Symetrie – XII Szkoła Matematyki Poglądowej, Miętne, 28.01.–01.02.1994

Symetrie w probabilistyce

Andrzej DĄBROWSKI, Wrocław

Zadanie o krzesłach – symetria rozwiązuje zadanie

Przy okrągłym stole, przy którym stoi n krzeseł, losowo zasiadło $k < n$ osób. Wśród nich pan Jan. Jaka jest oczekiwana liczba μ wolnych miejsc między panem Janem a jego sąsiadem?

Zadanie jest symetryczne względem obrotów i zależy jedynie od odstępów między zajętymi krzesłami. Oznaczmy krzesło, na którym siedzi pan Jan, numerem 1 i wybierzmy kierunek obrotu zgodny z ruchem wskazówek zegara. Niech X_i oznacza liczbę wolnych miejsc między osobą o numerze i a osobą o numerze $i+1$ (pan Jan jest również osobą o numerze $k+1$). Z symetrii wynika, że X_1 mają ten sam rozkład. Oczywiście, $X_1 + X_2 + \ldots + X_k = n - k$, stąd $n - k = E(X_1 + \ldots + X_k) = k\mu$. Oczekiwana liczba wolnych miejsc wynosi zatem $\mu = (n-k)/n$.

W zadaniu tym symetria, będąca synonimem losowości, znakomicie ułatwiła rozwiązanie zadania. Nie zawsze jednak, jak pokazuje paradoks Bertranda, można jednoznacznie wskazać, co jest losowe. Określenie losowości wymaga podania grupy symetrii, względem której prawdopodobieństwo jest niezmiennicze.

W fizyce statystycznej spotykamy się z zagadnieniem rozmieszczenia cząstek w komórkach przestrzeni.

W przestrzeni podzielonej na n komórek, znajduje się k cząsteczek; obliczyć prawdopodobieństwo, że utworzą one ustaloną konfigurację A.

Fizycy wyróżniają tu trzy modele:
- model Maxwella–Boltzmanna; każdy układ cząsteczek jest jednakowo prawdopodobny, cząsteczki są rozróżnialne;
- model Bosego–Einsteina; każdy układ cząsteczek jest jednakowo prawdopodobny, cząsteczki są nierozróżnialne;
- model Fermiego–Diraca; każdy układ cząsteczek jest jednakowo prawdopodobny, cząsteczki są nierozróżnialne, dwie cząsteczki nie mogą być umieszczone w tej samej komórce.

Rozmieszczenia cząsteczek w komórkach można opisać poprzez przekształcenie ze zbioru k elementowego w zbiór n elementowy. Model Maxwella–Boltzmanna jest symetryczny względem grupy, składającej się z przekształcenia tożsamościowego, model Bosego–Einsteina względem grupy przekształceń niezmienniczych na permutację zmiennych, natomiast model Fermiego–Diraca jest niezmienniczy względem tych przekształceń różnowartościowych, które są symetryczne na permutację zmiennych. Eksperymenty pokazały, że do żadnych znanych cząstek nie da się zastosować (ściśle) modelu Maxwella–Boltzmanna. Dla fotonów, nukleonów i atomów, zawierających parzystą liczbę cząstek elementarnych stosuje się model Bosego–Einsteina. Model Fermiego–Diraca stosuje się się do elektronów, protonów i neutronów.

Symetria może pojawić się również wtedy, gdy rozważamy ciąg zmiennych losowych. Zakładamy wtedy, że rozkład tych zmiennych jest niezmienniczy względem permutacji wskaźników. Tak się zdarzy, gdy zmienne są niezależne a rozkłady każdej z tych zmiennych są takie same. Najbardziej znanym przykładem skończonego ciągu o tej własności jest rozkład Bernoullego. Rozkład Bernoullego można traktować jako rozkład generatora liczb naturalnych: mówimy, że wygenerowano liczbę k spośród liczb $\{1, 2, \ldots, n\}$, jeśli uzyskano k sukcesów w schemacie n prób Bernoullego z prawdopodobieństwem sukcesu θ_n. Z kolei, gdy przyjmiemy założenie, że wraz ze wzrostem n stabilizuje się wielkość $n\theta_n$ i zbliża się ona do wartości λ, to otrzymamy generator losowy liczb naturalnych o wartości oczekiwanej λ, zwany rozkładem Poissona. Oba rozkłady

Można sobie wyobrazić, że w jednej jednostce czasu licznik zwiększa swoją wartość o 1 z prawdopodobieństwem θ_n (prawdopodobieństwo sukcesu), a więc θ_n oznacza też, jak często licznik zwiększa swoją wartość. Liczba $n\theta_n$ jest wtedy oczekiwaną liczbą zmian licznika w jednostce czasu.

