

Міністерство освіти і науки України
Чорноморський національний університет імені Петра Могили

П.І. Бідюк, І.О. Калініна, О.П. Гожий

БАЙЄСІВСЬКИЙ АНАЛІЗ ДАНИХ

Монографія

Миколаїв - 2021

УДК 519.226

Б 59

*Рекомендовано до друку
вченою радою Чорноморського національного університету імені Петра Могили
(протокол № 9 від 13.10.2021 р.)*

Рецензенти:

Крак Ю.В. д. фіз.-мат.н., професор, завідувач кафедри теоретичної кібернетики Київського національного університету імені Тараса Шевченка, член-кореспондент НАН України

Теленик С.Ф. д.т.н., професор, декан факультету інформатики та обчислювальної техніки Національного технічного університету України «КПІ імені Ігоря Сікорського»

Цюцюра С.В. д.т.н., професор, завідувача кафедри інформаційних технологій Київського національного університету будівництва та архітектури

Бідюк П.І.

Б 59 Байєсівський аналіз даних : монографія / П. І. Бідюк, І. О. Калініна, О. П. Гожий. – Херсон: Книжкове видавництво ФОП Вишемирський В.С., 2021. – 208 с.

ISBN 978-617-7941-52-0

В монографії з позицій байєсівського підходу представлені сучасні методи і технології аналізу даних. Представлено вступ до алгебри потенціалів і обчислення базових ймовірностей для випадкових змінних при використанні для моделювання ймовірнісно-статистичних методів, а також безпосередній зв'язок умовних ймовірностей з теоремою Байєса. Розглянуто можливі варіанти формулювання теореми Байєса для випадків дискретних даних і дискретних параметрів, неперервних даних і дискретного параметра, неперервних даних і неперервних параметрів. Детально розглянуті процедури оцінювання можливих типів апіорних розподілів, необхідних для подальшого застосування вибраного варіанту теореми Байєса. Досліджено питання використання спеціальних алгоритмів Монте-Карло для генерування псевдовипадкових послідовностей з прогнозних розподілів і процедуру генерування неперервних випадкових величин. Представлено процедуру створення ймовірнісно-статистичної моделі у формі нормального розподілу. Докладно представлені модифіковані алгоритми генерування псевдовипадкових величин Метрополіса-Хастінгса та Гіббса, а також приклади їх практичного застосування. Запропоновано практично корисну методичку аналізу, моделювання і прогнозування вказаних процесів. Для формального опису волатильності фінансових процесів запропонована модифікована порогова модель стохастичної волатильності, виконано докладний аналіз можливостей застосування цієї моделі. Кожний розділ монографії супроводжується чисельними прикладами моделювання.

Монографія призначена для наукових співробітників, які займаються розробкою та практичним аналізом даних, а також для студентів, аспірантів комп'ютерних спеціальностей.

УДК 519.226

ISBN 978-617-7941-52-0

© П. І. Бідюк, І. О. Калініна, О. П. Гожий, 2021
© ФОП Вишемирський В. С., 2021

Зміст

ПЕРЕЛІК СКОРОЧЕНЬ	5
ВСТУП	7
Розділ 1. ВСТУП ДО БАЙЄСІВСЬКОГО АНАЛІЗУ ДАНИХ	11
1.1 Вступ в ймовірнісний аналіз.....	11
1.2 Методологія байєсівського програмування.....	12
1.3 Схема байєсівського підходу до моделювання та прийняття рішень.....	15
1.4 Невизначеності у математичному моделюванні.....	17
Розділ 2. ЗАДАЧІ ТА ОСОБЛИВОСТІ	
ЙМОВІРНІСНО-СТАТИСТИЧНОГО АНАЛІЗУ ДАНИХ	19
2.1 Задачі, які розв'язують ймовірнісно-статистичними методами.....	19
2.2 Можливі помилки при використанні ймовірнісно-статистичних методів.....	21
2.3 Ймовірнісні та детерміновані моделі.....	23
2.4 Безумовна незалежність подій.....	28
2.5 Правила обчислення ймовірностей.....	28
2.6 Умовні ймовірності та їх зв'язок з теоремою (правилом) Байєса.....	32
2.7 Розширення правила додавання (другого правила обчислення ймовірностей).....	39
2.8 Правдоподібність і ймовірність.....	42
2.9 Обчислення ймовірностей для змінних.....	43
2.10 Вступ до алгебри потенціалів.....	54
2.11 Випадкові змінні.....	56
Розділ 3. ДЕЯКІ ВАРІАНТИ ТЕОРЕМИ БАЙЄСА	59
3.1 Функція і принцип правдоподібності.....	59
3.2 Теорема Байєса для довільних дискретних подій.....	65
3.2.1 Теорема Байєса для взаємодоповнюючих подій.....	66
3.2.2 Роль апіорних ймовірностей.....	66
3.2.3 Апостеріорні ймовірності.....	67
3.2.4 Шанс настання події.....	67
3.3 Теорема Байєса для дискретних даних та дискретних параметрів.....	71
3.3.1 Інтерпретація теореми Байєса для дискретних даних і дискретних параметрів.....	71
3.3.2 Теорема Байєса для дискретних даних і дискретних моделей.....	74
3.4 Теорема Байєса для неперервних даних і дискретного параметра.....	75
3.5 Теорема Байєса для дискретних даних і неперервного параметра.....	78
3.6 Теорема Байєса для неперервних даних і неперервного параметра.....	81



Розділ 4. ОЦІНЮВАННЯ (ВИБІР) АПРІОРНИХ РОЗПОДІЛІВ	87
4.1 Об'єктивні апріорні розподіли	87
4.2 Суб'єктивні апріорні розподіли.....	90
4.3 Спрощені (невизначені) апріорні розподіли для одного параметра	92
4.4 Природні (натуральні) спряжені апріорні розподіли	96
4.5 Апріорні розподіли на основі даних	101
4.6 Приклади розподілів на основі моделей з одним параметром	102
4.7 Ймовірнісний висновок для даних, що характеризуються заміщенням	109
4.8 Приклад застосування моделі Пуассона.....	115
Розділ 5. АЛГОРИТМИ МОНТЕ-КАРЛО ДЛЯ МАРКОВСЬКИХ ЛАНЦЮГІВ	121
5.1 Метод Монте-Карло.....	121
5.2 Апостеріорний висновок для довільних функцій.....	125
5.3 Генерування значень з прогнозного розподілу.....	126
5.4 Генерування неперервних випадкових величин	129
5.4.1 Рівномірний розподіл	129
5.4.2 Експоненціальний розподіл	130
5.4.3 Гамма розподіл	130
5.5 Модель у формі нормального розподілу	130
5.6 Висновок щодо середнього і дисперсії нормального розподілу	132
5.7 Спільний висновок щодо середнього і дисперсії нормального розподілу	138
5.8 Алгоритм Метрополіса-Хастінгса (<i>Metropolis-Hastings</i>)	143
5.9 Приклад застосування алгоритму М-Х (бінарні дані).....	148
5.10 Алгоритм Гіббса для генерування марковських ланцюгів	154
Розділ 6. ПРОГНОЗУВАННЯ ВОЛАТИЛЬНОСТІ ЗА СТОХАСТИЧНОЮ МОДЕЛЮ.....	157
6.1 Вступ	157
6.2 Порогова модель стохастичної волатильності.....	159
6.3 Байєсівський аналіз порогової моделі стохастичної волатильності	160
6.4 Дослідження моделі стохастичної волатильності	167
6.5 Аналіз чутливості моделі до припущень	172
6.6 Застосування методу до реальних даних	175
6.7 Умови імітаційного моделювання.....	184
Перелік посилань.....	187
Додаток А	195
Додаток Б	211

ПЕРЕЛІК СКОРОЧЕНЬ

АКФ	– автокореляційна функція
АР	– авторегресія
АРУГ	– авторегресія з умовною гетероскедастичністю
АФК	– адаптивний фільтр Калмана
БАД	– байєсівський аналіз даних
БК	– байєсівський класифікатор
БМ	– байєсівська мережа
БМД	– байєсівська мережа довіри
ВП	– випадковий процес
ДБМ	– динамічна байєсівська мережа
ДНК	– дезоксирибонуклеїнова кислота
ГФ	– гранулярний (particle) фільтр
ІАД	– інтелектуальний аналіз даних
ІМ	– імітаційна модель
МАП	– максимальна абсолютна похибка
МБ	– мережа Байєса
МГВА	– метод групового врахування аргументів
МКМЛ	– метод Монте-Карло для марковських ланцюгів
ММП	– метод максимальної правдоподібності
ММС	– максимізація математичного сподівання
МНК	– метод найменших квадратів
МП	– максимуму правдоподібності
МСВ	– модель стохастичної волатильності
М-Х алгоритм	– алгоритм Метрополіса-Хастінгса
НМНК	– нелінійний метод найменших квадратів
ПЗ	– програмне забезпечення
ПММ	– прихована марковська модель
ПМСВ	– порогова модель стохастичної волатильності
РКФ	– розширений фільтр Калмана
САП	– середнє абсолютне значення похибки
САПП	– середня абсолютна похибка у процентах



СКП	– середньоквадратична похибка
СППР	– система підтримки прийняття рішень
СРЙ	– спільний розподіл ймовірностей
ТБ	– теорема Байєса
ТЙ	– теорія ймовірностей
ТУІ (або ТУЙ)	– таблиця умовних імовірностей
УАРУГ	– узагальнена авторегресія з умовною гетероскедастичністю
УЙР	– умовний ймовірнісний розподіл
ФК	– фільтр Калмана
ФП	– функція правдоподібності
ЦГТ	– центральна гранична теорема
ЦФ	– цифровий фільтр
ЧАКФ	– часткова автокореляційна функція
ЩРЙ	– функція щільності розподілу ймовірностей
ВІС	– Bayesian information criterion (байєсівський інформаційний критерій)
ВН	– static Bayesian network (статична байєсівська мережа)
ДВН	– dynamic Bayesian network (динамічна байєсівська мережа)
ЕСС	– effective sample size (ефективний розмір вибірки)
НСІ	– Hang Seng Index (індекс цін на біржі Hang Seng Гонконгу)
МСМС	– Markov chain Monte Carlo (метод Монте-Карло для марковських ланцюгів)
ML model	– Markov localization model (модель марковської локалізації)
СПІ	– Standard&Poor 500 Index (фондовий індекс біржі США)

ВСТУП

Успішне виконання сучасних досліджень вимагає застосування самих різноманітних методів, які можуть бути використані для розв'язання задач в умовах наявності невизначеностей, відсутності необхідних об'ємів історичних даних стосовно розвитку досліджуваних процесів і т. ін. Зокрема, це стосується аналізу фінансових та інших типів ризиків, які зустрічаються практично в усіх галузях діяльності людини. Міжнародна фінансова криза 2008 року не була спрогнозована тому, що більшість фахівців з фінансового менеджменту користувались математичними моделями, які ґрунтувались на історичних даних, і не змогли адаптуватись до нових реалій. Тобто існуючі моделі використовувались для високоякісного короткострокового прогнозування фінансових процесів, але не змогли вийти за межі їх тодішніх можливостей.

Основною відмінністю методів байєсівського аналізу даних (БАД) є те, що вони не вимагають наявності значних об'ємів історичних даних, на яких можна було б побудувати необхідні моделі для їх подальшого використання. Фактично, БАД може ґрунтуватись на коротких вибірках, на окремих фактах, експертних оцінках, окремих вимірах і т. ін. Тобто на відміну від частотного підходу до аналізу даних, БАД не вимагає проведення багатьох експериментів з метою отримання необхідних для моделювання вибірок даних, виконання необхідних обчислень та усереднення результатів. Разом з тим, у методах БАД можна використовувати всю наявну інформацію з метою уточнення відомої апріорної інформації та оцінювання апостеріорних розподілів. Більше того, методи БАД, як і більшість методів ймовірно-статистичної обробки даних, можуть бути успішно використані для адаптування структури і параметрів моделей до нових даних з метою підвищення адекватності моделей, що будуються. Сама ідея байєсівського аналізу даних, що ґрунтується на використанні теореми Байєса, передбачає її повторне використання при появі нових вимірів, фактів та експертних оцінок. Таким чином, автоматично реалізується системна ідея структурно-параметричної



адаптації моделей при появі нової інформації про досліджувані процеси та об'єкти.

Коло задач, які можна розв'язувати байєсівськими методами, надзвичайно широке. В першу чергу це задачі менеджменту ризиків різної природи, оскільки за означенням ризику – це величина можливих втрат та їх ймовірність. Тобто при розв'язуванні таких задач необхідно оцінювати характеристики розподілів та умовні ймовірності відповідних подій, пов'язаних з аналізом впливу факторів ризику на можливі результати реалізації ризиків. Методи БАД успішно застосовують для розв'язування задач медичної, технічної та економічної діагностики в умовах наявності обмеженої інформації про досліджувані процеси та об'єкти. У тому числі відомі успішні застосування БАД у аналізі якості у системах сервісу, на виробництві та в освіті. При цьому паралельно розв'язуються задачі ідентифікації та врахування в обчислювальних процедурах можливих невизначеностей, тобто факторів негативного впливу на досліджувані процеси та відповідні дані. Ці фактори призводять до спотворення проміжних та остаточних результатів обчислень, прогнозів і, на кінець, погіршують якість рішень, що приймаються на основі відповідних прогнозів. Методи БАД, разом з нечіткою логікою, дають можливість враховувати невизначеності ймовірнісного та амплітудного типів і тим самим автоматично сприяють підвищенню якості рішень.

Ще один клас задач, що розв'язуються методами БАД, відноситься до теорії оцінювання. Тобто ймовірнісні байєсівські процедури успішно застосовують для оцінювання параметрів моделей різних структур і самих змінних. Сьогодні набирає популярність метод оцінювання параметрів, який отримав назву Монте-Карло для марковських ланцюгів (МКМЛ). Його використовують для параметричної ідентифікації лінійних та нелінійних моделей з довільними розподілами даних. Крім того, метод МКМЛ знаходить широке застосування в системах моделювання імітаційного типу, що робить його особливо корисним у випадку використання з моделями ймовірнісного типу. Байєсівські фільтри даних успішно застосовують



у системах супроводження рухомих об'єктів, робототехніці, прогнозуванні та підтримці прийняття рішень.

У монографії вказано на можливі помилки, яких можна припуститись при використанні для моделювання ймовірнісно-статистичних методів, а також на безпосередній зв'язок умовних ймовірностей з теоремою Байєса. Розглянуто такі базові поняття як правдоподібність і ймовірність, вступ до алгебри потенціалів і обчислення базових ймовірностей для випадкових змінних.

У другому розділі подано можливі варіанти формулювання теореми Байєса для випадків дискретних даних і дискретних параметрів, неперервних даних і дискретного параметра, неперервних даних і неперервних параметрів та інші. Це дає можливість користувачу оперативно і коректно вибрати необхідний варіант теореми для конкретного застосування.

Третій розділ присвячено оцінюванню можливих типів апіорних розподілів, необхідних для подальшого застосування вибраного варіанту теореми Байєса. При цьому розглянуто можливості вибору об'єктивного та суб'єктивного розподілів, оцінювання апіорного розподілу на основі наявних даних, вибір натурального типу розподілу та формулювання ймовірнісного висновку для даних, що характеризуються заміщенням. Подано відповідні числові приклади оцінювання апіорних розподілів.

Запропоновано використання спеціальних алгоритмів Монте-Карло для генерування псевдовипадкових послідовностей з прогнозних розподілів і процедуру генерування неперервних випадкових величин. Створено процедуру створення ймовірнісно-статистичної моделі у формі нормального розподілу. Докладно подано модифіковані алгоритми генерування псевдовипадкових величин Метрополіса-Хастінгса та Гіббса, а також приклади їх практичного застосування. Так, алгоритм Гіббса застосовано для генерування марковських ланцюгів, які широко використовуються у аналізі фінансових даних. Показано, що сучасні фінансові процеси характеризуються наявністю нестационарності та нелінійності, що



значно затрудняє виконання аналізу і вимагає застосування спеціальних ймовірнісно-статистичних моделей для формального опису процесів такого типу. Автори запропонували практично корисну методику аналізу, моделювання і прогнозування вказаних процесів.

Для формального опису волатильності фінансових процесів запропонована модифікована порогова модель стохастичної волатильності, виконано докладний аналіз можливостей застосування цієї моделі. Зокрема, досліджена чутливість цієї моделі до відповідних припущень, а також визначено умови для виконання імітаційного моделювання випадкових процесів з використанням запропонованих моделей. Подано відповідні чисельні приклади моделювання.

Монографія має суттєве теоретичне і практичне значення і може бути успішно використана фахівцями з менеджменту фінансових ризиків, аналізу та прогнозування розвитку сучасних фінансових процесів, а також у навчальному процесі студентами, аспірантами і викладачами відповідних спеціальностей.

Р О З Д І Л 1

ВСТУП ДО БАЙЄСІВСЬКОГО АНАЛІЗУ ДАНИХ

1.1 Вступ в ймовірнісний аналіз

Теорія ймовірностей і прикладна статистика – це мистецтво і наука збору, обробки даних і аналізу отриманих результатів з метою прийняття коректних обґрунтованих рішень технічного, ділового, політичного персонального або іншого характеру.

Ймовірнісно-статистичні методи необхідно розглядати як важливу трудомістку частину процесу прийняття рішень, яка дозволяє приймати обґрунтовані тактичні та стратегічні рішення. Ці методи ґрунтуються на глибокому аналізі наявної інформації, а також на знаннях, досвіді та інтуїції фахівців (експертні оцінки) [12, 33, 43].

Ймовірнісний аналіз процесів, подій, даних різних типів передбачає два підходи:

1. *частотний*, який ґрунтується на ймовірнісному аналізі за класичним підходом – дані накопичуються у процесі виконання експериментів (або збору статистики) і обробляються так званими частотними методами теорії ймовірностей;
2. *байєсівський*, в основу якого покладається той чи інший варіант теореми Байєса; цей підхід не виключає використання класичних методів, а інформація може бути подана у вигляді статистичних (експериментальних) даних, експертних оцінок, окремих фактів і т. ін.

Коректне використання ймовірнісно-статистичних методів надає значні переваги стосовно прийняття коректних рішень практично в усіх напрямках людської діяльності: в конкуренції за підвищення якості та продаж нової продукції, дає можливість отримати високоякісні оцінки прогнозів, обґрунтувати фінансові, макроекономічні, а також рішення стосовно раціонального ведення домашнього господарства і розв'язати багато інших задач [25, 32, 42].

Послідовність розв'язування задачі моделювання

Послідовність реалізації етапів розв'язування задачі моделювання і подальше застосування побудованої моделі при використанні ймовірнісно-



статистичних методів залишається такою ж як і при використанні інших методів, наприклад, регресійного аналізу.

Весь процес моделювання та прогнозування на основі побудованої моделі складається з таких етапів:

1. вивчення теоретичних основ можливих методів моделювання, придатних для розв'язування задач конкретного типу;
2. вибір методу (методів) моделювання;
3. збір даних і побудова моделей-кандидатів, що належать до вибраного класу моделей; вибір кращої моделі з множини побудованих кандидатів за статистичними критеріями адекватності;
4. обчислення оцінок прогнозів за допомогою побудованої моделі;
5. використання прогнозів для автоматичного керування або для підтримки прийняття управлінських, особистих та інших типів рішень.

1.2 Методологія байєсівського програмування

Сьогодні байєсівська методологія моделювання та формування ймовірнісного висновку (остаточного результату аналізу даних та експертних оцінок) на основі моделі отримала назву *байєсівського програмування*. Методологія означає, що використовується множина методів, які забезпечують розв'язання таких задач:

- побудову ймовірнісно-статистичних моделей різних типів (оцінювання структури і параметрів) з використанням статистичних даних і експертних оцінок;
- обчислення остаточних результатів на основі створеної моделі згідно з постановкою задачі: оцінок прогнозів, керуючих впливів, оцінок змінних і параметрів на виході фільтрів, розпізнавання образів, знаходження рішень стосовно управління досліджуваними процесами і об'єктами і т. ін.;
- аналіз коректності отриманих результатів за відповідними множинами статистичних критеріїв якості.

До методології байєсівського програмування відносять методи, подані нижче [13, 57, 59].

1. *Рекурсивне байєсівське оцінювання: фільтрація, прогнозування, згладжування* змінних. Основне рівняння оцінювання має такий вигляд:

$$P(S(k) | O(0) \dots O(k)) = P(O(k) | S(k)) \times \sum_{S(k-1)} [(P(S(k) | S(k-1)) (P(S(k-1) | O(0) \dots O(k-1)))]$$

де $S(0), \dots, S(k)$ – часовий ряд змінних стану; $O(0), \dots, O(k)$ – часовий ряд спостережень; $P(S(k) | S(k-1))$ – модель системи або модель переходів; $P(O(k) | S(k))$ – модель спостережень, яка показує яким буде спостереження у момент k , якщо система знаходиться у стані $S(k)$.

За цією моделлю можна знайти $P(S(k+l) | O(0) \dots O(k))$ – тобто, яким буде розподіл ймовірностей станів у момент $k+l$, маючи спостереження $O(0), \dots, O(k)$. Якщо $l = 0$, то реалізується процедура *фільтрації*; якщо $l > 0$, то реалізується процедура *прогнозування*; а при $l < 0$ відбувається *згладжування* – відновлення минулого стану на основі спостережень, зроблених до або після моменту згладжування.

2. *Приховані марковські моделі (ПММ)* – це модифікація байєсівського фільтра, у якому припускається, що дані дискретні; моделі переходів станів і спостережень задаються матрицями ймовірностей або *таблицями умовних ймовірностей*. Якщо спостережувані змінні неперервні, то такі моделі називають *напів-неперервними ПММ*.

3. *Оптимальні рекурсивні фільтри Калмана (КФ)*; змінні неперервні або дискретні; моделі переходів станів і спостережень задаються з використанням гаусівських процесів – зовнішніх випадкових збурень і похибок (шумів) вимірів. У випадку нелінійних моделей використовують розклад у ряд Тейлора, що дає можливість лінійно локальних моделей. Для одночасного оцінювання станів досліджуваних процесів та їх параметрів використовують *розширений фільтр Калмана (РКФ)*.

4. *Гранулярні (particle) фільтри (ГФ)*; розподіл ймовірностей станів описують такою моделлю: $P(S(k-1) | O(0), \dots, O(k-1))$ (тут $S(\cdot)$ – матриця станів; $O(\cdot)$ – матриця спостережень) апроксимується множиною гранул (particles), вагові коефіцієнти яких пропорційні ймовірностям їх появи. Для оновлення ймовірностей станів використовується рекурсивна процедура.

