Jest to tekst związany z odczytem wygłoszonym na XLVI Szkole Matematyki Poglądowej, *Podejście niestandardowe*, Warszawa–Miedzeszyn, styczeń 2011.

Kombinatoryka probabilistyczna

Andrzej GRZESIK, Kraków

W matematyce jest jak w życiu — czasem zupełnie nie wiadomo co zrobić, jaką drogę wybrać. W naszym „normalnym" życiu w takich przypadkach podejmujemy losową decyzję i patrzymy co z tego wyjdzie. Okazuje się, że takie samo losowe podejście można skutecznie zastosować w wielu problemach kombinatorycznych. W poniższym artykule postaramy się usystematyzować to podejście i opisać kilka użytecznych metod probabilistycznych.

Najpierw rozważmy podejście standardowe. Zdefiniujmy funkcję $m(n)$ jako najmniejszą liczbę l o tej własności, że istnieje l-elementowa rodzina $\mathcal{S}$ zbiorów n-elementowych taka, że przy każdym kolorowaniu dwoma kolorami elementów zbiorów, istnieje w $\mathcal{S}$ zbiór jednokolorowy.

Zadanie 1. *Udowodnić, że $m(n) \geq 2^{n-1}$, czyli że rodzinę $l < 2^{n-1}$ zbiorów można pokolorować bez wystąpienia zbioru jednokolorowego.*

Dowód. Oznaczmy przez B zbiór wszystkich elementów występujących w zbiorach w $\mathcal{S}$. Musimy pokazać, że ten zbiór można dobrze pokolorować, ale nie wiemy jak można to zrobić. Zatem pomalujmy go losowo. Jeśli uda nam się pokazać, że prawdopodobieństwo niewystąpienia żadnego zbioru jednokolorowego jest dodatnie, to będzie to oznaczać, że istnieje dobry dobór kolorów.

Niech A_i oznacza zdarzenie losowe, że i-ty zbiór w $\mathcal{S}$ jest jednokolorowy. Wtedy
$$\mathbf{P}\left(A_i\right) = \frac{1}{2^{n-1}}.$$
Stąd prawdopodobieństwo, że jakiś zbiór jest jednokolorowy wynosi
$$\mathbf{P}\left(\bigcup A_i\right) \leq \sum \mathbf{P}\left(A_i\right) = \frac{l}{2^{n-1}} < 1.$$
To oznacza, że prawdopodobieństwo, że żaden zbiór nie jest jednokolorowy, czyli $\mathbf{P}\left(\bigcap A_i\right)$ jest większe od 0. Czyli istnieje kolorowanie bez zbiorów jednokolorowych, co mieliśmy wykazać.

Podobnie zadanie to można rozwiązać używając tak zwanej metody pierwszego momentu. Jej zasada działania jest prosta — jeśli wartość oczekiwana pewnej zmiennej losowej jest równa a, to istnieje takie zdarzenie, że wartość tej zmiennej jest większa lub równa a oraz zdarzenie, w którym zmienna ma wartość mniejszą lub równą a.

W naszym zadaniu możemy zdefiniować zmienną $X = \sum X_i$, gdzie X_i jest równe 1 jeśli i-ty zbiór w $\mathcal{S}$ jest jednokolorowy, a 0 w przeciwnym przypadku. Innymi słowy, zmienna X mówi, ile zbiorów jest jednokolorowych. Wtedy, z liniowości wartości oczekiwanej, mamy
$$\mathbb{E}X = \sum \mathbb{E}X_i = l \cdot \mathbb{P}\left(X_i = 1\right) = l \cdot \frac{1}{2^{n-1}} < 1.$$
Więc, skoro zmienna X przyjmuje tylko wartości całkowite, istnieje kolorowanie, w którym liczba zbiorów jednokolorowych jest równa 0, co mieliśmy wykazać.