Prawdopodobieństwo, że otrzymamy liczbę k, wynosi $\lambda^k e^{-\lambda}/k!$; λ można tu interpretować jako przeciętną liczbę zmian licznika w jednostce czasu.

są symetryczne względem translacji na osi czasu: prawdopodobieństwo zmiany wskazań licznika jest takie samo w każdym momencie i zdarzenia „zmieni się czy też nie zmieni stan licznika?" są niezależne.

Zmienne losowe X_i w rozkładzie Bernoullego mają wartości 0 i 1, są niezależne i mają jednakowy rozkład. Wtedy

$$P\{X_1 = 1, X_2 = 1, \ldots, X_k = 1, X_{k+1} = 0, X_{k+2} = 0, \ldots, X_n = 0\} = \theta^k(1-\theta)^{n-k}$$

jest symetryczne względem permutacji wskaźników. Prawdopodobieństwo uzyskania k jedynek (k sukcesów) w ciągu n-elementowym wynosi $P\{S_n = k\} = \binom{n}{k}\theta^n(1-\theta)^{n-k}$, gdzie θ jest prawdopodobieństwem sukcesu.

Odrzućmy niezależność, pozostawiając jedynie założenie o niezmienniczości rozkładu zmiennych przy permutacji wskaźników. Takie zmienne nazywane są symetrycznie zależnymi. Twierdzenie, udowodnione w latach 30-tych przez de Finettiego pokazuje, że niewiele tu odbiegniemy od rozkładu Bernoullego.

Definicja: Zmienna losowe $X_1, X_2, \ldots, X_n$ są symetrycznie zależne jeżeli wszystkie permutacje tego ciągu mają ten sam rozkład. Zmienne ciągu nieskończonego $\{X_n\}$ są symetrycznie zależne jeśli $X_1, X_2, \ldots, X_n$ są symetrycznie zależne dla każdego n.

Dla nieskończonego ciągu zmiennych losowych $\{X_j\}$ symetrycznie zależnych, o wartościach 0 i 1,

$$P\{X_1 = 1, X_2 = 1, \ldots, X_k = 1, X_{k+1} = 0, X_{k+2} = 0, \ldots, X_n = 0\} = \int_0^1 \theta^k(1-\theta)^{n-k}\,\mathrm{d}F(\theta),$$

$$P\{S_n = k\} = \binom{n}{k}\int_0^1 \theta^k(1-\theta)^{n-k}\,\mathrm{d}F(\theta).$$

Twierdzenie de Finettiego pokazuje, że zmienne symetrycznie zależne można uzyskać według schematu zbliżonego do schematu Bernoullego: najpierw, według rozkładu F wybieramy (losujemy) prawdopodobieństwo sukcesu θ, a następnie wykonujemy zwykły schemat Bernoullego z takim prawdopodobieństwem sukcesu.

Twierdzenie to ma ciekawą interpretację w pewnym modelu epidemii:

Wylosowanie kuli czarnej w chwili t odpowiada zachorowaniu jednostki, kuli białej zaś temu, że w tej chwili nikt nie zachorował; zachorowanie jednostki oznacza zwiększenie szansy zachorowania w populacji, niezachorowanie oznacza tendencję do wygasania choroby.

