דגימה אקראית

מתוך: Boissonnat and Yvinec, *Algorithmic Geometry*

עבור הקורס נושאים מתקדמים בגיאומטריה חישובית

מרצה: גיל ברקת

מגישה: אסנת טל

הגדרות - אובייקט

נציג שיטה למידול בעיות גיאומטריות. כל בעיה תיוצג ע"י אובייקטים ואיזורים.

- **אובייקט** הוא איבר מ- $\mathcal{O}$.
  - בדר"כ $\mathcal{O}$ יהיה אינסופי.
  - בדר"כ נדבר על אלמנטים מהמרחב האוקלידי $\mathbb{E}_d$. (כמו נקודות, ישרים, חצאי מישורים וכו').

הגדרות – איזור

- **איזור** הוא איבר מ- $\mathcal{F}$.
- לכל איזור יש שתי קבוצות אובייקטים "המענינות" אותו:
  - קבוצת האובייקטים ש**מגדירה** אותו.
    גודל קבוצה זו חסום ע"י קבוע b (שנובע מאופי הבעיה).
  - קבוצת האובייקטים שב**קונפליקט** איתו.
    נקראת תחום ההשפעה של האיזור. בדר"כ קבוצה אינסופית.

הגדרות – המשך

תהא S קבוצת אובייקטים, ואיזור F ב- $\mathcal{F}$.

- F **מוגדר** מעל S אם קבוצת האובייקטים שמגדירה אותו מוכלת ב S.
- F **אינו בעל קונפליקט** מעל S אם תחום ההשפעה שלו לא מכיל איברים מ S.
- F הוא **בעל קונפליקטים** מעל S אם תחום ההשפעה שלו מכיל j אובייקטים מ- S.

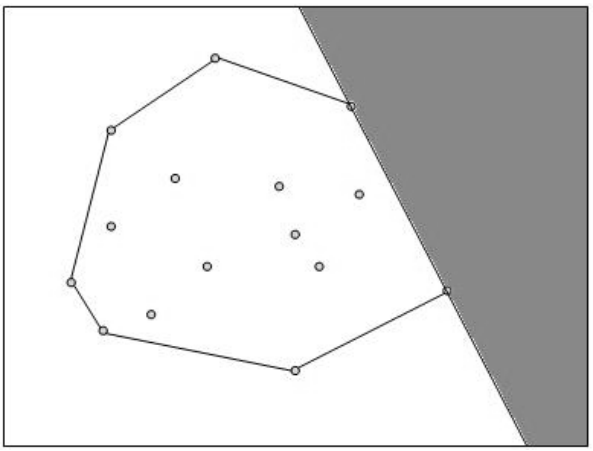
המטרה

- נחפש להציג את הבעיה הגיאומטרית כך שתהיה שקולה לחיפוש כל האיזורים המוגדרים מעל S, ואינם בקונפליקט מעליו.

דוגמא – מציאת Convex Hull

- תהא S קבוצה של n נקודות במרחב האוקלידי ה-d מימדי $\mathbb{E}_d$.
- כל קבוצה של d נקודות מגדירה על-מישור שעובר דרכה, אשר מחלק את המרחב לשני חצאי מרחבים פתוחים.
- נחפש את כל תת הקבוצות של S בעלות d נקודות, כך שאחד מ-2 חצאי המרחבים הפתוחים המוגדרים ע"י המישור שעובר דרכן, לא מכיל אף נקודה מ- S. – אלה הן הפיאות של $\mathrm{conv}(S)$.

דוגמא – מציאת Covex Hull - המשך



דוגמא – מציאת Covex Hull - המשך

- קבוצת האובייקטים S היא קבוצת n הנקודות.
  מוגדרת מעל $\mathcal{O}$ – כל הנקודות במרחב $\mathbb{E}_d$.
- קבוצת האיזורים – חצאי מרחבים פתוחים במרחב $\mathbb{E}_d$.
- כל d אובייקטים (נקודות) מגדירים 2 איזורים (חצאי מרחבים פתוחים).
- אובייקט (נקודה) הוא בקונפליקט עם איזור (חצי מרחב פתוח) אם הוא נמצא בשטחו.
- ע"מ למצוא את conv(S) עלינו למצוא את כל האיזורים המוגדרים וחסרי הקונפליקט מעל S.