5. *Статичні байєсівські мережі (static Bayesian networks (BN))* – це ймовірнісно-статистичні моделі для опису ймовірнісної і статистичної інформації в умовах наявності невизначеності. На змінні мережі практично не накладаються обмеження і немає спеціальної семантики для їх опису – тобто існує певна свобода вибору змінних для побудови мережі [92-95].

Байєсівська мережа (БМ) подається у формі спрямованого ациклічного графа, вершинами якого є змінні досліджуваного процесу, а дуги вказують на існуючі умовні залежності між змінними. Формально мережу описують трійкою: $BN = \{V, G, T\}$, де V – змінні (дані) для побудови мережі (база даних); G – спрямований ациклічний граф; T – таблиця умовних ймовірностей для вершин графа (змінних моделі).

Параметрами такої моделі є умовні ймовірності у таблицях умовних ймовірностей. Неперервні змінні подаються відповідними розподілами.



Якщо БМ містить дискретні і неперервні змінні, її називають гібридною. Неперервні змінні у таких випадках дискретизують.

Результатом побудови і використання БМ є ймовірнісний висновок у формі $P(X^i | \text{Known})$, де *Known* – підмножина інших змінних мережі, ймовірності станів яких відомі. Загалом ймовірнісний висновок в БМ полягає у поширенні ймовірностей і параметрів гаусівських законів розподілу по всій мережі залежно від отриманих *свідчень* (додаткова інформація про стани мережі). В основі процесу формування ймовірнісного висновку лежать досить складні математичні алгоритми.

6. *Динамічні байєсівські мережі (dynamic Bayesian networks (DBN))* створюються для того, щоб враховувати динаміку процесів (їх зміни у часі), а також можливі стохастичні впливи на їх протікання. Фактично, динамічна байєсівська мережа (ДБМ) – це розширення звичайних (статичних) мереж. Спочатку будується звичайна БМ для наявних змінних, структура якої передбачається інваріантною стосовно часу, тобто залишається сталою. Така структура повторюється для кожного наступного моменту часу з надходженням нових спостережень [12-13, 84].

Частину графа, що відповідає конкретному моменту часу t_k або просто k , називають часовим перерізом. Якщо приймається гіпотеза стосовно того, що стан поточного часового перерізу залежить тільки від попереднього, то таке припущення називають *марковським припущенням першого порядку*. Якщо структура всіх часових перерізів однакова, то таку ДБМ називають *стаціонарною*. У такому випадку модель, що відповідає одному часовому перерізу, називають *локальною* та *інваріантною стосовно часу* або *гомогенною*.

7. *Моделі марковської локалізації (Markov localization (ML) models)* – це моделі типу байєсівських фільтрів, які додатково включають керуючі змінні $\mathbf{u}(0), \dots, \mathbf{u}(k-1)$. Іноді їх називають ще прихованими марковськими моделями по входу-виходу.

У такій моделі ймовірності станів уточнюються за допомогою керуючих змінних таким чином: $P(S(k) | \mathbf{u}(k-1), S(k-1))$. Таку модель називають ще діючою моделлю (*action model*).

Форми таких моделей можуть бути різними. Загалом це матричні моделі; якщо вони подібні до тих, що використовуються у гранулярних фільтрах, то їх називають моделями Монте-Карло з марковською локалізацією (*Monte Carlo Markov Localization (MCML)*) [39, 44, 58].

Побудована модель призначена для того щоб давати відповідь на запитання стосовно ймовірності поточного стану досліджуваного об'єкта

$$P(S(k) | \mathbf{u}(0), \dots, \mathbf{u}(k-1), O(0), \dots, O(k))$$

на основі попередніх керуючих дій та спостережень за динамікою об'єкта. Термін «локалізація» пов'язаний із застосуванням у робототехніці, тобто локалізація робота у навколишньому середовищі.

Основне рівняння моделі подібне до основного рівняння фільтрації:

$$P(S(k) | \mathbf{u}(0), \dots, \mathbf{u}(k-1), O(0) \dots O(k)) = P(\mathbf{u}(k-1) | S(k)) \times \\ \times \sum_{S(k-1)} [(P(S(k) | \mathbf{u}(k-1), S(k-1))) (P(S(k-1) | \mathbf{u}(0), \dots, \mathbf{u}(k-2), O(0) \dots O(k-1)))]$$

8. *Байєсівські карти* – узагальнення моделей марковської локалізації, які також виникли у сфері управління роботами.

9. *Байєсівський метод обробки даних і прийняття рішень на основі ієрархічних моделей.*

10. *Байєсівська регресія*, узагальнені лінійні моделі.

1.3 Схеми байєсівського підходу до моделювання та прийняття рішень

Байєсівський підхід до побудови моделей та формування ймовірнісного висновку передбачає виконання етапів, поданих на рис. 1.1 [3]. У байєсівській статистиці передбачається, що інформація надходить з двох джерел: *апріорна інформація* від дослідника стосовно досліджуваної задачі і *статистичні* (експериментальні) *дані*, отримані в результаті виконання експериментів.

Апріорна інформація від дослідника – це результати попередніх досліджень; теоретичні основи досліджуваних явищ, процесів та об'єктів; додаткові неформальні дані, отримані з різних джерел. Загалом – це інформація, яка не повинна ґрунтуватись на експериментальних даних, що стосуються конкретної задачі.

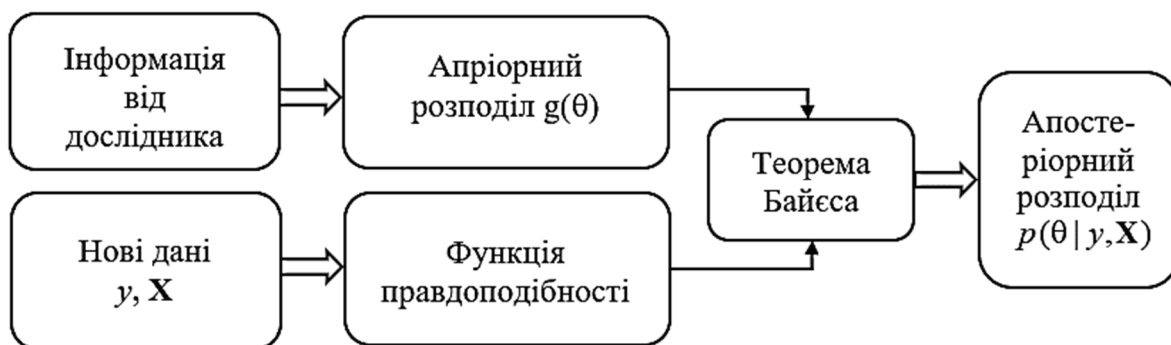


Рисунок 1.1 – Інформаційні потоки у системі моделювання на основі теореми Байєса

В центрі байєсівської методології знаходиться теорема Байєса (ТБ). Напишемо її з урахуванням позначень, що широко використовуються у економетричному аналізі даних:



$$p(\theta | y, \mathbf{X}) = \frac{g(\theta) p(y | \theta, \mathbf{X})}{p(y)} \propto g(\theta) p(y | \theta, \mathbf{X}), \quad (1.1)$$

де y – вектор вимірів основної (залежної) змінної досліджуваного процесу; \mathbf{X} – матриця вимірів незалежних змінних (*регресорів*), які визначають поведінку основної змінної; θ – вектор випадкових параметрів, які (разом з \mathbf{X}) визначають функцію щільності розподілу ймовірностей змінної y (у класичній регресійній моделі θ – вектор параметрів регресійної моделі); $p(\cdot)$ – функція щільності розподілу ймовірностей (ЩРЙ); $p(\theta)$ – апіорна щільність розподілу ймовірностей для значень випадкового параметра θ , яка ґрунтується на апіорних знаннях дослідника до використання даних (y, \mathbf{X}) ; $p(y | \theta, \mathbf{X})$ – умовна щільність даних y при конкретних значеннях θ і \mathbf{X} , іншими словами, це функція правдоподібності для даних y ; $p(y)$ – *маргінальна правдоподібність* для y (оскільки вилучено вплив незалежних змінних і параметрів).

Щільність $p(\theta | y, \mathbf{X})$ – апостеріорний розподіл θ , який ґрунтується на оновленні апіорного розподілу завдяки даним (y, \mathbf{X}) . Тобто ТБ показує як можна об'єднати інформацію, що надходить з двох джерел (апіорний розподіл і дані) з метою уточнення апіорного розподілу. Фактично, ТБ не тільки дає можливість об'єднувати інформацію з двох джерел, але будь-яке інше комбінування буде порушувати логічну (і математичну) сутність правил оперування з розподілами ймовірностей.

У виразі (1.1) перший варіант теореми записано як рівність, оскільки у ньому наявний знаменник $p(y)$ (безумовна щільність для y), який відіграє роль *нормуючої константи* і забезпечує те, що апостеріорна умовна щільність для θ є *належною* і інтегрується до 1 по області визначення параметрів. Другий варіант ТБ подано з точністю до константи пропорційності, тобто без нормування, що часто робиться для спрощення представлення виразів. При розв'язуванні практичних задач спочатку обчислюється чисельник, а потім, за необхідності, нормуюча константа. У багатьох прикладах існує необхідність у обчислення відношень результатів, а тому нормуюча константа скорочується.

Рисунок 1.1 ілюструє проходження двох паралельних потоків інформації через ТБ з метою обчислення апостеріорного розподілу параметра θ . Відносний вплив двох джерел інформації на остаточний результат залежить від їх точності. Наприклад, чим меншою буде дисперсія апіорного розподілу, тим більшу роль він відіграє у формуванні апостеріорного розподілу.

1.4 Невизначеності у математичному моделюванні

Невизначеність – у загальному випадку *фактор негативного впливу* на процес побудови математичної моделі, оцінювання прогнозу та генерування рішення на його основі, який призводить до погіршення або неможливості отримання проміжних та остаточних результатів [3].

Поява невизначеності зумовлена недостатністю або спотворенням інформації на будь-якому етапі згаданого процесу. Це може бути коротка вибірка даних, якої недостатньо для побудови адекватної моделі; спотворення вимірів шумовими складовими; некоректність встановлення типу розподілу даних і, як наслідок, неправильний вибір методу оцінювання параметрів моделі; помилки дослідника при обчисленні оцінок прогнозів або альтернативних рішень на їх основі.

Для боротьби з невизначеностями використовують нечітку логіку, ймовірно-статистичне моделювання, оптимальну і цифрову фільтрацію, методи заповнення пропусків і обробки екстремальних значень і т. ін.

Типи невизначеності, причини її виникнення та методи подолання подані у таблиці 1.1. При побудові математичних моделей на основі даних можна встановити наявність трьох основних типів невизначеності: *статистична, структурна і параметрична*.

Статистична невизначеність зумовлена самими даними, тобто наявністю пропусків, короткою вибіркою, наявністю екстремальних значень, впливом зовнішніх випадкових збурень на досліджуваний процес та похибками вимірів.

Структурна невизначеність виникає при оцінюванні структури моделі на основі даних. Так, наприклад, оцінити порядок авторегресії можна тільки наближено, оскільки значення автокореляційної функції (АКФ) – це випадкові величини. Це ж стосується оцінювання лагу (час запізнення) по входу, типу нелінійності, типу нестационарності. На кожному етапі аналізу даних ми маємо справу з оцінками, які є випадковими величинами, що вносить невизначеність в остаточний результат.

Параметрична невизначеність є наслідком наявності двох попередніх типів невизначеності. Наближеність оцінок структури моделі, наявність збурень і похибок вимірів, а також часто існує неможливість встановлення коректного типу розподілу даних – все це призводить до зміщеності (зсуву) оцінок параметрів моделі від точних значень та підвищення дисперсії цих оцінок. Тому завжди необхідно підходити обґрунтовано до вибору методу оцінювання параметрів на основі наявних даних.



Таблиця 1.1

Типи невизначеності і методи її подолання

№	Тип невизначеності	Причини виникнення	Методи подолання
1	Структурна невизначеність моделі	– неможливість встановлення всіх причинно-наслідкових зв'язків між змінними; – наближені значення елементів структури моделі.	– експертні методи; – застосування статистичних тестів; – застосування теорії перевірки гіпотез.
2	Статистичні невизначеності у процесі побудови моделей	– похибки вимірів; – стохастичні зовнішні збурення; – мультиколінеарність; – екстремальні значення; – пропуски вимірів;	– цифрові та оптимальні фільтри; – уточнення типів розподілів; – метод головних компонент; – теорія екстремальних значень; – методи заповнення пропусків.
3	Параметричні невизначеності	– некоректний вимір методу оцінювання; – короткі вибірки.	– забезпечення вибору альтернативних методів оцінювання параметрів; – розмноження вибірок.
4	Ймовірнісні невизначеності	– складні механізми виникнення причинно-наслідкових зв'язків; – відсутність детермінованості.	– статичні і динамічні мережі Байеса; – марковські моделі; – ймовірнісні фільтри; – умовні розподіли.
5	Невизначеність амплітудного типу	– наявність змінних, що не вимірюються.	– методи обробки нечіткої інформації.

Деякі методи подолання невизначеності подані в таблиці 1.1. У першу чергу це методи ймовірнісно-статистичного моделювання, нечітка логіка, оптимальна фільтрація, теорія обробки екстремальних значень та інші. Дана робота присвячено ймовірнісно-статистичному моделюванню, яке надає можливість описувати і враховувати невизначеності ймовірнісного і статистичного типів.

Р О З Д І Л 2

ЗАДАЧІ ТА ОСОБЛИВОСТІ ЙМОВІРНІСНО-СТАТИСТИЧНОГО АНАЛІЗУ ДАНИХ

Коректне використання ймовірнісно-статистичних методів надає значні переваги стосовно прийняття коректних рішень практично в усіх напрямках людської діяльності: в конкуренції за підвищення якості та продаж нової продукції, дає можливість отримати високоякісні оцінки прогнозів, обґрунтувати фінансові, макроекономічні, а також рішення стосовно раціонального ведення домашнього господарства і розв'язати багато інших задач [3, 16, 18-21, 45-47].

2.1 Задачі, які розв'язують ймовірнісно-статистичними методами

1. Виявлення причинно-наслідкових зв'язків у функціонуванні процесів різної природи з метою побудови коректних математичних моделей.
2. Аналіз і діагностика стану об'єктів різної природи і прийняття рішення стосовно коректності їх функціонування. Технічні системи, технологічні процеси, медична діагностика.
3. Прогнозування та управління соціально-економічними процесами і системами (регіональний, галузевий та державний рівні). Виявлення циклів в економіці та бізнесі. Використовують динамічні БМ.
4. Формування сценаріїв розвитку соціально-економічних процесів і систем. Використання побудованих сценаріїв для планування розвитку виробництва та різних видів бізнесу.
5. Виявлення катастрофічних явищ у природі, технічних системах, економіці.
6. Автоматичне керування технологічними процесами і технічними системами: побудова адекватних моделей, застосування методів статистичного керування, аналіз якості керування за допомогою статистичних параметрів.
7. Ймовірнісно-статистична підтримка прийняття експертних рішень.



8. Підтримка прийняття особистих рішень та рішень стосовно ведення домашнього господарства.
9. Контроль стану навколишнього середовища. Використовують три типи моделей:
 - мережні моделі (нейромережі та мережі Байєса);
 - моделі у вигляді диференціальних рівнянь в частинних похідних;
 - регресійні моделі: рівняння АР, АРКС, множинна регресія, змішана регресія і т. п.
10. Аналіз даних і прийняття рішень в генетиці, біології (існує напрям досліджень – *біостатистика*), психології (наприклад, сьогодні видається сучасний журнал: *Journal of Experimental Psychology* в США).
11. Розпізнавання образів.

Ймовірно-статистичні дослідження корисні особливо тоді, коли є великі групи людей, підприємств або інших об'єктів чи процесів (*генеральна сукупність*), які необхідно дослідити, але повне дослідження провести неможливо. Для того щоб отримати неідеальне, але практично корисне розуміння ситуації стосовно генеральної сукупності, можна відібрати невелику групу (*вибірку*) даних, яка складається із деяких, але не всіх, об'єктів генеральної сукупності [3].

Процес узагальнення результатів дослідження на всю генеральну сукупність називають *ймовірнісним* або *статистичним висновком* або рішенням.

Випадкова вибірка з генеральної сукупності є одним із кращих способів накопичення даних, оскільки генеральна сукупність, як правило, занадто велика, щоб її вивчати повністю.

Основою для прийняття рішення у ймовірно-статистичних дослідженнях є математична модель досліджуваного процесу, яка характеризується своєю структурою. Коректна оцінка структури дає можливість прийняти коректне рішення на основі такої моделі.

Поняття *структури математичної моделі* включає в себе такі елементи [1, 2, 9-11]:

1. розмірність моделі (кількість рівнянь, що утворюють модель);
2. порядок моделі (максимальний порядок різницевого або диференціального рівняння, що входить у модель);
3. нелінійність та її тип (нелінійність стосовно змінних або нелінійність стосовно параметрів);
4. час (або лаг) запізнення (по входу) та його оцінка;

5. збурення та його тип (детерміноване або випадкове, тип розподілу, параметри розподілу ймовірностей);
6. обмеження на змінні і параметри моделі.

2.2 Можливі помилки при використанні ймовірнісно-статистичних методів

Ймовірнісно-статистичні методи – надзвичайно могутній інструмент аналізу даних різної природи і їх силу можна використати коректно або некоректно. Є багато прикладів некоректного застосування статистичних методів, що проявилось у формулюванні висновків, протилежних до дійсності [3].

Наприклад, у 1936 році в США була допущена помилка журналом “*Literary Digest*”, який помилково визначив вибірку даних. Журнал виконав телефонне опитування своїх передплатників і спрогнозував, що *Альфред Лендон* легко переможе свого суперника *Франкліна Рузвельта* на виборах президента США. Однак, Рузвельт переміг Лендона в 46 із 48 штатів. Справа в тому, що 1936 рік – це один з років економічної депресії і дозволити собі телефон і передплату журналу могли тільки люди з хорошим матеріальним станом. Тобто вибірка, визначена журналом, не представляла всіх прошарків виборців у США – вона *не була представницькою*. Передплатники журналу у своїй більшості збирались голосувати за Лендона, в той час як основна частина генеральної сукупності громадян була за Рузвельта. Після виборів журнал “*Literary Digest*” швидко втратив свою популярність і досить швидко перестав існувати.

Звичайно, можна навести багато інших прикладів некоректного застосування ймовірнісно-статистичних методів і не тільки в минулому, а й сьогодні. На жаль, бувають випадки, коли некоректне застосування цих методів зумовлене не тільки поганими знаннями, але й цілеспрямованим навмисним формуванням некоректних висновків на основі зібраних даних з метою досягнення деякого запланованого (політичного) результату.

Очевидно, що майже всі статистичні процедури можна виконати некоректно як в результаті незнання, так і навмисно. Однак, як свідчить практика, навмисне спотворення методів статистичного висновку призводить тільки до негативних результатів. Якщо ми будемо стверджувати, що зростання ВВП України складає 13% на рік, а фактичне зростання складає 5% або ще менше, то добробут населення і довір'я до влади від цього не підвищаться.

Зазначимо, що невірні дані і некоректно виконаний ймовірнісно-статистичний аналіз можуть призвести до прийняття принципово неправильних тактичних і стратегічних рішень (висновків) з катастрофічними наслідками.



Випадки неправильного використання звичайних статистичних методів можна знайти в книзі американця Даррелла Хаффа “Як обманювати за допомогою статистики” [68].

Іноді говорять, що за допомогою статистики можна довести, що завгодно. Однак, якраз це “що завгодно” пов’язане з випадками некоректного застосування ймовірнісно-статистичних методів. Деякі компанії виконують дослідження, за допомогою яких доводять, що їхній продукт переважає за своїми якостями аналогічні продукти, що випускаються конкуруючими фірмами. Очевидно, що висновки будь-якого дослідження можна підігнати під бажаний результат за рахунок зміни техніки збору даних, підгонки моделі під поставлену мету або ж неправильної інтерпретації встановлених залежностей. У повсякденному житті кожний з нас зустрічається із статистичними даними та висновками, що знаходяться у протиріччі між собою. Тому немає нічого дивного в тому, що багато людей висловлюють недовіру до статистики, хоча не статистика винна у некоректних результатах.

Однак, у випадках зловживання ймовірнісно-статистичними методами *винні не цифри і не використані методи, а ті виконавці*, які мають схильність обманювати суспільство за їх допомогою. Чесні наміри і знання статистичних методів завжди дають корисні та змістовні результати.

Іноді можна зустріти заперечення проти застосування ймовірнісно-статистичних методів у зв’язку з тим, що отримувані за ними висновки носять узагальнений характер і не відображають індивідуальних якостей людей. Так, висновок стосовно того, що індивідууми з вищою освітою матимуть деякий рівень середньої зарплати, не сприймається деякими людьми. Вони вважають, що їх це не стосується і що висновок невірний. Так само як і у випадку голосування за деякого визначеного кандидата в президенти не всі представники однієї соціальної групи голосуватимуть за одного кандидата.

З цього приводу можна сказати, що ймовірнісно-статистичні висновки є узагальнюючими, але вони *не є абсолютними*. Загальне судження ніколи не залишається незмінним. Якщо ми будемо наводити приклади, які будуть знаходитись у протиріччі з висновком, то тим самим будемо надавати висновку *форму абсолютного судження*. Розв’язання цієї проблеми полягає в тому, що кожний індивідуум повинен сприймати справедливість ймовірнісно-статистичного судження, але в той же час повинен розуміти, що ніхто не забирає в нього свободи вибору за допомогою цього судження.

Загалом можна сказати, *що основною метою застосування ймовірнісно-статистичного аналізу є обґрунтоване прийняття рішень на основі аналізу даних та експертних оцінок*. Якщо ми хочемо, щоб рішення були об’єктивними і високоякісними, то необхідно знати і вміти користуватись методами ймовірнісно-статистичного аналізу та іншими методами прийняття рішень, що доповнюють їх. Доповнюючими методами

є методи оптимізації, чітка і нечітка логіка, нейромережі та багато інших методів. Інтегроване (спільне) застосування кількох методів прийняття рішень значно підвищує ймовірність отримання і вибір самих високоякісних рішень з множини можливих альтернатив.

2.3 Ймовірнісні та детерміновані моделі

Поняття невизначеності при обробці даних та експертних оцінок

Моделювання процесів різної природи пов'язане із *невизначеностями структурного, параметричного і статистичного* характеру. Велика перевага ймовірнісно-статистичного підходу до аналізу даних і прийняття рішень полягає у тому, що він дає можливість ідентифікувати і коректно боротись із такими невизначеностями, тобто враховувати їх при побудові моделей та формуванні альтернативних розв'язків задач [2, 3, 35, 36].

