УДК 681.3

**Д.Д. Плечистий, магістр**
*Житомирський інженерно-технологічний інститут*

# ОГЛЯД СУЧАСНОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ, ЩО РОЗВ'ЯЗУЄ ЗАДАЧІ ТЕОРІЇ ІМОВІРНОСТЕЙ ТА СТАТИСТИКИ

*(Представлено д.т.н., проф. Б.Б. Самотокіним)*

*Розглянуто сучасне програмне забезпечення, що розв'язує задачі теорії імовірностей та статистики. Сконцентровано увагу на таких аспектах, як реалізація інтерфейсу, архітектурна побудова. Наведена порівняльна характеристика розглянутих програмних прикладних систем за функціональним наповненням.*

Комп'ютер міцно тримає одну з ведучих позицій як інструмент наукового пошуку. Однією з важливих областей як для теоретичних досліджень, так і для практичного застосування є теорія імовірностей. Її закони та висновки використовуються математиками, конструкторами, лікарями, соціологами, психологами, економістами. Для практичних цілей особливе значення має статистика та статистичний аналіз.

За невеликий проміжок часу (80–90-ті роки XX сторіччя) була створена велика кількість програмного забезпечення як для дослідження теорії імовірностей, так і для практичних розрахунків, що пов'язані з нею так або інакше. В зв'язку з таким своєрідним поділом на теорію та практику програмне забезпечення (ПЗ) також можна поділити на два класи:

– універсальне математичне ПЗ – програми, що розв'язують різноманітні математичні задачі. Зазвичай такі пакети охоплюють якомога більше областей математики, хоча в разі необхідності використання не дуже відомих алгоритмів або методів їм бракує відповідної функціональності;

– спеціалізоване ПЗ – програми, що мають справу тільки з конкретними розділами тієї або іншої науки. В контексті теорії імовірностей така спеціалізація виділилась в спеціалізацію зі статистики. Тобто це програми, що за мету мають розв'язання суто практичних проблем.

Розглянуті автором в даному огляді програмні прикладні системи (ППС), що розгалужені за вказаною вище класифікацією, наведені на рис. 1.

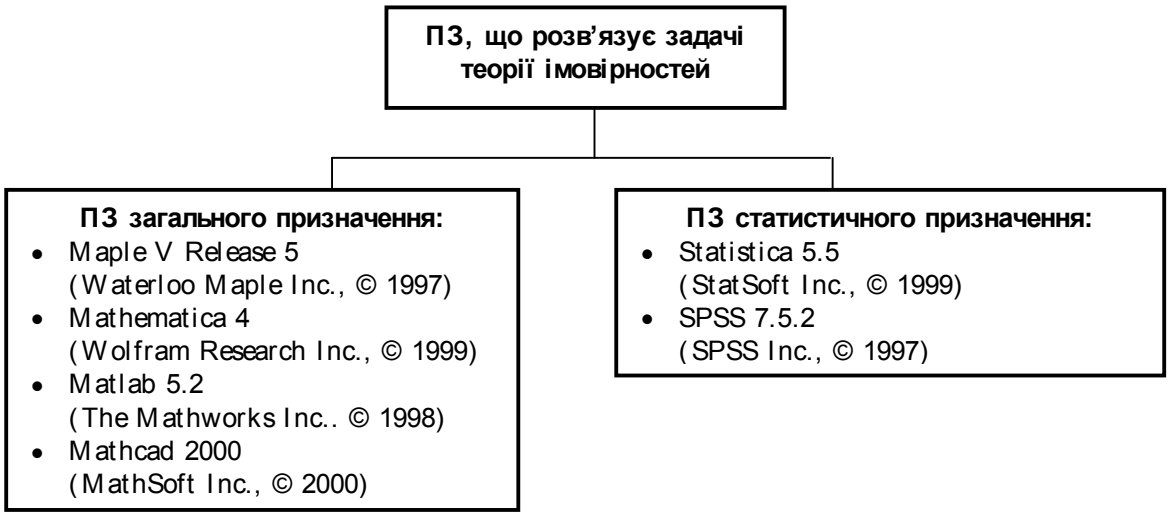


*Рис. 1. ПЗ, що розв'язує задачі теорії імовірностей*

Наведений список не претендує на повноту – для розгляду всіх програмних засобів знадобився би товстий підручник, але вказані вище програми на сьогоднішній день є найбільш відомими, розповсюдженими та функціонально повними.