הערות

- גודל קבוצת האובייקטים המגדירים איזור, חסום ע"י קבוע b הנובע מאופי הבעיה, אולם אין הכרח שכל האיזורים יוגדרו ע"י אותו מספר של אובייקטים.
  - בדוגמת ה-Convex Hull, כל האיזורים כן מוגדרים ע"י אותו מספר של אובייקטים, d, ובפרט $b = d$.
- איזור אינו בעל קונפליקט עם האובייקטים שמגדירים אותו.
  - בדוגמת ה-Convex Hull, הגדרנו לכן את שני חצאי המרחב כפתוחים, ע"מ שלא יכילו את הנקודות שהגדירו אותם.

הערות - המשך

- איזורים שמוגדרים ע"י קבוצות שונות של אובייקטים הם שונים.
  - בדוגמת ה-Convex Hull, אכן התקבלו חצאי מרחבים שונים עבור קבוצות שונות של נקודות.
- בדר"כ מספר האיזורים המוגדרים ע"י b (או פחות) אובייקטים חסום ע"י קבוע.

  במקרה כזה, מספר האיזורים המוגדרים על קבוצה בגודל n הוא $O(n^b)$.
  - בדוגמת ה Convex Hull, מספר האיזורים המוגדרים ע"י b אובייקטים הוא בדיוק 2.

סימונים

- $F(S)$ – קבוצת האיזורים המוגדרים מעל S.
- $F_j(S)$ – קבוצת האיזורים המוגדרים ובעלי j קונפליקטים מעל S (עבור j ב $[0,\ldots,n]$).
- $F_0(S)$ – קבוצת האיזורים המוגדרים וחסרי קונפליקט מעל S.
- $F_{\leq k}(S)$ – קבוצת האיזורים המוגדרים ובעלי k קונפליקטים לכל היותר מעל S.
- באופן דומה: $F^i(S)$, $F^i_j(S)$, $F^i_0(S)$, $F^i_{\leq k}(S)$ – קבוצות האיברים המוגדרות מעל S ע"י בדיוק i איברים.

הגדרה – דגימה אקראית

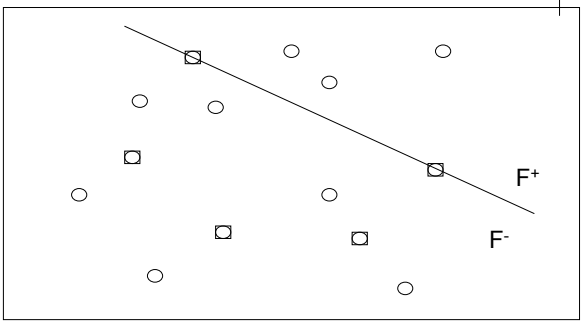
תהיה R תת קבוצה של S בגודל r.

R הינה **דגימה אקראית** של S אם האיברים בה נבחרו אקראית מתוך S. כלומר, כל תת קבוצה בגודל r יכולה להבחר בהסתברות שווה $\frac{1}{\binom{n}{r}}$.

סימונים

- באופן דומה לקודם, נגדיר קבוצות איזורים מעל R:
  $F(R)$, $F_j(R)$, $F_0(R)$, $F_{\leq k}(R)$.
  <u>נשים לב</u>: איזור הוא מוגדר ועם 0 או j או $\leq k$ קונפליקטים מעל R אם הוא מוגדר מעל R ועם 0 או j או $\leq k$ קונפליקטים <u>מעל R</u>.
- באופן דומה: $F^i(R)$, $F^i_j(R)$, $F^i_0(R)$, $F^i_{\leq k}(R)$ קבוצות האיברים המוגדרות מעל R ע"י בדיוק i איברים.

