PROBABILISTIC METHODS

DONGRYUL KIM

확률론적인 방법은 1950년대에 Erdős가 처음 사용하고 1970년대에 들어 Lovász, Spencer, Rödl 등 수많은 사람들의 노력으로 꽃피워 현대 조합론에 불가결한 존재가 되었다. 어떤 대상을 만들어갈 때 무작위적인 선택을 사용하기 때문에 우리가 직접 나서서 구체적으로 구조를 파헤칠 필요가 없다는 장점을 가지고 있어 특정한 조건을 만족하는 무언가의 존재성만 증명한다고 할 때 매우 강력한 도구로 사용될 수 있다. 하지만 한편으로는 우리가 만든 대상의 형태에 대한 정보를 얻기 힘들기 때문에 답답한 측면이 없지 않아 있다.

최근 들어 FKMO에도 확률론적 기법을 사용해야만 풀 수 있는 문제들이 출제되는 등 국내 올림피아드에서 확률론적 방법은 조합 문제를 푸는 기본적인 도구로 자리매김하고 있다. 물론 Lovász local lemma나 Rödl nibble과 같이 아주 어려운 형태의 기법들은 올림피아드에 출제되기 어렵겠지만, 확률변수를 설정하여 기댓값을 보는 것과 같은 기본적인 테크닉은 숙지하고 있는 것이 좋을 것이다. 아래 문제들을 풀어보며 확률론이 조합에서 어떻게 사용될 수 있는지 알아보자.

1. Basics of probability

확률변수라는 것은 쉽게 말해서 특정 값이 나올 확률이 이미 정해져 어떤 버튼을 누르기만 하면 결과가 나오는 상자 같은 것이다. 예를 들어

$$X = \text{“주사위를 던졌을 때 나오는 값”}$$

이라 하면 $X = 1$일 확률은 $1/6$인 것이다. 확률변수의 기댓값은 말 그대로 결과로 기대할 수 있는 값으로, 각 결과에 확률을 가중치로 붙여 평균 낸 값이다. 위의 예시에서 기댓값은

$$\mathbb{E}[X] = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3.5$$

로 계산할 수 있다. 당연한 이야기지만, $X \geq \mathbb{E}[X]$일 확률은 항상 0보다 크기 마련이다. 이 단순한 사실이 확률론적 기법의 핵심이 될 명제이다.

정리 1.1. *임의의 두 확률변수 X, Y에 대해 다음이 성립한다.*

(a) $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
(b) $\mathbb{E}[X^2]\mathbb{E}[Y^2] \geq \mathbb{E}[XY]^2$
(c) *X와 Y가 동립이라면 $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$이다.*

Date: July 3, 2016.

정리 1.2 (Markov inequality)**.** *항상 값이 0 이상인 확률변수 X와 $\lambda > 0$에 대해, 다음이 성립한다.*

$$P[X \geq \lambda] \leq \frac{\mathbb{E}[X]}{\lambda}$$

2. Looking at the expectation

이 절의 문제들은 모두 $P[X \geq \mathbb{E}[X]] \geq 0$을 이용하는 문제들이다. 부등식을 써내려가기는 쉽지만, X를 어떻게 설정하느냐가 관건이 되는 경우가 많으므로 어떤 상황에서 어떻게 확률변수를 잡는지 눈여겨보자.

예제 2.1. 변의 개수가 m인 그래프 G가 있다. G의 부분그래프 중에 이분그래프이며 변의 개수가 $m/2$개 이상인 것이 존재함을 보여라.

Proof. 각 점을 1/2의 확률로 집합 A에 넣고, 나머지 1/2의 확률로 B에 넣자. 이때 분할 $V(G) = A \cup B$이 만들어질 것이다. 여기서 A 안의 선분들과 B 안의 선분들을 모두 지워 그래프 G'를 만들자. 자명하게도 G'은 이분그래프가 될 것이다.

그러면 임의의 선분 $e = \{v, w\}$에 대해, e가 남아있을 확률은 $1/2 \cdot 1/2 + 1/2 \cdot 1/2 = 1/2$이다. 따라서

$$\mathbb{E}[|E(G')|] = \frac{|E(G)|}{2} = \frac{m}{2}$$

이다. 그러면 $G' \geq m/2$일 확률은 0보다 크게 되어 변의 개수가 $m/2$ 이상인 이분그래프가 존재함을 알 수 있다. □

두 번째 예제에서는 Turán의 정리를 일반화한 명제를 살펴볼 것이다. 이것을 왜 Turán의 정리의 일반화라 하였는지를 생각해보자.

