**ЛАБОРАТОРНА РОБОТА № 2**

**БАЙЄСІВСЬКИЙ АНАЛІЗ ДАНИХ**

**Теоретичні відомості**

Якщо у результаті досліду може відбутися декілька подій , то має місце **сукупність** або **група випадкових подій**. **Несумісними** (або взаємно виключеними) подіями є події, які одночасно відбуватися не можуть.

**Сумою двох подій**  і  називають подію , яка полягає в тому, що в результаті досліду відбувається хоча б одна (незалежно яка) з подій  або  чи  і . Це твердження є змістом **постулату адитивності ймовірності**.

Якщо події  і  – несумісні, то  – подія, що полягає в появі однієї з подій  або  (незалежно якої). Зауважимо, що таке означення суми двох подій повністю тотожне логічній операції ***АБО****.* Звідси випливає, що ймовірність появи однієї з двох несумісних подій  і  дорівнює сумі ймовірностей цих подій:

. (1)

Узагальнивши цю формулу для випадку  попарно несумісних подій , отримаємо

 (2)

Сукупність несумісних подій  утворює **повну групу**, якщо внаслідок досліду обов'язково відбувається одна з них, тобто якщо сума  є достовірною подією:

 (3)

Співвідношення (3) виражає зміст постулату **нормування ймовірності**.

Для **двох сумісних** події  і  з ймовірностями  і  та імовірністю їхньої сумісної появи , ймовірність суми подій  дорівнює сумі ймовірностей цих подій без імовірності їх спільної появи:

. (4)

**Добутком** подій  і  називають подію , яка полягає в тому, що в результаті досліду відбуваються як подія , так і подія . Означення добутку двох подій повністю співпадає з логічною операцією ***І*** *.*

Очевидно, що якщо  і  є несумісними, то  – неможлива подія.

Нехай при  випробуваннях подія  відбулася  разів, подія  –  разів, а у  випадках із  мали місце відразу обидві події  і . Тоді при  можна записати такі вирази для ймовірностей:

. (5)

Останній вираз у (5) визначає ймовірність сумісної реалізації подій  і .

Важливим є поняття **умовної ймовірності,** при якій реалізується подія  за умови, що відбулася (обов'язково!) подія :

.

Аналогічно можна записати й умовну ймовірність здійснення події  за умови, що подія  обов'язково має місце:

.

Оскільки , то між умовними і звичайними (тобто безумовними) ймовірностями має місце таке співвідношення:

. (6)

Умовну ймовірність називають **апостеріорною** (або «після дослідною» ймовірністю), а  – **апріорною** («перед дослідною»). З (6) отримаємо

. (7)

Дві події  і  називаються **незалежними**, якщо умовна ймовірність події  збігається з «безумовною» , тобто ймовірність появи події  не залежить від того, чи відбулася подія  (або навпаки ймовірність появи події  не залежить від того, чи відбулася подія ). При цьому з (6) випливає, що ймовірність добутку незалежних подій є добутком ймовірностей цих подій

. (8)

У загальному випадку для довільного числа незалежних подій множення ймовірностей визначається співвідношенням

 (9)

**Повна ймовірність. Формула Байєса.**

Виведемо вираз для обрахунку **повної ймовірності**. Вважатимемо, що подія  може відбутися тільки з однією із  несумісних подій , які утворюють повну групу і розглядаються як **гіпотези**, які пов’язані з появою події .

Оскільки події  та  при є несумісними, то подію  можна розглядати як суму подій , тому з теореми додавання ймовірностей (3) отримаємо:

. (10)

З теореми множення (8) випливає також, що , тому співвідношення (10) перепишеться у вигляді:

. (11)

Отримана формула повної ймовірності дозволяє порахувати ймовірність  події , якщо є відомими ймовірності  кожної гіпотези  та умовна ймовірність  події  за умови, що гіпотеза є правдивою.

Знаючи ймовірність  та умовні ймовірності  з теореми множення можна розрахувати умовні ймовірності  гіпотез :

.

З останнього виразу отримуємо:

.

Підставляючи в отриману формулу вираз (11) для , отримаємо відому так звану **формулу Байєса**:

 (12)

Як зазначалося вище, формула Байєса дає можливість оцінити апостеріорні ймовірності подій, тобто подій, які відбудуться з урахуванням попередньо проведених дослідів.

Проілюструємо застосування формули Байєса на прикладі. Розглянемо три урни, у яких є по 10 кульок білого та чорного кольору. Нехай у першій урні є 8 білих та 2 чорні кулі, у другій урні – 5 білих та 5 чорних куль, а у третій – 2 білі і 8 чорних куль. Далі проводимо експеримент, який полягає у тому, що випадково вибирається одна з урн, а потім із вибраної урни вибирається дві кульки. Завдання полягає у тому, щоб за результатами проведення експерименту встановити, яка з трьох урн була вибрана.

У розглядуваному прикладі існує три гіпотези , які полягають у тому, що вибрано одну з трьох урн. Апріорні ймовірності таких подій є рівними .

Нехай подія  полягає у тому, що вибрано 2 чорні кулі.

Тоді , , .

Тому для апостеріорних ймовірностей отримуємо такі значення:



Таким чином, з ймовірністю, рівною приблизно 0,72 можна стверджувати, що справедливою є третя гіпотеза, тобто була вибрана третя урна.

Аналогічним способом можна порахувати ймовірності інших подій , які, наприклад, полягають у тому, що було вибрано білу і чорну кульки (незалежно від порядку, подія ), вибрано 2 білі кульки ( подія ) і т.д. Оцінивши для цих подій апостеріорні ймовірності, можемо зробити висновок про ймовірність реалізації тої чи іншої події .