Далі кожна з ППС розглянута окремо з точки зору архітектури, реалізації інтерфейсу та інших особливостей, після чого наведена узагальнююча порівняльна характеристика ППС за функціональним наповненням (тут і далі в огляді функціональне наповнення розглядається в контексті теорії імовірностей та відповідних алгоритмів аналізу).

**Програмна система "Maple V Release 5"**

Як математичний пакет загального призначення Maple V має як потужну обчислювальну здатність, так і велику кількість алгоритмів.

© Д.Д. Плечистий, 2001

В основу інтерфейсу користувача програмного засобу покладена так звана концепція “записника” – користувач за допомогою мови Maple V описує задачу, її умови та отримує відповідь, що записана тією ж мовою.

Мова програми дуже насичена, але для нормальної роботи користувачу потрібно вивчити її, щоб вміти записати та прочитати необхідну інформацію. Варто згадати незручне чергування квадратних та круглих дужок, які використовуються при викликах функцій, що призводить до неабиякої плутанини. Ось приклад виклику функції в Maple V:

P2:= random[ normald ]('generator[2]'): 'P2()'$11;

В архітектурному плані весь спектр математичних функцій, закладених в Maple V, розподілено на частини-модулі. Таким чином, крім знання конкретної необхідної функції, ще потрібно знати, в якому модулі вона знаходиться. Великі модулі, в свою чергу, можуть розгалужуватися на декілька підмодулів.

До позитивних рис Maple V можна віднести:

- Стабільність, стійкість роботи;
- Відмінну довідкову систему, яка дозволяє швидко та зручно відшукати потрібну інформацію;
- Широкий діапазон задач, що розв’язуються;
- Наявність панелі інструментів, створеної для спрощення введення математичних виразів.

До недоліків даного продукту можна віднести:

- Надто складну мову;
- Помилки, що іноді зустрічаються в довідковій системі.

Приклад роботи Maple V наведено на рис. 2.

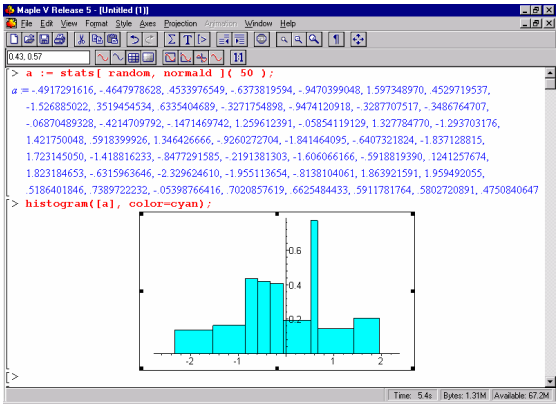


*Рис. 2. Приклад роботи ППС “Maple V”*

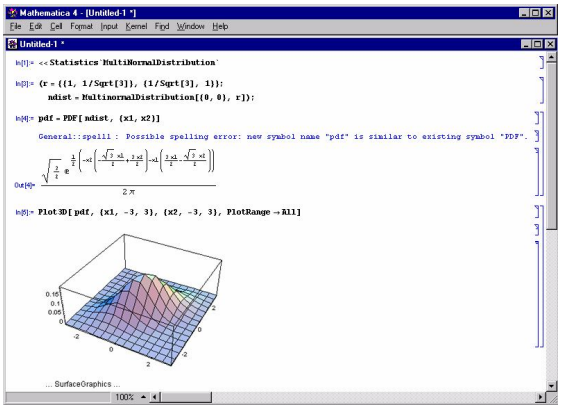


*Рис. 3. Приклад роботи ППС “Mathematica”*

**Програмна система “Mathematica 4”**

У ППС Mathematica інтерфейс користувача за принципом такий самий, як і в Maple V – “записник”. Однією з відмінностей є “equation builder”, що значно спрощує роботу користувача і також надає запису задачі вигляд, схожий з математичною нотацією.