W urnie jest n kul, c czarnych i b białych. Po wylosowaniu kuli z urny wrzucamy ją z powrotem wraz z d kulami tego samego koloru. $X_t = 1$ oznacza, że w chwili t wylosowano kulę czarną. Zmienne losowe X_t są w sposób oczywisty zależne; są one też symetrycznie zależne. Twierdzenie de Finettiego orzeka iż prawdopodobieństwo tego, że do chwili t będzie k zachorowań jest równe prawdopodobieństwu k sukcesów w t próbach Bernoullego z prawdopodobieństwem sukcesu (zachorowania) wylosowanym z pewnego rozkładu (rozkładu beta), którego wartość oczekiwana wynosi $c/(c+b)$, a więc jest równa szansie zachorowania w chwili 0. Różnica tej sytuacji i zwykłego schematu Bernoullego polega na tym, że w nim prawdopodobieństwo sukcesu jest z góry znane, a w przypadku zmiennych symetrycznie zależnych należy je wylosować.

Twierdzenie nie jest prawdziwe dla zależnych i symetrycznie zależnych zmiennych losowych, jeśli jest ich skończona liczba.

Przykład (zadanie o kapeluszach)

W szatni wisi n kapeluszy. Z przyjęcia wyszli ich właściciele i losowo włożyli kapelusze na głowę. $X_n = 1$, gdy n-ta osoba założy właściwy kapelusz.

$P\{X_1{=}1, X_2{=}1\} = 1/(n(n-1))$,
$P\{X_1{=}1\} = 1/n$,
$\mu_k = P\{X_{i_1}{=}1, X_{i_2}{=}1, \ldots, X_{i_k}{=}1\} = (n-k)!/n!$
Łatwo przekonać się, że zmienne te są zależne: dla $n = 2$
$P\{X_2{=}1|X_1{=}1\} = 1$, $P\{X_2{=}1\} = 0{,}5$.

Zmienne te są zależne symetrycznie i nie są niezależne. Można jednak wykazać, że nie odpowiada im twierdzenie de Finettiego.

Miara Lebesgue'a, jako niezmiennicza względem grupy obrotów i przesunięć, jest dobrym modelem losowości w przestrzeniach euklidesowych. Niestety, nie można określić, zgodnie z aksjomatyką Kołmogorowa, miary probabilistycznej, jeśli zbiór zdarzeń elementarnych ma nieskończoną miarę Lebesgue'a. Można jednak podać taki system aksjomatów, uogólniający aksjomatykę przestrzeni probabilistycznej Kołmogorowa, że w każdym zbiorze B o skończonej mierze Lebesgue'a da się określić warunkowy rozkład jednostajny wzorem

Definicja ta dotyczy również dowolnej miary, na przykład miary liczącej. Definicję tę podał węgierski matematyk Alfred Renyi w 1955 roku.

$P(A|B) = |A \cap B|/|B|$, będący modelem losowości. Definicja ta idealnie pasuje do sytuacji, kiedy mamy do czynienia z losowymi układami obiektów geometrycznych (punkty, proste, obroty itp.).

Dla skończonego zbioru N punktów w przestrzeni p-wymiarowej możemy określić warunkowy rozkład jednostajny opierając się na Np-wymiarowej mierze Lebesgue'a. Ciekawym i bardzo użytecznym, choć mało używanym wynikiem jest twierdzenie Croftona z 1885 roku:

Oznacza to, że zdarzenie dotyczące układu punktów zależy tylko od ich wzajemnego położenia.

Losowo rozmieszczamy N punktów w p-wymiarowym, ograniczonym i domkniętym obszarze D. Niech zdarzenie A opisane będzie przez podzbiór przestrzeni R^{Np} symetryczny względem ponumerowania punktów. Wtedy $dP/dv = N(P_1 - P)/v$, gdzie v jest miarą obszaru D, P – poszukiwanym prawdopodobieństwem, że zajdzie zdarzenie A pod warunkiem, że $(N-1)$ punktów leży w obszarze D, a jeden z nich leży na brzegu D.

W tym równaniu różniczkowym P rozpatrujemy jako funkcję v – miary wszystkich możliwych zbiorów, powstałych z D przez wybrane przez nas przekształcenie jednoparametrowe, zmieniające miarę tych zbiorów. Zazwyczaj wybiera się rodzinę podobieństw.

Przykład

Jakie jest prawdopodobieństwo tego, że dwa punkty, rozmieszczone losowo w odcinku o długości a, są odległe nie więcej niż o u $(u < a)$?

