^8	Análisis de Error
Lovász y Simonovits (1993)	n^7	Varias mejoras
Kannan y cols. (1997)	n^5	Isotropía y paseo acelerado
Lovász y Vempala (2003)	n^4	Recocido Simulado
Lovász y Vempala (2006b)	n^4	Integración y análisis local
Cousins y Vempala (2014)	n^3	Enfriamiento Gaussiano

2.3. Recocido Simulado

Para estimar $\int f$, consideremos una secuencia $f_0, \dots, f_m = f$ con $\int f_0$ fácil de calcular (i.e., una función constante sobre una bola). Siguiendo el mismo esquema del muestreo anterior, queremos expresar la integral mediante el producto de ratios de integrales fáciles de calcular:

$$\int f = \int f_0 \frac{\int f_1}{\int f_0} \cdots \frac{\int f_m}{\int f_{m-1}}.$$

Cada fracción puede ser estimada mediante muestreo: 1. Muestreamos una variable aleatoria X con una densidad proporcional a f_i . 2. Calculamos $Y = f_{i+1}(X)/f_i(X)$. Entonces, $\mathbb{E}(Y) = \int \frac{f_{i+1}(X)}{f_i(X)}$

Debemos escoger funciones f_i para que el muestreo sea óptimo. Una elección posible y con la cual obtendremos buenas garantías (LV'06) es escoger $f_i(X) = \exp(-a_i\|X\|)$ para $X \in K$, y cero en caso contrario, tomando

$$a_0 = 2R, \quad a_{i+1} = \frac{a_i}{\left(1 + \frac{1}{\sqrt{n}}\right)}, \quad a_m = \frac{\epsilon}{2R},$$

de tal forma que la primera función esté muy concentrada dentro de una pequeña bola, mientras que la última función está muy esparcida en el espacio, siendo uniforme (casi constante). El punto interesante es que la tasa a la que cambiamos la función es más rápido a como lo hacíamos con el algoritmo anterior. En total necesitamos solo $m \approx \sqrt{n} \log(2R/\epsilon)$. Como son m fases y m puntos en cada fase, la ganancia de este algoritmo con respecto al anterior es de orden n , es decir, $O^*(n^4)$.

El estimado final podría ser $n^{\Omega(n)}$, es decir, cada ratio podría ser de orden $n^{\sqrt{n}}$ o mayor. Esto es contraintuitivo ya que estamos estimando una diferencia grande con una cantidad constante muestras. Sin embargo, lo importante tenemos una cota para la varianza que varianza y el siguientes asegura iente lema nos da una cota para ella:

Lema 2.6 (Lovász y Vempala (2006b)). $\text{Var}(Y) < 4\mathbb{E}(Y)^2$

Demostración. Sea $f(a_i, x) = \exp(-a_i\|x\|)$. Luego, $\mathbb{E}(Y) = \int f(a_{i+1}, x) / \int f(a_i, x)$ y entonces

$$\begin{aligned} \frac{\mathbb{E}(Y^2)}{\mathbb{E}(Y)^2} &= \int \frac{f(a_{i+1}, x)^2}{f(a_i, x)^2} \frac{f(a_i, x)}{\int f(a_i, x)} dx \cdot \left(\frac{\int f(a_{i+1}, x)}{\int f(a_i, x)} \right)^2 \\ &= \frac{\int f(2a_{i+1} - a_i, x) \int f(a_i, x)}{(\int f(a_{i+1}, x))^2} \\ &= \frac{F(a_{i+1}(1 - \alpha))F(a_{i+1}(1 + \alpha))}{F(a_{i+1})^2}, \end{aligned}$$

donde $F(a) = \int f(a, x)dx$ y $\alpha = 1/\sqrt{n}$. Notemos que si F fuera log-cóncava, tendríamos que el cociente anterior está acotado por 1, lo que no puede ocurrir ya que el segundo momento de una variable aleatoria siempre es más grande que el primer momento al cuadrado. Tenemos que encontrar otra forma de acotarla.

Aunque F no es log-cóncava, la función $Z(a) = a^n \int f(a, x)dx$ sí lo es, y podemos usar esto para acotar la expresión anterior, ya que

$$(a^n F(a))^2 \geq (a(1 - \alpha))^n F(a(1 - \alpha)) (a(1 + \alpha))^n F(a(1 + \alpha))$$

y por lo tanto

$$\begin{aligned} \frac{F(a_{i+1}(1 - \alpha))F(a_{i+1}(1 + \alpha))}{F(a_{i+1})^2} &\leq \left(\frac{1}{(1 - \alpha)(1 + \alpha)} \right)^n \\ &\leq 4. \end{aligned}$$

□

Notemos que la planificación del recocido (la forma en que escogemos la sucesión a_i) podría haber sido escogido de otra manera para obtener otras garantías. En el caso anterior, teníamos $O^*(\sqrt{n})$ fases y $O^*(\sqrt{n})$ muestras por fase, pero podemos pensar en otras formas de planificar el recorrido que nos den otras garantías.