Власна мова присутня, вона так само різноманітна та її так само складно вивчити. Як і у Maple V, мова Mathematica надає користувачу одразу всі конструкції та вирази, незалежно від того, яку проблему він розв’язує. Це є перевагою для досвідченого користувача і недоліком – для починаючого.

На архітектурному рівні Mathematica так само побудована на основі модулів, кожний з яких відповідає за той чи інший розділ математики.

Позитивні риси Mathematica:

- Зручна, відмінно систематизована довідкова система;
- Деякі найбільш часто використовувані функції віддубльовані в окремих модулях – очевидно, з метою запобігання завантаження непотрібних функцій та економії пам’яті;
- Широкий діапазон задач, що розв’язуються; один із найбільш функціонально повних пакетів.

Негативні риси Mathematica:

- Нестійка робота програми – якщо припуститися помилки при описі задачі, то трапляються випадки, коли, навіть якщо виправити її, для продовження роботи доводиться створювати новий документ;
- Громіздка мова;
- Незручно реалізований “записник” – без детального ознайомлення з посібником користувача складно інтуїтивно порозумітися з інтерфейсом та мовою.

Приклад роботи Mathematica 4 наведено на рис. 3.

**Програмна система "Matlab 5.2"**

Даний програмний продукт в значній мірі відрізняється від вже розглянутих. Для того, щоб розв'язати задачу за допомогою Matlab, необхідно створити програму на мові Matlab у вигляді так званого М-файлу. Декілька М-файлів можна зібрати в колекцію. На основі колекцій створені деякі з інструментів, що написані за допомогою Matlab та інстальовані разом з ним як приклади (наприклад, імовірнісний калькулятор). Як і у вже розглянутих ППС, в Matlab мова дає користувачу одразу весь набір інструментів для розв'язання будь-яких задач, але ж для недосвідченого користувача це зайве.

Розробники архітектурно ППС розбивають Matlab на набори інструментів (toolbox). Так для теорії імовірностей та статистики є відповідний statistics toolbox.

Інтерфейс програми є гранично простим – користувач бачить перед собою командну строку, де він вводить послідовність команд на виконання. Результати виконання можуть бути виведені як у тому самому, так і у окремому вікні.

Ізюминкою Matlab є можливість конвертування М-коду (програми на мові Matlab) в код мови С – можливо, так розробники Matlab намагаються компенсувати недоліки власної мови.

Переваги Matlab:

- Достатнє для розв'язку загальних типів задач функціональне наповнення;
- Можливість конвертації М-коду в мову С;
- Простий, що добре підходить для такої організації процесу опису задачі та її розв'язку, інтерфейс.

Недоліки Matlab:

- Вкрай незручна довідкова система – відсутність виділення головних термінів, гіпертексту, звичної системи навігації;
- Громіздка мова;
- Повна перевага програмування, відсутність в самому Matlab будь-яких діалогових засобів, що спрощують розв'язок задач (хоча їх можна створити за допомогою даного продукту).

Приклад використання Matlab 5.2 наведено на рис. 4.

**Програмна система "Mathcad 2000"**

Дана ППС використовує систему інтерфейсу "робочий аркуш" (spreadsheet). На одному документі розміщені текст, розрахунки, графіка з широкими можливостями форматування. Таким чином, в Mathcad можна не тільки виконувати розрахунки, але й створювати готовий документ для статті, роботи чи презентації.

Опис та розв'язання задач в даній ППС організовано простим та зрозумілим для користувача способом – введено набір діалогових засобів, які дозволяють інтуїтивно записати умови задачі у вигляді, що є схожими до математичної нотації. Звичайно, своя мова у Mathcad є, але вона несе менше навантаження, ніж у попередніх пакетах. Іноді виявляється такий недолік, як невідповідність між інформацією, що введена безпосередньо з клавіатурі, та тією, що введена за допомогою діалогових засобів.

В цілому даний пакет справляє гарне враження і поєднує в собі всі найкращі якості комп'ютерної індустрії, але його слабким місцем є повнота в контексті алгоритмів та методів розв'язку задач теорії імовірностей – присутній лише невеликий набір (в порівнянні з наведеними раніше ППС) найбільш загальних методів.

До позитивних рис Mathcad 2000 можна віднести:

- Зручний, інтуїтивно зрозумілий інтерфейс користувача;
- Зручну довідкову систему.