У цьому полягає суть **Байєсівського аналізу даних,** який ґрунтується на проведенні певного числа дослідів, за результатами яких обраховуються апостеріорні ймовірності гіпотез.

**Приклад з дайсом (dice)**

Припустимо, маємо 3 гральні кубики (далі **дайси**). Перший кубик має 4 грані, другий – 6 граней, третій – 8 граней. Вибираємо один дайс випадково, котимо і отримуємо результат – випала грань із числом 2. **Завдання полягає у тому, щоб вгадати, який кубик було вибрано.** Для цього потрібно обчислити ймовірність того, що дайс був з 4-ма гранями (4d дайс), з 6-ма (6d дайс) чи з 8-ма гранями (8d дайс).



Оскільки маємо 3 дайси, то *початкова ймовірність P(Вi)* вибору кожного з них однакова і дорівнює 1/3.

Зауважимо:

* якщо число, що випало у результаті кочення випадково вибраного дайсу, більше за розмірність дайсу, то ймовірність його появи на дайсі рівна 0 (тобто, якщо випало число 6, то ймовірність появи цього числа на 4d-дайсі нульова);
* якщо число, що випало, менше за розмірність дайсу, то ймовірність його появи на дайсі складає 1/(розмірність дайсу) (тобто, якщо випало число 6, то ймовірність появи цього числа на 8d-дайсі дорівнює 1/8).

Наше спостереження (*подія А*) – випало число 2. Якщо ми припускаємо, що кидали 4d-дайс, то ймовірність випадання числа 2 (*P(A|Bi))* дорівнює 1/4. Якщо 6d – то 1/6. Якщо 8d – то 1/8.

Перемножуємо отримані ймовірності:

* для 4d дайса - ймовірність вибору дайса 1/3, ймовірність випадання числа 2 - 1/4. Добуток дорівнює 1/3 \* 1/4 = 1/12.
* для 6d дайса: 1/3 \* 1/6 = 1/18.
* для 8d дайса: 1/3 \* 1/8 = 1/24.

Нормалізуємо результати. Сумарна ймовірність *P(А)* вибору одного з 3-х дайсів і випадання числа 2 складає:

1/12 + 1/18 + 1/24 = 13/72.

Це число менше за 1, бо ймовірність випадання числа 2 менша 1. Проте ми знаємо, що число 2 вже випало. Тому треба поділити ймовірність для кожного дайса на 13/72, щоб сума усіх ймовірностей для усіх дайсів була рівна 1. Цей процес відомий як *нормалізація*.

Після нормалізації знаходимо ймовірність *P(Bi|A)* того, що дайс є одним з вибраних:

4d дайс: (1/12) / (13/72) = (1\*72) / (12\*13) = 6/13

6d дайс: (1/18) / (13/72) = (1\*72) / (18\*13) = 4/13

8d дайс: (1/24) / (13/72) = (1\*72) / (24\*13) = 3/13

Отже, на початку розв’язання цієї задачі ми припустили, що ймовірність вибору кожного дайса дорівнює 33.3%. Після кочення одного з дайсів і випадання числа 2 отримали результат: ймовірність вибору 4d дайса - 46.1%, 6d дайса – 30.8%, 8d дайса – 23.1%.

Результати занесено у *таблицю:*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Дайс | Початкова ймов., *P(Вi)* | Ймовірність випадання числа 2, *P(A|Bi)* | *P(Вi)*\**P(A|Bi)* | *P(Bi|A)* |
| 4d | 1/3 | 1/4 | 1/12 | 6/13 |
| 6d | 1/3 | 1/6 | 1/18 | 4/13 |
| 8d | 1/3 | 1/8 | 1/24 | 3/13 |
| Сумарна ймовірність, *P(А)* | | | 13/72 |

**Завдання.**

1. Використовуючи описаний вище приклад із 4d-, 6d- і 8d-дайсами, складіть імовірнісну таблицю для випадання числа 5 (замість 2). Порівняйте отримані результати із прикладом для числа 2.
2. Прорахуйте варіант випадання чисел 2-5-4 у процесі трьох кочень випадково вибраного дайсу. Складіть імовірнісну таблицю, врахувавши те, що, оскільки спершу випало число 2, то в колонці *P(Вi)* нової таблиці для числа 5 стоятимуть числа з колонки *P(Bi|A)* таблиці для числа 2.
3. Порівняйте отримані у п.1 і п.2 результати, зробіть висновки.
4. Розгляньте випадок 6-ти дайсів: 4d, 6d, 8d, 10d, 12d та 20d. Згенеруйте випадкові результати 10 кочень випадково вибраного дайсу і знайдіть ймовірності вибору кожного конкретного дайсу. (Для обчислень можна використати Excel, щоб створити у цій програмі ймовірнісну таблицю).
5. Результати обчислень занесіть у таблицю:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **Дайс** → | 4d | 6d | 8d |  |
| Номер кидання  ↓ | Число, що спостерігаємо  ↓ | *P(Вi)* → | 1/6 | 1/6 | 1/6 |  |
| 1 |  | *P(B1|A)* → |  |  |  |  |
|  |  |  |  |  |  |  |
| 10 |  |  |  |  |  |  |

1. Проаналізуйте отримані результати. Зобразіть графічно залежності ймовірності вибору дайса *P(Bi|A)* від порядкового номера кочення (1÷10).
2. Зробіть висновки. Оформіть звіт.