Ejercicio 3: Encuentre una planificación del recocido (cómo definir las variables a_i) para que se requieran $O^*(n)$ fases y $O^*(1)$ muestras por fase.

2.4. Optimización por Recocido

Ya que $f(x) = \exp(-g(x))$ es una función log-cóncava cuando g es convexa, el problema de maximizar una función log-cóncava generaliza el problema de minimizar una función convexa. Esto nos motiva a atacar el problema de la optimización de una función log-cóncava y usaremos el mismo método de recocido modificando la forma en que escogemos la sucesión de funciones.

Consideremos $f_i(x) = f(x)^{a_i}$ con

$$a_0 = \frac{\epsilon}{2R}, \quad a_{i+1} = a_i \left(1 + \frac{1}{\sqrt{n}}\right), \quad a_m = \frac{2n}{\epsilon}.$$

Notemos como a diferencia del recocido que utilizamos para estimar las integrales, ahora empezamos con un valor pequeño de a_i y vamos creciendo. Lo anterior hace que las distribuciones comiencen por ser cercanas a uniformes y se vayan volviendo más concentradas a medida que progresa el algoritmo. Esto es ya que en el caso de optimización, queremos comenzar con un muestreo que abarque toda la función y a medida que avanzamos nos concentramos en un solo punto (máximo aproximado).

3. Muestreo y Probabilidad

3.1. Paseos Aleatorios Geométricos

Ahora sí responderemos la pregunta sobre como realizar el muestreo de todos los algoritmos aleatorizados que hemos hablado hasta ahora. Para ello, consideraremos paseos aleatorios dentro del conjunto K convexo tal que su distribución estacionaria asemeja la distribución de la cual queremos samplear. En particular, estudiaremos dos tipos: el paseo de bola y el paseo Pega-y-Corre.

En el paseo de bola, definimos un δ y cada vez que nos encontramos un punto x , nos movemos al azar a un punto $y \in B^n(x, \delta)$ siempre y cuando este punto esté en K . Si no está en K , seguimos iterando.

En el paseo de Pega-y-Corre, cada vez que nos encontramos en un punto x escogemos una cuerda al azar que pasa por x y de esa cuerda, escogemos un punto y al azar que se encuentre dentro del conjunto. Para encontrar una cuerda aleatoria que pase por x y se encuentre dentro del conjunto, podemos escoger una dirección aleatoria y utilizar búsqueda binaria para determinar los puntos de la cuerda que están dentro del convexo.

Antes de continuar, repasemos un par de conceptos relacionados con las cadenas de Markov. El espacio de estados será el conjunto K y para cada punto u en K , se asocia una ley de probabilidad $P_u(A)$ la cual representa la probabilidad de pasar en un paso de u al conjunto A . Recordemos además la propiedad markoviana que nos dice que la probabilidad de un paso condicional a la historia depende solo del estado actual y no del pasado. Si la cadena de Markov posee una distribución estacionaria Q , podemos definir el flujo ergódico como

$$\Phi(A) = \int_A P_u(K \setminus A) dQ(u).$$

Es claro que este flujo cumple que $\phi(A) = \phi(K \setminus A)$ ya que Q es estacionaria.

Ejercicio 4: Demuestre que la distribución estacionaria de Pega-y-Corre es uniforme en el conjunto convexo.

3.2. Conductancia y Convergencia

La conductancia de un conjunto A está definida por

$$\phi(A) = \frac{\Phi(A)}{\min\{Q(A), Q(K \setminus A)\}}.$$

Podemos pensar en la conductancia de un conjunto como la probabilidad de escape condicional a encontrarse dentro del conjunto. Con esto, la conductancia de la cadena queda definida por

$$\phi = \inf_{A \subseteq K} \phi(A).$$

Ya que toma $1/\phi$ pasos salir de un conjunto, es intuitivo pensar que la tasa de mezcla de la cadena no puede ser mayor a $1/\phi$, por lo que la conductancia impone un límite superior sobre la velocidad de convergencia. Es natural preguntarse si la conductancia puede ser utilizada para encontrar un cota superior. El siguiente resultado sobre cadenas de Markov a espacio discreto da ciertas luces al respecto.

Teorema 3.1 (Sinclair y Jerrum (1989)). *Consideremos una cadena de Markov discreto y sea λ el segundo valor propio más grande de la matriz de transición, entonces*

$$\frac{\phi^2}{2} \leq 1 - \lambda \leq 2\phi^2.$$

Por lo tanto, la tasa de mezcla está acotada por $2/\phi^2$ y, en general, esta cota no puede ser mejorada.

Más formalmente, podemos encontrar expresiones que acotan la distancia de la distribución de nuestro proceso a paso t , Q_t con respecto a la distribución estacionaria Q .