1) *Структурні невизначеності* пов'язані з неможливістю (у багатьох випадках) знайти прийнятну за точністю оцінку структури моделі внаслідок відсутності необхідної інформації. Тобто оцінити з належною точністю ті елементи, які входять у поняття структури математичної моделі.

2) *Параметричні невизначеності* зумовлені труднощами отримання незміщених, консистентних і ефективних оцінок параметрів (коефіцієнтів) моделі в умовах низької інформативності даних, наявності пропусків, екстремальних значень та похибок (шумів) вимірів.

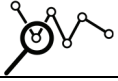
3) *Невизначеності статистичного характеру* зумовлені відсутністю належної інформації (статистичних характеристик) стосовно збурюючих впливів на стан об'єкта, шумів вимірів, відсутністю належної інформації стосовно розподілів інших випадкових змінних, які є у моделі.

Очевидно, що всі три зазначені типи невизначеностей взаємопов'язані між собою. Так, в усіх трьох випадках велику роль стосовно величини похибок оцінок структури і параметрів моделі відіграє неповна інформація стосовно початкового і поточного стану досліджуваного процесу. Коректний підхід до побудови моделі потребує мінімізації впливу невизначеностей усіх типів на якість результуючої моделі.

У подальшому викладенні матеріалу будемо характеризувати досліджувані об'єкти і процеси множиною *кількісних* (числових) або *якісних* (символьних, лінгвістичних) змінних:

$$\mathbf{x} = \{X_1, X_2, \dots, X_n\},$$

де кожна змінна X_j має свою область визначення D_j .



Детермінована модель

Детермінована модель представляє собою детерміновану функцію або множину правил стосовно взаємного зв'язку між змінними $X_i, i=1, \dots, n$ [3]. Наприклад, моделлю ідеального гармонічного коливання може бути така:

$$y_1(t) = A_1 \sin(\omega_1 t + \varphi_1), \quad (2.1)$$

або

$$y_2(t) = A_2 \cos(\omega_2 t + \varphi_2). \quad (2.2)$$

Прикладом детермінованого правила може бути таке:

$$\text{якщо } X_1 = \text{"людина"}, \text{ то } X_2 = \text{"розумна"}, \quad (2.3)$$

де X_1 – змінна, яка відноситься до множини живих організмів; X_2 – змінна, яка характеризує когнітивні можливості деякого організму (тобто його здатність до мислення).

Якщо кількість змінних і кількість значень, які вони можуть приймати, великі, то об'єкт може перебувати у одному із множини його можливих станів.

Ймовірнісна модель

Ймовірнісне моделювання передбачає, що кожний стан (або множина станів) об'єкта може наступити із деякою ймовірністю [70]. При використанні такого підходу кожний стан характеризується спільним розподілом ймовірностей (СРЙ) множини змінних

$$\mathbf{x} = \{ X_1, X_2, \dots, X_n \},$$

який позначимо так: $P(X_1, X_2, \dots, X_n)$. Областю визначення функції спільного розподілу ймовірностей є Декартів добуток:

$$D_1 \times D_2 \times \dots \times D_n,$$

а значення, які може приймати СРЙ, знаходяться в інтервалі $[0; 1]$.

Переваги ймовірнісного моделювання

1. Ймовірнісні моделі точніше (повніше) представляють функціонування об'єктів і процесів у будь-яких галузях нашої діяльності. Стосовно можливостей опису станів досліджуваного об'єкта ймовірнісне моделювання набагато потужніше ніж детерміноване. Детерміновану

модель можна розглядати як окремий випадок ймовірнісної [3]. Наприклад, можна стверджувати, що

$$\langle \text{Миші сірі} \rangle, \quad (2.4)$$

або

$$\langle x\% \text{ мишей сірі} \rangle, \quad (2.5)$$

де $x \rightarrow 100$. Вербальна модель (2.4) – це окремий випадок моделі (2.5), оскільки остання дещо *складніша* і має *ймовірнісний характер*. Модель (2.5) можна записати у формалізованому вигляді з використанням умовної ймовірності так:

$$p(x_2 = \langle \text{сірий} \rangle | x_1 = \langle \text{мишка} \rangle) = y. \quad (2.6)$$

У даному випадку імовірнісна модель точніше представляє колір мишей, оскільки у природі є білі миші, але їх кількість дуже незначна.

Іншим прикладом може бути детермінована і стохастична модель у просторі станів, яка широко використовується у проектуванні систем керування:

$$\mathbf{x}(k) = \mathbf{A}(k, k-1)\mathbf{x}(k-1) + \mathbf{B}(k, k-1)\mathbf{u}(k-1) + \mathbf{w}(k-1), \quad (2.7)$$

де $\mathbf{x}(k)$ – n -вимірний вектор станів системи; $\mathbf{u}(k-1)$ – m -вимірний вектор детермінованих вхідних величин (сигнали керування); $\mathbf{w}(k-1)$ – n -вимірний вектор випадкових зовнішніх збурень; $\mathbf{A}(k, k-1)$ – $(n \times n)$ матриця динаміки системи (вона містить коефіцієнти, що характеризують динаміку, тобто швидкість зміни станів у часі); $\mathbf{B}(k, k-1)$ – $(n \times m)$ матриця коефіцієнтів керування. Подвійний часовий аргумент у вигляді $(k, k-1)$ означає, що величина з цим аргументом використовується в момент k , але її значення ґрунтується на попередніх даних, які відомі на момент $k-1$ включно. Далі будемо записувати для простоти матриці \mathbf{A} і \mathbf{B} з одним аргументом, тобто $\mathbf{A}(k)$ та $\mathbf{B}(k)$. Очевидно, що стаціонарна система описується матрицями з постійними коефіцієнтами, які записують просто \mathbf{A} і \mathbf{B} .

У класичній постановці задачі оптимального оцінювання і керування послідовність зовнішніх збурень $\mathbf{w}(k)$ задовольняє властивостям білого гаусового шуму з нульовим середнім значенням і коваріаційною матрицею \mathbf{Q} , тобто статистики шуму мають вигляд:

$$\begin{aligned} E[\mathbf{w}(k)] &= 0, \quad \forall k, \\ E[\mathbf{w}(k)\mathbf{w}^T(j)] &= \mathbf{Q}(k)\delta_{kj}, \end{aligned} \quad (2.8)$$



де δ_{kj} – дельта-функція Кронекера, що визначається як $\delta_{kj} = \begin{cases} 0 & \text{для } k \neq j \\ 1 & \text{для } k = j \end{cases}$;

$\mathbf{Q}(k)$ – додатне визначена коваріаційна матриця зовнішніх збурень розмірності $(n \times n)$. Діагональні елементи матриці представляють собою дисперсії компонент вектора збурень $\mathbf{w}(k)$. Очевидно, що для застосування моделі (2.7) необхідно знати параметри нормального розподілу процесу $\mathbf{w}(k)$

Окремим випадком моделі (2.7) є спрощена детермінована модель у просторі станів:

$$\mathbf{x}(k) = \mathbf{A}(k, k-1)\mathbf{x}(k-1) + \mathbf{B}(k, k-1)\mathbf{u}(k-1). \quad (2.9)$$

Вона простіша, але і грубіша одночасно, оскільки не враховує важливі зовнішні впливи на процес [71-73].

2. Друга перевага ймовірнісної моделі полягає у тому, що її побудова вимагає систематичного дослідження усіх можливих значень кожної змінної (розподілу змінної), тобто кожної підмножини D_j , а також усіх можливих комбінацій змінних (конфігурацій об'єкта) [3]. Тобто необхідно розглядати кожний елемент з декартового добутку:

$$D_1 \times D_2 \times \dots \times D_n.$$

Такий підхід зменшує вплив на процес побудови моделі факторів, пов'язаних із культурними і психологічними особливостями індивідуума, а також зменшується ризик залишити поза увагою деякі аспекти функціонування об'єкта.

Таким чином, ймовірнісний підхід до моделювання надає можливість максимально точного представлення функцій об'єкта, оскільки при цьому розглядаються всі його теоретично можливі стани (конфігурації). Кожному стану ставиться у відповідність один елемент із нескінченної множини значень в інтервалі $[0; 1]$.

3. Третім позитивним фактором використання ймовірнісних моделей є те, що їх застосування досить часто є простішим ніж детермінованих.

Приклад 2.1. При виборі майбутньої спеціальності випускникам шкіл бажано мати інформацію стосовно ймовірності подальшого працевлаштування. У такому випадку не обов'язково потрібно займатись збором відповідної статистики, що вимагає багато часу. Можна звернутись

до декількох випускників вибраної спеціальності і отримати потрібну інформацію. Вона може бути отримана у вигляді:

«Близько 80% випускників даної спеціальності знаходять роботу відразу після закінчення університету.»

Звичайно, що це наближена інформація, але її може виявитись цілком достатньо для прийняття рішення абітурієнтом.

Недоліки ймовірнісного моделювання

1. У багатьох випадках *потрібно мати апріорну інформацію про процес* у вигляді оцінок початкового стану, типів розподілів змінних, попередніх оцінок параметрів розподілів і т. ін. Очевидно, що ця задача не завжди розв'язується просто і швидко, але той чи інший розв'язок практично завжди існує [60, 67, 74].

2. Основні труднощі виникають із *побудовою та використанням спільних розподілів ймовірностей* для змінних, що характеризують поведінку процесу. При цьому виникають такі проблеми:

- Графічне представлення СРЙ можливе тільки у тих випадках, коли кількість змінних не перевищує 2. Приклад представлення СРЙ у самому простому випадку ($n=2$; $m=2$, де n – кількість змінних; m – кількість значень, які можуть приймати змінні X_1 і X_2) наведено у таблиці 2.1.
- Використання СРЙ призводить до виникнення проблеми комбінаторного «вибуху» – великої кількості комбінацій значень змінних. Так, якщо кожна змінна приймає m різних значень ($m \geq 1$), то спільний розподіл ймовірностей описується m^n потенціально можливими конфігураціями (станами) об'єкта. Наприклад, якщо $n=m=10$, то отримаємо 10^{10} (10 мільярдів) значень.

Таблиця 2.1

Спільний розподіл ймовірностей для $n=m=2$

X_1, X_2	$x_{2,1}$	$x_{2,2}$
$x_{1,1}$	$p(X_1 = x_{1,1}; X_2 = x_{2,1})$	$p(X_1 = x_{1,1}; X_2 = x_{2,2})$
$x_{1,2}$	$p(X_1 = x_{1,2}; X_2 = x_{2,1})$	$p(X_1 = x_{1,2}; X_2 = x_{2,2})$



2.4 Безумовна незалежність подій

Згідно з пропозицією Декарта, множину n змінних можна розділити на декілька підмножин меншого розміру і аналізувати окремо [3, 123].

Припустимо, наприклад, що множину n змінних можна розділити на дві підмножини розмірністю j і $(n-j)$ змінних, відповідно:

$$p(X_1, X_2, \dots, X_n) = p(X_1, \dots, X_j) \cdot p(X_{j+1}, \dots, X_n). \quad (2.10)$$

Тепер задача моделювання спрощується. Спочатку можна обчислити спільну ймовірність для підмножини змінних X_1, \dots, X_j , а потім для підмножини X_{j+1}, \dots, X_n . Остаточний результат отримаємо за виразом (2.10).

Рівність (2.10) означає, що підмножини X_1, \dots, X_j і X_{j+1}, \dots, X_n – незалежні або *безумовно незалежні*. Це означає, що будь-яка інформація стосовно підмножини X_1, \dots, X_j , наприклад, $X_1 = 7$ або $X_1 + X_2 > 5$, не змінить розподіл ймовірностей другої підмножини X_{j+1}, \dots, X_n .

Однак, для реальних об'єктів і процесів безумовна незалежність цих двох підмножин мало ймовірна. Якщо така ситуація має місце, то це означає, що початкова постановка задачі стосовно моделювання такого об'єкта (процесу) з високою ймовірністю була некоректною. У такому випадку є смисл будувати дві моделі, тобто описувати математично кожен підмножину окремо.

2.5 Правила обчислення ймовірностей

Ймовірність – це ступінь нашої віри у те, що деяка подія відбудеться.

Існує також поняття „свідчення” стосовно попереднього твердження; свідчення проявляється (знаходить своє вираження) через інші твердження. Приклади розглянемо нижче.

Вибірковий простір експерименту

Множину можливих результатів деякого експерименту (або послідовності подій) називають *вибірковим простором* цього експерименту. Будемо припускати, що вибірковий простір досліджуваного експерименту містить всі можливі результати цього експерименту і що кожна пара результатів представляє собою взаємовиключні події. Такі припущення гарантують, що експеримент закінчиться точно одним із визначених результатів у вибірковому просторі [3, 75].

Наприклад, при підкиданні костей вибіркового простір буде такий:

$$S = \{1, 2, 3, 4, 5, 6\},$$

а для опису результату футбольного матчу можна скористатись таким простором:

$$S = \{winner, looser\}.$$

Підмножина вибіркового простору складається із однієї або декількох *подій*. Наприклад, подію, пов'язану з тим, що ми отримаємо значення три («3») або вище при використанні 6-гранних костей, відповідає такій підмножині:

$$\{3, 4, 5, 6\} \subseteq \{1, 2, 3, 4, 5, 6\}.$$

Така подія матиме місце якщо в результаті підкидання кості отримаємо елемент, який належить вказаній підмножині.

В загальному випадку ми говоримо, що подія $A = \text{''істина''}$, тобто мала місце у деякому експерименті, якщо результат експерименту є елементом події A . Якщо подія містить тільки один елемент, то його також вважають результатом.

Для того щоб виміряти ступінь невизначеності стосовно результату експерименту використовують ймовірність події $p(A)$, $\forall A \subseteq S$. Ці ймовірності повинні задовольняти трьом аксіомам, які подані нижче [23, 28].

Аксіома 1. *Ймовірність того, що результат деякого експерименту буде належати його вибіркового простору дорівнює одиниці: $p(S) = 1$.*

Іншими словами: ступінь віри (довіри) у те, що деяка подія A відбудеться, характеризується дійсним числом в інтервалі від нуля до одиниці:

$$0 \leq p(A) \leq 1. \quad (2.11)$$

Аксіома 2. Будь-яка подія A повинна мати невід'ємну ймовірність:

$$\text{Для всіх } A \subseteq S \text{ виконується умова, що } p(A) \geq 0.$$

Аксіома 3. Якщо дві події A і B не перетинаються (рис. 2.1а), тобто не залежать одна від одної, то ймовірність комбінованої події визначається сумою ймовірностей двох окремих подій:

$$\text{Якщо } A \subseteq S, B \subseteq S \text{ і } A \cap B = \emptyset, \text{ то } p(A \cup B) = p(A) + p(B).$$

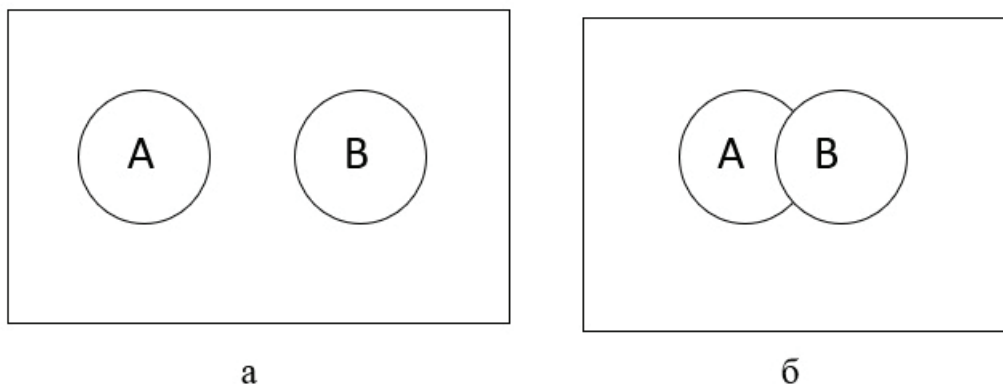


Рисунок 2.1 – Незв'язані події (А) і зв'язані події (Б)

Наприклад, якщо подія B стосується числа 3 на грані кості, тобто $B=\{3\}$; а подія A пов'язана з тим, що випаде парне число, тобто: $A=\{2, 4, 6\}$, то події A і B – не перетинаються (незалежні). А тому ймовірність настання однієї з цих подій буде такою:

$$p(A \cup B) = p(A) + p(B) = \frac{1}{6} + \frac{3}{6} = \frac{4}{6}.$$

Якщо події перетинаються (рис. 1.1б), можна легко показати, що

$$p(A \cup B) = p(A) + p(B) - p(A \cap B), \quad (2.12)$$

де $A \cap B$ – перетин подій A і B , який представляє подію, яка означає *появу обох подій* A і B .

Приклад 2.2. Розглянемо приклад з гральними картами (52 карти у колоді). Позначимо через A подію, яка означає, що із колоди витягли будь-яку карту *пікової* масті, а через B – подію, яка означає, що із колоди витягли короля (будь-якої масті). Очевидно, що ці події залежні (пов'язані між собою), оскільки їх перетин означає, що ми витягли *пікового короля*, тобто

$$A \cap B = \{\text{піковий король}\}.$$

Таким чином, ймовірність того, що ми витягнемо з колоди короля або будь-яку карту *пікової* масті складе:

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) = \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52}.$$

Пояснення: у колоді з 52 карт міститься 13 карт *пікової* масті (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, валет, дама, король і туз), 4 королі і 1 *піковий король*.

Примітка: Іноді будемо підкреслювати, що значення ймовірності знайдено частотним методом (а не експертним суб'єктивним); таке значення ймовірності будемо позначати через $p^\#$. Якщо подія A описується тільки

одним результатом a , то ми будемо використовувати позначення $p(a)$, а у інших випадках $p(\{a\})$.

Правило додавання ймовірностей можна поширити на випадок будь-якої кількості *взаємовиключних* подій A_1, A_2, \dots, A_n :

$$p(A_1 \text{ or } A_2 \text{ or } \dots \text{ or } A_n) = p(A_1) + p(A_2) + \dots + p(A_n). \quad (2.13)$$

Взаємовиключними є подія A та її логічне заперечення \bar{A} , які утворюють повну множину станів для деякого об'єкта із двома станами:

$$p(\bar{A}) = 1 - p(A).$$

Наприклад, це справедливо для підкидання монети із двома можливими результатами: *орел і решка*. Третій стан – балансування на боку – практично неможливий. Інший приклад утворюють стани електричної лампочки: *робочий* (нормальний), коли лампочка може давати світло під дією електричного струму; і *неробочий* – спіраль перегоріла.

При оцінюванні ймовірностей подій завжди використовують додаткову інформацію, навіть якщо вона не задана у явному вигляді. Наприклад, нехай відомо, що деяка монета завжди падає орлом вгору після підкидання. Очевидно, що у даному випадку існує додаткова інформація (або *свідчення*) про те, що монета незбалансована, а тому завжди падає на один і той же бік.

Необхідно розрізняти „свідчення” і „додаткову (так звану *фонову*) інформацію”, оскільки деякі твердження необхідно розглядати як свідчення, а деякі – як додаткову інформацію.

Наприклад, нехай виконується аналіз дезоксирибонуклеїнової кислоти (ДНК). Припустимо, що методика аналізу відповідає теоретичним вимогам і її повністю дотримано у процесі аналізу. Всі ці *припущення складають додаткову інформацію* або її частину, оскільки це обов'язкові елементи виконання аналізу, яких необхідно притримуватись.

Свідчення, які мають відношення до виконання аналізу, можуть стосуватися лише опису результатів аналізу, а також статистичних даних, які характеризують надійність свідчень.

Додаткову інформацію можна ввести у означення умовної ймовірності; так, позначення $p(A|I)$ означає:

- „ймовірність виникнення події A (тобто $A=істина$) за умови наявності додаткової інформації I ”.

A позначення $p(A|B,I)$ означає:



- „ймовірність виникнення події A за умови, що відбулась подія B і за наявності додаткової інформації I ”.

При цьому A не залежить від B тільки у тому випадку, якщо

$$p(A|B, I) \neq p(A|I). \quad (2.14)$$

Поняття ймовірнісної залежності

Існує також поняття „ймовірнісної залежності” між подіями та „ймовірнісної незалежності”. Це практично неформалізовані поняття і наявність того чи іншого випадку необхідно розглядати у кожному випадку окремо [22, 28, 38].

Наприклад, нехай монетка підкидається два рази. Чи будуть пов’язані між собою результати першого і другого підкидання? Іншими словами, чи впливає на визначення ймовірності результату другого підкидання знання стосовно результату першого підкидання?

Очевидно, що у випадку гарно збалансованої монетки відповіддю буде „ні”. Однак, якщо невідомо, що монета добре збалансована, то відповіддю буде „так”.

Звичайно, що одного підкидання недостатньо для встановлення взаємної залежності (чи незалежності) результатів. Однак, якщо розглянути множину дослідів, то можна встановити різницю між двома ситуаціями – збалансована і незбалансована монета. Якщо невідомо, чи монета добре збалансована, то можна виконати ряд дослідів для того щоб встановити наскільки якісно виготовлена монета. Результати підкидання не будуть у даному випадку незалежними.

2.6 Умовні ймовірності та їх зв’язок з теоремою (правилом) Байєса

Коли ми висловлюємо твердження стосовно ймовірності події A , тобто $p(A)$, то неявно висловлюємо умови, пов’язані з іншими відомими факторами [24]. Наприклад, „ймовірність того, що при підкиданні кості випаде число 6 дорівнює $\frac{1}{6}$ ”, як правило, передбачає, що кості виготовлені правильно (збалансовані). Тобто попереднє твердження має починатись так: „за умови, що кості збалансовані, ймовірність того ...”. Ймовірності такого типу називають умовними, а загальна форма твердження у таких випадках має вигляд:

„За умови відомої (ймовірності) події B , ймовірність події A складає p .”

Твердження такого типу формально позначаються так: $p(A|B)=c$. Однак, необхідно зазначити, що запис $p(A|B)=c$ не означає, що у всіх випадках, коли B ="істина", то ймовірність A дорівнює c . Це означає тільки те, що коли B настає і більше нічого не впливає на A , тоді ймовірність події A дорівнює c .

Властивість 2.1 (умовна ймовірність). Якщо настає (відома) подія B , то результат (експерименту) визначається елементами підмножини B , а результати, які відповідають події A , визначаються перетином: $A \cap B$. Тобто, якщо відома інформація про B , то необхідно визначити ймовірність $A \cap B$. Це дає можливість сформулювати властивість умовної ймовірності таким чином:

$$p(A|B) = \frac{p(A \cap B)}{p(B)}, \quad p(B) > 0.$$

Наприклад, умовна ймовірність того, що при підкиданні кості ми отримаємо 4, за умови, що ми отримуємо парне число, визначається так:

$$p(A = 4 | B = \{2, 4, 6\}) = \frac{p(4)}{p(\{2, 4, 6\})} = \frac{1/6}{3/6} = \frac{1}{3},$$

за умови, що кості виготовлені якісно (збалансовані).