דוגמא



הנקודות המסומנות בריבוע (□) הן הנקודות ב-R.

- F^+ שייך ל- $F_0(R)$ ול- $F_3(S)$
- F^- שייך ל- $F_3(R)$ ול- $F_9(S)$

סימונים

- תהא $g(R)$ פונקציה על דגימה אקראית R מ S, אזי נגדיר $g(r, S)$ – התוחלת של $g(R)$ עבור דגימה אקראית בגודל r.

למשל:

- $f_j(R) = |F_j(R)|$ גודל הקבוצה המוגדרת ובעלת j קונפליקטים מעל R.
- $f_j(r, S)$ – התוחלת של $f_j(R)$.
- באופן דומה $f'_j(R)$ והתוחלת $f'_j(r, S)$.

משפט הדגימה האקראית

- משפט הדגימה האקראית נותן חסם עליון למספר האיזורים המוגדרים ובעלי k קונפליקטים לכל היותר מעל S.

  החסם תלוי בתוחלת מספר האיזורים המוגדרים וללא קונפליקטים מעל דגימה אקראית של S, בגודל $\lfloor n/k \rfloor$.

רעיון המשפט: ככל שלאיזור יש פחות אובייקטים שהוא בקונפליקט איתם, כך גדלה ההסתברות שאיזור זה חסר קונפליקטים בדגימה אקראית.

משפט הדגימה האקראית

תהיה S קבוצת n אובייקטים, ויהיה k מספר שלם כך ש

$$2 \le k \le \frac{n}{b+1}$$

אזי:

$$\left|F_{\le k}(S)\right| \le 4(b+1)^b k^b f_0\left(\left\lfloor n/k \right\rfloor, S\right)$$

כך ש b הוא חסם עליון למספר האובייקטים המגדירים איזור, $\left|F_{\le k}(S)\right|$ הוא מספר האיזורים המוגדרים ובעלי לכל היותר k קונפליקטים מעל S, ו- $f_0\left(\left\lfloor n/k \right\rfloor, S\right)$ הוא התוחלת של מספר האיזורים המוגדרים וללא קונפליקט מעל דגימה אקראית בגודל $\left\lfloor n/k \right\rfloor$.

משפט הדגימה האקראית – למה 4.2.1

עבור הוכחת משפט הדגימה האקראית נשתמש בשתי הלמות הבאות.

<u>למה 4.2.1</u>

תהיה S קבוצה של n אובייקטים ויהי F איזור בקונפליקט עם j אובייקטים ב S המוגדר ע"י i אובייקטים ב S.

אם R היא דגימה אקראית של S בגודל r, אז ההסתברות $p^i_{j,k}(r)$ ש F מוגדר ובעל k קונפליקטים ב-R היא

$$p^i_{j,k}(r) = \frac{\binom{j}{k}\binom{n-i-j}{r-i-k}}{\binom{n}{r}}$$

משפט הדגימה האקראית – למה 4.2.1

<u>הסבר</u>

מספר כל הדגימות האקראיות בגודל r הוא $\binom{n}{r}$

הדגימות החוקיות מבחינתנו, הן אלה שכל התנאים הנ"ל מתקיימים בהן:

- מתוך *j* האובייקטים שבקונפליקט עם *F*, יבחרו *k*.
- כל *i* האובייקטים המגדירים את *F* יבחרו.
- מתוך *n-i-j* האובייקטים הנותרים (שאינם בקונפליקט ואינם מגדירים) יבחרו *r-i-k*

משפט הדגימה האקראית – למה 4.2.2

<u>למה 4.2.2</u>

תהיה *S* קבוצה בת *n* איברים, ו- *R* דגימה אקראית שלה.
התוחלת $f^i_k(r, S)$ של איזורים המוגדרים ע"י *i* אובייקטים מ- *R* ובעלי *k* קונפליקטים ב- *R* הינה:

$$f_k^i(r,S) = \sum_{j=0}^{n-i} \left|F_j^i(S)\right| \frac{\binom{j}{k}\binom{n-i-j}{r-i-k}}{\binom{n}{r}}$$