Teorema 3.2 (Lovász y Simonovits (1993)). *Sea Q_0 la distribución inicial de la cadena y*

$$M_1 = \sup_{A \subseteq K} \frac{Q_0(A)}{Q(A)}, \quad M_2 = \mathbb{E}_{Q_0} \left(\frac{Q_0(x)}{Q(x)} \right),$$

entonces

$$d_{tv}(Q_t, Q) \leq \sqrt{M_1} \left(1 - \frac{\phi^2}{2} \right)^t, \quad d_{tv}(Q_t, Q) \leq \varepsilon + \sqrt{\frac{M_2}{\varepsilon}} \left(1 - \frac{\phi^2}{2} \right)^t.$$

Es necesario limitar la conductancia, pero en el caso del paseo en bola, esta puede ser arbitrariamente pequeña a nivel local. La conductancia local $l(x) = \text{vol}(x + \delta B^n \cap K) / \text{vol}(\delta B^n)$.

Tenemos dos alternativas. Podemos suavizar el conjunto K para evitar las esquinas, o podemos comenzar de un punto (casi) aleatorio y esperar no partir de un punto inconveniente. Para suavizar el conjunto consideramos $K' = K + \alpha B^n$ con un $\alpha \geq \delta\sqrt{n}$ que asegure que la conductancia local sea grande.

Ahora que hemos acotado la conductancia local, debemos hacer lo mismo con la conductancia pero a nivel de conjuntos. Queremos que para un $S \subset K$ la probabilidad de salir de S sea grande. Para esto, necesitamos que los puntos que no se crucen estén suficientemente lejos y que si dos subconjuntos están lejos, entonces el resto del conjunto es grande. La idea es que si las distribuciones de dos puntos u y v están alejadas, las bolas alrededor de estos puntos tendrán poca intersección lo que implica que u y v están alejados.

Lema 3.3. Sean $u, v \in K$ con conductancias locales $l(u), l(v) \geq l$ en el paseo aleatorio en la bola de paso δ . Si

$$d(u, v) \leq \frac{t\delta}{\sqrt{n}}$$

entonces $d(P_u, P_v) \leq 1 + t - l$.

Sea S un subconjunto de K , y consideremos la siguiente partición: definamos por S_3 todos los puntos de S y $K \setminus S$ que se encuentran cerca entre ellos, S_1 el resto de los puntos de S y S_2 el resto de los puntos de $K \setminus S$. Para acotar la conductancia necesitamos que los puntos cercanos tengan distribuciones de un paso parecidas y la condición geométrica extra de que los conjuntos grandes tengan fronteras grandes. A esto último lo llamaremos isoperimetría, sobre lo cual discutiremos más adelante. Por ahora, mencionaremos un par de resultados interesantes que nos serán útiles para acotar la conductancia.

Teorema 3.4 (Lovász y Simonovits (1990), Dyer-Frieze Dyer y Frieze (1991)). Sea S_1, S_2, S_3 una partición de un convexo K con diámetro D . Entonces,

$$\text{vol}(S_3) \geq \frac{2d(S_1, S_2)}{D} \min\{\text{vol}(S_1), \text{vol}(S_2)\}.$$

Se extiende también a distribuciones log-cóncavas π_f :

$$\pi_f(S_3) \geq \frac{2d(S_1, S_2)}{D} \min\{\pi_f(S_1), \pi_f(S_2)\},$$

con D el diámetro del soporte.

Teorema 3.5. La conductancia de la bola es al menos $\frac{l^2\delta}{16\sqrt{n}D}$.

Usando $l = 1/n$ y $\delta = l/\sqrt{n} = 1/n\sqrt{n}$, tenemos que $\phi \geq C/(n^2D)$ y por lo tanto la tasa de mezcla es $O(n^4D^2)$.

Notamos que si S es convexo, entonces el paseo en bola con una partida M -warm alcanza un punto (independiente) casi aleatorio en tiempo polinomial en n, D y M .

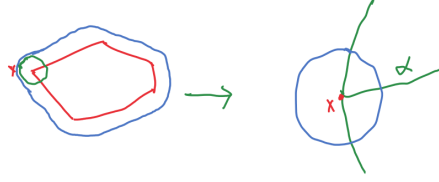


Figura 1: Suavizar el conjunto con el paseo en bola.

En realidad, esto no es una mezcla *rápida*. Aún queda por determinar como obtenemos el primer punto y si es posible obtener una mejor dependencia del parámetro D . Ahora es cuando es importante mencionar que aunque el paseo en bola puede tener malas partidas, esto no ocurre con Pega-y-Corre: incluso si el punto está en una esquina, al moverse a un punto aleatorio dado por una cuerda que siempre estará contenida en K (ya que es convexo), se logra escapar con suficiente rapidez. Para ello, debemos considerar otra noción de distancia: la distancia de ratio-cruzado entre dos puntos está dada por

$$d_K(u, v) = \frac{|u - v||p - q|}{|p - u||v - q|}$$

con p y q los puntos de intersección de la frontera del convexo K con la recta que pasa por u y v .

