Conteo Aproximado

El Problema Fundamental de Conteo

Pregunta

Dado un conjunto S definido por propiedades combinatorias complejas, ¿cuál es su cardinalidad, $|S|$?

Ejemplos de Problemas de Conteo Difíciles (#P-hard)

- ¿Cuántas configuraciones "hard-core" factibles existen en un grafo?
- ¿Cuántas q-coloraciones válidas existen para un grafo G?
- ¿De cuántas maneras se puede embaldosar un tablero de ajedrez con dominós?

Freddy Hernández-Romero, UNAL-Bogotá

El Problema Fundamental de Conteo

Freddy Hernández-Romero, UNAL-Bogotá

Dificultad

La enumeración exhaustiva es computacionalmente inviable. Por ejemplo, en el problema de las q-coloraciones, para $q = 5$ colores y $k = 50$ vértices, hay $5^{50} \approx 10^{34}$ configuraciones totales que revisar.

Tipos de Algoritmos de Aproximación

Esquema de Aproximación en Tiempo Polinomial (PTAS)

Un algoritmo que, para cualquier $\epsilon > 0$, encuentra una respuesta entre $(1-\epsilon)N$ y $(1+\epsilon)N$ (donde N es la respuesta real) en un tiempo polinomial en el tamaño del problema k.

Esquema Aleatorizado de Aproximación en Tiempo Polinomial (FPRAS)

Un **FPRAS** (Fully Polynomial Randomized Approximation Scheme) es un algoritmo que:

1. Con alta probabilidad (e.g., $\geq 3/4$), da una respuesta en el rango $[(1-\epsilon)N, (1+\epsilon)N]$.
2. Dado $\epsilon > 0$, su tiempo de ejecución es polinomial en el tamaño del problema k.

Freddy Hernández-Romero, UNAL-Bogotá

Idea 1: Conteo con Muestreo de Monte Carlo

Un Algoritmo Ingenuo

Para contar el número de q-colorings válidos, $Z_{G,q}$:

1. Consideremos el espacio de **todas** las q^k configuraciones posibles (válidas o no).
2. La probabilidad de que una configuración elegida al azar sea un q-coloring válido es $p = \frac{Z_{G,q}}{q^k}$.
3. Podemos estimar p generando n configuraciones al azar y contando la proporción $\hat{p}$ que son válidas.
4. Nuestra estimación de $Z_{G,q}$ sería $\hat{Z}_{G,q} = \hat{p} \cdot q^k$.

Freddy Hernández-Romero, UNAL-Bogotá

Idea 1: Conteo con Muestreo de Monte Carlo

¿Cuál es el problema con esta idea?

La probabilidad p es **exponencialmente pequeña**.

Para un grafo con k vértices, se puede mostrar que

$$p \leq \left(\frac{q-1}{q}\right)^{k-1}.$$

Para obtener una estimación precisa, necesitaríamos un número de muestras n exponencialmente grande, lo que hace que el método sea inútil.

Freddy Hernández-Romero, UNAL-Bogotá

Anotaciones

k vertices y q colores

$$Z_{G,q} \leq q(q-1)(q-1)\ldots(q-1) = q(q-1)^{k-1}$$

$\{Y_i\}_{i=1}^{\infty}$ muestra iid de configuraciones (no necesariamente q-coloraciones) al azar

$$\mathbb{P}\left(\text{en } \{Y_j\}_{i=1}^{n} \text{ haya al menos una } q\text{-coloración}\right)$$

Freddy Hernández-Romero, UNAL-Bogotá

Anotaciones

$$\leq \sum_{i=1}^{n} \mathbb{P}(Y_i \text{ sea una } q\text{-coloración})$$

$$\leq \sum_{i=1}^{n} \frac{Z_{6,q}}{q^k} \leq \sum_{i=1}^{n} \frac{q(q-1)^{k-1}}{q^k}$$

$$= n\left(\frac{q-1}{q}\right)^{k-1}$$

Si queremos que

$$\frac{1}{2} \leq \mathbb{P}(\text{en } \{Y_i\}_{i=1}^{n} \text{ haya al menos una } q\text{-coloración})$$