До негативних рис можна віднести:

- Невелику кількість доступних методів та алгоритмів розв'язку задач;
- Неповна відповідність між інформацією, що вводиться з клавіатури і через діалогові засоби.

Приклад роботи в Mathcad 2000 наведено на рис. 5.

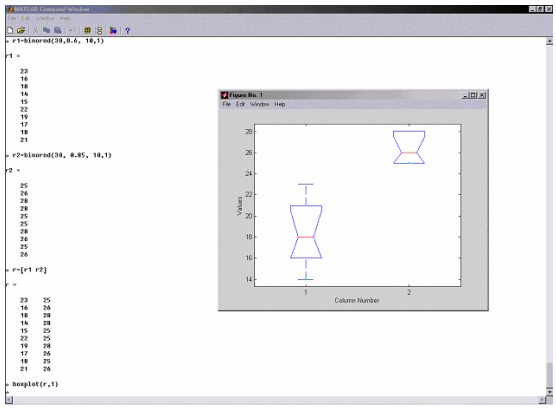

*Рис. 4. Приклад роботи ППС "Matlab"*

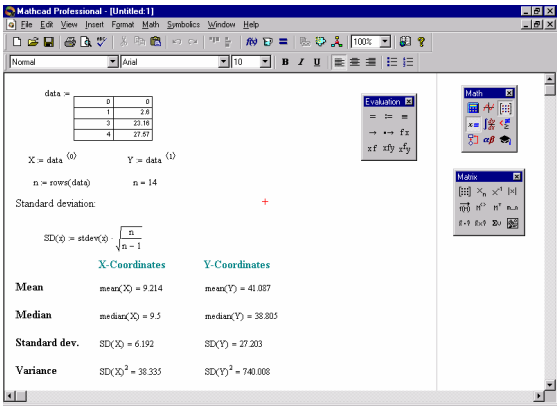

*Рис. 5. Приклад роботи ППС "Mathcad"*

**Програмна система "STATISTICA 5.5"**

Даний продукт є націленим лише на статистичну обробку даних. Джерелом даних і основним документом є таблиця, над стовпцями та рядками якої і виконуються аналіз та розрахунки.

Приємною рисою даної програми є постійний супровід користувача – при запуску він обирає розділ статистики, з яким він буде працювати, і надалі постійно працює з тими чи іншими діалогами, тобто програма не дає користувачу "заблукати". Хоча є можливість і працювати без такого супроводу. Для користувачів, які, незважаючи на всі доступні діалогові засоби, бажають програмувати, присутня внутрішня мова Statistica.

Згадавши вибір розділу статистики при запуску програми, можна зрозуміти, що в архітектурному плані використовується розмежування на певні модулі, але для користувача воно є прозорим.

Довідкова система робить і без того нескладну роботу з програмою абсолютно зрозумілою – для кожного діалогу є детальний опис. До речі, для різних методів та алгоритмів аналізу також існують доступні пояснення з практичними прикладами.

Без сумнівів, як інструмент статистичного аналізу, Statistica виконує поставлені завдання та заслуговує своєї назви.

Позитивні риси Statistica 5.5:

- Дружній, інтуїтивно-зрозумілий інтерфейс;
- Зручна довідкова система, яка дозволяє порозумітися як з особливостями праці програми, так і з суттю тих чи інших методів аналізу;
- Велика кількість методів аналізу, статистичних графіків.

Негативні риси Statistica 5.5:

- Якщо у вже запущеній програмі обрати перехід до іншого розділу, то буде створена дочірня копія програми, і тоді закрити основну програму, не закривши всіх дочірніх копій, буде неможливо.

Приклад роботи в Statistica 5.5 наведено на рис. 6.

**Програмна система "SPSS 7.5.2"**

Дана ППС концептуально не відрізняється від Statistica. Вона також є спеціалізованим статистичним пакетом, а тому як інтерфейс, так і функціональне наповнення у неї схожі з Statistica.

В програмі не можна виділити якогось чіткого розмежування на модулі, всі функції та методи аналізу доступні водночас, незалежно від задачі, що розв'язується тому очевидно, що всі алгоритмічні засоби завантажуються одним блоком.