Con esto, podemos concluir que

$$\pi_f(S_3) \geq d_K(S_1, S_2) \pi_f(S_1) \pi_f(S_2).$$

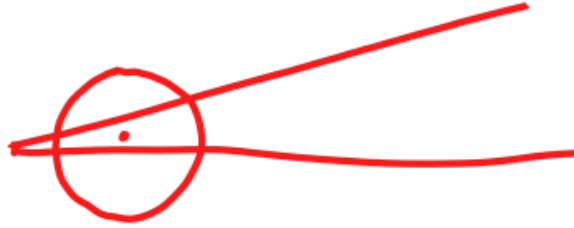


Figura 2: El paseo en bola puede tener malas partidas (esquinas), pero el paseo Pega-y-Corre logra escapar.

Teorema 3.6 (Lovász y Vempala (2006a)). *El paseo Pega-y-Corre mezcla en tiempo polinomial desde cualquier punto de partida en un conjunto convexo. La conductancia es $\Omega(1/nD)$ y bajo una transformación isotrópica, el algoritmo de muestreo da $O^*(n^3)$.*

En el Cuadro 2 resume los avances en proceso de disminuir la complejidad del problema de muestreo a través de algoritmos aleatorizados.



Figura 3: Conductancia: puntos cercanos tienen distribuciones sobrelapadas. Los conjuntos grandes tienen fronteras grandes.

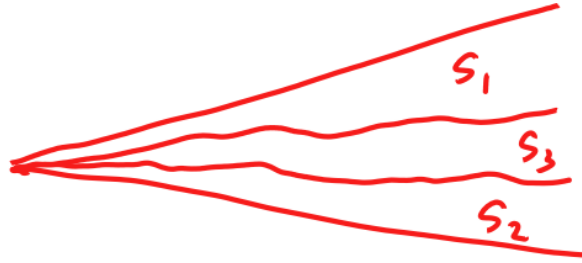


Figura 4: La distancia isoperimétrica es pequeña.

Cuadro 2: Avance de complejidades en el problema de muestreo.

Referencia	Complejidad	Nuevo Ingrediente
Dyer y cols. (1989)	n^{20}	Todo
Lovász y Simonovits (1990)	n^{13}	Lema de Localización
Applegate y Kannan (1991)	n^7	Muestreo log-cóncavo
Lovász (1990)	n^7	Paseo en Bola
Lovász y Simonovits (1993)	n^5	Varias mejoras
Kannan y cols. (1997)	n^3	Isotropía y paseo acelerado
Lovász y Vempala (2006a)	n^3	Paseo Pega-y-Corre

Se conjetura que la cota inferior para el problema de muestreo es n^2 .

4. Geometría

4.1. Isoperimetría

Dado un conjunto K se define la isoperimetría (o constante de Cheeger) por

$$\psi_K = \left[\inf_{S \subseteq K} \frac{\text{vol}^{n-1}(\partial S)}{\min\{\text{vol}^n(S), \text{vol}^n(K \setminus S)\}} \right]^{-1}.$$

Notemos la similitud entre este concepto y la conductancia. Sin embargo, la isoperimetría es una propiedad netamente geométrica. Aún así, podemos relacionarla con los paseos aleatorios en \mathbf{R}^n .

Teorema 4.1. *El tiempo de mezcla del paseo de bola es $O^*(n^2\psi_K^2)$.*

El problema de la isoperimetría es que incluso para conjuntos convexos, esta puede ser grande. Por ejemplo, un cilindro de sección transversal 1 y largo L tiene isoperimetría $L/2$, valor que puede ser arbitrariamente grande (ver Figura 5). Sin embargo, si aplicamos una transformación isotrópica al cuerpo es de esperar que podamos controlar la isoperimetría. La siguiente conjetura refleja justamente esta intuición:

Conjetura 1 (Kannan-Lovász-Simonovits). *La isoperimetría de cualquier convexo isotrópico K es $\psi_K = O(1)$.*

Notemos que si fuera verdad que la isoperimetría es constante, entonces el paseo en bola sería $O^*(n^2)$.

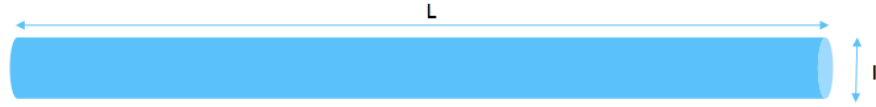


Figura 5: La isoperimetría de un cilindro puede ser arbitrariamente grande.