Tu $N = 2$, $p = 1$. Rozważamy rodzinę odcinków podobnych o skali podobieństwa λ. Wtedy $v = \lambda$, P_1 jest prawdopodobieństwem, że odległość dwóch punktów jest mniejsza od u, gdy jeden z nich jest na brzegu. Wynosi ono u/v. Należy rozwiązać równanie $P' = 2(u/v - P)/v$. Rozwiązanie ma postać $P = 2uv^{-1} + cv^{-2}$, gdzie c jest stałą. Stałą c wyznaczamy z warunku: $P = 1$ gdy $v = u$, czyli $c = -u^2$, stąd $P = 2uv^{-1} - u^2v^{-2}$. Podstawiając $\lambda = 1$ uzyskamy odpowiedź: $P = 2ua^{-1} - u^2a^{-2}$.

Wystarczy zauważyć, że równanie ma postać $(v^2P)' = 2u$.

Mówiliśmy dotąd o losowych konfiguracjach skończonej liczby punktów. Często jednak spotykamy się z sytuacją, kiedy mówi się o losowym układzie punktów i nie jest określona ich liczba. Takim problemem jest oszacowanie gęstości losowych punktów na płaszczyźnie, na przykład gęstości roślin czy drzew na danym obszarze.

Będziemy mówić, że punkty są rozmieszczone losowo na płaszczyźnie, gdy w każdym obszarze o mierze Lebesgue'a A ich liczba ma rozkład Poissona o średniej λA. Parametr λ nazywany jest gęstością punktów na płaszczyźnie.

Chcąc otrzymać przeciętną liczbę punktów w obszarze należy gęstość λ pomnożyć przez A.

Niech O będzie dowolnym, ustalonym punktem na płaszczyźnie. Będziemy obserwować zachowanie się zmiennej D, określającej odległość do najbliższego, losowego punktu płaszczyzny. Zdarzenie $\{D > x\}$ oznacza, że w kole o środku O i promieniu x nie ma żadnego punktu losowego. Prawdopodobieństwo tego zdarzenia wynosi

$$\{D > x\} = \Lambda(x)^0 e^{-\Lambda(x)}/0! = e^{-\Lambda(x)},$$

gdzie $\Lambda(x) = \lambda(\pi x^2)$ i λ jest gęstością punktów na płaszczyźnie.

A więc dystrybuanta D ma postać $F_D(x) = P\{D < x\} = 1 - e^{-\Lambda(x)}$, a jej gęstość $f_D(x) = 2\lambda\pi x e^{-\Lambda(x)}$. Przeciętna odległość od losowego punktu wynosi $1/(2\sqrt{\lambda})$. Wynika stąd metoda oszacowania gęstości losowych punktów na płaszczyźnie:

$E(D) = \int x f_D(x)\,dx$

- wybierz punkt obserwacyjny i zmierz odległość od tego punktu do najbliższego punktu losowego,
- powtórz tę operację wielokrotnie, wybierając za każdym razem inny punkt obserwacyjny,
- oblicz średnią odległość $\overline{D}$,
- oszacowaniem λ będzie $(2\overline{D})^{-2}$, bo $\lambda = (2E(D))^{-2}$.

Zasada niezmienniczości i reguły decyzyjne w statystyce

Statystyka jest narzędziem w opracowywaniu wyników eksperymentów. Celem badania statystycznego jest bądź oszacowanie parametrów w populacji na podstawie próbki, bądź rozstrzygnięcie, która z postawionych hipotez jest prawdziwa.

Na przykład, porównujemy nowy lek na obniżenie ciśnienia ze starym. Obserwujemy dwie grupy chorych i w każdej grupie podajemy inny lek. Chorzy są tak wybrani, aby ich stan zdrowia był podobny, a jedyne zróżnicowanie wynikało z ich osobistych (genetycznych), nie podlegających analizie cech. O skuteczności leku świadczyć więc mogą średnie wartości ciśnienia μ_i po zakończeniu kuracji w całej populacji chorych. Interesuje nas, który lek po zakończeniu kuracji jest skuteczniejszy.

μ_1 jest średnią wartością ciśnienia w grupie leczonych nowym lekiem, μ_2 jest średnią wartością ciśnienia w grupie leczonych starym lekiem.