W powyższym zadaniu było dość łatwo domyślić się, co należy losować. Niestety nie zawsze tak jest. Ale często w takim przypadkach możemy coś zmodyfikować dodatkowo, by osiągnąć cel. Zobrazujemy to następnym zadaniem.

Zadanie 2. *Na konferencję przyjechało n osób, wśród których każda znała co najmniej d innych osób. Chcemy wybrać spośród nich delegację taką, żeby każda osoba na konferencji znała co najmniej jednego delegata. Jaka liczność delegacji będzie zawsze wystarczająca?*

Jeśli będziemy losować ludzi do delegacji, to ciężko nam oszacować możliwość zajścia wymaganego warunku. Drugiem podejściem byłoby losowanie spośród zbiorów osób, dla których wymagany warunek jest spełniony, ale w takim wypadku ciężko nam jest szacować liczebność tych zbiorów.

Wylosujmy zatem dowolny zbiór osób (wybierając każdą osobę z prawdopodobieństwem p), a następnie dodajmy do delegacji wszystkich tych, którzy nie znają nikogo z wylosowanego zbioru. Niech więc R będzie tym losowym zbiorem, a Z zbiorem osób nie znających nikogo z R. Wtedy

$$\mathbb{E}|R| = np,$$
$$\mathbb{E}|Z| \leq n(1-p)^{d+1} \leq ne^{-p(d+1)},$$

gdyż, żeby ktoś nie znał nikogo wylosowanego do R, to w szczególności nie mogli zostać wylosowani jego znajomi, których jest co najmniej d, ani on sam.

Stąd możemy oszacować liczność delegacji spełniającej warunki zadania

$$\mathbb{E}|R \cup Z| \leq n\left(p + e^{-p(d+1)}\right).$$

Minimalizując po p, dostajemy, że dla $p = \frac{\ln(d+1)}{d+1} \leq 1$

$$\mathbb{E}|R \cup Z| \leq \frac{n(\ln(d+1)+1)}{d+1}.$$

Zatem istnieje taki wybór co najwyżej $\frac{n(\ln(d+1)+1)}{d+1}$ delegatów, że każdy zna co najmniej 1 delegata.

Innym bardzo ciekawym zastosowaniem metody probabilistycznej jest tzw. podnoszenie lokalnych własności. Opiera się na takim pomyśle: jeśli umiemy coś pokazać lokalnie, czyli że coś musi zachodzić dla każdego podzbioru, to musi to też zachodzić dla zbioru losowego. Korzystając z tego można dostać lepsze własności globalne. Dobrze zilustruje to następne zadanie.

Zadanie 3. *Ile co najmniej samoprzecięć musi mieć graf o n wierzchołkach i e krawędziach narysowany na papierze?*

Zauważmy, że jak narysujemy graf, który nie ma samoprzecięć (graf planarny), to musi być spełniony warunek Eulera, czyli

$$f - e + n = 2,$$

gdzie f to liczba ścian.

Skoro każda krawęd należy do dwóch ścian, a każda ściana ma co najmniej trzy krawędzie, to $2e \geq 3f$, więc $e \leq 3n - 6$.

Wemy teraz dowolny graf narysowany na papierze. Niech c będzie liczbą samoprzecięć. Możemy tak usunąć tyle krawędzi ile jest samoprzecięć, aby dostać graf planarny, zatem

$$(e - c) \leq 3n - 6,$$

a stąd

$$c \geq e - 3n + 6.$$

To oszacowanie nie jest dobre. Zauważmy, że jeśli liczba krawędzi jest kwadratowa ze względu na liczbę wierzchołków, to oszacowanie daje nam liczbę samoprzecięć wielkości n^2. Natomiast w rzeczywistości będzie ona dużo wyższa — rzędu n^4.