La siguiente desigualdad es corolario del Teorema 3.4 cuando π es una distribución que proviene de una densidad log-cóncava:

$$\pi(\partial S) \geq \frac{2}{D} \min\{\pi(S), \pi(K \setminus S)\}$$

lo que es equivalente a

$$\psi_K \leq \frac{D}{2}.$$

El problema del diámetro (o del soporte) es que puede ser muy grande. En particular, si el objeto es un cono, el diámetro crece con la dimensión de forma lineal! Si

$$A = \mathbb{E}_\pi((x - \bar{x})(x - \bar{x})^t)$$

es la matriz de covarianza para la distribución π , consideramos R tal que

$$R^2 = \mathbb{E}_\pi(\|x - \bar{x}\|^2) = \text{Tr}(A) = \sum_i \lambda_i(A).$$

Entonces, el siguiente resultado afina los resultados anteriores.

Teorema 4.2 (Kannan y cols. (1995)). *Sea π una distribución log-cóncava. Luego,*

$$\pi(S_3) \geq \frac{c}{R} \min\{\pi(S_2), \pi(S_3)\}.$$

Para distribuciones isotrópicas se tiene que $\psi = O(R) = O(\sqrt{n})$, ya que sus valores propios son todos iguales a uno.

La conjetura KLS (Conjetura 1) se puede entonces reexpresar como

Conjetura 2 (KLS). *Para una distribución isotrópica, $\psi = O(\sqrt{\lambda_1(A)}) = O(1)$.*

A continuación describiremos algunas conjeturas relacionadas con la conjetura KLS y para las cuales una respuesta afirmativa de la conjetura KLS implicaría que son ciertas.

Conjetura 3 (Hiperplano KLS). *La mínima expansión/isoperimetría está dado por un hiperplano hasta un valor constante.*

Conjetura 4 (Cascarón Delgado). *Si X una v.a. con densidad log-cóncava, entonces*

$$\text{Var}(\|X\|^2) = O(n), \quad \mathbb{E}((\|X\|^2 - \sqrt{n})^2) = O(1).$$

En otras palabras, casi toda la masa de una distribución isotrópica está contenida en un cascarón de grosor constante.

La mejor cota hasta ahora es de $O(n^{1/3})$, debido a Guedon y Milman.

La conjetura anterior se puede verificar para una distribución Gaussiana.

Conjetura 5 (Slicing o Hiperplano). *Todo convexo de volumen unitario tiene una sección cuyo volumen (en una dimensión menos) es constante (independiente de la dimensión).*

Equivalentemente, para toda densidad log-cóncava isotrópica f , $\max f^{1/n} = O(1)$.

La mejor cota hasta ahora es de $O(n^{1/4})$, debido a Guedon y Milman.

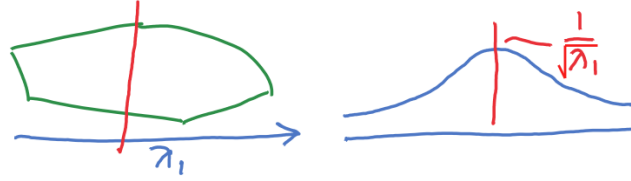


Figura 6: El slice o hiperplano que se forma es log-cóncavo.

Conjetura 6. *Sea f una densidad log-cóncava, Lipschitz con $\mathbb{E}(f) = 0$,*

$$\mathbb{P}(f(x) > t) = \exp(-\Omega(t/\psi)).$$

Para KLS, la mejor cota que se conoce es $n^{1/3} \sqrt{\log(n)}$. Tanto KLS, como Cascarón Delgado e Hiperplano se conjeturan $O(1)$ para distribución log-cóncavas. Aunque aún no se tiene un resultado definitivo, hay acercamientos promisorios.

Teorema 4.3 (Cousins y Vempala (2013)). *Las conjeturas anteriores son ciertas cuando la densidad es proporcional al producto de una función log-cóncava y una función Gaussiana estándar.*

El último resultado que enunciaremos en esta sección es uno muy reciente.

Teorema 4.4 (Lee y Vempala (2016a)). *Para una densidad log-cóncava, $\psi = O(n^{1/4})$. Esto implica que todas las constantes tienen el mismo límite.*

Otra forma de enunciar lo anterior es decir que

$$\psi \leq C \left(\sum_i \lambda_i(A)^2 \right)^{\frac{1}{4}} = C(\text{Tr}(A^2))^{\frac{1}{4}}.$$

4.2. Localización

Para probar la desigualdad isoperimétrica desarrollaremos una técnica llamada *localización*, la cual consiste en reducir un problema en el que está involucrada una integral en n dimensiones a uno de una sola dimensión.