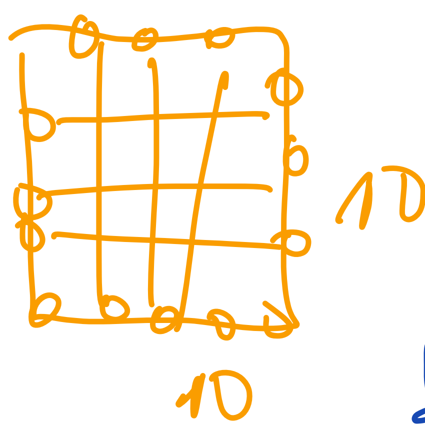
Freddy Hernández-Romero, UNAL-Bogotá

Anotaciones

$$\Rightarrow \quad \frac{1}{2} \leq n \left(\frac{q-1}{q}\right)^{k-1}$$

$$\Rightarrow \quad \frac{1}{2} \left(\frac{q}{q-1}\right)^{k-1} \leq n$$

Ejemplo:



3 colores

$$1 \times 10^{17} \approx \frac{1}{2}\left(\frac{3}{2}\right)^{99} \leq n$$

Idea 2: La Solución MCMC (Producto Telescópico)

Descripción General del Algoritmo

La idea clave es no estimar $Z_{G,q}$ directamente, sino descomponerlo en un producto de ratios más fáciles de estimar.

1. Se construye una secuencia de grafos $G_0, G_1, \ldots, G_l = G$ añadiendo una arista a la vez. G_0 es el grafo sin aristas.
2. Sea Z_i el número de q-colorings válidos para el grafo G_i.
3. Escribimos $Z_{G,q} = Z_l$ como un producto telescópico:

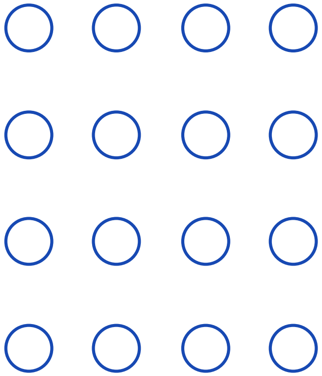
$\rho_{G,q}$: dist uniforme

$$Z_l = \frac{Z_l}{Z_{l-1}} \cdot \frac{Z_{l-1}}{Z_{l-2}} \cdots \frac{Z_1}{Z_0} \cdot Z_0 \rightarrow q^k$$

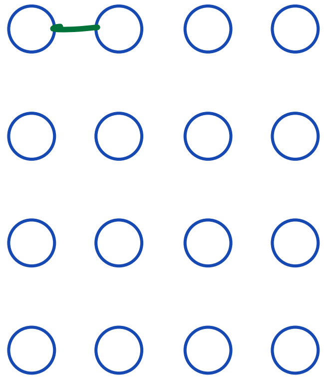
4. Sabemos que $Z_0 = q^k$ (sin aristas, cualquier coloreado es válido).
5. Cada ratio $\frac{Z_i}{Z_{i-1}}$ se puede interpretar como la probabilidad de que un coloreado válido para G_{i-1} también sea válido para G_i.

$$\frac{Z_i}{Z_{i-1}} = P_{\rho_{G_{i-1},q}}(\text{un coloreado es válido para } G_i)$$

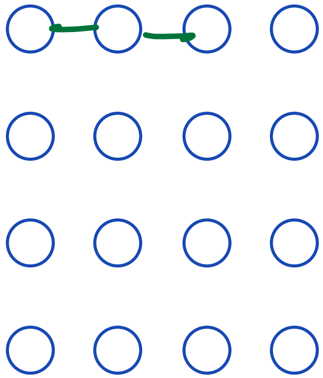
Freddy Hernández-Romero, UNAL-Bogotá



G_0

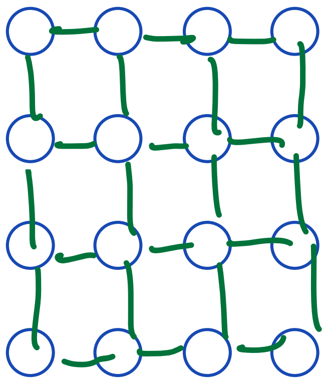


G_1



G_2

. . .