Очевидно, що у розрахунках умовних ймовірностей умова може стосуватись більше однієї події, наприклад:

$$p(A|B \cap C) = \frac{p(A \cap B \cap C)}{p(B \cap C)}.$$

Останній вираз можна переписати так, щоб отримати фундаментальне правило обчислення (умовних) ймовірностей.

Теорема 2.1 (фундаментальне правило):

$$p(A|B) \cdot p(B) = p(A \cap B). \quad (2.15)$$

Ця формула інтерпретується так: ймовірність того, що настануть обидві події A і B , визначається ймовірністю A за умови B і ймовірністю B .

Введення умови стосовно ще однієї події приводить до виразу:

$$p(A|B \cap C) \cdot p(B|C) = p(A \cap B|C).$$



Оскільки $p(A \cap B) = p(B \cap A)$, а також $p(A \cap B | C) = p(B \cap A | C)$, то на основі (1.6.1) можна записати:

$$p(A | B) \cdot p(B) = p(A \cap B) = p(B | A) \cdot p(A).$$

Це дає можливість записати правило (теорему) Байєса.

Теорема 2.2 (правило Байєса):

$$p(A | B) = \frac{p(B | A) p(A)}{p(B)}. \quad (2.16)$$

Правило Байєса – це метод оновлення ймовірності події A за умови, що наявна інформація стосовно події B . З цієї причини введено такі пояснення:

- $p(A)$ – називають апріорною ймовірністю події A ;
- $p(A | B)$ – апостеріорна ймовірність події A за умови, що мала місце подія B ;
- ймовірність $p(B | A)$ називають *правдоподібністю події A за умови (настання) події B* (пояснюючий приклад стосовно цієї назви наведемо нижче).

Таким чином, теорему Байєса можна інтерпретувати так:

Апостеріорна ймовірність \propto *Правдоподібність* \times *Апріорна ймовірність*, де символ „ \propto ” означає пропорційність.

Правило Байєса у контексті наявності ще однієї події C можна записати так:

$$p(A | B, C) = \frac{p(B | A, C) p(A | C)}{p(B | C)}.$$

Приклад 2.3. Нехай дві хвороби a_1 і a_2 мають один і той же симптом b ; і нехай $p(b | a_1) = 0,9$, а $p(b | a_2) = 0,3$. Припустимо, що апріорні ймовірності для обох хвороб однакові: $p(a_1) = p(a_2)$.

Якщо з'являється симптом b , то за правилом Байєса:

$$p(a_1 | b) = \frac{p(b | a_1) p(a_1)}{p(b)} = 0,9 \frac{p(a_1)}{p(b)};$$
$$p(a_2 | b) = \frac{p(b | a_2) p(a_2)}{p(b)} = 0,3 \frac{p(a_2)}{p(b)}.$$

Незважаючи на те, що ми не можемо завершити обчислення апостеріорних ймовірностей, можна зробити висновок, що ймовірність хвороби a_1 у три рази перевищує ймовірність a_2 , за умови наявності симптому b .

Якщо a_1 і a_2 – це єдині причини виникнення симптому b , то, припускаючи, що ймовірність захворіти двома хворобами одночасно дорівнює нулю, можна записати, що:

$$p(a_1 | b) + p(a_2 | b) = 1,$$

а звідси:

$$0,9 \frac{p(a_1 | b)}{p(b)} + 0,3 \frac{p(a_2 | b)}{p(b)} = 1.$$

Тепер, якщо апріорні ймовірності захворювання першою і другою хворобами однакові, тобто: $p(a_1) = p(a_2)$, то

$$\frac{p(a_1 | b)}{p(b)} = \frac{p(a_2 | b)}{p(b)} \Rightarrow \frac{p(a_1 | b)}{p(b)} (0,9 + 0,3) = 1,$$

а звідси маємо:

$$\frac{p(a_1 | b)}{p(b)} = \frac{p(a_2 | b)}{p(b)} = \frac{1}{0,9 + 0,3} = \frac{1}{1,2},$$

що дає можливість розрахувати ймовірності захворювання кожною хворобою за умови наявності симптому b :

$$p(a_1 | b) = \frac{p(b | a_1) p(a_1)}{p(b)} = 0,9 \frac{p(a_1)}{p(b)} \approx 0,9 \frac{p(a_1 | b)}{p(b)} = 0,9 \cdot \frac{1}{1,2} = 0,75;$$

$$p(a_2 | b) = \frac{p(b | a_2) p(a_2)}{p(b)} = 0,3 \frac{p(a_2)}{p(b)} \approx 0,3 \frac{p(a_2 | b)}{p(b)} = 0,3 \cdot \frac{1}{1,2} = 0,25.$$

Таким чином, ми виконали тут розрахунки без інформації стосовно $p(B)$; ця ймовірність відіграє тільки роль нормуючої величини.

Умовна незалежність

Якщо інформація стосовно події B не змінює ступінь нашої впевненості стосовно появи події A , то говорять, події A і B – незалежні.

Означення 2.1 (незалежність). Події A і B незалежні, якщо

$$p(A | B) = p(A).$$



Наприклад, якщо ми підкидаємо одна за одною дві добре збалансованих (виготовлених) кості, то випадання числа 2 на одній з них не впливає на число, яке випадає на іншій.

Коли дві події незалежні одна від одної, то *фундаментальне правило обчислення ймовірностей* можна записати так:

$$p(A \cap B) = p(A|B) p(B) = p(A) \cdot p(B).$$

Тобто ймовірність того, що обидві події матимуть місце, визначається добутком ймовірностей окремих подій. (Вище, фундаментальне правило (2.15) ми записали так: $p(A|B) \cdot p(B) = p(A \cap B)$.)

Симетричність незалежності

Поняття незалежності симетричне, тобто, якщо A не залежить від B , то B не залежить від A :

$$p(B|A) = \frac{p(A \cap B)}{p(A)} = \frac{p(A|B) p(B)}{p(A)} = \frac{p(A) p(B)}{p(A)} = p(B),$$
$$p(A) > 0.$$

З іншого боку, якщо подія A неможлива, то ми її просто не розглядаємо, тобто умова $p(A) > 0$ не має значення.

Концепція незалежності у випадку, коли умова стосується декількох подій.

Так, якщо інформація стосовно події B не змінює нашу ступінь упевненості у настанні події A , але відома інформація про подію C (яка може впливати на A), то говорять, що A і B *умовно незалежні* при наявності інформації стосовно події C .

Означення 2.2 (умовна незалежність). Дві події A і B *умовно незалежні* за наявності інформації стосовно події C , якщо:

$$p(A|B \cap C) = p(A|C).$$

Твердження стосовно умовної незалежності (так само як і твердження стосовно умовної залежності) симетричне. Якщо A умовно не залежить від B за умови наявності інформації стосовно події C , то B умовно не залежить від A за умови наявності інформації стосовно події C :

$$p(B|A \cap C) = \frac{p(A \cap B|C) p(C)}{p(A|C) p(C)} = \frac{p(A|B \cap C) p(B|C)}{p(A|C)} =$$

$$= \frac{p(A|C) p(B|C)}{p(A|C)} = p(B|C).$$

Ймовірність виникнення обох умовно незалежних подій можна обчислити за допомогою розглянутого вище правила множення:

$$p(A \cap B | C) = p(A|C) \cdot p(B|C).$$

Зазначимо, що випадок, коли події незалежні, – це окремий випадок умовної незалежності при $C = \emptyset$.

Приклади умовної незалежності

Стосовно реальних об'єктів концепція умовної незалежності є реалістичнішою. Розглянемо її на простому прикладі.

Приклад 2.4. Припустимо, що студенту необхідно у понеділок здавати іспит. Для того щоб визначити ступінь ризику не здати іспит внаслідок поганого самопочуття розглянемо такі ситуації:

- 1 – перед іспитом студент спав не менше 7 годин і почувається добре;
- 2 – перед початком іспиту студент відчуває значну втому.

У цьому прикладі існує причинний зв'язок між тривалістю сну студента перед іспитом, ступенем його втоми і ризиком не здати іспит успішно. Припустимо, що перед іспитом студент стомлений, відчуває себе погано. *Для оцінювання ризику не має значення, яка причина його втоми – неспокійний сон перед іспитом чи інша причина.*

Якщо ж студент перед початком іспиту почувається добре, то якість його сну у ніч з неділі на понеділок не має значення для оцінювання ризику не здати іспит. Такий попередній аналіз дає можливість стверджувати, що ризик не здати іспит є умовно незалежним від якості сну студента перед іспитом, за умови, що він почувається добре.

Для формального опису ситуації введемо три бінарні змінні:

- X_1 – характеризує якість сну студента перед іспитом;
- X_2 – стан студента перед іспитом (добре або погане самопочуття);
- X_3 – студент здасть іспит чи ні.

Тобто X_3 не залежить від X_1 при будь-якому значенні X_2 (для встановлення значення X_3 досить знати X_2), що можна записати так:

$$p(X_3 | X_1 \text{ і } X_2) = p(X_3 | X_1 \cap X_2) = p(X_3 | X_2). \quad (2.17)$$



Таким чином, для того щоб визначити ймовірність успішної задачі іспиту студентом досить знати значення X_2 .

У даному випадку немає необхідності вводити функцію трьох змінних:

$$p(X_1, X_2, X_3) = p(X_1) \cdot p(X_2 | X_1) \cdot p(X_3 | X_2), \quad (2.18)$$

яка послідовно описує процес оцінювання ймовірності того, що студент успішно здасть екзамен. Тобто якість сну студента впливає на самопочуття, а самопочуття визначає ступінь ризику не здати екзамен:

$$\text{якість сну} \rightarrow \text{самопочуття} \rightarrow \text{ступінь ризику не здати іспит} \quad (2.19)$$

Приклад 2.5. Розглянемо тестування спортсменів на допінг. Як правило, для цього виконується аналіз крові, A , і аналіз сечі, B . Ці два аналізи виконуються незалежно один від другого у різних лабораторіях.

Очевидно, що результати цих аналізів характеризуються залежними змінними. Якщо аналіз A – позитивний, то аналіз B також, скоріше всього, буде позитивним.

Ці аналізи можна розглядати як незалежні події, оскільки їх роблять у різних лабораторіях за різними методиками. Тобто результати обох аналізів є умовно незалежними для будь-якого стану спортсмена.

Для формального опису ситуації із аналізами можна ввести три бінарні змінні:

- X_1 – характеризує стан спортсмена (приймав допінг чи ні);
- X_2 – результат аналізу крові (позитивний чи негативний);
- X_3 – результат аналізу сечі (позитивний або негативний).

Для цих змінних можна записати:

$$p(X_2 | X_1, X_3) = p(X_2 | X_1). \quad (2.20)$$

Для того щоб оцінити ймовірність того, що аналіз сечі буде позитивним, досить знати, що спортсмен прийняв допінг. А рівність (2.20) має такий смисл: „для того щоб оцінити ймовірність того, що аналіз крові буде позитивним, досить знати, що спортсмен прийняв допінг”.

Розширення правила множення на довільну кількість подій

Закон множення також можна поширити на будь-яку кількість незалежних подій:

$$p(A_1, A_2, \dots, A_n | I) = p(A_1 | I) \cdot p(A_2 | A_1, I) \cdot \dots \cdot p(A_n | A_1, A_2, \dots, A_{n-1}, I),$$

де I – додаткова інформація, наприклад, стосовно дотримання умов експерименту.

Нехай A_1, A_2, \dots, A_n – така послідовність незалежних подій, пов'язаних із підкиданням монети:

- результат першого підкидання – „орел”;
- результат другого підкидання – „орел”;
- результат третього підкидання – „решка”;
- ...

Якщо монета добре збалансована, то результати підкидань від 1-го до $(n - 1)$ -го не будуть впливати на майбутні результати (майбутні результати не залежать від попередніх), в тому числі і на результат n -го підкидання, що можна формально записати так:

$$p(A_1, A_2, \dots, A_n) = p(A_1 | I) \cdot p(A_2 | I) \cdot \dots \cdot p(A_n | I).$$

2.7 Розширення правила додавання (другого правила обчислення ймовірностей)

Для будь-яких пропозицій A і B ступінь довіри до того, що A – істинна, за наявності додаткової інформації I , визначається сумою (правилом) таких ймовірностей:

$$p(A | I) = p(A, B | I) + p(A, \bar{B} | I), \quad (2.21)$$

де B – подія з двома станами.

Важливість цього правила підтверджується тим фактом, що деякій пропозиції легше присвоїти ступінь довіри, якщо вона розглядається у контексті інших пропозицій, які мають стосунок до неї.

Застосовуючи до наведеного правила закон множення ймовірностей, отримаємо:

$$p(A | I) = p(A | B, I) \cdot p(B | I) + p(A | \bar{B}, I) \cdot p(\bar{B} | I). \quad (2.22)$$

Вираз (2.22) – це **розширення другого правила (закону) (додавання ймовірностей взаємовиключних подій)**, яке мало вигляд (2.13):

$$p(A_1 \text{ or } A_2 \text{ or } \dots \text{ or } A_n) = p(A_1) + p(A_2) + \dots + p(A_n).$$

Повернемось до прикладу з підкиданням монети. Нехай A – це така пропозиція: „результатом першого підкидання монети є орел”.

Загальну ступінь довіри до A (ймовірність настання A) можна легко визначити за допомогою виразу (2.22). Правило (2.21) легко розширюється на будь-яку кількість пропозицій, за умови що вони є взаємовиключними і



утворюють повну множину. У прикладі з монетою існує тільки три пропозиції:

- $B_1 =$ „монета якісна”;
- $B_2 =$ „монета має два орли (герби)”;
- $B_3 =$ „монета має дві решки” (лицеві сторони).

Таким чином, ймовірність того, що результатом першого підкидання буде орел, можна обчислити за виразом:

$$p(A|I) = p(A|B_1, I) \cdot p(B_1|I) + p(A|B_2, I) \cdot p(B_2|I) + p(A|B_3, I) \cdot p(B_3|I)$$

Отриманий вираз можна узагальнити на n взаємовиключних пропозицій (подій) – це знаменник теореми Байєса:

$$p(A|I) = \sum_{i=1}^n p(A|B_i, I) \cdot p(B_i|I). \quad (2.23)$$

Насамкінець, на основі виразу (2.22) можна стверджувати, що події A і B зв’язані між собою, якщо

$$p(A|B, I) \neq p(A|\bar{B}, I). \quad (2.24)$$

Дійсно, якщо подія B не має відношення до A , то

$$p(A|B, I) = p(A|\bar{B}, I).$$

Оскільки $p(\bar{B}|I) = 1 - p(B|I)$, то права частина виразу (2.22) тепер тотожно дорівнює лівій частині незалежно від значення $p(B|I)$:

$$\begin{aligned} p(A|I) &= p(A|B, I) \cdot p(B|I) + p(A|\bar{B}, I) \cdot [1 - p(B|I)] = \\ &= p(A|B, I) \cdot p(B|I) + p(A|\bar{B}, I) - p(A|\bar{B}, I) \cdot p(B|I) = \\ &= p(A|B, I), \end{aligned}$$

оскільки $p(A|B, I) = p(A|\bar{B}, I)$.

Поняття апіорної умовної ймовірності

Отримання нової інформації (свідчення) стосовно деякої події можна розглядати як двокроковий процес у часі – моменту t_0 і моменту t_1 . У момент часу t_0 починається пошук свідчення B , яке пов’язане з подією A . В цей момент ймовірність того, що $A =$ ”істина”, за умови, що $B =$ ”істина”, можна обчислити за виразом (2.21). Якщо позначити через $p_0(\cdot)$ –

ймовірність події у момент t_0 , то правило (2.10) можна переписати у вигляді:

$$p_0(A|B,I) = \frac{p_0(B|A,I) \cdot p_0(A|I)}{p_0(B|A,I) \cdot p_0(A|I) + p_0(B|\bar{A},I) \cdot p_0(\bar{A}|I)}, \quad (2.25)$$

що впливає із третього правила (закону) обчислення ймовірностей (закон множення) та розширення другого закону – виразу (2.22). У виразі (2.25) $p_0(A|I)$, $p_0(\bar{A}|I)$ і $p_0(A|B,I)$ – апіорні (початкові) ймовірності відповідних подій.

Ймовірність $p_0(B|A,I)$ називають *правдоподібністю* A за умови B (у момент t_0). Аналогічно, ймовірність $p_0(B|\bar{A},I)$ – це правдоподібність \bar{A} за умови B (у момент t_0). Поняття „*правдоподібності*” – це ще один із елементів процедури формулювання гіпотез з подальшою метою прийняття рішення.

Оцінювання правдоподібності пов’язане із відповіддю на запитання:

„Якщо гіпотеза A вірна, (або невірна), то з яким ступенем довіри буде вірною гіпотеза B ?

Нехай у момент часу t_1 виявляється, що гіпотеза B – вірна. Якою буде ймовірність того, що A – також вірна у цей же момент t_1 , тобто $p_1(A|I) = ?$

Очевидно, що необхідно врахувати відомий факт, що B – вірна; це частина додаткової інформації у момент t_1 . Тому загальна ймовірність події A у момент t_1 повинна враховувати те, що мала місце подія B , тобто:

$$p_1(A|I) = p_0(A|B,I). \quad (2.26)$$

Ситуація не зміниться якщо у момент t_1 з’являється інформація стосовно гіпотези B (B – вірна), але до цього моменту подія B не розглядалась. Встановлення відношення події B до A , а також ступеня впливу цього відношення на ймовірності подій вимагає розгляду правдоподібностей і початкових ймовірностей у момент t_0 . Тобто, яка ймовірність того, що подія B – вірна за наявності деякої додаткової інформації і за припущення, що гіпотеза A – вірна (або невірна)? Якою була б апіорна ймовірність того, що A – вірна за наявності деякої додаткової інформації, не враховуючи те, що B – вірна?



2.8 Правдоподібність і ймовірність

Правдоподібність і ймовірність відіграють значну роль у імовірнісних дослідженнях, а тому важливо знати різницю між правдоподібністю події A при даному (відомому) B і ймовірністю A , яке залежить від B .

Розглянемо такий *приклад*: випадковий свідок бачив, що Іванов втікав від будинку, де було скоєно злочин. Нехай пропозиція B звучить так:

B : „Іванов втікав від місця злочину у той час, коли він був скоєний”;

а пропозицію A сформульовано так:

A : „Іванов скоїв злочин”.

Цілком логічно припустити, що ймовірність $p(B|A,I)$ є високою, але ймовірність $p(A|B,I)$ не обов’язково повинна бути високою.

Якщо Іванов дійсно скоїв злочин, то можна очікувати, що він постарается швидше покинути місце злочину. Гіпотеза стосовно того, що він винен – це хороше пояснення до свідчення (тобто втікав). Однак, той факт, що він біг від будинку ще не дає високої ймовірності того, що саме він скоїв злочин. Цьому факту можна дати також інше пояснення.

Якщо між правдоподібністю гіпотези A , за наявності свідчення B , і ймовірністю цієї ж гіпотези за умови B (тобто при існуванні залежності між A і B) ставлять знак рівності, то це називають *софізмом переміни умов* (*fallacy of the transposed conditions*).

Він виникає у тому випадку, коли із факту: якщо мало місце A , то B виникне з великою ймовірністю, помилково роблять такий висновок: якщо мало місце B , то A виникне із високою ймовірністю. Найкращим засобом боротьби з таким софізмом є теорема Байєса у формі (2.16):

$$p(A|B,I) = \frac{p(B|A,I) \cdot p(A|I)}{p(B|A,I) \cdot p(A|I) + p(B|\bar{A},I) \cdot p(\bar{A}|I)}.$$

Члени у знаменнику не обов’язково близькі один до одного. У другій складовій знаменника правдоподібність \bar{A} за наявності B також може бути великою, тобто може існувати інше хороше пояснення для свідчення B , навіть якщо A = „невірно”. Більше того, необхідно прийняти до уваги ступінь довіри до A у випадку, якщо відома тільки додаткова (другорядна) інформація.

Очевидно, що наведену вище теорему Байєса можна розширити на випадок n взаємовиключних пропозицій A_i , $i = 1, \dots, n$, які утворюють повну множину подій і представляють собою альтернативні пояснення для B :



$$p(A_i | B, I) = \frac{p(B | A_i, I) \cdot p(A_i | I)}{\sum_{i=1}^n p(B | A_i, I) \cdot p(A_i | I)}.$$

У процесі дослідження поставленої задачі може з'явитись нова інформація у вигляді експертних оцінок та статистичних даних (вимірів), які дадуть можливість уточнити прийняте рішення. У такому випадку раніше отриманим результатом можна скористатись як апіорною інформацією. Тобто ТБ можна застосовувати повторно для розв'язання однієї і тієї ж задачі. Це нагадує адаптивний підхід до побудови математичних моделей, але є відмінності у критеріях і обчислювальних процедурах.

2.9 Обчислення ймовірностей для змінних

До цих пір ми розглядали обчислення ймовірностей для простих змінних та результатів настання подій у деякому вибіркового просторі. Тепер розширимо розглянуті вище концепції стосовно обчислення ймовірностей на *змінні* (множини вибіркового просторів). Результатом кожного експерименту є *стан* (процесу, об'єкта), який характеризується значенням змінної [3, 29, 30].

Множину станів, пов'язаних із змінною A , будемо позначати через

$$sp(A) = (a_1, a_2, \dots, a_n).$$

За аналогією до вибіркового простору подій ці стани повинні бути *взаємно виключними* та утворювати *скінченну повну множину*. Останнє припущення гарантує, що змінна (процес) перебуває у одному із своїх станів (хоча ми можемо не знати у якому), а перше припущення означає, що у кожний момент часу змінна перебуває тільки у одному стані.

Наприклад, якщо позначити через D змінну, яка характеризує результат підкидання кубиків, то її простір станів буде таким:

$$sp(D) = (1, 2, 3, 4, 5, 6).$$

У подальшому змінні будемо позначати великими (прописними) літерами, а їх стани – малими.

Невизначеність стосовно станів (a_1, a_2, \dots, a_n) змінної A будемо виражати через розподіл ймовірностей $P(A)$ цих станів:

$$P(A) = (x_1, \dots, x_n); \quad x_i \geq 0; \quad \sum_{i=1}^n x_i = x_1 + \dots + x_n = 1,$$



де x_i – ймовірності перебування змінної A у стані a_i . Розподіл називають *рівномірним (uniform, even)*, якщо всі ймовірності однакові.

Примітка. У загальному випадку ймовірність перебування A у стані a_i позначають так: $p(A = a_i)$, або просто $p(a_i)$, якщо змінна, що аналізується, очевидна із контексту.