Статистичні дані представлені у вигляді таблиці, за стовпцями та рядками якої виконуються різні види аналізу. Проводити обчислення можливо як через діалогові засоби, так і за допомогою внутрішньої мови SPSS. Результати обчислень наводяться в окремому вікні та згруповані у вигляді звіту. Необхідно згадати також, що, згідно з документацією SPSS, деякі з операцій неможливо виконати діалоговими засобами, а лише при використанні мови.

До позитивних рис SPSS 7.5.2 можна віднести:

- Велику кількість методів аналізу, статистичних графіків;
- Достатньо простий інтерфейс.

До негативних рис SPSS 7.5.2 можна віднести:

- Незручну навігацію в довідковій системі.

Приклад роботи в SPSS 7.5.2 наведено на рис. 7.

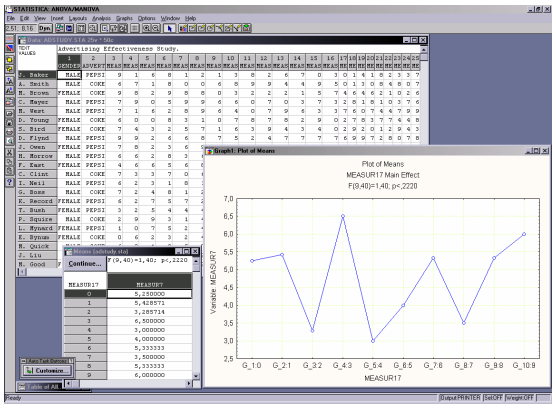


*Рис. 6. Приклад роботи ППС "Statistica"*

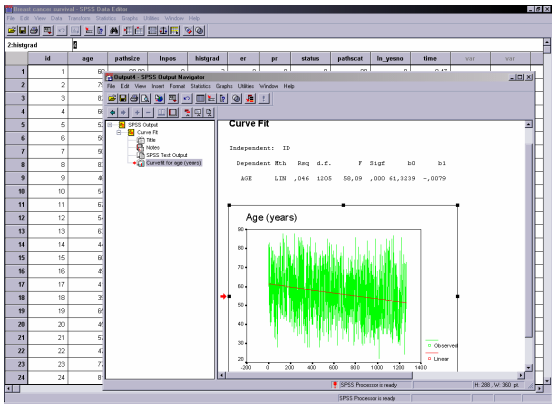


*Рис. 7. Приклад роботи ППС "SPSS"*

**Порівняльна характеристика**

Отже, вище були розглянуті 6 програмних систем, що розв'язують задачі теорії імовірностей, а саме: 4 системи загального призначення (Maple V Release 5, Mathematica 4, Matlab 5.2, Mathcad 2000) та 2 спеціалізовані статистичні системи (Statistica 5.5 та SPSS 7.5.2). Порівняльна характеристика різноманітних методів та алгоритмів аналізу, доступних в розглянутих ППС, наведена в табл. 1. Завдяки тому, що системи загального призначення та спеціалізовані статистичні системи сильно відрізняються як з боку проблем, для розв'язання яких вони створювалися, так і з боку присутнього в них функціонального наповнення, характеристика їм надається окремо, хоча й в одній таблиці.

*Таблиця 1*

*Порівняльна характеристика програмного забезпечення, що розв'язує задачі теорії імовірностей та статистики*

| **Програмні системи загального призначення** | | | | |
|---|---|---|---|---|
| Алгоритми аналізу | Програмні системи | | | |
| | 1 Maple V | 2 Mathematica | 3 Matlab | 4 Mathcad |
| Розрахунок імовірностей випадкових подій | | | | |
| Розрахунок загальних імовірнісних характеристик (мат. очікування, дисперсія, ексцес, асиметрія тощо) | √ | √ | √ | √ |
| Розрахунок моментів, медіан, процентилей, квартилей тощо. | √ | √ | √ | |
| Лінійна регресія | √ | √ | √ | |
| Аналіз дисперсії | √ | √ | √ | |
| Кількість розподілів, що підтримуються | 19 | 29 | 20 | 18 |