G_{24}

Freddy Hernández-Romero, UNAL-Bogotá

Estimando los Ratios con MCMC

El Algoritmo en la Práctica

1. Para cada $i = 1, \ldots, l$:
   - Queremos estimar el ratio $r_i = \frac{Z_i}{Z_{i-1}}$.
   - Usamos un algoritmo MCMC (como el Muestreador de Gibbs con barrido sistemático) para generar “muchas” muestras $\{X_n^{(i-1)}\}$ de la distribución uniforme de colorings válidos para G_{i-1}.
   - Estimamos r_i con el promedio: $\hat{r}_i = \frac{1}{N} \sum_{n=1}^{N} \mathbb{I}_{\{X_n^{(i-1)} \text{ es válido para } G_i\}}$.
2. La estimación final es $\hat{Z}_{G,q} = (\hat{r}_l \cdot \hat{r}_{l-1} \cdots \hat{r}_1) \cdot \underbrace{q^k}_{r_0}$.

Freddy Hernández-Romero, UNAL-Bogotá

Estimando los Ratios con MCMC

Freddy Hernández-Romero, UNAL-Bogotá

¿Por qué funciona esto?

La condición clave ($q > 2d^2$) asegura dos cosas:

1. **Convergencia rápida de MCMC**: Podemos generar muestras de $\rho_{G_{i-1},q}$ en tiempo polinomial.
2. **Los ratios no son demasiado pequeños**: La probabilidad $\frac{Z_i}{Z_{i-1}}$ no es exponencialmente pequeña, por lo que podemos estimarla con un número polinomial de muestras.

Idea 2: La Solución MCMC (Producto Telescópico)

Teorema 9.1

Para contar q-colorings en grafos con grado máximo d, si $q > 2d^2$, existe un **FPRAS**.

Freddy Hernández-Romero, UNAL-Bogotá

Anotaciones

Primera Parte de la demostración:

$G=(V,E)$ donde

$V=\{v_1,v_2,\dots,v_k\}$ $E=\{e_1,\cdots,e_\ell\}$

Construimos los grafos está fijo

$$G_i := (V, E_i) \quad i=0,1,2,\dots,\ell$$

donde $E_i=\{e_1,\dots,e_i\}$ $E_0=\emptyset$

usaremos las notaciones

$$Z_i := Z_{G_i,q} \quad i=0,\dots,\ell$$

Freddy Hernández-Romero, UNAL-Bogotá

$e_i := \{x_i, y_i\}$

$i = 0, \ldots, l$

finalmente

$$Z_{G,q} = \frac{Z_l}{Z_{l-1}} \cdot \frac{Z_{l-1}}{Z_{l-2}} \cdot \ldots \cdot \frac{Z_1}{Z_0} \cdot Z_0$$

además

$$Z_0 = q^k.$$

Freddy Hernández-Romero, UNAL-Bogotá

Anotaciones

Recordemos que

$\rho_{G,q}$: Probabilidad uniforme sobre q-coloraciones de G;

y observemos que

color en el vértice x_i

color en el vértice y_i

$$\rho_{G_{i-1},q}\left(\{\xi \in \{1,\dots,q\}^V : \xi(x_i) \neq \xi(y_i)\}\right) = \frac{Z_i}{Z_{i-1}}$$

Luego, por la Ley de los grandes números

$$\{X_n^{(i)}\}_{n\geq 0} \text{ iid } \sim \rho_{G_{i-1},q}$$

Freddy Hernández-Romero, UNAL-Bogotá

Anotaciones

$$W_n^{(i)} = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{q\text{-Coloración de } G_i\}}(X_j^{(i)}) \xrightarrow[n\to\infty]{} \frac{Z_i}{Z_{i-1}}$$

$$\frac{1}{n}\sum_{j=1}^{n} Y_j \longrightarrow E[Y]$$

$$Y_j = \mathbb{1}_{\{q\text{-Coloración de } G_i\}}(X_j^{(i)})$$

La idea es usar el "muestreador sistemático de Gibbs" para simular

$$(X_1^{(i)}, X_2^{(i)}, \dots, X_n^{(i)})$$ las muestras

aproximar $\frac{Z_i}{Z_{i-1}}$ y finalmente aproximar $Z_{G,q}$ (con el producto)