Так само як і для звичайних подій, розглянемо *умовні ймовірності* для змінних. Якщо A має стани (a_1, \dots, a_n) а змінна B має стани (b_1, b_2, \dots, b_m) , то $P(A|B)$ містить $n \times m$ умовних ймовірностей $p(a_i | b_j)$, які визначають ймовірність стану a_i за умови b_j . Тобто умовна ймовірність для змінної за умови наявності інформації про іншу змінну, являє собою множину ймовірностей (які представляють, як правило, у табличному вигляді). При цьому кожне число таблиці – це ймовірність деякої конфігурації (*комбінації*) станів змінних (табл. 2.2).

Крім того, оскільки $P(A|B)$ визначає розподіл ймовірностей для кожної події $B = b_j$, то згідно з аксіомою 1 сума ймовірностей по всіх значеннях змінної A повинна дорівнювати 1 для кожного стану змінної B :

$$\sum_{i=1}^n p(A = a_i | B = b_j) = 1, \text{ для кожного } b_j, j = 1, \dots, m. \quad (2.27)$$

Приклад таблиці умовних ймовірностей $P(A|B)$ для бінарної змінної A за умови тернарної змінної B наведено у таблиці 2.2. Для кожного стану змінної B сума ймовірностей станів A дорівнює 1.

Таблиця 2.2

Умовні ймовірності $P(A|B)$ для бінарної змінної A
за умови тернарної змінної B

	b_1	b_2	b_3
a_1	0,4	0,3	0,6
a_2	0,6	0,7	0,4

Спільний розподіл ймовірностей

Ймовірність появи декількох станів можна виразити *спільною ймовірністю* для двох або більше змінних. Для кожної конфігурації змінних (a_i, b_j) , A і B , $P(A, B)$ визначає ймовірність настання обох станів: $A = a_i$ і $B = b_j$. Таким чином розподіл $P(A, B)$ складається з $n \times m$ чисел (приклад наведено у табл. 2.3).

Таблиця 2.3

Розподіл спільних ймовірностей $P(A, B)$ для бінарної змінної A за умови тернарної змінної B

	b_1	b_2	b_3
a_1	0,16	0,12	0,12
a_2	0,24	0,28	0,08

Зазначимо, що оскільки простори станів обох змінних A і B є взаємно виключними і повними множинами, то всі комбінації їх станів (Декартів добуток) також будуть взаємно виключними і повними множинами, а тому їх можна розглядати як вибірковий простір (за аналогією з простими подіями). Таким чином, згідно з аксіомою 1, маємо:

$$P(A, B) = \sum_{i=1}^n \sum_{j=1}^m p(A=a_i, B=b_j) = 1.$$

Приклад таблиці розподілу спільних ймовірностей $P(A, B)$ для бінарної змінної A за умови тернарної змінної B наведено у таблиці 2.3. Зазначимо, що сума всіх значень ймовірностей у таблиці дорівнює 1.

Коли фундаментальне правило $p(A|B) \cdot p(B) = p(A \cap B)$ застосовується до змінних A і B , то таку процедуру обчислення ймовірностей необхідно застосувати до кожної пари (a_i, b_j) , а всього таких пар маємо $n \times m$:

$$p(a_i | b_j) p(b_j) = p(a_i, b_j).$$

Це означає, що для побудови таблиці $P(A|B)$ кожену ймовірність з множини $P(A|b_j)$ необхідно помножити на $p(b_j)$, що дасть значення $P(A, b_j)$. Виконуючи такі обчислення для кожного b_j , отримаємо $P(A, B)$.

Таблицю 2.3 можна отримати в результаті застосування фундаментального правила до таблиці 2.2 при $P(B) = (0,4; 0,4; 0,2)$ (відповідна ілюстрація наведена у таблиці 2.4). Таким чином, у таблиці 2.4 представлені спільні ймовірності $P(A, B)$, отримані як добуток $P(B) = (0,4; 0,4; 0,2)$ на значення $P(A|B)$ з таблиці 2.2.

Таблиця 2.4

Спільні ймовірності $P(A, B)$

	b_1	b_2	b_3
a_1	$0,4 \cdot 0,4$	$0,3 \cdot 0,4$	$0,6 \cdot 0,2$
a_2	$0,6 \cdot 0,4$	$0,7 \cdot 0,4$	$0,4 \cdot 0,2$



У випадку застосування до змінних фундаментальне правило визначається теоремою 2.3.

Теорема 2.3 (фундаментальне правило для змінних) – обчислення ймовірностей одночасної появи відповідних станів змінних A і B :

$$P(A, B) = P(A|B) \cdot P(B),$$

а після введення умови стосовно третьої змінної отримаємо:

$$P(A, B|C) = P(A|B, C) \cdot P(B|C).$$

Обчислення маргінальних ймовірностей (маргіналізація, marginalization)

Використовуючи таблицю спільних ймовірностей, можна розрахувати розподіл значень $P(A)$, розглядаючи ті значення B , які можуть мати місце разом із кожним станом a_i змінної A . Тобто для кожного стану a_i змінної A необхідно розрахувати m різних взаємно виключних значень: $(a_i, b_1), \dots, (a_i, b_m)$.

Таким чином, згідно з аксіомою 3 (правило обчислення ймовірностей) маємо:

$$p(a_i) = \sum_{j=1}^m p(a_i, b_j).$$

Такі обчислення називають *маргіналізацією*, тобто говорять, що змінна B маргіналізується з розподілу $P(A, B)$ (що приводить до $P(A)$). Процедура маргіналізації позначається так:

$$P(A) = \sum_B P(A, B) = \underset{B}{\text{mar}} P(A, B).$$

Маргіналізуючи B з таблиці розподілу спільних ймовірностей 2.3, отримаємо:

$$P(A) = (0,16 + 0,12 + 0,12; 0,24 + 0,28 + 0,08) = (0,4; 0,6).$$

Тобто $P(A) = (0,4; 0,6)$ – це і є результуючий розподіл.

Маргіналізація A з таблиці 2.3 дає такий результат для ймовірностей станів B :

$$P(B) = (0,16 + 0,24; 0,12 + 0,28; 0,12 + 0,08) = (0,4; 0,4; 0,2).$$

Таким чином, процедура маргіналізації дає можливість видалити змінні із спільного розподілу ймовірностей і знайти таким чином остаточний розподіл, який нас цікавить.

Узагальнення маргінальної щільності на розподіл ймовірностей

Досить часто необхідно знайти так звану маргінальну (остаточну) щільність розподілу ймовірностей, яку позначають через $p(\mathbf{y})$. Таку ЩРЙ можна отримати таким чином:

$$p(\mathbf{y}) = \int_{R_{\theta}} p(\theta, \mathbf{y}) d\theta = \int_{R_{\theta}} p(\mathbf{y}|\theta) p(\theta) d\theta. \quad (2.28)$$

Тобто (2.28) показує, що маргінальна ЩРЙ – це середня умовна ЩРЙ $p(\mathbf{y}|\theta)$, отримана за допомогою апіорної ЩРЙ $p(\theta)$, яка відіграє роль вагової функції.

Приклад 2.6. Нехай y_1 – вибірка даних, отримана з нормально розподіленої генеральної сукупності з невідомим математичним сподіванням μ і відомим середнім квадратичним відхиленням $\sigma = \sigma_0$. Розподіл для y_1 має вигляд:

$$p(y_1|\mu, \sigma = \sigma_0) = \frac{1}{\sigma_0 (2\pi)^{1/2}} \exp\left[-\frac{1}{2\sigma_0^2} (y_1 - \mu)^2\right].$$

Виберемо апіорну щільність для μ також нормальною:

$$p(\mu) = \frac{1}{\sigma_a (2\pi)^{1/2}} \exp\left[-\frac{1}{2\sigma_a^2} (\mu - \mu_a)^2\right], \quad -\infty < \mu < \infty,$$

де μ_a і σ_a – апіорні математичне сподівання і середнє математичне відхилення, відповідно. Таким чином, маргінальна ЩРЙ для y_1 має вигляд:

$$\begin{aligned} p(y_1) &= \int_{-\infty}^{\infty} p(y_1|\mu, \sigma = \sigma_0) p(\mu) d\mu = \\ &= \frac{1}{2\pi \sigma_a \sigma_0} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2} \left(\frac{(y_1 - \mu)^2}{\sigma_0^2} + \frac{(\mu - \mu_a)^2}{\sigma_a^2}\right)\right] d\mu. \end{aligned}$$

Після виділення повного квадрату стосовно μ та інтегрування отримаємо:

$$p(y_1) = \frac{1}{[2\pi(\sigma_a^2 + \sigma_0^2)]^{1/2}} \exp\left[-\frac{(y_1 - \mu_a)^2}{2(\sigma_a^2 + \sigma_0^2)}\right].$$



Таким чином, маргінальна ЩРЙ для y_1 нормальна з апіорним математичним сподіванням μ_a і дисперсією $\sigma_a^2 + \sigma_0^2$. Оскільки параметри μ_a , σ_a^2 і σ_0^2 вважаються відомими, то розподіл $p(y_1)$ можна використовувати для того щоб формулювати ймовірнісні твердження стосовно y_1 , що часто досить корисно робити до фактичного спостереження y_1 .

Теорема Байєса для змінних

Теорема 2.4 (правило Байєса для змінних):

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{P(A,B)}{\sum_B P(A,B)},$$

а після введення додаткової умови за змінною C отримаємо:

$$P(B|A,C) = \frac{P(A|B,C)P(B|C)}{P(A|C)} = \frac{P(A,B|C)}{\sum_B P(A,B|C)}.$$

Дві наведені рівності можливі завдяки, по-перше, фундаментальному правилу обчислення ймовірностей і, по-друге, описаній вище процедурі маргіналізації [3, 99].

У таблиці 2.5а наведено приклад обчислення розподілу $P(B|A)$, використовуючи $P(A)$, $P(B)$ і $P(A|B)$ з наведених вище таблиць. Умовні ймовірності $P(B|A)$, обчислені шляхом застосування правила Байєса до $P(A|B)$ у таблиці 2.2: $P(A)=(0,4; 0,6)$ і $P(B)=(0,4; 0,4; 0,2)$. Зазначимо, що сума ймовірностей по B дорівнює 1 для кожного стану A .

Таблиця 2.4а

Приклад обчислення розподілу $P(B|A)$

	a_1	a_2
b_1	$(p(a_1 b_1) \cdot p(b_1)) / p(a_1)$	$(p(a_2 b_1) \cdot p(b_1)) / p(a_2)$
b_2	$(p(a_1 b_2) \cdot p(b_2)) / p(a_1)$	$(p(a_2 b_2) \cdot p(b_2)) / p(a_2)$
b_3	$(p(a_1 b_3) \cdot p(b_3)) / p(a_1)$	$(p(a_2 b_3) \cdot p(b_3)) / p(a_2)$

Таблиця 2.5б

Приклад обчислення розподілу $P(B|A)$ з підстановкою числових значень

	a_1	a_2
b_1	$(0,4 \cdot 0,4) / 0,4$	$(0,6 \cdot 0,4) / 0,6$
b_2	$(0,3 \cdot 0,4) / 0,4$	$(0,7 \cdot 0,4) / 0,6$
b_3	$(0,6 \cdot 0,2) / 0,4$	$(0,4 \cdot 0,2) / 0,6$

Таблиця 2.5б містить результати підстановки числових значень у таблицю 2.5а. Умовні ймовірності $P(B|A)$, обчислені шляхом застосування правила Байєса до $P(A|B)$ у таблиці 1.1, $P(A)=(0,4; 0,6)$ і $P(B)=(0,4; 0,4; 0,2)$. Зазначимо, що сума ймовірностей по B дорівнює 1 для кожного стану A .

Результати обчислень за таблицею 2.5б наведені у таблиці 2.5в. Значення $p(a_1|b_1) \dots p(a_i|b_i)$ взяті з таблиці 2.2.

Таблиця 2.5в

Результати обчислення розподілу $P(B|A)$

	a_1	a_2
b_1	0,4	0,4
b_2	0,3	0,47
b_3	0,3	0,13
	$\sum p(b_i a_1)=1$	$\sum p(b_i a_2)=1$

Концепція умовної незалежності для змінних

Означення 2.3 (умовна незалежність для змінних). Дві змінні A і C умовно незалежні при відомій змінній B , якщо:

$$p(a_i|c_k, b_j) = p(a_i|b_j)$$

для кожного $a_i \in sp(A)$, $b_j \in sp(B)$ і $c_k \in sp(C)$.

У скороченому записі: $P(A|C, B) = P(A|B)$.

Це означає, що за відомого стану змінної B ніяка інформація стосовно стану змінної C не змінить ймовірності A . Зазначимо, що твердження незалежності виконується для кожного стану B ; якщо множина, яка задає умову (тобто B), порожня, то говорять, що A і C маргінально незалежні або просто незалежні ($P(A|C) = P(A)$).



Якщо дві змінні A і C умовно незалежні за наявності третьої (зв'язаної з ними) змінної B , то фундаментальне правило для змінних (теорема 2.3) спрощується до вигляду:

$$P(A, C | B) = P(A | B, C) \cdot P(C | B) = P(A | B) \cdot P(C | B).$$

Згідно з цим виразом необхідно перемножити дві таблиці умовних ймовірностей, які відносяться до своїх областей визначення. Таке множення є безпосереднім (прямим) розширенням тих операцій, які розглянуто у попередніх прикладах:

$$p(a_i, c_k | b_j) = p(a_i | b_j) \cdot p(c_k | b_j).$$

Наприклад, перемножуючи $P(A | B)$ на $P(C | B)$, значення яких наведені у таблицях 2.2 і 2.6, відповідно, отримаємо таблицю 2.7. Тобто спочатку заповнюємо таблицю 2.6 для $P(C | B)$ (у даному випадку вона просто задається, а не розраховується, а саме розрахунок можна зробити за формулою Байєса: $P(C | B) = \frac{P(B | C) P(C)}{P(B)} = \frac{P(C, B)}{\sum_C P(C, B)}$). Результати розрахунків використовуються для обчислень у таблицях 2.2 і 2.6.

Приклад значень умовних ймовірностей $P(C | B)$ для тернарної змінної C за умови тернарної змінної B наведено у табл. 2.6. Для кожного стану змінної B сума ймовірностей станів C дорівнює 1.

Таблиця 2.6

Умовні ймовірності $P(C | B)$ для тернарної змінної C
за умови тернарної змінної B

	b_1	b_2	b_3
c_1	0,2	0,9	0,3
c_2	0,05	0,05	0,2
c_3	0,75	0,05	0,5

Використовуючи наведені нижче вирази, обчислимо елементи таблиці умовних ймовірностей (таблиця 2.7а, 2.7б і 2.7в) для умовно незалежних змінних за наявності третьої (зв'язаної з ними) змінної B , тобто:

$$P(A, C | B) = P(A | B, C) \cdot P(C | B) = P(A | B) \cdot P(C | B)$$

$$p(a_i, c_k | b_j) = p(a_i | b_j) \cdot p(c_k | b_j) =$$

Приклад підстановки числових значень умовних ймовірностей наведено у табл. 2.7б ($p(c_1 | b_1) = 0,2$ – взято з таблиці 2.6, а $p(a_1 | b_1) = 0,4$ – з таблиці 2.2).

Таблиця 2.7а

Обчислення умовних ймовірностей у символній формі

	b_1	b_2	b_3
c_1	$(p(a_1 b_1) \cdot p(c_1 b_1);$ $p(a_2 b_1) \cdot p(c_1 b_1))$	$(p(a_1 b_2) \cdot p(c_1 b_2);$ $p(a_2 b_2) \cdot p(c_1 b_2))$	$(p(a_1 b_3) \cdot p(c_1 b_3);$ $p(a_2 b_3) \cdot p(c_1 b_3))$
c_2	$(p(a_1 b_1) \cdot p(c_2 b_1);$ $p(a_2 b_1) \cdot p(c_2 b_1))$	$(p(a_1 b_2) \cdot p(c_2 b_2);$ $p(a_2 b_2) \cdot p(c_2 b_2))$	$(p(a_1 b_3) \cdot p(c_2 b_3);$ $p(a_2 b_3) \cdot p(c_2 b_3))$
c_3	$(p(a_1 b_1) \cdot p(c_3 b_1);$ $p(a_2 b_1) \cdot p(c_3 b_1))$	$(p(a_1 b_2) \cdot p(c_3 b_2);$ $p(a_2 b_2) \cdot p(c_3 b_2))$	$(p(a_1 b_3) \cdot p(c_3 b_3);$ $p(a_2 b_3) \cdot p(c_3 b_3))$

Таблиця 2.7б

Підстановка числових значень умовних ймовірностей

	b_1	b_2	b_3
c_1	$(0,2 \cdot 0,4; 0,2 \cdot 0,6)$	$(0,9 \cdot 0,3; 0,9 \cdot 0,7)$	$(0,3 \cdot 0,6; 0,3 \cdot 0,4)$
c_2	$(0,05 \cdot 0,4; 0,05 \cdot 0,6)$	$(0,05 \cdot 0,3; 0,05 \cdot 0,7)$	$(0,2 \cdot 0,6; 0,2 \cdot 0,4)$
c_3	$(0,75 \cdot 0,4; 0,75 \cdot 0,6)$	$(0,05 \cdot 0,3; 0,05 \cdot 0,7)$	$(0,5 \cdot 0,6; 0,5 \cdot 0,4)$

Таблиця 2.7в

Результати обчислень умовних ймовірностей

	b_1	b_2	b_3
c_1	$(0,08; 0,12)$	$(0,27; 0,63)$	$(0,18; 0,12)$
c_2	$(0,02; 0,03)$	$(0,015; 0,035)$	$(0,12; 0,08)$
c_3	$(0,3; 0,45)$	$(0,015; 0,035)$	$(0,3; 0,2)$

Результати обчислень умовних ймовірностей за виразом $P(A, C | B) = P(A | B, C) \cdot P(C | B) = P(A | B) \cdot P(C | B)$ наведено у таблиці 2.7в. Суми усіх значень умовних ймовірностей для кожного стовпчика дорівнюють 1.

Приклад виконання обчислень значень таблиць умовних ймовірностей за наявності додаткових свідчень (тобто наявності значень окремих станів)

Для ілюстрації розглянутих вище теорем розглянемо процедуру обчислення умовних ймовірностей для трьох змінних A, B і C . Для цих змінних дана таблиця 2.8 числових значень спільного розподілу ймовірностей. Нехай до цієї відомої таблиці додатково поступають свідчення $A = a_2$ і $C = c_1$. Три числа, які дані у кожній комірці таблиці, відповідають станам c_1, c_2 і c_3 .

Задача полягає у тому, щоб обчислити значення таблиці умовних ймовірностей $P(B | a_2, c_1)$.



Таблиця 2.8

Спільний розподіл ймовірностей для змінних A , B і C

	b_1	b_2	b_3
a_1	(0; 0,05; 0,05)	(0,05; 0,05; 0)	(0,05; 0,05; 0,05)
a_2	(0,1; 0,1; 0)	(0,1; 0; 0,1)	(0,2; 0; 0,05)

Спочатку розглянемо ту частину таблиці 2.8, яка відповідає $A = a_2$ і $C = c_1$; аналізуючи таблицю, отримаємо:

$$P(a_2, B, c_1) = (0,1; 0,1; 0,2), \quad (2.29)$$

тобто це перші елементи множин значень у клітинках другого рядка. Перші – тому що $C = c_1$, а другого рядка, тому що $A = a_2$.

Для того щоб обчислити значення $P(B|a_2, c_1)$, скористаємось теоремою 2.4 (правило Байєса для змінних):

$$P(B|a_2, c_1) = \frac{P(a_2, B, c_1)}{P(a_2, c_1)} = \frac{P(a_2, B, c_1)}{\sum_B P(a_2, B, c_1)}. \quad (2.30)$$

Маргіналізуючи B з рівняння (2.29) (маргіналізація у даному випадку означає обчислення ймовірності спільної появи станів (a_2, c_1) , тобто це знаменник (2.30)), отримаємо:

$$P(a_2, c_1) = \sum_B P(a_2, B, c_1) = 0,1 + 0,1 + 0,2 = 0,4.$$

Тепер виконаємо операцію ділення (нормування) у рівнянні (2.30):

$$P(B|a_2, c_1) = \left(\frac{0,1}{0,4}; \frac{0,1}{0,4}; \frac{0,2}{0,4} \right) = (0,25; 0,25; 0,5).$$

Іншими словами можна сказати, що ми ставили задачу перетворення $P(a_2, B, c_1)$ у розподіл ймовірностей. Оскільки сума значень ймовірностей в (2.29) не дорівнює 1, то ми нормуємо розподіл шляхом ділення кожного числа на суму їх усіх значень.

Припустимо тепер, що ми отримали тільки свідоцтво $A = a_2$ і необхідно обчислити $P(B|a_2, C)$. Обчислення елементів цієї таблиці ймовірностей виконується так само, як і у попередньому випадку, за винятком того, що тепер обчислення будуть ґрунтуватись на таблицях.

Як і у попередньому випадку, зосередимось на тій частині таблиці ймовірностей 2.8 $P(A, B, C)$, яка відповідає $A = a_2$; в результаті отримаємо таблицю 2.9.

Приклад таблиці ймовірностей $P(a_2, B, C)$, яка відповідає тій частині таблиці 2.8, що обмежена значенням $A = a_2$ наведено у таблиці 2.9.

Таблиця 2.9

Таблиця ймовірностей $P(a_2, B, C)$

	b_1	b_2	b_3
c_1	0,1	0,1	0,2
c_2	0,1	0	0
c_3	0	0,1	0,05

Тобто для побудови цієї таблиці вибрані елементи нижнього рядка (який відповідає a_2) і трьом станам змінних B і C .

Для того щоб обчислити розподіл $P(B|a_2, C)$, скористаємось виразом:

$$P(B|a_2, C) = \frac{P(a_2, B, C)}{P(a_2, C)} = \frac{P(a_2, B, C)}{\sum_B P(a_2, B, C)}. \quad (2.31)$$

Ймовірність $P(a_2, C)$ обчислимо шляхом маргіналізації B із таблиці 2.9:

$$P(a_2, C) = (0,1 + 0,1 + 0,2; 0,1 + 0 + 0; 0 + 0,1 + 0,05) = (0,4; 0,1; 0,15), \quad (2.32)$$

а отриманий результат підставимо у рівняння (2.31). Проміжний та остаточний результати наведено у таблицях 2.10а і 2.10б. Для отримання результатів було використано $P(a_2, B, C)$ (по табл. 2.9) і $P(a_2, C)$ (по рівнянню (2.32)).