Recordemos que la desigualdad isoperimétrica nos dice que para una partición de un convexo en tres conjuntos, se tiene que

$$\pi_f(S_3) \geq \frac{2d(S_1, S_2)}{D} \min\{\pi_f(S_1), \pi_f(S_2)\}$$

con π_f es una distribución de una densidad f log-cóncava y D su soporte. Sin pérdida de generalidad, podemos escribir la desigualdad anterior como dos desigualdades:

$$\pi_f(S_1) \leq \pi_f(S_2) \text{ y } \pi_f(S_3) \geq \psi \pi_f(S_1).$$

Si definimos funciones g y h tales que

$$g(x) = f(x)(\mathbb{1}_{S_2}(x) - \mathbb{1}_{S_1}(x)) \text{ y } h(x) = f(x)(\psi \mathbb{1}_{S_1}(x) - \mathbb{1}_{S_3}(x)),$$

entonces nos basta probar que

$$\int g \geq 0 \Rightarrow \int h \leq 0.$$

Procedamos por contradicción y supongamos que existen S_1, S_2, S_3 que cumplen que $\int g \geq 0$ y $\int h > 0$. El esquema de la demostración será en dos pasos: primero probaremos que no existe un contraejemplo en una dimensión, y luego probaremos que si existiera un contraejemplo en alguna dimensión, entonces también existiría en una dimensión.

Lema 4.5 (Lema de Localización (Kannan y cols. (1995))). Sean $g_i : \mathbf{R}^n \rightarrow \mathbf{R}$, $i = 1, 2$, funciones integrables semicontinuas inferiores. Supongamos que $\int g, \int h > 0$. Entonces existe un intervalo $[a, b] \in \mathbf{R}^n$ y una función lineal $l : [a, b] \rightarrow \mathbf{R}_+$ tal que

$$\int_0^1 g_i((1-t)a + tb)l(t)^{n-1} dt > 0, \forall i = 1, 2.$$

Para continuar, definimos $Z_i = \{t \in [0, 1] : (1-t)a + tb \in S_i\}$ y $F(t) = f((1-t)a + tb)l(t)^{n-1}$. Como F es log-cóncavo, podemos plicar el lema de localización a las funciones g y h implicando que

$$\int_{Z_1} F \leq \int_{Z_2} F, \text{ pero } \int_{Z_3} F < \psi \int_{Z_1} F,$$

y por lo tanto, si hay un contraejemplo en alguna dimensión, debe haber uno en una dimensión.

Nos reducimos entonces a probar la misma desigualdad de isoperimetría pero para $n = 1$:

$$\int_{S_3} F \geq \frac{2d(S_1, S_2)}{D} \min \left\{ \int_{S_1} F, \int_{S_2} F \right\}.$$

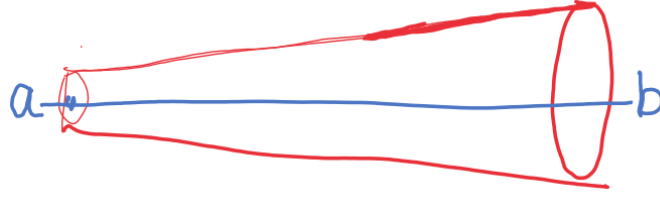


Figura 7: El problema se puede reducir a uno de una dimensión en un intervalo $[a, b]$. Integramos a través del cono.

Se puede probar que basta verificar lo anterior solo para 3 intervalos y que esto se puede extender a cualquier partición de conjuntos. La desigualdad anterior (sin el factor 2) sigue por unimodalidad. Sigue que la misma desigualdad isoperimétrica se tiene en \mathbb{R}^n .

La técnica de localización tiene otras aplicaciones fuera de la isoperimetría. Por ejemplo, ha sido usada en cotas de concentración de polinomios (Cranberry-Wright), o para isoperimetría con otras métricas (ratio-cruzado, distancia de Dikin, etc), así también para el análisis de algoritmos en general.

Para tener una mejor idea de como funciona el lema de localización, podemos hacer un bosquejo de la demostración. Primero encontramos un semiespacio que bisecta alguna de las funciones (se puede garantizar su existencia por el teorema del valor medio). Segundo, probamos que el soporte del límite de las bisecciones es un intervalo o un punto. Finalizamos notando que la función límite tiene un perfil cóncavo, y reducimos a un perfil lineal.

Es necesario aplicar la bisección hasta que el soporte contenga un plano en dos dimensiones con punto interior racional. La secuencia de soportes es convexa y decreciente, por lo que la función límite es un punto o intervalo.

Para comprobar que el perfil es de la forma que queremos, considere la función de radio de la sección transversal. Reemplazamos cada sección transversal por una bola del mismo volumen. La desigualdad de Brunn-Minkowski implica que esta función es cóncava. Cada una de estas funciones está acotada y su límite es una función cóncava. Concluir linealidad toma un poco más de trabajo.

4.3. Enfriamiento Gaussiano

Por el Teorema 4.3 sabemos que la conjetura KLS se cumple para funciones Gaussianas restringidas a conjuntos convexos (usando la desigualdad de Brascamp-Lieb). Con lo anterior, podemos establecer el siguiente resultado:

Teorema 4.6 (Cousins y Vempala (2013)). *Sea g una función log-cóncava y*

$$f(x) = g(x)e^{-\frac{\|x\|^2}{2\sigma^2}}$$

entonces

$$\pi(S_3) \geq \frac{c}{\sigma} d(S_1, S_2) \min\{\pi(S_1), \pi(S_2)\}$$

con π una distribución dada por una densidad proporcional a f .