משפט הדגימה האקראית – למה 4.2.2

<u>הוכחה</u>

נרוץ על מספר הקונפליקטים האפשרי לאיזור, ונסכום את כל האיזורים בעלי מספר זה של קונפליקטים ב-*S*, כפול ההסתברות שאיזור זה יופיע גם ב-*R*, אבל עם *k* קונפליקטים.

$$f_k^i(r,S) \quad = \quad \sum_{j=0}^{n-i} \left|F_j^i(S)\right| p_{j,k}^i(r)$$

ע"י הצבה ב $p^i_{j,k}(r)$ את מה שקבלנו בלמה הקודמת, נקבל את המשוואה שטענו לקבל.

□

משפט הדגימה האקראית - הוכחה

- תזכורת למשפט: $\left|F_{\le k}(S)\right| \le 4(b+1)^b k^b f_0\left(\left\lfloor n/k \right\rfloor, S\right)$
- <u>שלב א'</u> - נוכיח לכל $1 \le i \le b$ את המשוואה הבאה:

$$\left|F_{\le k}^i(S)\right| \le 4(b+1)^i k^i f_0^i\left(\left\lfloor n/k \right\rfloor, S\right)$$

יהיה *k* מספר שלם כך ש $2 \le k \le \frac{n}{b+1}$

ותהיה *R* דגימה אקראית של *S* בגודל $\left\lfloor n/k \right\rfloor$

משפט הדגימה האקראית - הוכחה

בעזרת 4.2.2 נסיק

$$f_0^i(r,S) \;=\; \sum_{j=0}^{n-i}\left|F_j^i(S)\right| \frac{\binom{n-i-j}{r-i}}{\binom{n}{r}} \;\geq\; \left|F_{\leq k}^i(S)\right| \frac{\binom{n-i-k}{r-i}}{\binom{n}{r}}$$

ניתן להראות ע"י פיתוח אלגברי שעבור כל $2 \leq k \leq \frac{n}{b+1}$ שלם,
ודגימה בגודל $\lfloor n/k \rfloor$ מתקיים:

$$\frac{\binom{n-i-k}{r-i}}{\binom{n}{r}} \;\geq\; \frac{1}{4(b+1)^i k^i}$$

משפט הדגימה האקראית - הוכחה

ומכאן,

$$f_0^i(r,S) \;\geq\; \left|F_{\leq k}^i(S)\right| \frac{\binom{n-i-k}{r-i}}{\binom{n}{r}} \;\geq\; \frac{\left|F_{\leq k}^i(S)\right|}{4(b+1)^i k^i}$$

ומוכח שלב א':

$$\left|F_{\leq k}^i(S)\right| \leq 4(b+1)^i k^i f_0^i\left(\lfloor n/k \rfloor, S\right)$$

- <u>שלב ב'</u> – נסכום על כל $1 \leq i \leq b$, ונקבל:

$$\left|F_{\leq k}(S)\right| \;=\; \sum_{i=1}^{b}\left|F_{\leq k}^i(S)\right| \;\leq\; \sum_{i=1}^{b} 4(b+1)^i k^i f_0^i\left(\lfloor n/k \rfloor, S\right)$$

$$\leq\; 4(b+1)^b k^b \sum_{i=1}^{b} f_0^i\left(\lfloor n/k \rfloor, S\right) \;=\; 4(b+1)^b k^b f_0\left(\lfloor n/k \rfloor, S\right)$$

□

הערות

1. משפט הדגימה האקראית מדבר רק עבור $2 \leq k \leq \frac{n}{b+1}$
עבור *k=0,1*, מתקיים:

$$|F_0(S)| \leq |F_{\leq 1}(S)| \leq |F_{\leq 2}(S)| \leq 4(b+1)^b 2^b f_0\left(\left\lfloor n/2 \right\rfloor, S\right)$$

כאשר *n ≥ 2(b+1)*.