Lema 1: dado $\epsilon \in (0,1)$, sea $l \in \mathbb{N}$ y $a_1, \dots, a_l$ y $b_1, \dots, b_l$ números positivos tales que

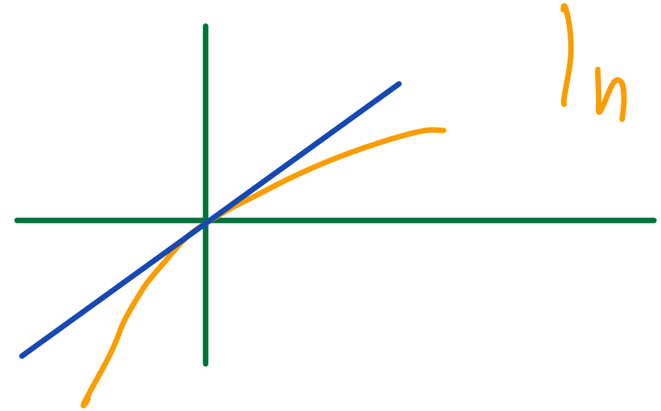
$$\left(1 - \frac{\epsilon}{2l}\right) \leq \frac{a_j}{b_j} \leq \left(1 + \frac{\epsilon}{2l}\right)$$

luego

$$1 - \epsilon \leq \prod_{j=1}^{l} \frac{a_j}{b_j} \leq 1 + \epsilon$$

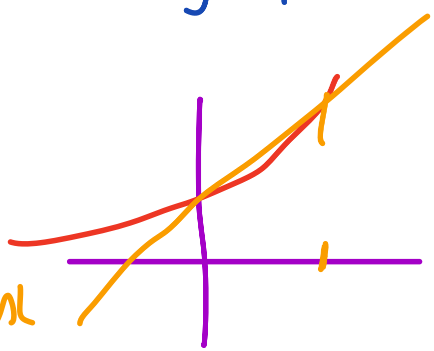
Anotaciones

i) $\dfrac{a_j}{b_j} \le \left(1+\dfrac{\epsilon}{2l}\right) \Rightarrow \ln\left(\dfrac{a_j}{b_j}\right) \le \ln\left(1+\dfrac{\epsilon}{2l}\right) \le \dfrac{\epsilon}{2l}$



$\ln(1+x)$

$$\Rightarrow \sum_{j=1}^{l} \ln\left(\frac{a_j}{b_j}\right) \le \frac{\epsilon}{2} \Rightarrow \prod_{j=1}^{l} \frac{a_j}{b_j} \le e^{\epsilon/2} \le 1+\epsilon$$



$1+2x$

$e^{x} \le 1+2x \qquad 0 \le x \le 1$

Anotaciones

ii) $1 - \frac{\epsilon}{2\ell} \le \frac{a_j}{b_j} \Rightarrow \left(1 - \epsilon/2\ell\right)^{\ell} \le \prod_{j=1}^{\ell} \frac{a_j}{b_j}$

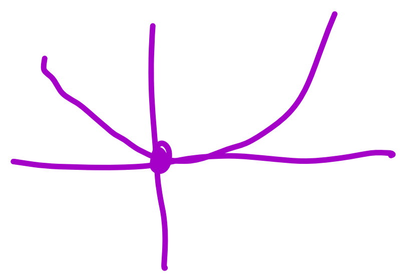
$$\Rightarrow 1 - \epsilon/2 \le 1 - \ell\left(\frac{\epsilon}{2\ell}\right) + \sum_{j=2}^{\ell} \binom{k}{j} (-1)^{\ell-j} \left(\frac{\epsilon}{2\ell}\right)^{j} = \left(1 - \frac{\epsilon}{2\ell}\right)^{\ell}$$

$$f(x) = (1-x)^{\ell} - (1 - \ell x) \geq 0$$

$$f'(x) = -\ell(1-x)^{\ell-1} + \ell$$

$$f''(x) = \ell(\ell-1)(1-x)^{\ell-2} \geq 0 \qquad x \in (0,1)$$

$$\Rightarrow f(0) = 0, \quad f'(0) = 0$$



Anotaciones

Lema 2: $\overset{\text{grado máximo}}{d \geq 2}$, $q > 2d^2$. $x, y \in V$ $x \neq y$

