

ЗАСТОСУВАННЯ ТЕОРЕМИ БАЙЄСА В ОБРОБЦІ МЕДИЧНИХ ДАНИХ

¹ Вінницький національний медичний університет ім. М. І. Пирогова
² Вінницький національний технічний університет

Анотація

Теорія ймовірностей завжди присутня у повсякденній практиці лікаря і є фундаментом доказової медицини. Для нівелювання можливості похибки методу медичного дослідження, одержання хибнопозитивного чи хибнонегативного результату може прислугуватись теорема Байєса, яка дозволяє перетворити результати дослідження в ймовірності подій. Дана робота присвячена можливості застосування теореми Байєса в обробці медичних даних.

Ключові слова: доказова медицина, ймовірність, онкомаркери, подія, дослідження, теорема Байєса, умовна ймовірність.

Abstract

The theory of probabilities is always present in the everyday practice of a doctor and is the foundation of evidence-based medicine. To eliminate the possibility of an error in the medical research method, obtaining a false-positive or false-negative result, Bayes' theorem can be used, which allows you to convert the research results into the probability of events. This work is devoted to the possibility of applying Bayes' theorem in the processing of medical data.

Key words: evidence-based medicine, probability, tumor markers, event, research, Bayes theorem, conditional probability.

Вступ

Розвиток ефективних методів лікування (ведення протоколів лікування), кількісний підхід до терапії багатьох захворювань, вимагав достовірних методів оцінки їх ефективності [1, 2]. Із середини 80-х років минулого століття почало розвиватися нове поняття – «Доказова медицина», яке передбачає використання математичних методів для опису біомедичних процесів. Ці методи призначені для виявлення закономірностей, притаманних біомедичним об'єктам, пошуку подібності і відмінності між окремими групами об'єктів, оцінки впливу на них різноманітних факторів і т. п.

Практично всі процеси, що вивчаються різноманітними науками, можна розглядати, як ймовірнісні [3]. Причина цього криється у тому, що при вивченні тих чи інших закономірностей важливу роль можуть відігравати фактори, що не є підконтрольними для дослідника і про їх поведінку він може тільки здогадуватися або якось оцінити. Особливо це проявляється у дослідженнях, пов'язаних із людським фактором, зокрема у медицині.

Теорія ймовірностей завжди присутня у повсякденній практиці лікаря. Він збирає статистичні дані і користується даними, зібраними іншими; робить висновки і ставить діагноз, обирає протоколи лікування і т. ін. – все це є похідними доказової медицини, в основі якої лежить доказова база, яка спирається на теорію ймовірностей.

Результати дослідження

Результати будь-якого медичного дослідження – це не випробування. Більше того, часто ці методи можуть виявити те, чого немає (хибнопозитивний результат), і не виявляють те, що є насправді (хибнонегативний результат). З іншого боку, при діагностиці використовують результати дослідження самі по собі і не враховують похибки методу. Для нівелювання вказаних недоліків може прислугуватись теорема Байєса [4], яка дозволяє перетворити результати дослідження в ймовірність подій:

- якщо нам відома ймовірність події і ймовірність хибнопозитивних і хибнонегативних результатів, то можна виправити похибки вимірювань;
- теорема співвідносить ймовірність події з ймовірністю певного результату. Ми можемо порівняти ймовірність події A , якщо відомий результат X (умовна ймовірність

$p(A|X)$) та ймовірність результату X , якщо дано подію A (умовна ймовірність $p(X|A)$).

Припустимо, що ми досліджуємо метод діагностики раку яєчника за допомогою аналізу крові на онкомаркери – специфічні білки/антигени. У нормі вони виділяються клітинами в незначній кількості, однак при наявності злоякісного процесу концентрація цих речовин різко зростає – онкомаркер виділяється як патологічними раковими клітинами, так і здоровими клітинами, що оточують новоутворення. Аналізи крові на онкомаркери мають досить високу інформативність при дотриманні рекомендацій перед забором крові. Недотримання цих рекомендацій може дати хибнопозитивний або хибнонегативний результат. Наявність запалення також може призвести до хибнопозитивного результату (8,4%).

Нехай маємо такі дані:

- 5% усіх жінок хворіють на рак яєчника (відповідно 95% не хворіють);
- 75% аналізів на онкомаркери виявляють захворювання, коли воно є (відповідно 25% не виявляють);
- 8,4% досліджень виявляють рак тоді, коли його немає (і, відповідно, 91,6% вірно визначають негативний результат).

За початковими даними оформимо таблицю (див. табл. 1)

Таблиця 1. Початкові дані щодо діагностики раку яєчника за допомогою аналізу крові на онкомаркери

	Хворіють (5%)	Не хворіють (95%)
Позитивний результат методу	75%	8,4%
Негативний результат методу	25%	91,6%

Як працювати з цими даними? Якщо у пацієнтки виявили захворювання, то дивимось у другий стовпець табл. 1: є 75% ймовірності того, що метод дав вірний результат, 25% ймовірність того, що результат дослідження невірний (хибнонегативний). Якщо у пацієнтки захворювання не виявили, тоді: з ймовірністю 8,4% можна сказати, що позитивний результат дослідження невірний, з ймовірністю 91,6% можна стверджувати, що пацієнтка дійсно здорова. Яка ж ймовірність того, що пацієнтка дійсно хвора?