Таблиця 2.10а

Проміжний результат обчислень $P(B|a_2, C)$

	b_1	b_2	b_3
c_1	0,1/0,4	0,1/0,4	0,2/0,4
c_2	0,1/0,1	0/0,1	0/0,1
c_3	0/0,15	0,1/0,15	0,05/0,15

Таблиця 2.10б

Остаточний результат обчислень $P(B|a_2, C)$

	b_1	b_2	b_3
c_1	0,25	0,25	0,5
c_2	1	0	0
c_3	0	2/3	1/3



2.10 Вступ до алгебри потенціалів

Розглянемо властивості алгебри, яка стосується множення і маргіналізації таблиць. Такі таблиці не обов'язково повинні бути таблицями (умовних) ймовірностей, а в загальному випадку їх називають *потенціалами* [3, 61].

Потенціал ϕ – це дійсна функція, визначена над областю скінченних значень змінних χ :

$$\phi : sp(\chi) \rightarrow \mathfrak{R}.$$

Область визначення потенціалів (*домен*) позначається так: $dom(\phi)$. Наприклад, областю визначення потенціалу $P(A, B | C)$ є $dom(P(A, B | C)) = \{A, B, C\}$.

Потенціали можна *перемножувати*; ця операція позначається крапкою (яка може не наводитись у явному вигляді). Множення потенціалів має такі властивості:

1. Область визначення добутку потенціалів:

$$dom(\phi_1 \phi_2) = dom(\phi_1) \cup dom(\phi_2).$$

2. Комутативний закон для добутку потенціалів:

$$\phi_1 \cdot \phi_2 = \phi_2 \cdot \phi_1.$$

3. Асоціативний закон для добутку потенціалів:

$$(\phi_1 \cdot \phi_2) \phi_3 = \phi_1 (\phi_2 \cdot \phi_3).$$

4. Існування одиничного потенціалу: одиничним називають потенціал, який містить тільки одиниці („1”) і який визначений на будь-якій області, такій, що $\mathbf{1} \cdot \phi = \phi$ для всіх потенціалів ϕ .

Визначений вище оператор маргіналізації (підрозділ 2.9) можна узагальнити на потенціали таким чином: $\sum_A \phi$ – це потенціал над областю $dom(\phi) \setminus \{A\}$.

Далі: операція маргіналізації *комутативна*:

$$\sum_A \sum_B \phi = \sum_B \sum_A \phi.$$

5. Для потенціалів типу $P(A|\vartheta)$, де ϑ – множина змінних, маємо:

властивість одиничного потенціалу: $\sum_A P(A|\vartheta) = 1$.

6. Розподільний закон:

якщо $A \notin \text{dom}(\phi_1)$, то $\sum_A \phi_1 \phi_2 = \phi_1 \sum_A \phi_2$.

Розподільний закон відомий у формі: $ab + ac = a(b + c)$, а попередня формула (у пункті 6) – це такий же закон, але застосований до таблиць. Для перевірки його дії розглянемо обчислення, наведені у таблицях 2.11 – 2.15. Бачимо, що таблиці 2.13 і 2.15 – однакові і відповідають лівій і правій частинам закону розподілу.

Приклад результату множення потенціалів $\phi_1(A, B)$ і $\phi_2(C, B)$ наведено у таблиці 2.12. Добутки $(x_i y_j, x_i y_l)$ відповідають станам c_1 і c_2 .

Таблиця 2.11

Потенціали $\phi_1(A, B)$ і $\phi_2(C, B)$

$B \setminus A$	a_1	a_2	$B \setminus C$	c_1	c_2
b_1	x_1	x_2	b_1	y_1	y_2
b_2	x_3	x_4	b_2	y_3	y_4

Таблиця 2.12

Результат множення потенціалів $\phi_1(A, B)$ і $\phi_2(C, B)$.

$B \setminus A$	a_1	a_2
b_1	$(x_1 y_1, x_1 y_2)$	$(x_2 y_1, x_2 y_2)$
b_2	$(x_3 y_3, x_3 y_4)$	$(x_4 y_3, x_4 y_4)$

Результат маргіналізації по C таблиці 2.12 наведено у таблиці 2.13. Це продовження ілюстрації розподільного закону $\sum_C \phi_1(A, B) \cdot \phi_2(C, B)$.

Таблиця 2.13

Маргіналізація по C таблиці 2.12.

$B \setminus A$	a_1	a_2
b_1	$x_1 y_1 + x_1 y_2$	$x_2 y_1 + x_2 y_2$
b_2	$x_3 y_3 + x_3 y_4$	$x_4 y_3 + x_4 y_4$



Результат маргіналізації по C потенціалу $\phi_2(C, B)$ наведено у таблиці 2.14. Це продовження ілюстрації розподільного закону: $\sum_C \phi_2(C, B)$.

Процедуру маргіналізації будемо називати також *проекцією*. Наприклад, якщо A і B маргіналізуються з $\phi(A, B, C)$, то можна сказати, що ϕ проектується на C ; ця операція позначається так: $\phi^{\downarrow C}$. Використовуючи це позначення, властивості маргіналізації можна представити таким чином (тут V і W – множини змінних), як представлено у таблиці 2.15. Це продовження ілюстрації розподільного закону: $\phi_1(A, B) \sum_C \phi_2(C, B)$.

Таблиця 2.14

Маргіналізація по C потенціалу $\phi_2(C, B)$

B	
b_1	$y_1 + y_2$
b_2	$y_3 + y_4$

Таблиця 2.15

Ілюстрація розподільного закону

$B \setminus A$	a_1	a_2
b_1	$x_1(y_1 + y_2)$	$x_2(y_1 + y_2)$
b_2	$x_3(y_3 + y_4)$	$x_4(y_3 + y_4)$

7. Комутативний закон для проекції:

$$(\phi^{\downarrow V})^{\downarrow W} = (\phi^{\downarrow W})^{\downarrow V}.$$

8. Дистрибутивний закон для проекції:

$$\text{якщо } \text{dom}(\phi_1) \subseteq V, \text{ то } (\phi_1, \phi_2)^{\downarrow V} = \phi_1(\phi_2^{\downarrow V}).$$

2.11 Випадкові змінні

Розглянемо вибіркового простір S . *Випадкова змінна* визначається дійсною функцією $V : S \rightarrow \mathfrak{R}$. Наприклад, якщо при підкиданні кубиків вважається, що випадання значень $\{1, 2, 3\}$ означає програш однієї гривні, а випадання $\{4, 5, 6\}$ означає виграш однієї гривні, то випадкова змінна є функцією із значеннями:

"-1" на області $\{1, 2, 3\}$ і "+1" на області $\{4, 5, 6\}$.

Середнє значення μ випадкової величини V у вибірковому просторі S визначається так:

$$\mu(V) = \sum_{s \in S} V(s) p(s). \quad (2.33)$$

Наприклад, для збалансованого грального кубика середнє значення випадкової змінної, пов'язаної із виграш або програш однієї гривні, складає:

$$(-1) \cdot \frac{1}{6} + (-1) \cdot \frac{1}{6} + (-1) \cdot \frac{1}{6} + (+1) \cdot \frac{1}{6} + (+1) \cdot \frac{1}{6} + (+1) \cdot \frac{1}{6} = 0.$$

Середнє називають ще *очікуваним значенням* (або сподіванням).

Мірою розсіювання випадкової змінної є дисперсія σ^2 , яка визначається як середнє суми квадратів відхилень значень змінної від її середнього:

$$\sigma^2(X) = \sum_{s \in S} [X(s) - \mu(X)]^2 p(s). \quad (2.34)$$

Для наведеного вище прикладу:

$$\sigma^2 = 3(-1-0)^2 \frac{1}{6} + 3(1-0)^2 \frac{1}{6} = 1.$$

Вступ до неперервних розподілів

Розглянемо експеримент із стрільбою стріл у мішень одиничного розміру: $[1,0 \times 1,0]$. Результатами стрільби будуть попадання у точки мішені з координатами: (x, y) на одиничній площі. Оскільки для будь-якого конкретного результату (координат) ймовірність попадання складає 0, то розподіл ймовірностей присвоюється до підмножин одиничної площі. Цю процедуру (оцінювання розподілу ймовірностей) можна розглядати як розподіл "1" –ї ймовірності на вибірковому просторі. Наприклад, можна присвоїти ймовірність попадання стріли у квадратик (елемент) площею $[x, x+\varepsilon] \times [y, y+\varepsilon]$. Для більшої систематизації цього процесу виберемо деяке натуральне число n і розділимо одиничну площу мішені на невеликі квадратики:

$$\left[\frac{i}{n}, \frac{i+1}{n} \right] \times \left[\frac{j}{n}, \frac{j+1}{n} \right].$$



Тепер присвоїмо ймовірності попадання стріли в ці квадратики:

$$P\left(\left[\frac{i}{n}, \frac{i+1}{n}\right] \times \left[\frac{j}{n}, \frac{j+1}{n}\right]\right),$$

які мають площу $\frac{1}{n^2}$. Якщо $P\left(\left[\frac{i}{n}, \frac{i+1}{n}\right] \times \left[\frac{j}{n}, \frac{j+1}{n}\right]\right) = x$, то значення ймовірності x розподілене на невеликій площі із середньою густиною $x/(1/n^2) = n^2 x$.

Це дає можливість визначити *функцію густини (щільності)* розподілу (її називають також *частотною функцією*) $f(x, y)$ у вигляді:

$$f(x, y) = \lim_{n \rightarrow \infty} n^2 P\left(\left[x, x + \frac{1}{n}\right] \times \left[y, y + \frac{1}{n}\right]\right).$$

Якщо S – неперервний вибірковий простір, то функція щільності – це невід’ємна дійсна функція f на просторі S , така, що для будь-якої підмножини $A \subseteq S$ виконується рівність:

$$\int_A f(s) ds = p(A).$$

Зокрема,

$$\int_S f(s) ds = 1.$$

Якщо S визначена на інтервалі $[a, b]$ (можливо нескінченному) і результатами експерименту є дійсні числа (наприклад, висота або вага), а нас цікавить середнє (середня висота або вага), його можна визначити так:

$$\mu = \int_a^b x f(x) dx,$$

а дисперсія визначається за виразом:

$$\sigma^2 = \int_a^b (\mu - x)^2 f(x) dx.$$

Математично середнє і дисперсія – це середнє і дисперсія одиничної функції $I(x) = x$, але ми будемо використовувати термін „середнє і дисперсія розподілу”.

РОЗДІЛ 3

ДЕЯКІ ВАРІАНТИ ТЕОРЕМИ БАЙЄСА

3.1 Функція і принцип правдоподібності

Функцією правдоподібності (ФП) для множини вимірів $[x_1, \dots, x_n]$ є їх спільна щільність розподілу ймовірностей, яка розглядається як функція невідомого параметра θ , який задає (ідентифікує) розподіл, з якого згенеровані значення $x_i, i = 1, \dots, n$. ФП будемо позначати так:

$$L(\theta; x_1, \dots, x_n).$$

Таке позначення підкреслює той факт, що функція правдоподібності є функцією параметра θ , і що при побудові алгоритму оцінювання на основі максимізації правдоподібності, або при використанні ФП для формування байєсівського висновку цільовою змінною є параметр θ , а не змінні $x_i, i = 1, \dots, n$. Загалом, правдоподібність означає збіжність істинного (невідомого) розподілу, з якого фактично згенеровані дані, з розподілом, який ми оцінюємо. Ця збіжність досягається шляхом пошуку такого значення параметра розподілу, який максимізує функцію правдоподібності [3, 22-24].

Функція щільності розподілу ймовірностей позначається так:

$$f(x_i, \dots, x_n | \theta).$$

Це функція від даних за умови, що вони відповідають розподілу з параметром θ . Математично, з точністю до постійного множника, можна записати:

$$L(\theta; x_1, \dots, x_n) = f(x_1, \dots, x_n | \theta)$$

для всіх допустимих значень θ і $x_i, i = 1, \dots, n$. Функція правдоподібності визначається однозначно тільки до постійного множника. При використанні алгоритму на основі максимуму правдоподібності (МП) константа пропорційності не грає фундаментальної ролі. Оцінка, обчислена за методом максимальної правдоподібності, не залежить від константи пропорційності, тобто при прирівнюванні похідної нулю



$$c \frac{\partial f(\cdot)}{\partial \theta} = 0$$

остаточний результат не залежить від константи. У загальному випадку ФП – суб’єктивна, оскільки залежить від суб’єктивно зібраних та оброблених даних.

Як ілюстрацію розглянемо біноміальний розподіл для кількості успішних випадків (результатів): $r = \sum_{i=1}^n x_i$ з n незалежних експериментів (Бернуллі), де

$$x_i = \begin{cases} 1 & \text{з ймовірністю } \theta, \\ 0 & \text{з ймовірністю } 1 - \theta. \end{cases}$$

Функція щільності розподілу має вигляд:

$$f(r|\theta) = \binom{n}{r} \theta^r (1-\theta)^{n-r} = C_n^r \theta^r (1-\theta)^{n-r}.$$

Ігноруючи біноміальний коефіцієнт $\binom{n}{r}$, функцію правдоподібності для θ можна записати таким чином:

$$L(\theta; r) = \theta^r (1-\theta)^{n-r}.$$

У такому записі підкреслюється, що ключовим елементом ФП є параметр θ , а біноміальний коефіцієнт не відіграє фундаментальної ролі для формування байєсівського висновку.

Зазначимо, що функції $L(\theta; r)$ і $f(r|\theta)$ відрізняються тільки константою пропорційності $\binom{n}{r}$, і що оцінка параметра за методом максимальної правдоподібності не залежить від значення константи пропорційності.

Приклад 3.1. Функція правдоподібності для розподілу Пуассона.

Розподіл Пуассона використовують для опису випадкової кількості подій, які відбуваються протягом деякого визначеного часового інтервалу. Він має один параметр θ , який характеризує швидкість появи випадкової події, тобто скільки в середньому відбувається подій за одиницю часу; розподіл описується виразом:

$$p(X = k) = \frac{\theta^k}{k!} e^{-\theta}, \quad k = 0, 1, 2, \dots \quad (3.1)$$

Припустимо, що необхідно дослідити щорічну кількість дефолтів американських корпорацій (США), які представлені на фондових біржах. Дані стосовно дефолтів зібрано за період з 1986-го по 2005-й рік. Для опису даних вибираємо розподіл Пуассона. Позначимо наявні спостереження x_1, x_2, \dots, x_{20} і запишемо для них функцію правдоподібності:

$$\begin{aligned} L(\theta | x_1, x_2, \dots, x_{20}) &= \prod_{i=1}^{20} p(X=x_i | \theta) = \prod_{i=1}^{20} \frac{\theta^{x_i}}{x_i!} e^{-\theta} = \\ &= \frac{\theta^{\sum_{i=1}^{20} x_i}}{\prod_{i=1}^{20} x_i!} \exp(-20\theta). \end{aligned} \quad (3.2)$$

Як буде показано нижче, досить часто у виразах для функції правдоподібності і для розподілів ймовірностей залишають тільки ті члени, які містять невідомі параметри. Тобто члени, які представляють собою константи стосовно параметрів, не записують. Таким чином, вираз (3.2) можна записати так:

$$L(\theta | x_1, x_2, \dots, x_{20}) \propto \theta^{\sum_{i=1}^{20} x_i} \exp(-20\theta), \quad (3.3)$$

де символ « \propto » означає *пропорційність*. Очевидно, що для заданої вибірки даних вирази (3.2) і (3.3) пропорційні один одному і містять однакову інформацію про θ . Максимізація кожного з них по θ (похідна по θ прирівнюється нулю) дасть оцінку цього параметра розподілу Пуассона, тобто вибіркоче середнє.

Позначимо $\sum_{i=1}^{20} x_i = a$ і знайдемо похідну від функції правдоподібності по θ і прирівняємо її нулю:

$$\begin{aligned} \frac{\partial(\theta^a e^{-20\theta})}{\partial \theta} &= a \theta^{a-1} e^{-20\theta} + \theta^a (-20e^{-20\theta}) = \\ &= \theta^a e^{-20\theta} (a \theta^{-1} - 20) = 0 \end{aligned}$$

а звідси маємо: $a \theta^{-1} - 20 = 0$, або $\theta^{-1} a = 20$, і

$$\hat{\theta} = \bar{x} = \frac{\sum_{i=1}^{20} x_i}{20} = 51,6.$$

Таким чином, для 20 спостережень стосовно щорічної кількості дефолтів корпорацій (США) значення вибіркового середнього складає 51,6.

Приклад 3.2. Функція правдоподібності для нормального розподілу.

Нормальний (*гаусівський*) розподіл – найшироковживаний і важливий розподіл завдяки відносній простоті його застосування та



наявності відповідних теоретичних розробок. Так, існує центральна гранична теорема, яка стверджує (за виконання деяких умов), що суми незалежних випадкових змінних прямують до нормального розподілу при нескінченному зростанні їх кількості. Цей розподіл задається двома параметрами: – *середнім* μ , яке визначає, також, положення розподілу на осі дійсних чисел, і *стандартним відхиленням* σ (параметр масштабу).

Функція щільності (або просто щільність) нормального розподілу має вигляд:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(y(k) - \mu)^2}{2\sigma^2}\right], \quad (3.4)$$

де y і μ – можуть приймати будь-які дійсні значення, а дисперсія σ^2 – тільки додатні дійсні значення; k – номер виміру. Скорочено нормальний розподіл змінної Y записують так: $Y \sim N(\mu, \sigma)$. Графік нормального розподілу симетричний стосовно середнього.

Припустимо, що зібрані значення ціни біржового активу за період 02.01.1998 – 31.12.2003 (1548 значень), які незалежні (одне від одного) і мають однаковий розподіл. Припустимо, що ці значення мають нормальний розподіл. У такому випадку функція правдоподібності має вигляд:

$$\begin{aligned} L(\mu, \sigma | y(1), y(2), \dots, y(1548)) &= \prod_{k=1}^{1548} f[y(k)] = \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{1548} \exp\left[-\sum_{k=1}^{1548} \frac{(y(k) - \mu)^2}{2\sigma^2}\right] \propto \\ &\propto \sigma^{-1548} \exp\left[-\sum_{k=1}^{1548} \frac{(y(k) - \mu)^2}{2\sigma^2}\right]. \end{aligned} \quad (3.5)$$

Оскільки функція правдоподібності – це функція двох аргументів, то її можна представити у вигляді тривимірного графіка (рис. 3.1а). Також доцільно розглянути контурний багаторівневий графік (рис. 3.1б), який ілюструє так звані контури правдоподібності, які можна отримати шляхом введення горизонтальної січної площини на різних рівнях правдоподібності.

Кожний контур на рис. 3.1б відповідає парі конкретних значень параметрів розподілу. При цьому ймовірніші значення цих параметрів відповідають більшим значенням правдоподібності (по вертикалі). Так, пара значень $(\mu_1, \sigma_1) = (-0,23E-03; 0,31E-03)$ є правдоподібнішою на рівні правдоподібності 0,6, ніж пара $(\mu_2, \sigma_2) = (0,096E-03; 0,33E-03)$ на рівні правдоподібності 0,1.

Принцип правдоподібності

Нехай $\mathbf{x} = (x_1, \dots, x_n)^T$ і $\mathbf{y} = (y_1, \dots, y_n)^T$ – дві множини спостережень і $f(x_1, \dots, x_n | \theta)$ – щільність розподілу ймовірностей для значень $x_i, i = 1, \dots, n$. Припустимо, що $L(\theta; \mathbf{x}) \propto L(\theta; \mathbf{y})$, тобто функція правдоподібності для вимірів змінної x пропорційна ФП для вимірів змінної y ; при цьому константа пропорційності не залежить від θ . Таким чином, ці функції правдоподібності відрізняються лише мультиплікативною константою. Але це означає, що оцінка параметра θ , обчислена за методом МП (ММП) на основі вибірки значень змінної x , буде такою ж, як і оцінка, обчислена за вибіркою змінної y за однакових початкових умов (апостеріорного розподілу). Цей факт називають принципом правдоподібності.

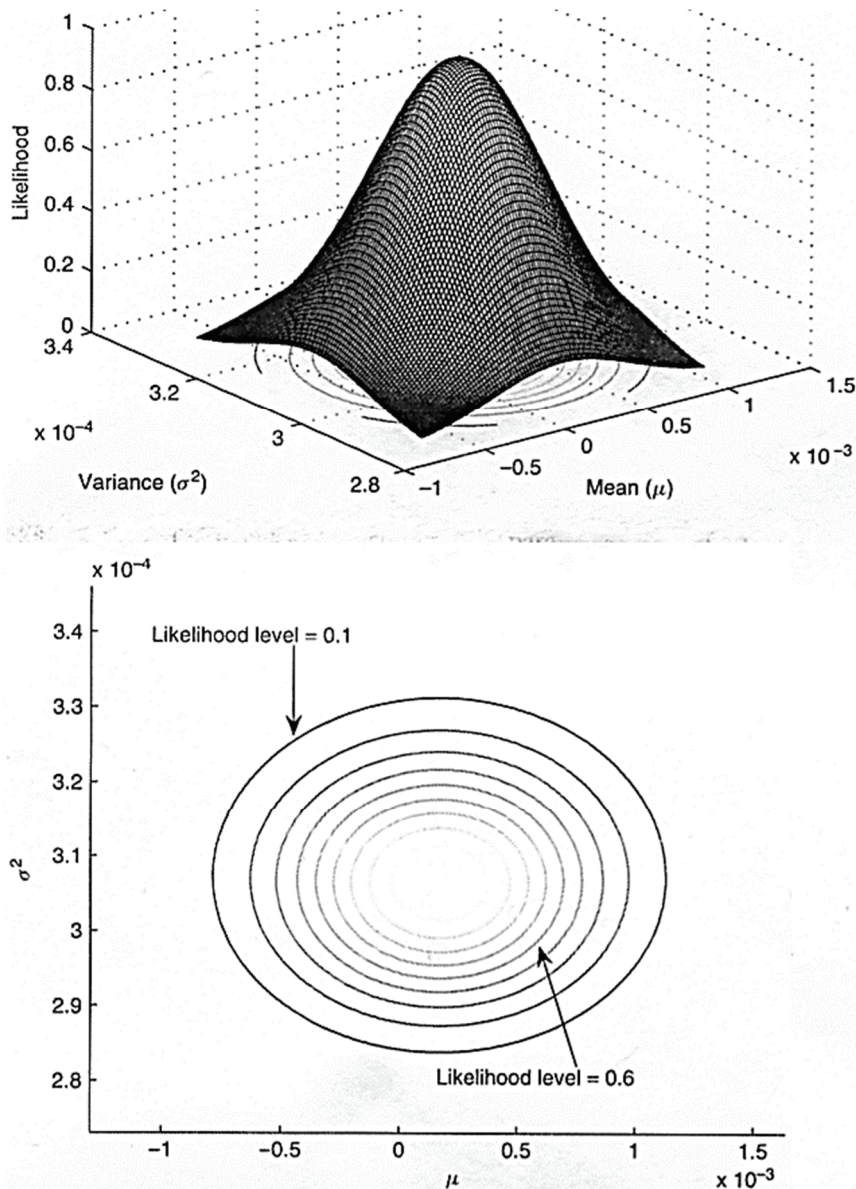
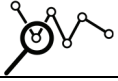


Рисунок 3.1 – Функція правдоподібності нормального розподілу



Оцінка параметра θ , обчислена за методом максимальної правдоподібності (ММП) на основі вибірки значень змінної x , буде такою ж, як і оцінка, обчислена за вибіркою змінної y . Цей факт називають *принципом правдоподібності*.