Entonces

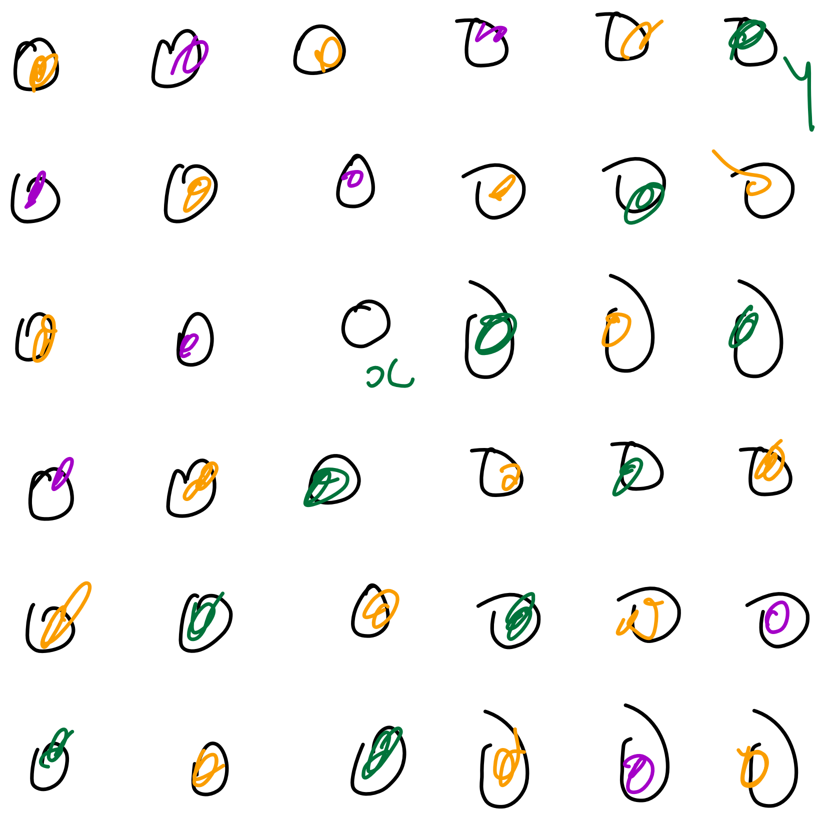
$$\rho_{G,q}\big(\underbrace{\xi(x)}_{\text{color en } x} \neq \underbrace{\xi(y)}_{\text{color en } y}\big) \geq 1/2$$

fijemos $x, y \in V$, $x \neq y$

$$\rho_{G,q}\big(\xi(x) = \xi(y)\big) = \sum_{\eta:\, q\text{-coloración}} \rho_{G,q}\big(\xi(x) = \xi(y), \xi(z) = \eta(z)\ \forall z \neq x\big)$$

$$= \sum_{\eta:\, q\text{-coloración}} \rho_{G,q}\big(\xi(x) = \xi(y) \mid \xi(z) = \eta(z), z \neq x\big)\, \rho_{G,q}\big(\xi(z) = \eta(z), z \neq x\big)$$

$$\rho_{G,q}\left(\{\xi \in \{1,\dots,q\}^V : \xi(x) \neq \xi(y)\}\right) \approx 1/2$$



7-coloración

$$\leq \frac{1}{q-d} \sum_{\eta:\, q\text{-coloración}} P_{\xi_0,q}\left(\xi(z)=\eta(z),\ z\neq x\right)$$

$$\leq \frac{1}{q-d}$$

$$P_{\xi_0,q}\left(\xi(x)\neq\xi(y)\right) = 1 - P_{\xi_0,q}\left(\xi(x)=\xi(y)\right)$$

$$q > 2d^2 > 2d \quad \geq 1 - \frac{1}{q-d} \geq 1 - \frac{1}{2d-d}$$

$$d \geq 2 \qquad = 1 - \frac{1}{d} \geq 1 - \frac{1}{2} = \frac{1}{2}$$

Lema 3: Sea $p \in [0,1]$ y $n \in \mathbb{N}$.