Ale zauważmy, że to oszacowanie działa także dla każdego podgrafu, stąd też dla podgrafu losowego. Stwórzmy taki podgraf H wybierając wierzchołki z prawdopodobieństwem p. Wartość oczekiwana liczby wierzchołków będzie równa np, liczby krawędzi — ep^2, natomiast liczby samoprzecięć — cp^4, gdyż żeby było samoprzecięcie, musimy wylosować końce obu krawędzi, które go tworzą. Udowodniony warunek daje nam

$$cp^4 \geq ep^2 - 3np + 6 \geq ep^2 - 3np,$$

czyli

$$c > \frac{ep - 3n}{p^3}.$$

Minimalizując po p dostajemy, że jeśli $e > 4n$, to dla $p = \frac{4n}{e}$

$$c > \frac{e^3}{64n^2},$$

co jest znacznie lepszym oszacowaniem dla dużej liczby krawędzi w grafie.

Na koniec, warto zaprezentować metodę, która spowodowała bardzo duży skok w rozwoju kombinatoryki i obecnie jest uważana za jedno z najbardziej podstawowych narzędzi w dowodach kombinatorycznych, czyli Lemat Lovásza. Idea opiera się na następującej obserwacji. Jeśli mamy zbiór zdarzeń niezależnych o prawdopodobieństwie mniejszym niż 1, to z pewnością prawdopodobieństwo, że żadne z tych zdarzeń nie zajdzie jest dodatnie. Natomiast co jeśli mamy zbiór wielu zdarzeń, które zależą od siebie, ale tylko lokalnie, tzn. każde zdarzenie wpływa tylko na niewielką część wszystkich zdarzeń? Okazuje się, że przy odpowiednim założeniu o lokalności, wtedy również da się osiągnąć dodatnie prawdopodobieństwo, że żadne zdarzenie nie zajdzie.

Twierdzenie 1 (Lemat Lovásza)**.** *Niech A_i będzie rodziną złych zdarzeń, których prawdopodobieństwo jest nie większe niż p. Niech każde A_i będzie niezależne od wszystkich A_j oprócz co najwyżej d z nich. Wtedy, jeśli $ep(d+1) < 1$, to $\mathbf{P}\left(\bigcap \overline{A_i}\right) > 0$, czyli możliwe jest, że wszystkie zdarzenia nie zajdą.*

Zobrazujemy możliwości zastosowania tego twierdzenia na przykładzie.

Zadanie 4. *Mamy rodzinę zbiorów 12-elementowych takich, że każdy element jest w co najwyżej 46 zbiorach. Czy da się tak pomalować elementy dwoma kolorami, by nie było zbioru jednokolorowego?*

Dowód. Tu mamy problem polegający na tym, że nie posiadamy żadnej kontroli nad tym, jak zbiory na siebie nachodzą. Nie wiemy, jak zmiana koloru jednego elementu wpływa na to, co się dzieje w różnych zbiorach. Na szczęście Lemat Lovásza może nam pomóc.

Pomalujmy elementy losowo z prawdopodobieństwem 1/2. Niech A_i to zdarzenie, że i-ty zbiór jest jednokolorowy. Wtedy
$$\mathbf{P}(A_i) = \frac{1}{2^{11}} = \frac{1}{2048}.$$
Zdarzenie A_i jest zależne od co najwyżej $45 \cdot 12 = 540$ innych zdarzeń, czyli żeby skorzystać z Lematu Lovásza musimy mieć $e \cdot \frac{1}{2048} \cdot 541 < 1$, a to akurat zachodzi. Zatem da się tak pomalować.

Oczywiście, istnieją tez bardziej szczegółowe wersje tego lematu, np. gdy prawdopodobieństwa złych zdarzeń są różne i wspólne ograniczenie jest za słabym założeniem, by coś udowodnić.

Jak widać na przykładzie przedstawionych zadań, metody probabilistyczne dostarczają użytecznych narzędzi do pomocy w sytuacjach, gdy mamy za mało informacji, by próbować coś dowodzić konstruktywnie. Co warto zapamiętać z tego artykułu? To co wielu kombinatoryków powtarza — gdy nie wiadomo co robić, należy losować.