Еквівалентним твердженням є таке: всю інформацію про параметр θ , що міститься в експериментальних (статистичних) даних, і яка необхідна для отримання значення оцінки цього параметра, можна отримати з функції правдоподібності для цих даних. Відповідно до принципу правдоподібності, інформація стосовно умов постановки експерименту (або інша інформація про експеримент) не має відношення до формування висновку (обчислення оцінок), оскільки її немає у функції правдоподібності.

Формування (обчислення) класичного статистичного висновку ґрунтується на багаторазовому повторенні експерименту. Але згідно з принципом правдоподібності висновок повинен ґрунтуватись тільки на тих спостереженнях, які отримані в одному виконаному експерименті. Тобто результат формується тільки на основі даних поточного експерименту.

Наприклад, розглянемо можливості аналізу радіоактивно забруднених зразків ґрунту і рослин з Чорнобильської зони у Франції і Англії. В обох країнах існують відповідні лабораторії високого рівня компетентності. Внаслідок простого жеребкування вибрали лабораторію у Англії, де і було виконано аналіз зразків ґрунту і рослин. Постає запитання: чи потрібно приймати для використання результати аналізу, зробленого у Англії, або ж посилати ґрунт та рослини додатково на аналіз у Францію, оскільки вибір Англії був випадковим? *Принцип правдоподібності говорить*: «ні», оскільки для отримання достовірного результату достатньо одного експерименту, виконаного у одному місці.

Зв'язок правдоподібності з байєсівським висновком

Існують різні варіанти формулювання теореми Байєса. Нехай невідомий параметр θ – це неперервна величина, яку необхідно оцінити на основі вибірки даних $[x_1, \dots, x_n]$. Для такого випадку теорема Байєса дає можливість знайти (обчислити) апостеріорну функцію щільності розподілу ймовірностей (ЩРЙ) для θ такого вигляду:

$$h(\theta | x_1, \dots, x_n) = \frac{f(x_1, \dots, x_n | \theta) g(\theta)}{\int_{\theta} f(x_1, \dots, x_n | \theta) g(\theta) d\theta}$$

де $f(\cdot)$ – спільна функція щільності розподілу ймовірностей для даних; $g(\theta)$ – апіорна ЩРЙ для параметра θ ; $h(\theta | x_1, \dots, x_n)$ – апостеріорна ЩРЙ для цього параметра. Необхідно підкреслити, що $h(\theta | x_1, \dots, x_n)$ залежить від даних тільки через ЩРЙ для $f(\cdot)$. Більше того, якщо ми замінимо $f(\cdot)$ функцією правдоподібності $L(\theta; x_1, \dots, x_n)$, яка має такий же вигляд, як і $f(\cdot)$, за винятком константи пропорційності c , наприклад такий:

$$h(\theta | x_1, \dots, x_n) = \frac{c L(\theta; x_1, \dots, x_n) g(\theta)}{\int_{\theta} c L(\theta; x_1, \dots, x_n) g(\theta) d\theta},$$

то видно, що константа скорочується. Таким чином, апостеріорна ЩРЙ для параметра залежить від даних тільки через функцію правдоподібності $L(\cdot)$. Оскільки байєсівський висновок у цілому формується на основі апостеріорного розподілу, то він також залежить від даних тільки через функцію правдоподібності.

Звідси очевидно, яку важливу фундаментальну роль відіграє функція правдоподібності у байєсівському висновку. Вона повинна бути максимально наближеною до даних, тобто відображати їх характеристики. Функція правдоподібності може забезпечувати належну апроксимацію тільки за умови наявності вибірок великої потужності. Якщо використовуються вибірки недостатнього розміру, то виникає параметрична невизначеність стосовно оцінки параметра θ , зумовлена неточністю апіорного розподілу і функцією правдоподібності [65, 66].

3.2 Теорема Байєса для довільних дискретних подій

Теорема Байєса (ТБ) стосується умовних ймовірностей подій загального характеру. Нехай $\mathbf{A} = [A_1, \dots, A_n]$ – множина взаємовиключних подій, які повністю описують деякий процес. Розглянемо деяку подію A_j , виникнення якої зв'язане з подією B . Теорема Байєса дає можливість визначити умовну ймовірність настання події A_j при умові, що мала місце подія B , використовуючи при цьому інформацію щодо умовної ймовірності настання події B (при умові, що мала місце подія A_j). Виходячи з наведеного формулювання призначення теореми Байєса, можна сказати, що вона стосується так званої “інверсної ймовірності”. Теорема Байєса для довільних подій записується так:

$$p(A_j | B) = \frac{p(B | A_j) p(A_j)}{\sum_{j=1}^n p(B | A_j) p(A_j)} = \frac{p(B | A_j) p(A_j)}{p(B)}, p(B) \neq 0. \quad (3.6)$$



Зазначимо, що інтерпретація події A_j має конкретний (персональний) характер. Тобто ймовірність настання події A_j (до появи B , яка може вплинути на A) визначається апіорно і називається *апіорною ймовірністю* події A_j . Аналогічно, $p(A_j|B)$ в рівнянні (3.6) є *апостеріорною ймовірністю* події A_j , обчисленою за умови, що мала місце подія B .

Таким чином, можливою інтерпретацією теореми (3.6) є така: вона являє собою *механізм оновлення ступеня упевненості в тому, що відбудеться A_j в світлі нової інформації B* , яка має відношення до A_j . Первинна (апіорна) інформація $p(A_j)$ замінюється уточненою $p(A_j|B)$. З цієї точки зору теорему можна розглядати як *правило навчання на основі минулого досвіду*. Оновлення інформації відбувається від початкового ступеня довіри (або ймовірності події) $p(A_j)$ до оновленої ймовірності $p(A_j|B)$.

3.2.1 Теорема Байєса для взаємодоповнюючих подій

Окремим випадком теореми (3.6) є її формулювання для взаємодоповнюючих подій. Позначимо через \bar{C} подію, яка є доповненням для C . Тобто, C і \bar{C} – взаємовиключні події, які повністю описують деякий процес. Для будь-якої іншої події B ($p(B) \neq 0$) теорема Байєса для таких подій має вигляд:

$$p(C|B) = \frac{p(B|C)p(C)}{p(B|C)p(C) + p(B|\bar{C})p(\bar{C})}. \quad (3.7)$$

3.2.2 Роль апіорних ймовірностей

Апіорна ймовірність $p(A_j)$ необхідна для застосування теореми Байєса для подій (3.6). *Апіорні ймовірності* – це ступені віри (довіри), які доступні аналітику до появи та аналізу даних, які мають відношення до проблеми. У тих випадках, коли подальші дані щодо проблеми отримати неможливо, то апіорними ймовірностями можна скористатись для формулювання висновку (прийняття рішень). Тобто, ТБ у таких випадках не використовується [28, 105]. Наприклад, у політиці більшість рішень приймаються без чисельних даних, вони приймаються на основі суджень із врахуванням наявної (різномірної) інформації. Все що є у розпорядженні аналітика в даному випадку – це апіорна інформація. З цього випливає, що

до визначення апіорних ймовірностей потрібно підходити дуже виважено.

В науці, бізнесі, праві, техніці і медицині висновки та прийняття рішень стосовно невідомих величин приймаються на основі навчання, попереднього досвіду та розуміння теорії, яка лежить в основі явищ та процесів. Якщо у процесі виконання дослідження задачі вдається зібрати досить великий (достатній для формулювання достовірного висновку) об'єм інформації щодо цієї задачі, то апіорні ймовірності, як правило, опускаються з розгляду і для подальшого аналізу використовуються тільки та інформація, яка була зібрана у процесі виконання дослідження. Однак, якщо об'єм зібраної інформації невеликий (недостатній), то при формулюванні висновку апіорна інформація може відігравати значну роль.

3.2.3 Апостеріорні ймовірності

Апостеріорні ймовірності – це ймовірності, які ми отримуємо в результаті застосування теореми Байєса. Для того щоб розв'язок задачі був коректним, сума апостеріорних ймовірностей взаємовиключних подій, які повністю описують деякий процес, повинна дорівнювати 1. Якщо після виконання аналізу даних за допомогою ТБ з'явилися нові дані, то апостеріорні ймовірності, отримані на основі попередніх даних, необхідно використати як апіорні ймовірності для нових даних [108-110].

3.2.4 Шанс настання події

Відношення доповнюючих ймовірностей – це альтернативне представлення апіорних або апостеріорних ймовірностей. Так, якщо p – ймовірність події E , то шанс *проти* появи події E визначається відношенням: $\frac{1-p}{p}$, а шанс *на користь* появи події E визначається відношенням $\frac{p}{1-p}$.

Приклад 3.3. Застосування теореми Байєса до аналізу простих бінарних подій: розпізнавання відбитків пальців. Знімання відбитків пальців запропоновано Френсісом Гальтоном в 1890-х роках. Цей метод швидко знайшов широке застосування для розпізнавання (ідентифікації) людей за відбитками їх пальців, які не змінюються протягом всього життя.

Існує також метод ідентифікації особи, який ґрунтується на визначенні її групи крові. Наприклад, до групи A відносяться особи, які мають гени типу A ; до групи B – ті, що мають гени типу B ; до групи AB



– особи, які мають гени обох типів – A і B і до групи O відносять тих, що не мають генів жодної з груп – A і B . Однак, виникає запитання: що робити у випадках, коли аналіз крові виконати неможливо?

Пізніше було встановлено, що люди сильно відрізняються один від одного за будовою ланцюжків дезоксирибонуклеїнової кислоти (ДНК), які можна знайти у кожній клітині тіла. Тобто в окремих випадках для ідентифікації особи можна скористатись дослідженням клітин тіла замість відбитків пальців або аналізів крові. Ця методика була розроблена в Англії професором А. Джеффри в університеті Лісестера в 1984 році. Згідно з цією методикою, для ідентифікації особи необхідно порівняти кілька різних характеристик ДНК для окремо взятих зразків. Розглянемо основні принципи розпізнавання за цією методикою.

Нехай на місці злочину знайдено артефакти (матеріали), які містять інформацію щодо ДНК *злочинця* (наприклад, краплі крові, пучок волосся, сім'я, зідрану шкіру і т. ін.). Припустимо, слідчі коректно вважають, що (*підозрювана*) особа, ДНК якої виявлено на місці злочину, дійсно вчинила цей злочин. Таким чином, ДНК підозрюваного необхідно співставити з ДНК дійсного злочинця. Застосуємо для розв'язку цієї задачі теорему Байєса.

Введемо в розгляд такі події:

- C – підозрюваний вчинив злочин;
- \bar{C} – підозрюваний не вчинив злочин;
- M – ДНК підозрюваного та ДНК, визначене за допомогою знайдених артефактів (частинок тіла), збігаються;
- \bar{M} – ДНК підозрюваного та ДНК, визначене за допомогою знайдених частинок тіла, не збігаються.

Нехай дослідження підозрюваного показало, що його ДНК збігається з ДНК, встановленого за допомогою частинок тіла, знайдених на місці злочину. В результаті *декларується збіжність* двох ДНК. Чи дійсно у такому випадку підозрюваний вчинив злочин? За теоремою Байєса (3.7) *апостеріорна ймовірність* того, що підозрюваний вчинив злочин за умови збіжності типів ДНК:

$$p(C|M) = \frac{p(M|C)p(C)}{p(M|C)p(C) + p(M|\bar{C})p(\bar{C})}. \quad (3.8)$$

Введемо позначення: p_{ca} – апіорна ймовірність події C , тобто, апіорну ймовірність події $p(C)$ і запишемо теорему у вигляді:

$$p(C|M) = \frac{p(M|C)p_{ca}}{p(M|C)p_{ca} + p(M|\bar{C})(1 - p_{ca})} = \frac{p(M|C)p_{ca}}{p(M)}.$$

Оскільки

$$p(\bar{C} | M) = \frac{p(M | \bar{C}) (1 - p_{ca})}{p(M | \bar{C}) (1 - p_{ca}) + p(M | C) p_{ca}}, \quad (3.9)$$

то апостеріорний шанс того, що підозрюваний не вчинив злочин має вигляд:

$$\frac{p(\bar{C} | M)}{p(C | M)} = \frac{p(M | \bar{C})}{p(M | C)} \cdot \frac{(1 - p_{ca})}{p_{ca}}, \quad (3.10)$$

де $[(1 - p_{ca}) / p_{ca}]$ – апіорний шанс того, що підозрюваний не вчинив злочин. З цього прикладу видно, що перевагою використання відношення шансів є те, що немає необхідності обчислювати безумовну ймовірність $p(M)$, тобто, немає необхідності обчислювати *константу пропорційності* (знаменник) в теоремі Байєса [111, 113].

Фактор Байєса

Фактор Байєса – це апостеріорний шанс того, що підозрюваний скоїв злочин.

Оскільки $p(M | C) = 1 - p(\bar{M} | C)$, то можна записати вираз для фактора Байєса у вигляді:

$$r = \frac{p(M | C)}{p(M | \bar{C})} = \frac{p(M | C)}{1 - p(\bar{M} | C)}, \quad (3.11)$$

який залежить тільки від даних. На основі виразу (3.10) отримаємо:

$$\frac{p(C | M)}{p(\bar{C} | M)} = \frac{p(M | C)}{p(M | \bar{C})} \left[\frac{p_{ca}}{1 - p_{ca}} \right] = r \left[\frac{p_{ca}}{1 - p_{ca}} \right], \quad (3.12)$$

де r – *фактор Байєса*, який являє собою апостеріорний шанс того, що підозрюваний скоїв злочин. Таким чином, незалежно від значення p_{ca} , чим більшим буде фактор Байєса, тим більшим буде шанс на користь події C (підозрюваний скоїв злочин). Зазначимо, що значення r залежить тільки від даних. В подальшому фактор Байєса буде використовуватись при перевірці гіпотез.

Тепер необхідно оцінити апіорну ймовірність p_{ca} того, що підозрюваний вчинив злочин. Підозрюваного бачили біля місця злочину приблизно у той же час, коли було скоєно злочин. Для простоти обчислень прийемо, що апіорна ймовірність $p_{ca} = 0,5$.



У даному прикладі є ймовірності припуститись двох помилок:

1. $p(\bar{M} | C)$ – ймовірність прийти до висновку, що немає збіжності ДНК, якщо вона повинна бути;
2. $p(M | \bar{C})$ – ймовірність того, що є збіжність ДНК, коли її не має бути (припустимо, що підозрюваний – це випадкова особа, яка не є злочинцем).

Ці ймовірності визначаються поточним рівнем розвитку технології визначення ДНК, припустимо (за наявними статистичними даними), що $p(M | \bar{C}) = 0,02$, а $p(\bar{M} | C) = 0,001$.

Обчислення цих значень досить складне і залежить від результатів порівняння ДНК підозрюваного з ДНК осіб, які мають таке ж етнічне походження (з відповідної бази даних), що і підозрюваний. Воно залежить також від результатів порівняння декількох незалежних характеристик цих двох зразків ДНК (оцінених за різними зразками).

Використовуючи наведені вище значення для $p(M | \bar{C})$, $p(\bar{M} | C)$, а також те, що $p(M | C) = 1 - p(\bar{M} | C) = 1 - 0,001 = 0,999$, обчислимо ймовірність того, що підозрюваний вчинив злочин:

$$\begin{aligned} p(C | M) &= \frac{p(M | C) p_{ca}}{p(M | C) p_{ca} + p(M | \bar{C})(1 - p_{ca})} = \\ &= \frac{0,999 \cdot 0,5}{0,999 \cdot 0,5 + 0,02 \cdot 0,5} = 0,9804, \end{aligned} \quad (3.13)$$

або біля 98%. З (3.13) знайдемо, що фактор Байєса $r = \frac{0,9804}{1 - 0,9804} = 50,02$.

Апостеріорний шанс того, що підозрюваний не скоїв злочин складає:

$$\frac{1 - p(C | M)}{p(C | M)} = \frac{1 - 0,9804}{0,9804} = 0,02002 \approx \frac{2}{100} = \frac{1}{50},$$

або апостеріорне свідчення на користь того, що підозрюваний скоїв злочин складає біля 50:1.

Таким чином, збіжність ДНК є, у даному випадку, переконливим доведенням того, що підозрюваний скоїв злочин. У випадку, коли апріорний шанс того, що підозрюваний скоїв злочин, відрізняється від $|1|$, то необхідно пожити його на r для того щоб отримати новий фактор Байєса.

3.3 Теорема Байєса для дискретних даних та дискретних параметрів

Нехай X_1, \dots, X_n – взаємно незалежні однаково розподілені спостережувані випадкові змінні з функціями щільності розподілу $f(x|\Theta=\theta)$, де Θ – невідома і неспостережувана дискретна випадкова змінна, можливі значення якої $\theta_1, \dots, \theta_m$ визначають конкретний розподіл $(X|\Theta=\theta)$. Необхідно визначити ймовірність (шанс) того, що $\Theta=\theta$. Теорема Байєса для даного випадку має такий вигляд:

$$p\{\Theta=\theta|x_1, \dots, x_n\} \equiv h(\theta|\text{дані}) = \frac{f(x_1|\theta)\dots f(x_n|\theta)g(\theta)}{\sum_{i=1}^m f(x_1|\theta_i)\dots f(x_n|\theta_i)g(\theta_i)}, \quad (3.14)$$

де $h(\cdot)$ – апостеріорна щільність розподілу ймовірностей для Θ ; $g(\cdot)$ – апріорна щільність розподілу ймовірностей для Θ . Очевидно, що знаменник виразу (3.14) не повинен дорівнювати нулю [3].

3.3.1 Інтерпретація теореми Байєса для дискретних даних і дискретних параметрів

Зазначимо, що X і Θ мають дискретні розподіли ймовірностей. При цьому значення X_i , отримані в результаті конкретного експерименту, повинні нести інформацію щодо ймовірностей тих чи інших значень Θ . Знаменник виразу (3.14) – це константа, яка залежить тільки від даних X і не залежить від θ . Таким чином, цей знаменник є константою по відношенню до Θ . Це дає можливість записати рівняння (3.14) у еквівалентній формі:

$$h(\theta|x_1, \dots, x_n) = c \cdot f(x_1|\theta)\dots f(x_n|\theta)g(\theta) \propto f(x_1|\theta)\dots f(x_n|\theta)g(\theta), \quad (3.15)$$

де c – коефіцієнт пропорційності; \propto – символ пропорційності. Константу c завжди можна відновити, оскільки її необхідно визначати через

$$\sum_{i=1}^m p\{\Theta=\theta_i|x_1, \dots, x_n\} = \sum_{i=1}^m h(\theta_i|x_1, \dots, x_n) = 1. \quad (3.16)$$

Вираз $f(x_1|\theta)\dots f(x_n|\theta)$ в (2.3.2) зазвичай називають функцією правдоподібності для даних. Таким чином, з теореми Байєса для дискретних даних і дискретних параметрів у формі (3.15) випливає, що апостеріорний умовний розподіл (щільність) ймовірностей для Θ є пропорційним добутку функції правдоподібності для даних і апріорної щільності ймовірностей для параметрів.



Цей результат є справедливим як для одновимірних, так і для багатовимірних даних. При цьому елементи даних X_i не повинні бути взаємно незалежними; вони можуть бути корельованими, що, як правило, має місце у стохастичних процесах. Однак, у такому випадку в рівнянні (3.15) необхідно використовувати правдоподібність у формі: $f(x_1, \dots, x_n | \theta)$.

Приклад 3.4. Аналіз якості на виробництві: дискретні дані і дискретні параметри (висновок щодо пропорційності).

Компанія виробляє компоненти комп'ютерів, у тому числі модулі оперативної пам'яті (оперативні запам'ятовуючі пристрої або ОЗП) для персональних комп'ютерів. Виробництво модулів пам'яті сягає сотень тисяч одиниць на місяць. Позначимо через θ – ймовірність виготовлення дефектного модуля пам'яті для серії з n одиниць. Для спрощення подальшого аналізу задачі приймемо, що θ може приймати три можливих значення:

- 0,25 – хороший результат;
- 0,50 – прийнятний результат;
- 0,75 – поганий (неприйнятний) результат.

Статистика виробничого процесу компанії *MEM* свідчить, що протягом п'яти попередніх років на виробництві ОЗП утримувались такі показники якості процесу:

- протягом 60% часу вироблялись пристрої з ймовірністю дефекту $\theta = 0,25$;
- протягом 30% часу вироблялись пристрої з ймовірністю дефекту $\theta = 0,50$;
- протягом 10% часу з попередніх 5 років вироблялись пристрої з ймовірністю дефекту $\theta = 0,75$.

Компанія вирішила скористатись цими результатами як апіорними ймовірностями для того щоб спрогнозувати рівень якості продукції в майбутньому. Значення цих апіорних ймовірностей подані в таблиці 3.1.

Після виробництва 10000 пристроїв оперативної пам'яті компанія вирішила перевірити цю партію для того щоб встановити – чи підтримується якість на попередньому рівні?

Виявилось, що з трьох пристроїв пам'яті, випадково вибраних для перевірки, два пристрої мають дефекти. Який висновок щодо якості продукції необхідно зробити, тобто, яким є апостеріорний розподіл для θ ?

У даному випадку дані мають дискретний характер і спочатку потрібно вибрати (наприклад, оцінити на основі досвіду) тип розподілу для

даних. Припустимо, що вони мають біноміальний розподіл з параметром θ :

$$f(\theta, n, r) = \binom{n}{r} \theta^r (1 - \theta)^{n-r} = C_n^r \theta^r (1 - \theta)^{n-r},$$

де
$$\binom{n}{r} = \frac{n!}{r!(n-r)!}.$$

Таблиця 3.1.

Апріорні ймовірності

Параметри	Якість продукції		
	Хороша	Прийнятна	Погана
Ймовірність дефекту (θ)	0,25	0,50	0,75
Щільність ймовірностей $\theta(p(\theta))$	0,60	0,30	0,10

Кількість успішних подій $r = 2$; успішними вважаємо події, пов'язані з випуском дефектних пристроїв з трьох можливих, тобто $n = 3$.