Se lanzan n monedas i.i.d con probabilidad p de salir cara. Entonces el número de caras obtenidas H satisface

$$\mathbb{P}(|H - np| \geq a) \leq \frac{n}{4a^2}$$

Freddy Hernández-Romero, UNAL-Bogotá

Anotaciones

$\{X_k\}_{k\geq 1}$ iid $\sim$ Bernoulli(p)

$H = \sum_{k=1}^{n} X_k \qquad E[X] = p \qquad E[X^2] = p \Rightarrow Var[X] = p - p^2$

$= p(1-p)$

$\mathbb{P}(|H - np| \geq a) \leq \frac{1}{a^2} Var[H]$

Chebyshev

$= \frac{n}{a^2} Var[X] = n\,\frac{p(1-p)}{a^2} \leq \frac{n}{4a^2}$

Freddy Hernández-Romero, UNAL-Bogotá

Demostración (del Teorema)

$$W_n^{(j)} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{q\text{-Coloración de } G_j\}}(X_i^{(j)})$$

en vista del Lema 1 basta probar que

$$\left(1 - \frac{\epsilon}{2l}\right)\frac{Z_j}{Z_{j-1}} \leq W_n^{(j)} \leq \left(1 + \frac{\epsilon}{2l}\right)\frac{Z_j}{Z_{j-1}}$$

para $j = 1, 2, \ldots, l$ o equivalentemente

$$\left| W_n^{(j)} - \frac{Z_j}{Z_{j-1}} \right| \leq \frac{\epsilon}{2l}\frac{Z_j}{Z_{j-1}}$$

Freddy Hernández-Romero, UNAL-Bogotá

Anotaciones

Como $\frac{z_j}{z_{j-1}} = \rho_{G_{j-1},f}(\xi(x_j) \neq \xi(y_j))$

el Lema 2 implica que $\frac{z_j}{z_{j-1}} \geq 1/2$

y por lo tanto basta probar

$$\left| W_n^{(j)} - \frac{z_j}{z_{j-1}} \right| \leq \frac{\epsilon}{4\ell}$$

$j = 1, 2, \ldots, \ell.$

Freddy Hernández-Romero, UNAL-Bogotá

Anotaciones

Tenemos que controlar 2 fuentes de error de aprox

i) El "Gibbs Sampler" $\quad X_n \xrightarrow[n\to\infty]{dist} \rho_{G,q}$

ii) $W_m^{(i)} = \frac{1}{m} \sum_{j=1}^{m} \mathbb{1}_{\{q\text{-Coloración de } G_i\}}(X_j^{(i)}) \xrightarrow[m\to\infty]{} \frac{Z_i}{Z_{i-1}}$

Freddy Hernández-Romero, UNAL-Bogotá

Anotaciones

$$\left| W_\ell^{(j)} - \rho_{G_{j-1},q}\left(\xi(x_i) \neq \xi(y_i)\right) \right|$$

$$\leq \left| W_\ell^{(j)} - \mu^{(n)}\left(\xi(x_i) \neq \xi(y_i)\right) \right|$$

$$+ \left| \mu^{(n)}\left(\xi(x_i) \neq \xi(y_i)\right) - \rho_{G_{j-1},q}\left(\xi(x_i) \neq \xi(y_i)\right) \right|$$

$$\leq \left| W_\ell^{(j)} - \mu^{(n)}\left(\xi(x_i) \neq \xi(y_i)\right) \right| + d_{TV}\left(\mu^{n}, \rho_{G_{j-1},q}\right)$$