Esto nos permite obtener mejores cotas para el muestreo.

Teorema 4.7 (Cousins y Vempala (2013)). *El paseo de bola aplicado a una Gaussiana $\mathcal{N}(0, \sigma^2 I)$, restringido a un convexo, se mezcla en $O^*(n^2 \max\{\sigma^2, 1\})$ pasos desde una partida warm.*

Para aplicar lo anterior en el problema de muestreo, aplicaremos una técnica llamada *enfriamiento Gaussiano*, que no es más que un recocido simulado adaptativo. Comenzamos con funciones $f_i(x) = \exp(-\|x\|^2/(2\sigma_i^2))$, donde $\sigma_0^2 = 1/n$ y $\sigma_m^2 = O(n)$.

Al igual que en el método de recocido original, estimamos $\int f_{i+1}/\int f_i$ con muestras de acuerdo a f_i , pero la planificación será distinta y dependerá del valor de σ_i . Si $\sigma_i^2 \leq 1$, entonces utilizamos la planificación habitual de aumentar la varianza por un factor $(1 + 1/\sqrt{n})$. Si $\sigma_i^2 > 1$, entonces cambiamos la planificación por un factor $(1 + \sigma_i/\sqrt{n})$. Como el valor de σ_i va aumentando a través de los pasos, las últimas etapas del enfriamiento gaussiano ocurren más rápidamente. Este proceso continúa hasta que la varianza es lineal en n , donde se puede saltar a la distribución uniforme en el último paso.

Analizaremos el algoritmo por parte. En los pasos en los que $\sigma_i \leq 1$, el tiempo de muestreo es de orden n^2 , mientras que el número de fases y la cantidad de muestreos por fases es de orden \sqrt{n} , dando un total de n^3 . Ahora, cuando $\sigma_i > 1$, el análisis debe ser más fino. Aunque toma un tiempo $\sigma^2 n^2$, se puede compensar ya que la cantidad de fases que se requiere para doblar σ es \sqrt{n}/σ , y la cantidad de muestras también. Luego, el tiempo total también es n^3 .

Resta probar que es posible acelerar el recocido en los últimos pasos. Para ello, el siguiente lema nos entrega una cota sobre la varianza del estimador del ratio.

Lema 4.8. *Sea*

$$Y = \frac{f(\sigma^2, x)}{f\left(\frac{\sigma^2}{1+\alpha}, x\right)}$$

donde $f(\sigma^2, x) = \exp(-\|x\|^2/(2\sigma^2)) \mathbb{1}_{x \in K}$ y $\alpha = O(\sigma/\sqrt{n})$. Entonces,

$$\frac{\mathbb{E}(Y^2)}{\mathbb{E}(Y)^2} = O(1).$$

La demostración utiliza el lema de localización para reducirse a una desigualdad en una dimensión.

Cabe la pregunta de si no se puede acelerar el proceso de recocido aún más para tener una mejor garantía. La respuesta es negativa: $1 + \sigma/\sqrt{n}$ es la mejor tasa posible.

Un cuerpo se dice "bien redondeado" si contiene una bola de radio uno y está contenido en una bola de radio $O^*(\sqrt{n})$. Cerramos la sección presentando el teorema que resume el resultado anterior.

Teorema 4.9 (Cousins y Vempala (2015)). *El volumen de un cuerpo convexo bien redondeado puede ser estimado usando $O^*(n^3)$ llamadas al oráculo de membresía.*

La utilidad del recocido adaptativo no es solo para el problema de cálculo de volumen. Para una gran cantidad de problemas discretos, permite reducir complejidad de conteo a muestreo (Stefankovic, Vempala, y Vigoda (2007)).

5. Problemas Abiertos

Los siguientes son preguntas que, a la fecha de la publicación de este apunte, aún están sin respuesta.

Geometría

1. Probar o desmentir la conjetura KLS.

Probabilidades

2. Sabemos que el paseo en bola se puede quedar atrapado en malas esquinas, pero ¿qué ocurre que si comenzamos de un punto conveniente (por ejemplo, el centroide)?
3. Una vez que comenzamos el muestreo, nos gustaría tener reglas estadísticas sobre cuando parar para asegurar tener un punto casi aleatorio o estar cerca de la probabilidad estacionaria. ¿Es suficiente verificar que las medidas de todos los semiespacios han convergido?

Algoritmos y Paseos

4. ¿Cómo mejorar la velocidad de redondeo de $O^*(n^4)$?
5. Probar otros tipos de paseos. Por ejemplo, Pega-y-Corre por Coordenadas. Este paseo es más eficiente que el Pega-y-Corre normal por un factor de n , ya que solo hay que mirar una coordenada. Otro ejemplo es el Paseo Geodesico Lee y Vempala (2016b).
6. ¿Qué tan bien podemos aproximar un polítopo P dado que solo tenemos muestras aproximadas?
7. Si tenemos un polítopo explícito $Ax \leq b$, ¿podemos estimar su volumen en tiempo polinomial pero sin aleatoriedad?