2. עבור ערכי *k* הקרובים ל- *n*, קיים החסם הטריוויאלי:

$$|F_{\leq k}(S)| \leq |F(S)| = O\left(n^b\right)$$

במידה שמספר האיזורים שמגדירים *b* (או פחות) אובייקטים, חסום ע"י קבוע.

מסקנה 4.2.4

תהיה *S* קבוצה בת *n* אובייקטים, כך ש *n ≥ 2(b+1)*.
לכל מספר שלם *r* המקיים *2(b+1) ≤ r ≤ n* מתקיים:

$$f_1(r,S) \leq \beta f_0\left(\left\lfloor r/2 \right\rfloor, S\right)$$
$$f_2(r,S) \leq \beta f_0\left(\left\lfloor r/2 \right\rfloor, S\right)$$

כאשר $f_j(r, S)$ היא התוחלת של מספר האיזורים המוגדרים ובעלי *j* קונפליקטים מעל דגימה אקראית בגודל *r* מעל *S*.
ו- β הוא הקבוע $\beta = 4(b+1)^b 2^b$

הוכחת מסקנה 4.2.4

תהיה R דגימה אקראית של S בגודל r, כך שמתקיים $r \le 2(b+1)$.

אם נדגום מתוך הדגימה R, דגימה אקראית בגודל $r/2$, אזי ע"פ הערה 1 למשפט הדגימה האקראית מתקבל:

$$|F_1(R)| \le 4(b+1)^b 2^b f_0\left(\left\lfloor r/2 \right\rfloor, R\right)$$

נשים לב שהתוחלת של $|F_1(R)|$ היא $f_1(r,S)$

וש- $f_0\left(\left\lfloor r/2 \right\rfloor, R\right) = f_0\left(\left\lfloor r/2 \right\rfloor, S\right)$

ונקבל: $f_1(r,S) \le \beta f_0\left(\left\lfloor r/2 \right\rfloor, S\right)$

באופן דומה, נראה גם את $f_2(r,S) \le \beta f_0\left(\left\lfloor r/2 \right\rfloor, S\right)$

□

מומנט - הגדרה

יהא k מספר שלם קטן או שווה ל n.

המומנט מסדר k של R ביחס ל S הינו:

$$m_k(R,S) = \sum_{F \in F_0(R)} \binom{|S(F)|}{k}$$

$F_0(R)$ – קבוצת האיזורים המוגדרים וחסרי הקונפליקט מעל R.

$S(F)$ – קבוצת האובייקטים ב S שיש ביניהם ובין האיזור F קונפליקט.

מומנט

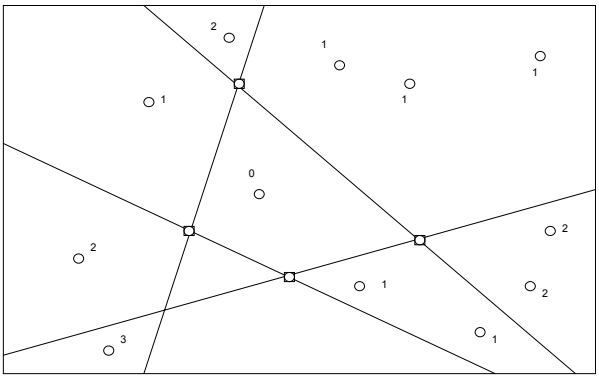
- המומנט מסדר 0, $m_0(R,S)$, הוא מספר האיזורים המוגדרים וחסרי הקונפליקטים מעל R.

$$m_0(R,S) = \left|F_0(R)\right|$$

- המומנט מסדר 1, הוא סך כל מספר הקונפליקטים בין האובייקטים ב S לבין האיזורים המוגדרים וללא קונפליקטים מעל R.

$$m_1(R,S) = \sum_{F \in F_0(R)} \left|S(F)\right|$$

מומנט מסדר 1 - דוגמא



בדוגמת ה convex hull, בציור זה 17 = $m_1(R,S)$

מומנט

- תוחלת המומנט $m_k(R,S)$ עבור דגימה אקראית R בגודל r של S, תסומן $m_k(r,S)$.