Należy zdecydować, czy hipoteza

$$H: \mu_1 \leq \mu_2,$$

jest prawdziwa, czy też prawdziwa jest hipoteza konkurencyjna

$$K: \mu_1 > \mu_2.$$

Prawdziwość hipotezy H oznacza, że nowy lek jest nie gorszy od starego, gdyż przeciętne ciśnienie w grupie leczonych tym lekiem jest niższe.

Decyzję podejmiemy na podstawie dwóch próbek $\{X_{11}, X_{12}, \ldots, X_{1n}\}$ i $\{X_{21}, X_{22}, \ldots, X_{2n}\}$. O tym, którą hipotezę bezpieczniej wybrać, powinna rozstrzygać funkcja decyzyjna o argumentach zależnych od obserwacji w próbkach. Hipotezy są symetryczne (niezmiennicze) względem wszystkich przekształceń rosnących i ciągłych w dziedzinie parametru μ. Symetrii tej odpowiada analogiczna symetria w przestrzeni próbek: każda rosnąca i ciągła funkcja nałożona na obserwacje ciśnienia w obu grupach chorych nie powinna zmieniać naszej decyzji o tym, który lek wybrać jako skuteczniejszy. Funkcja decyzyjna winna być więc niezmiennicza względem tej rodziny funkcji. Uwzględnienie specyficznej dla danego problemu symetrii pozwala zredukować liczbę możliwych funkcji decyzyjnych, co ma niebagatelne znaczenie praktyczne.

Ogólnie, gdy mamy do rozstrzygnięcia hipotezy

$$H: \theta \in \Omega_H$$
$$K: \theta \in \Omega_K$$

Ω_H (Ω_K) jest zbiorem parametrów w populacji spełniających warunek wyrażony w treści hipotezy H (K).

na podstawie próbki X,

to prawdziwe są następujące fakty:

- każdemu przekształceniu g różnowartościowemu i „na" w przestrzeni próbek odpowiada indukowane przez g przekształcenie g^* w $\Omega_H \cup \Omega_H$; Ω_H jest niezmiennicze względem g, gdy $g^*\Omega_H = \Omega_H$;
- niech G będzie najmniejszą klasą przekształceń 1-1 przestrzeni próbek, taką że są one niezmiennicze w Ω_H i Ω_K, zamkniętą względem superpozycji i brania odwrotności; wtedy klasa indukowanych przez nią przekształceń g^* w $\Omega_H \cup \Omega_K$ jest grupą homomorficzną z grupą G.

Funkcję T nazywamy maksymalnym niezmiennikiem względem grupy G gdy jest ona stała na orbitach każdego punktu w przestrzeni próbek i różna na różnych orbitach.

Interesujące będzie teraz wyznaczenie maksymalnych niezmienników grupy G. Powiedzą one bowiem, na jakie cechy próbki wystarczy zwrócić uwagę, aby rozstrzygnąć, która hipoteza jest prawdziwa.

Rangą obserwacji x_i nazywamy liczbę obserwacji w próbce nie większych od niej.

W naszym przykładzie maksymalnym niezmiennikiem jest zbiór rang obserwacji w próbce.

To znaczy, gdy zaobserwujemy wielkości $x_1, x_2, \ldots, x_n$ przyporządkowujemy im taki wektor $[x_{(1)}, x_{(2)}, \ldots, x_{(n)}]$, że $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$.

- Dla grupy permutacji, to znaczy gdy hipotezy są symetryczne względem ponumerowania obserwacji, maksymalnym niezmiennikiem jest zbiór uporządkowanych współrzędnych dla obserwacji.
- Dla grupy przesunięć na prostej (hipotezy nie zależą od ustalenia początku skali) maksymalnym niezmiennikiem dla próbki $\{x_1, x_2, \ldots, x_n\}$ jest $\{x_2 - x_1, \ldots, x_n - x_1\}$.
- Dla grupy przekształceń ortogonalnych (obrotów) maksymalny niezmiennik jest odległością tych punktów od początku układu współrzędnych.

Literatura

W. Feller, *Wstęp do rachunku prawdopodobieństwa*, PWN, Warszawa 1966 (t. I), 1969 (t. II)

M. Kendall, P. Moran, *Geometriceskije verojatnosti*, Nauka, Moskwa 1972.

E.L. Lehmann, *Testowanie hipotez statystycznych*, PWN, Warszawa 1968.