예제 2.2. 점 개수가 n인 그래프 G가 있다. 각 점의 차수들을 $d_1, d_2, \ldots, d_n$이라 하면, G에서 임의의 두 점도 이웃하지 않게

$$\frac{1}{d_1+1} + \frac{1}{d_2+1} + \cdots + \frac{1}{d_n+1}$$

개 이상의 점을 뽑을 수 있음을 증명하여라.

Proof. 그래프 G의 점들을 $V = \{v_1, \ldots, v_n\}$라 하자. V의 순열 $\Pi = (v_{\sigma(1)}, \ldots, v_{\sigma(n)})$을 무작위적으로 선택하자. 여기서

$$S(\Pi) = \{v : v\text{와 이웃한 점들은 모두 }\Pi\text{에서 }v\text{ 뒤에 위치한다}\}$$

를 정의하자. 정의상 $S(\Pi)$의 임의의 두 점은 서로 이웃할 수 없음을 확인할 수 있다.

이제 $|S(\Pi)|$의 기댓값을 구해보자. 점 $v_i \in V$에 대해, $v_i \in S(\Pi)$일 확률은 v_i가 이웃한 다른 점들보다 앞에 위치할 확률이므로 $1/(d_i+1)$이다. 따라서

$$\mathbb{E}[|S(\Pi)|] = \frac{1}{d_1+1} + \cdots + \frac{1}{d_n+1}$$

이 되어 크기가 그보다 크며 임의의 두 점도 서로 이웃하지 않은 집합이 존재함을 알 수 있다. □

연습문제 2.3. 복소수 $z_1, \dots, z_n \in \mathbb{C}$가 있을 때, 어떤 $a_1, \dots, a_n \in \{-1, 1\}$가 존재하여 다음 부등식이 성립함을 보여라.

$$\left|\sum_{i=1}^{n} a_i z_i\right|^2 \geq \sum_{i=1}^{n} |z_i|^2$$

연습문제 2.4 (IMO Shortlist 1999 C4)**.** 양의 정수 N과 N개의 $\mathrm{mod} N^2$ 잉여들로 이루어진 집합 A가 주어져 있다. 이때 N개의 $\mathrm{mod} N^2$ 잉여들로 이루어진 집합 B가 존재하여

$$A + B = \{a + b : a \in A, b \in B\}$$

가 $\mathrm{mod} N^2$으로 절반 이상의 잉여들을 가지게 됨을 증명하여라.

연습문제 2.5. 평면 위에 10개의 점이 있다. 10개의 서로 겹치지 않는 단위원을 잘 그려 주어진 10개의 점이 모두 덮이도록 할 수 있음을 보여라.

연습문제 2.6 (Russia 1999)**.** 남학생들과 여학생들로 이루어진 반이 있다. 임의의 두 학생은 서로 아는 관계이거나 서로 모르는 관계이고, 임의의 남학생은 아는 여학생이 적어도 한 명 존재한다고 한다. 이때 반에서 절반 이상의 학생을 선택하여, 임의의 선택된 남학생은 홀수명의 선택된 여학생을 알도록 할 수 있음을 보여라.

연습문제 2.7 (Taiwan 1997 9)**.** 정수 $n \geq k \geq 3$에 대해, $X = \{1, \dots, n\}$이라 하자. F_k는 X의 크기 k 부분집합들의 모임이고, F_k의 서로 다른 임의의 두 원소는 최대 $k-2$개의 원소를 공통으로 가진다고 한다. 이때 원소의 개수가 $\log_2 n$ 이상이며 F_k의 원소들을 부분집합으로 가지지 않는 집합 $M_k \subseteq X$가 존재함을 보여라.

연습문제 2.8 (Crossing lemma)**.** 평면 상에 점이 v개이고 변이 e인 단순그래프를 그렸다. 만약 $e > 4v$라면, 두 변 e_1과 e_2가 교차하게 되는 $\{e_1, e_2\}$의 개수는 $e^3/64v^2$ 이상임을 증명하여라.