*Закінчення таблиці 1*

| | | | | |
|---|---|---|---|---|
| Генерація випадкових чисел за заданим законом розподілу | √ | √ | √ | √ |
| Розрахунок статистичних функцій (функція щільності імовірності, функція розподілу імовірності, зворотна функція розподілу імовірності) | √ | √ | √ | √ |
| Графічне подання, аналіз даних | √ | √ | √ | √ |
| Розрахунок довірчих інтервалів | | √ | √ | |
| Трансформація даних (розрахунок таблиці частот, нормалізація тощо) | √ | √ | √ | |
| Згладжування даних | | √ | | √ |
| Порівняльний аналіз даних (кореляція, коваріація) | √ | √ | √ | √ |
| Перевірка гіпотез | | √ | √ | |

*Таблиця 2*

Легенда: √ – елемент присутній; пуста клітина – елемент відсутній; * – можливості недостатньо документовані.

| **Спеціалізовані програмні системи** | | |
|---|---|---|
| Алгоритми аналізу | Програмні системи | |
| | 1<br>Statistica | 2<br>SPSS |
| Загальна статистика (мат. очікування, дисперсія, процентилі, таблиці частот тощо) | √ | √ |
| Кореляція, коваріація, кростабуляція | √ | √ |
| Аналіз дисперсії | √ | √ |
| Часові ряди, передбачення даних | √ | √ |
| Кластерний аналіз | √ | √ |
| Факторний аналіз | √ | √ |
| Канонічний аналіз | √ | |
| Дискримінантний аналіз | √ | √ |
| Багатовимірні трансформації даних | √ | * |
| Класифікаційні дерева | √ | |
| Аналіз відповідності | √ | √ |
| Виявлення лінійних взаємозв'язків між даними | √ | |
| Аналіз надійності | | √ |
| Аналіз виживання | √ | √ |
| Регресія | √ | √ |
| Порівняння даних та розподілів | √ | √ |
| Графічне подання, аналіз даних | √ | √ |

**Висновок**

Розглянуто сучасне програмне забезпечення, що розв'язує задачі теорії імовірностей та статистики. Проведений аналіз показав, що жодна із ППС загального призначення не є водночас зручною у використанні та функціонально повною. Розглянуті ППС статистичного призначення і зручні, і потужні, але їх ціна досить висока. Тому сьогодні на ринку існує потреба в якісному, зручному, потужному та порівняно недорогому програмному забезпеченні, здатному розв'язувати означені задачі.

**ЛІТЕРАТУРА:**


1. *Потемкин В.Г.* Система MATLAB 5 для студентов. – М.: ДИАЛОГ-МИФИ, 1998. – 314 с.
2. *John H. Mathews, Kurtis Fink.* Numerical Methods Using MATLAB. 3rd Edition, Pearson Education, 1999. – 680 p.
3. *Gaylord R.J., Kamin S.N. and Wellin P.R.* Programming with Mathematica. 2nd Edition, Springer-Verlag, TELOS, Santa Clara, 1996.
4. *Stephen Wolfram.* Mathematica Book, 4th Edition. Cambridge University Press, 1999. – 1470 p.
5. *Hecol K.M., Hansen M.L., Rickard K.M.* Maple V. Learning Guide. Springer-Verlag, 1996. – 274 p.
6. *Monagan M.B., Geddes K.O., Heal K.M., Labahn G., Vorkoetter S.M.* Maple V. Programming Guide. Springer-Verlag, 1996. – 383 p.
7. StatSoft, Inc. (1995). Statistica for WINDOWS [Computer program manual].
8. MATHCAD 6.0 PLUS. Финансовые, инженерные и научные расчеты в среде Windows 95: Пер. с англ. – М.: Информационно-издательский дом "Филинъ", 1996. – 712 с.
9. *Тюрин Ю.Н., Макаров А.А.* Статистический анализ данных на компьютере. – М.: ИНФРА-М, 1998. – 528 с.
10. *Колодницкий Н.М., Левицкий В.Г.* Обзор основных программных средств для моделирования математических задач // САПР и графика. – 1999. – № 10. – С. 56–65.


ПЛЕЧИСТИЙ Дмитро Дмитрович – магістр комп'ютерних наук, випускник Житомирського інженерно-технологічного інституту.

Наукові інтереси:

– комп'ютерно-інформаційні технології.

Подано 20.09.2001