Referencias

- Adamczak, R., Litvak, A., Pajor, A., y Tomczak-Jaegermann, N. (2010). Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. *J. Amer. Math. Soc.*, 23, 535–561.
- Applegate, D., y Kannan, R. (1991). Sampling and integration of near log-concave functions. En *Stoc* (pp. 156–163).
- Cousins, B., y Vempala, S. (2013). Volume computation of convex bodies. *MATLAB File Exchange*. (<http://www.mathworks.com/matlabcentral/fileexchange/43596-volume-computation-of-convex-bodies>)
- Cousins, B., y Vempala, S. (2014). A cubic algorithm for computing Gaussian volume. En *Soda* (pp. 1215–1228).
- Cousins, B., y Vempala, S. (2015). Bypassing KLS: gaussian cooling and an $\tilde{O}(n^3)$ volume algorithm. En *Proceedings of the forty-seventh annual ACM on symposium on theory of computing, STOC 2015, portland, or, usa, june 14-17, 2015* (pp. 539–548). Descargado de <http://doi.acm.org/10.1145/2746539.2746563> doi: 10.1145/2746539.2746563
- Dadush, D., y Vempala, S. (2012). Deterministic construction of an approximate m-ellipsoid and its application to derandomizing lattice algorithms. En *Soda*.
- Dyer, M. E., y Frieze, A. M. (1988). On the complexity of computing the volume of a polyhedron. *SIAM J. Comput.*, 17(5), 967–974. doi: 10.1137/0217060
- Dyer, M. E., y Frieze, A. M. (1991). Computing the volume of a convex body: a case where randomness provably helps. En *Proc. of ams symposium on probabilistic combinatorics and its applications* (pp. 123–170).
- Dyer, M. E., Frieze, A. M., y Kannan, R. (1989). A random polynomial time algorithm for approximating the volume of convex bodies. En *Stoc* (pp. 375–381).
- Elekes, G. (1986). A geometric inequality and the complexity of computing volume. *Discrete & Computational Geometry*, 289–292.
- Furedi, Z., y Barany, I. (1986). Computing the volume is difficult. En *Stoc '86: Proceedings of the eighteenth annual acm symposium on theory of computing* (pp. 442–447). New York, NY, USA: ACM. doi: 10.1145/12130.12176
- Kannan, R., Lovász, L., y Simonovits, M. (1995). Isoperimetric problems for convex bodies and a localization lemma. *Discrete & Computational Geometry*, 13, 541–559.
- Kannan, R., Lovász, L., y Simonovits, M. (1997). Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Structures and Algorithms*, 11, 1–50.
- Lee, Y. T., y Vempala, S. S. (2016a). Eldan's stochastic localization and the KLS hyperplane conjecture: An improved lower bound for expansion. *CoRR*, abs/1612.01507. Descargado de <http://arxiv.org/abs/1612.01507>
- Lee, Y. T., y Vempala, S. S. (2016b). Geodesic walks on polytopes. *CoRR*, abs/1606.04696. Descargado de <http://arxiv.org/abs/1606.04696>
- Lovász, L. (1990). How to compute the volume? *Jber. d. Dt. Math.-Verein, Jubiläumstagung 1990*, 138–151.
- Lovász, L., y Simonovits, M. (1990). Mixing rate of Markov chains, an isoperimetric inequality, and computing the volume. En *Rocs* (pp. 482–491).
- Lovász, L., y Simonovits, M. (1993). Random walks in a convex body and an improved volume algorithm. En *Random structures and alg.* (Vol. 4, pp. 359–412).

- Lovász, L., y Vempala, S. (2003). Simulated annealing in convex bodies and an $O^*(n^4)$ volume algorithm. En *Focs* (pp. 650–659). doi: 10.1109/SFCS.2003.1238237
- Lovász, L., y Vempala, S. (2006a). Hit-and-run from a corner. *SIAM J. Computing*, 35, 985–1005.
- Lovász, L., y Vempala, S. (2006b). Simulated annealing in convex bodies and an $O^*(n^4)$ volume algorithm. *J. Comput. Syst. Sci.*, 72(2), 392–417.
- Lovász, L., y Vempala, S. (2007). The geometry of logconcave functions and sampling algorithms. *Random Structures and Algorithms*, 30(3), 307–358.
- Sinclair, A., y Jerrum, M. (1989). Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation*, 82, 93–133.
- Stefankovic, D., Vempala, S., y Vigoda, E. (2007). Adaptive simulated annealing: A near-optimal connection between sampling and counting. En *Focs* (pp. 183–193). Washington, DC, USA: IEEE Computer Society. doi: 10.1109/FOCS.2007.8