Freddy Hernández-Romero, UNAL-Bogotá

Anotaciones

$$\boxed{d_{VT}\left(\mu^n, \rho_{\mathcal{G}_{j-1},f}\right) < \frac{\epsilon}{8\ell}}$$

Por el Teorema B.1 (notas anteriores)

$$n \geq k\left(\frac{\log(k) + \log\left(\frac{8\ell}{\epsilon}\right) - \log(d)}{\log\left(q/2d^2\right)} + 1\right)$$

$$\Rightarrow \quad d_{VT}\left(\mu^n, \rho_{\mathcal{G}_{j-1},f}\right) \leq \frac{\epsilon}{8\ell}$$

Anotaciones

$\mathbb{P}\left(\left| W_t^{(j)} - \mu^{(n)}(\xi(x_i) \neq \xi(y_i)) \right| > \frac{\epsilon}{8l} \right) \leq \frac{1}{3}$ Por el Lema 3

$$\mathbb{P}\left(\left| W_t^{(j)} - p \right| > \frac{\epsilon}{8l} \right) \leq \frac{t}{4\left(\frac{\epsilon t}{8l}\right)^2} = \frac{16 l^2}{\epsilon^2 t}$$

donde

$$p = \mu^{(n)}(\xi(x_i) \neq \xi(y_i)).$$

Luego

$$t \geq \frac{48 l^3}{\epsilon^2} \Rightarrow \mathbb{P}\left(\left| W_t^{(j)} - p \right| > \frac{\epsilon}{8l} \right) \leq \frac{1}{3l}$$

Freddy Hernández-Romero, UNAL-Bogotá

Anotaciones

Como $l \leq dk$

$$t \geq \frac{48 d^3 k^3}{\epsilon^2} \Rightarrow \mathbb{P}\left(\left| W_t^{(j)} - p \right| > \frac{\epsilon}{8l} \right) \leq \frac{1}{3l}$$

Freddy Hernández-Romero, UNAL-Bogotá

Anotaciones

Let us summarize: The algorithm has $\bar{k}$ factors Y_j to compute. Each one is obtained using no more than $\frac{48d^3k^3}{\varepsilon^2}$ simulations, and, by (76), each simulation requires no more than $k\left(\frac{2\log(k)+\log(\varepsilon^{-1})+\log(8)}{\log\left(\frac{q}{2d^2}\right)}+1\right)$ steps of the Gibbs sampler. The total number of steps needed is therefore at most

$$dk \times \frac{48d^3k^3}{\varepsilon^2} \times k\left(\frac{2\log(k)+\log(\varepsilon^{-1})+\log(8)}{\log\left(\frac{q}{2d^2}\right)}+1\right)$$

which is of the order $Ck^5\log(k)$ as $k \to \infty$ for some constant C that does not depend on k. This is less than Ck^6, so the total number of iterations in the Gibbs sampler grows no faster than polynomially. Since, clearly, the running times of all other parts of the algorithm are asymptotically negligible compared to these Gibbs sampler iterations, Theorem 9.1 is established. $\square$

Freddy Hernández-Romero, UNAL-Bogotá

Resumen del Algoritmo FPRAS

Análisis de Complejidad

- El algoritmo tiene l (el número de aristas) etapas.
- En cada etapa, estimamos un ratio. Esto requiere un número de muestras polinomial en l y $1/\epsilon$.
- Cada muestra se genera con MCMC, lo que requiere un número de pasos polinomial en k (el número de vértices).

El resultado final es un algoritmo cuyo tiempo de ejecución total es **polinomial en k y $1/\epsilon$**, lo que cumple la definición de un FPRAS.

Freddy Hernández-Romero, UNAL-Bogotá

Resumen del Algoritmo FPRAS

El Gran Avance

Esta técnica (a menudo llamada "MCMC acoplado" o "path coupling") fue un gran avance en la ciencia de la computación teórica. Permitió resolver, de manera aproximada, problemas de conteo que se consideraban intratables, conectando profundamente la teoría de Cadenas de Markov con la complejidad computacional.

Freddy Hernández-Romero, UNAL-Bogotá

Anotaciones

Freddy Hernández-Romero, UNAL-Bogotá