Ймовірність істинно позитивного результату дорівнює добутку ймовірності захворювання на ймовірність того, що аналіз дійсно виявив захворювання. Аналогічний підхід можна застосувати і для визначення ймовірності хибнопозитивного результату. Результати перерахунків можна подати у вигляді таблиці (див. табл. 2).

Таблиця 2. Оброблені дані діагностики

	Хворіють (5%)	Не хворіють (95%)
Позитивний результат методу	Істинно позитивний: $\frac{5\% \cdot 75\%}{(100\%)^2} = 0,0375$	Хибнопозитивний: $\frac{95\% \cdot 8,4\%}{(100\%)^2} = 0,0798$
Негативний результат методу	Хибнонегативний: $\frac{5\% \cdot 25\%}{(100\%)^2} = 0,0125$	Істинно негативний: $\frac{95\% \cdot 91,6\%}{(100\%)^2} = 0,8702$

Яка ймовірність того, що пацієнтка дійсно хвора, якщо отримано позитивний результат аналізу крові на онкомаркери? Ймовірність події – це відношення кількості сприятливих подій результатів до загальної кількості всіх можливих результатів.

Ймовірність істинно позитивного результату 0,0375. Ймовірність позитивного результату дорівнює сумі ймовірностей істинно позитивного результату та хибнопозитивного: $0,0375+0,0798=0,1173$. Таким чином, ймовірність захворювання за позитивного результату аналізу крові обчислюється так: $\frac{0,0375}{0,1173} \approx 0,32$.

Тобто позитивний результат аналізу означає тільки те, що ймовірність наявності захворювання 32%, а не 75% (остання величина – передбачувана точність методу). Такий результат здається спочатку незрозумілим і дивним, але потрібно врахувати, що метод дає хибнопозитивний результат у 8,4% випадків, тому у вибірці буде досить багато хибнопозитивних результатів. Для рідкісної патології більшість позитивних результатів будуть хибнопозитивними.

Тепер опишемо хід наших міркувань за теоремою Байєса. Ця теорема дозволяє виправити результати дослідження у відповідності із спотворенням, яке вносять хибнопозитивні результати:

$$p(A|X) = \frac{p(X|A)p(A)}{p(X|A)p(A) + p(X|\bar{A})p(\bar{A})}. \quad (1)$$

Поясними значення ймовірностей, задіяних у формулі (1).

- $p(A|X)$ – ймовірність захворювання (A) за позитивного результату (X). Це саме те, що ми хочемо знати: яка ймовірність події у випадку позитивного результату. В нашому варіанті вона дорівнює 32%.
- $p(X|A)$ – ймовірність позитивного результату (X) у випадку, коли пацієнтка дійсно хвора (A). В нашому випадку це величина істинно позитивних 75%.
- $p(A)$ – ймовірність захворіти (5%).
- $p(\bar{A})$ – ймовірність не захворіти (95%).
- $p(X|\bar{A})$ – ймовірність позитивного результату дослідження у випадку, якщо захворювання немає.

Це величина хибнопозитивних – 8,4%.

Таким чином, щоб отримати ймовірність події, потрібно ймовірність істинно позитивного результату розділити на ймовірність усіх позитивних результатів. Можна спростити формулу (1):

$$p(A|X) = \frac{p(X|A)p(A)}{p(X)}. \quad (2)$$

$p(X)$ – це константа нормалізації, ймовірність будь-якого позитивного результату, чи буде діагноз справжнім при обстеженні хворих пацієнток (5%) чи хибнопозитивним при обстеженні здорових пацієнток (95%).

Висновки

Лікарі проводять медичні дослідження, щоб з'ясувати реальний стан речей (діагноз). Якщо дослідження досконалі та точні, тоді ймовірності дослідження і ймовірності подій співпадуть. Всі позитивні результати будуть дійсно позитивними, а негативні – негативними. Теорема Байєса враховує спотворені результати, виправляє помилки, відтворює генеральну сукупність та знаходить ймовірність істинно позитивного результату.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Мінцер О.П. Оброблення клінічних і експериментальних даних у медицині: навч. посібник / О.П. Мінцер, Ю.В. Вороненко, В.В. Власов - К.: Вища шк., 2003. - 350 с.
2. Медичні знання та прийняття рішень в медицині. Збірник методичних рекомендацій до практичних занять з медичної інформатики /Під ред. Проф. І. І. Хаїмзона. – Вінниця, 2007. – 180 с.
3. Личковський Е. І. Вища математика. Теорія наукових досліджень у фармації та медицині: підручник / Е. І. Личковський, П. Л. Свердан. – К.: Знання, 2021. – 476 с.
4. Сачанюк-Кавецька Н.В. Елементи теорії ймовірностей та математичної статистики. Ч. 1. Навчальний посібник /Н.В. Сачанюк-Кавецька, Л.І. Педорченко, Н.Б. Дубова. – Вінниця: ВНТУ, 2008. – 108 с.

Кавецька Анастасія Вячеславівна, Вінницький національний медичний університет, 1-й медичний факультет, 3-й курс, група 13 А, ankavecka@gmail.com

Сачанюк-Кавецька Наталія Василівна, к. т. н., доцент, Вінницький національний технічний університет, кафедра вищої математики, skn1901@gmail.com

Kavetska Anastasia V. – student, medical faculty, group 13-A, Vinnytsia National Medical University, Vinnytsia, e-mail: ankavecka@gmail.com

Sachaniuk-Kavetska Natalia V. Candidate of Technical Sciences, Associate Professor, Department of Higher Mathematics, Vinnytsia National Technical University, skn1901@gmail.com