- בפרט, מתקיים $m_0(r,S) = f_0(r,S)$

למה 4.2.5

<u>למה 4.2.5</u>

$$m_k(r,S) = \sum_{i=1}^{b}\sum_{j=0}^{n-i}\left|F_j^i(S)\right|\binom{j}{k}p_{j,0}^i(r)$$

<u>הוכחה</u>

ניזכר כי $p^i_{j,0}(r)$ היא ההסתברות שאיזור נתון F שהינו $F^i_j(S)$, הינו מוגדר וחסר קונפליקטים מעל דגימה אקראית בגודל r של S.

לכן:

$$m_k(r,S) = \sum_{i=1}^{b}\sum_{j=0}^{n-i}\sum_{F\in F_j^i(S)}\binom{j}{k}p_{j,0}^i(r)$$

ומכאן נובעת הלמה.

□

משפט המומנט

תהיה S קבוצת אובייקטים בגודל n.

התוחלת $m_k(r,S)$ של המומנט מסדר k של דגימה אקראית של S בגודל r, קשורה לתוחלת $f_k(r,S)$ של מספר האיזורים המוגדרים ובעלי k קונפליקטים מעל דגימה אקראית של S בגודל r בקשר זה:

$$m_k(r,S) \quad \le \quad f_k(r,S)\frac{(n-r+k)!}{(n-r)!}\frac{(r-b-k)!}{(r-b)!}$$

כאשר כל איזור מוגדר ע"י b אובייקטים לכל היותר.

משפט המומנט - הוכחה

לפי למה 4.2.5, ולמה 4.2.1 נובע:

$$m_k(r,S) \quad = \quad \sum_{i=1}^{b}\sum_{j=0}^{n-i}\left|F_j^i(S)\right|\binom{j}{k}\frac{\binom{n-i-j}{r-i}}{\binom{n}{r}}$$

ע"י פיתוחים אלגבריים ובעזרת שימוש נוסף בלמה 4.2.1 מתקבל:

$$m_k(r,S) \quad = \quad \frac{(n-r+k)!}{(n-r)!}\frac{(r-b-k)!}{(r-b)!}\sum_{i=1}^{b}\sum_{j=0}^{n-i}\left|F_j^i(S)\right|p_{j,k}^i(r)$$

כאשר $p^i_{j,k}(r)$ היא ההסתברות שאיזור F ב- $F^i_j(S)$ שייך ל $F^i_k(R)$

משפט המומנט – המשך הוכחה

ולכן מתקבל

$$m_k(r,S) \quad \le \quad f_k(r,S)\frac{(n-r+k)!}{(n-r)!}\frac{(r-b-k)!}{(r-b)!}$$

□

משפט 4.2.7

תהיה S קבוצת אובייקטים בגודל n.

קיים מספר קבוע ממשי γ ומספר שלם r_0, בלתי תלויים ב - n, כך שעבור כל $n \ge r \ge r_0$ מתקיים:

$$m_1(r,S) \le \gamma \frac{n-r}{r} f_0\left(\left\lfloor r/2 \right\rfloor, S\right)$$

$$m_2(r,S) \le \gamma \frac{(n-r)^2}{r^2} f_0\left(\left\lfloor r/2 \right\rfloor, S\right)$$

כאשר $m_k(r,S)$ היא התוחלת של המומנט מסדר k על דגימה אקראית של S בגודל r, ו- $f_0(r,S)$ היא התוחלת של מספר האזורים המוגדרים וחסרי הקונפליקט מעל דגימה אקראית של S בגודל r.

משפט 4.2.7 - הוכחה

עבור $k=1$ משפט המומנט אומר

$$m_1(r,S) \quad \le \quad f_1(r,S)\frac{n-r+1}{r-b}$$

בעזרת משפט 4.2.4 הנותן חסם עליון ל $f_1(r,S)$ נקבל את תוצאת המשפט.

באופן דומה ניתן להוכיח עבור $m_2(r,S)$.

□

תודה רבה ☺