연습문제 2.9. 점의 개수가 n이고 각 점의 차수가 γ 이상인 그래프 G가 있다. 이때 원소의 개수가

$$\frac{n(1 + \log(1 + \gamma))}{1 + \gamma}$$

이상인 점들의 집합 A가 존재하여, 임의의 점 v에 대해 $v \in A$이던가 A의 원소 중 v와 인접한 점이 존재하게 됨을 증명하여라.

연습문제 2.10 (KMO Winter 2015 4)**.** 원소의 개수가 n이 집합 S에 다음 조건들을 만족하는 연산 $\oplus$가 주어져 있다.

- 임의의 $x, y \in S$에 대해 $x \oplus y = y \oplus x \in S$
- 임의의 $x \in S$에 대해 $x \oplus x = x$
- 임의의 서로 다른 $x \neq y \in S$에 대해 $x \oplus y \neq x$

이때 임의의 서로 다른 $x \neq y \in X$에 대해 $x \oplus y \notin X$이며 $|X| \geq \sqrt{n}/2$인 부분집합 $X \subseteq S$가 존재함을 증명하여라.

연습문제 2.11 (Kömal A.309)**.** 점 개수가 n인 그래프 G가 있다. 각 점의 차수들이 $0 < d_1, d_2, \ldots, d_n$이라 한다면, G에서

$$\frac{2}{d_1+1} + \frac{2}{d_2+1} + \cdots + \frac{2}{d_n+1}$$

개 이상의 점을 선택하여 선택한 점들로 이루어진 회로가 없게 할 수 있음을 증명하여라.

연습문제 2.12 (FKMO 2016 6)**.** 삼각형 m개의 집합 U가 있다. 이때 아래 두 조건을 모두 만족하는 U의 부분집합 W가 존재함을 증명하여라.

(i) W에 속한 삼각형의 개수는 $0.45m^{4/5}$개 이상이다.

(ii) 6개의 삼각형 ABC, BCD, CDE, DEF, EFA, FAB가 모두 W에 속하게 되는 서로 다른 6개의 점 A, B, C, D, E, F는 존재하지 않는다.

연습문제 2.13. 홀수 $N \geq 3$에 대해, $N \times N$ 크기의 격자의 중앙 칸에 애벌레가 앉아있고 그 칸을 제외한 다른 칸에는 각각 서로 다른 양의 정수가 하나씩 적혀있다. 애벌레는 격자를 탈출하고자 하지만, 항상 이웃한 칸으로 이동해야 하고, 수 k가 적힌 칸으로 이동하려면 $1/k$ 만큼의 음식을 먹어야 한다. 애벌레는 1 만큼의 음식밖에 먹을 수 없다고 할 때, 수가 어떻게 쓰여있든 애벌레는 격자를 탈출할 수 있음을 보여라.

연습문제 2.14. $n \times n$ 크기의 체스판의 각 칸에 전구가 달려있고, 몇 개는 켜져있는 상태이다. 각 행과 열에 스위치가 하나씩 달려 있으며, 스위치를 누르면 해당하는 행 또는 열에 있는 전구들의 점멸 상태가 바뀐다고 한다. 이때 스위치를 잘 조작함으로써 켜져있는 전구의 수와 꺼져있는 전구의 수의 차가 $\sqrt{n^3/2}$ 이상이 되도록 할 수 있음을 보여라.

연습문제 2.15 (LYM inequality)**.** 집합 $S_1, S_2, \ldots, S_m \subseteq \{1, \ldots, n\}$이 임의의 $i \neq j$에 대해 S_i와 S_j가 부분집합 관계가 아니라는 성질을 가지고 있을 때, 다음 부등식을 증명하여라.

$$\sum_{i=1}^{m} \frac{1}{\binom{n}{|S_i|}} \leq 1$$

연습문제 2.16 (Erdős)**.** 어떤 집합 X에 $a+b=c$인 $a, b, c \in X$가 없으면 X를 *sum-free*라 부른다. 임의의 양의 정수들의 유한집합 S에 대해, X가 sum-free이며 $3|X| \geq |S|$인 부분집합 $X \subseteq S$가 존재함을 보여라.

연습문제 2.17 (Erdős-Ko-Rado theorem)**.** 양의 정수 n과 r이 $2r \leq n$을 만족한다. 크기 r인 부분집합 $S_1, \ldots, S_m \subset \{1, \ldots, n\}$들 중 임의의 두 집합이 공통 원소를 가진다고 할 때, $m \leq \binom{n-1}{r-1}$임을 증명하여라.