Таким чином, функція правдоподібності має вигляд:

$$L(\theta) = \binom{3}{2} \theta^2 (1 - \theta) = C_3^2 \theta^2 (1 - \theta),$$

де $\theta = [0,25; 0,5; 0,75]$ – вектор значень дефектності. Рівняння (3.15) має у даному випадку такий вигляд:

$$h(\theta | r, n) \propto L(r | \theta, n) g(\theta).$$

Обчислимо правдоподібність $L(\theta)$:

- якщо $\theta = 0,25$, то $L(\theta) = \frac{1 \cdot 2 \cdot 3}{1 \cdot 2} (0,25)^2 (1 - 0,25) = 0,140625$;
- якщо $\theta = 0,50$, то $L(\theta) = \frac{1 \cdot 2 \cdot 3}{1 \cdot 2} (0,50)^2 (1 - 0,50) = 0,375000$;
- якщо $\theta = 0,75$, то $L(\theta) = \frac{1 \cdot 2 \cdot 3}{1 \cdot 2} (0,75)^2 (1 - 0,75) = 0,421875$.

Апріорні та апостеріорні щільності ймовірності для h наведені в таблиці 3.2.

Елементи в стовпчику h знайдені в результаті ділення значень третього стовпчика ((апріорна ймов.) \times (правдоподібн.)) на загальну суму для цього (передостаннього) стовпчика з метою нормування остаточного результату до 1, тобто:



$$p(\Theta = 0,25 | r, n) \equiv h(0,25 | 2,3) = \frac{0,60 \cdot 0,140625}{0,6 \cdot 0,14625 + 0,3 \cdot 0,375 + 0,1 \cdot 0,421875} =$$

$$= \frac{0,084375}{0,2390625} = 0,35294;$$

$$p(\Theta = 0,50 | r, n) \equiv h(0,50 | 2,3) = \frac{0,30 \cdot 0,375}{0,6 \cdot 0,14625 + 0,3 \cdot 0,375 + 0,1 \cdot 0,421875} =$$

$$= \frac{0,084375}{0,2390625} = 0,47059;$$

$$p(\Theta = 0,75 | r, n) \equiv h(0,75 | 2,3) = \frac{0,10 \cdot 0,421875}{0,6 \cdot 0,14625 + 0,3 \cdot 0,375 + 0,1 \cdot 0,421875} =$$

$$= \frac{0,0421875}{0,2390625} = 0,17647.$$

Таблиця 3.2.

Апріорні та апостеріорні щільності ймовірності

Ймовірність дефекту (θ)	Апріорні ймовірн. для θ	Правдоподібність $L(\theta)$	(апріорна ймов.) x (правдоподібн.)	h = апостер. щільність для θ
0,25	0,60	0,140625	0,0843750	0,35294
0,50	0,30	0,375000	0,1125000	0,47059
0,75	0,10	0,421875	0,0421875	0,17647
Загальна сума:			0,2390625	0,99999

З таблиці 3.2 видно, що найбільш ймовірним значенням θ , на основі аналізу якості останньої партії пристроїв пам'яті, є $\theta = 0,5$, тобто, виробництво знаходиться на "прийнятному" рівні щодо якості готової продукції.

Очевидно, що перед тим як зробити висновок щодо модифікації виробничого процесу з метою підвищення якості продукції, компанія *MEM* повинна тестувати більшу кількість блоків пам'яті ніж $n = 3$; наприклад, $n = 100$, 500 або 1000 .

3.3.2 Теорема Байєса для дискретних даних і дискретних моделей

Припустимо, що ставиться задача порівняти декілька статистичних моделей (наприклад, у формі щільності розподілу ймовірностей) M_1, M_2, \dots, M_k з метою визначення – яка з них є найкращою (правдоподібною). У такому випадку ситуація подібна тій, що мала місце у

прикладі 3.2, де було три альтернативи для ступеня дефектності θ пристроїв оперативної пам'яті. Тобто ми можемо розглядати можливі значення θ як показники якості моделей, кожна з яких є претендентом на пояснення деякого явища.

Можна мати деяке уявлення щодо ймовірності наближення кожної моделі до досліджуваного явища, що розглядається, але для цього потрібно знайти обґрунтування. Для обґрунтування можна скористатись експериментальними даними, які дадуть можливість порівняти апостеріорні ймовірності, що характеризують наближення (адекватність) моделей до цих даних. Теорему Байєса для дискретного числа моделей можна записати у вигляді:

$$p(M_i | X = x) = \frac{f(x | M_i) p(M_i)}{\sum_{j=1}^k f(x | M_j) p(M_j)}, \quad (3.17)$$

де $i=1, \dots, k$; $f(x | M_i)$ – правдоподібність дискретних даних щодо моделі M_i , яка визначається через щільність розподілу.

3.4 Теорема Байєса для неперервних даних і дискретного параметра

Припустимо, що X – спостережувана, абсолютно неперервна змінна з функцією розподілу $f(x | \theta)$. Позначимо через Θ неспостережуваний невідомий дискретний параметр розподілу змінної X , а через $g(\theta)$ – апіорну щільність розподілу Θ . У такому контексті теорема Байєса для апостеріорної щільності розподілу θ визначається таким виразом:

$$h(\theta | x_1, \dots, x_n) = \frac{L(x_1, \dots, x_n | \theta) g(\theta)}{\sum_{i=1}^k L(x_1, \dots, x_n | \theta_i) g(\theta_i)}, \quad (3.18)$$

де $L(x_1, \dots, x_n | \theta)$ – функція правдоподібності для θ .

Форма (3.18) еквівалентна наступній:

$$h(\theta | x) \propto L(x_1, \dots, x_n | \theta) g(\theta).$$

Інтерпретація теореми Байєса для неперервних даних і дискретного параметра

Фактичні дані можуть бути багатовимірними і можуть бути взаємно залежними; але вони повинні бути неперервними, а знаменник (3.18) повинен бути відмінним від нуля [3, 23].



Приклад 3.5. Група студентів складається з трьох підгруп. Для тестування необхідно вибрати студента із однієї з цих трьох підгруп. **(Неперервні дані і дискретний параметр.)**

Таким чином, для тестування необхідно вибрати одного студента з групи, яка складається з трьох підгруп. Задача полягає у тому, щоб визначити: *з якої підгрупи правильно вибрати студента для тестування?*

Спочатку розглянемо яка апіорна інформація може бути використана. Припустимо, що у минулому році всі три підгрупи навчалися з тими ж викладачами, що і поточного року; іспит був таким же, а результати тестування студентів в деякий момент часу були такими, як наведені у таблиці 3.3. З цієї таблиці видно, що оцінки студентів мають нормальний розподіл із середніми значеннями m_i та дисперсією $S_i^2 = 225$, $i = 1, 2, 3$.

Таблиця 3.3

Статистичні моделі минулорічних оцінок студентів

Модель	Середнє	Модель	Апіорна ймовірн. $p\{M\}$
M_1	$\theta_1 = 74$	$N(74, 225)$	0,25
M_2	$\theta_2 = 81$	$N(81, 225)$	0,50
M_3	$\theta_3 = 68$	$N(68, 225)$	0,25

Для того щоб допомогти передбачити оцінки, які будуть отримані поточного року, необхідно присвоїти апіорні ймовірності оцінкам поточного року на основі інформації щодо минулорічних результатів та поточної інформації стосовно студентів. Випадково виберемо студента A , який має хороші поточні оцінки. Також відомо, що кращою серед трьох підгруп є *друга підгрупа*, яка характеризується моделлю M_2 . На основі наявної інформації ми присвоюємо апіорні ймовірності трьом статистичним моделям (розподілам ймовірностей) так, як показано в 4-у стовпчику таблиці 3.3.

Правдоподібність (щільність ймовірностей) для нормального розподілу визначається за виразом:

$$L(x|\theta_i) = \frac{1}{15\sqrt{2\pi}} \exp\left\{-0,5\left(\frac{x-\theta_i}{15}\right)^2\right\},$$

$$i = 1, 2, 3.$$

Оскільки у поточному році студент A набрав 76% від максимальної оцінки, то для обчислення функцій правдоподібності скористаємось

оцінкою $x = 0,76$. Знайдемо z -оцінки (нормовані або стандартизовані випадкові змінні) за виразом:

$$z_i = \frac{x - \mu}{\sigma} = \frac{x - \theta_i}{15}.$$

Оскільки $z_i \sim N(0, 1)$ із стандартною нормальною щільністю розподілу

$$\Phi(z_i) \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_i^2},$$

то значення правдоподібності можна визначити за допомогою таблиці площ стандартного нормального розподілу. Відповідні результати наведені в таблиці 3.4.

Таблиця 3.4
Апостеріорні ймовірності для моделей оцінок студентів

Модель	z_i	$\Phi(z_i)$	Правдоподібність $\Phi(z_i)/15$	Апріорні ймовірн.	(Правдопод.) \times (апріорній мов.)	Апостеріорні	
M_1	0,133	0,395	0,02633	0,25	0,0065825	0,264	
M_2	-0,333	0,377	0,02513	0,50	0,0125650	0,504	
M_3	0,533	0,346	0,02307	0,25	0,0057675	0,231	
Загальна сума						0,0249150	0,999

З таблиці 3.4 видно, що модель M_2 має найвищу апостеріорну ймовірність з трьох моделей. Таким чином, робимо висновок, що випадково вибраний для тестування студент навчається в третій групі.

Зазначимо, що в даному прикладі ми скористались тільки одним експериментальним значенням. Можна було б поставити задачу визначення підгрупи для кількох студентів, які належать до однієї підгрупи. Це збільшило б об'єм експериментальних даних. Більше того, оскільки вибірка складалась тільки з одного значення (оцінка студента A), то це мало несуттєвий вплив на зміну апріорних ймовірностей, тобто, апріорні та апостеріорні ймовірності майже однакові. Якби вибірка даних була більшою, то апріорні ймовірності не мали б такого значного впливу на результат.



3.5 Теорема Байєса для дискретних даних і неперервного параметра

Теорема Байєса для дискретних даних і неперервного параметра θ з щільністю розподілу $g(\theta)$ має вигляд:

$$h(\theta | x_1, \dots, x_n) = \frac{L(x_1, \dots, x_n | \theta) g(\theta)}{\int_{\theta} L(x_1, \dots, x_n | \theta) g(\theta) d\theta}, \quad (3.19)$$

де інтегрування виконується по усіх значеннях θ ; $L(x_1, \dots, x_n | \theta)$ – функція правдоподібності; а $g(\theta)$ – апіорна щільність ймовірності для значень параметра. Апостеріорний розподіл неперервний, тобто h – щільність розподілу.

Приклад 3.6. Аналіз якості на виробництві: дискретні дані і неперервний параметр.

Розглянемо ще раз приклад 3.4, пов'язаний з аналізом якості на виробництві. Однак тепер θ – неперервний параметр (ймовірність виробництва пристрою пам'яті з дефектом), який має обмеження: $0 \leq \theta \leq 1$. В прикладі 3.4 з трьох випадково вибраних пристроїв два виявились з дефектом. У цьому випадку також скористаємось біноміальним розподілом (для оцінювання правдоподібності), але θ вже не обмежується трьома значеннями. Необхідно вибрати апіорний розподіл для θ .

Випадок 1: рівномірний розподіл. Припустимо, що фірма-виробник не має статистики щодо дефектних пристроїв пам'яті за минулі роки і на момент виконання статистичного аналізу немає інформації – чи всі партії виробів зовсім без дефектів, чи більшість виробів – дефектні, чи якість посередня. В подібній невизначеній ситуації для θ необхідно вибрати рівномірний (неінформативний) апіорний розподіл на одиничному інтервалі:

$$g(\theta) = \begin{cases} 1, & 0 < \theta < 1, \\ 0, & \text{інакше.} \end{cases}$$

Апостеріорна щільність розподілу для θ (по відношенню до міри Лебега) визначається виразом:

$$h(\theta | r, n) = \frac{L(r | \theta, n) g(\theta)}{\int_0^1 L(r | \theta, n) g(\theta) d\theta}$$

Для конкретних значень $r = 2$, $n = 3$ маємо:

$$\begin{aligned} h(\theta | r = 2, n = 3) &= \frac{C_3^2 \theta^2 (1 - \theta)}{\int_0^1 C_3^2 \theta^2 (1 - \theta) d\theta} = \frac{\theta^2 (1 - \theta)}{\int_0^1 \theta^2 (1 - \theta) d\theta} = \\ &= \frac{\theta^2 (1 - \theta)}{B(3, 2) \int_0^1 \frac{1}{B(3, 2)} \theta^{3-1} (1 - \theta)^{2-1} d\theta} = \frac{1}{B(3, 2)} \theta^{3-1} (1 - \theta)^{2-1} \end{aligned}$$

де

$$B(a, b) \equiv \int_0^1 y^{a-1} (1 - y)^{b-1} dy; \quad (3.20)$$

$$B(a, b) = \frac{\Gamma(a) \Gamma(b)}{\Gamma(a + b)}, \quad (3.21)$$

де $\Gamma(a)$ – гамма-функція з аргументом a . Бета-розподіл випадкової змінної X з параметрами a і b ($a > 0$, $b > 0$) визначається за виразом:

$$f(x | a, b) = \begin{cases} \frac{1}{B(a, b)} x^{a-1} (1 - x)^{b-1}, & 0 < x < 1, \\ 0, & \text{інакше.} \end{cases} \quad (3.22)$$

Таким чином, видно, що апостеріорною щільністю розподілу для θ є бета-розподіл з параметрами 3 і 2. Графіки рівномірного апріорного та апостеріорного бета-розподілів для θ показані на рис. 3.2.

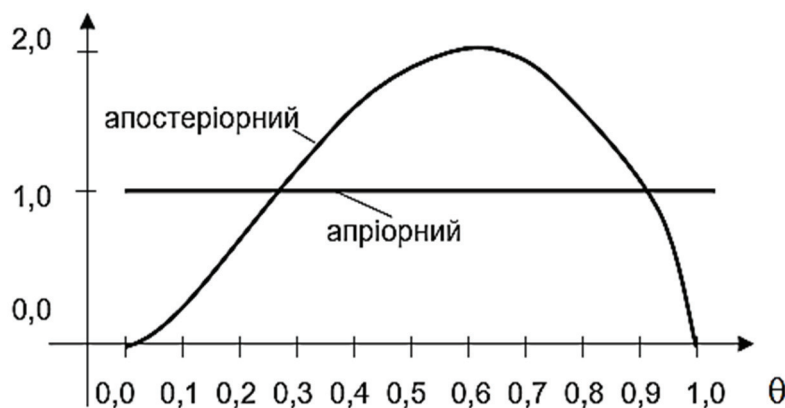


Рисунок 3.2 – Рівномірна апріорна щільність і апостеріорна щільність бета (3,2): θ – ймовірність виготовлення дефектного блока пам'яті

Випадок 2: за апріорний вибираємо бета-розподіл. Тепер вважаємо, що є більше інформації щодо θ , тобто приймаємо, що апріорним розподілом для θ є бета-розподіл. Апостеріорна щільність розподілу для θ



визначається за аналогією з першим випадком, але тепер $g(\theta)$ – бета-розподіл [3].

Відповідно, вибираємо наступну функцію бета-розподілу:

$$g(\theta) = \begin{cases} \frac{1}{B(a,b)} \theta^{a-1} (1-\theta)^{b-1}, & 0 < \theta < 1, \\ 0, & \text{інакше.} \end{cases} \quad (3.23)$$

де a і b значення параметрів, які конкретизують вигляд розподілу.

Визначимо апостеріорну щільність розподілу для θ :

$$h(\theta | r=2, n=3) = \frac{[C_3^2 \theta^2 (1-\theta)] \left[\frac{1}{B(a,b)} \theta^{a-1} (1-\theta)^{b-1} \right]}{\int_0^1 [C_3^2 \theta^2 (1-\theta)] \left[\frac{1}{B(a,b)} \theta^{a-1} (1-\theta)^{b-1} \right] d\theta}$$

Отриманий вираз можна спростити таким чином:

$$\begin{aligned} h(\theta | r=2, n=3) &= \frac{\theta^{a+1} (1-\theta)^b}{\int_0^1 \theta^{a+1} (1-\theta)^b d\theta} = \\ &= \frac{\theta^{(a+2)-1} (1-\theta)^{(b+1)-1}}{B(a+2, b+1) \int_0^1 \frac{\theta^{(a+2)-1} (1-\theta)^{(b+1)-1}}{B(a+2, b+1)} d\theta}, \end{aligned}$$

і оскільки інтеграл в знаменнику останнього виразу повинен дорівнювати 1, то апостеріорну щільність для θ можна записати у вигляді:

$$h(\theta | r=2, n=3) = \frac{1}{B(a+2, b+1)} \theta^{(a+2)-1} (1-\theta)^{(b+1)-1}, \quad 0 < \theta < 1.$$

Зазначимо, що так само як і у першому випадку апостеріорна щільність розподілу знову належить до сімейства бета-розподілу, але з іншими параметрами. Використовуючи апріорну інформацію, необхідно знайти параметри a і b .

Відомо, що статистичні параметри бета-розподілу (3.23) визначаються за виразами:

$$E(\theta) = \frac{a}{a+b}; \quad Var(\theta) = \frac{ab}{(a+b)^2 (a+b+1)}.$$

Припустимо що наявна інформація щодо виробничої діяльності компанії дала можливість визначити, що $E(\theta) = 0,2$ і $Var(\theta) = 0,01$.

Розв'язок наведених виразів щодо a і b дає такий результат: $a = 3$; $b = 12$. Таким чином, можемо конкретизувати апіорну щільність розподілу:

$$g(\theta) = \frac{1}{B(3, 12)} \theta^{3-1} (1-\theta)^{12-1},$$

а підстановка значень параметрів в апостеріорну щільність приводить до виразу:

$$h(\theta | r = 2, n = 3) = \frac{1}{B(5, 13)} \theta^{5-1} (1-\theta)^{13-1},$$

$$0 < \theta < 1.$$

Зазначимо, що апіорна та апостеріорна щільності розподілів цілком схожі між собою (рис. 3.3). Це можна пояснити незначним об'ємом експериментальних даних, $n = 3$, які не змінили суттєво апіорну інформацію.

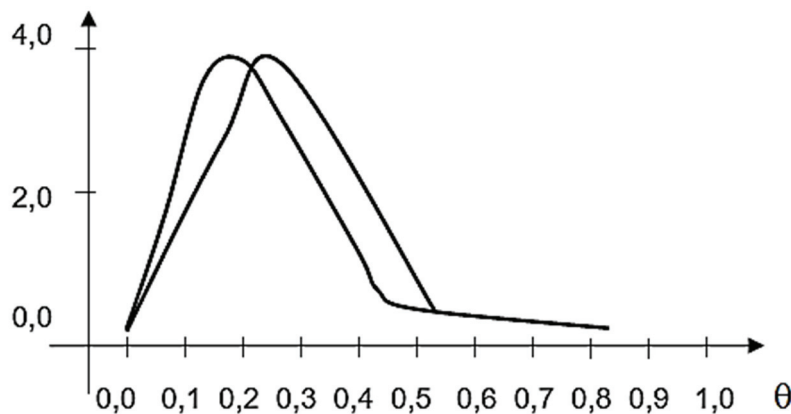


Рисунок 3.3 – Апіорна щільність бета (3,12) і апостеріорна щільність бета (5,13): θ – ймовірність виготовлення дефектного блока пам'яті

3.6 Теорема Байєса для неперервних даних і неперервного параметра

Теорема Байєса для неперервних даних і неперервного параметра записується так:

$$h(\theta | x_1, \dots, x_n) = \frac{L(x_1, \dots, x_n | \theta) g(\theta)}{\int_{\theta} L(x_1, \dots, x_n | \theta) g(\theta) d\theta},$$

де $L(\cdot)$ – функція правдоподібності для даних; $g(\theta)$ – апіорна щільність розподілу ймовірностей (ЩРЙ) для параметра θ ; $h(\theta | x_1, \dots, x_n)$ – апостеріорна ЩРЙ для цього параметра. Необхідно підкреслити, що $h(\theta | x_1, \dots, x_n)$ залежить від даних тільки через функцію правдоподібності



для $L(\cdot)$ для даних. Інтеграл у знаменнику береться по всіх значеннях θ і не повинен дорівнювати нулю. Дані можуть бути багатовимірними і корельованими, а θ може бути вектором параметрів.

Приклад 3.7. Випадок нормально розподілених даних з відомою дисперсією і невідомим середнім.

Сучасні маркетологи повинні мати глибокі знання у галузі маркетингу для просування товарів на ринок. Тому час від часу вони проходять курси підвищення кваліфікації із залученням провідних маркетологів та фахівців з інформаційних технологій. Тому велика рекламна компанія вибрала n діючих маркетологів для того щоб перевірити їх професійні знання. Спочатку їх знання перевірили перед проходженням курсів підвищення кваліфікації, а потім перевірили знання після проходження курсів.

Позначимо через x_i зміну (приріст) значення скорингового тесту, застосованого до кожного i -го маркетолога ($i=1, \dots, n$) в результаті проходження курсів підвищення кваліфікації. Значення скорингу (для кожного маркетолога окремо) мають нормальний розподіл з відомою дисперсією і невідомим середнім: $\{x_i\} \sim N(\theta, 25)$. Дисперсія може бути відомою завдяки багаторічному досвіду досліджень, протягом яких цей статистичний параметр залишався сталою величиною, а середнє – волатильна величина, яка залежить від багатьох поточних факторів, наприклад, стану макроекономіки, поточної ситуації на ринку і т. ін. Очевидно, що рекламна компанія, де працюють маркетологи, зацікавлена у зростанні середнього значення (тобто середнього скорингу), що свідчить про підвищення кваліфікації її працівників в результаті проходження курсів підвищення кваліфікації.

Випадок 1. Наявна незначна апіорна інформація, апіорний розподіл параметра рівномірний. Припускаємо, що інформація про ефективність курсів підвищення кваліфікації досить невизначена, а тому значення θ можуть знаходитись у широкому інтервалі дійсних чисел з однаковою ймовірністю. Тому приймається рівномірний розподіл для можливих значень θ .

Рівномірний розподіл називають *неналежним апіорним розподілом*, тобто це константа на всьому інтервалі визначення цього параметра. Таку *апіорну щільність називають неналежною*, оскільки в результаті її інтегрування можна отримати не одиницю, а довільне число. З іншого боку, такий простий вибір апіорного розподілу сприяє спрощенню процедури розв'язання задачі.

Вибір апіорної ЩРЙ у вигляді константи на всьому інтервалі визначення дійсних чисел називають по-різному, наприклад, так:

- плоский розподіл;
- невизначений розподіл;
- неінформативний розподіл;
- дифузійний;
- розподіл за замовчуванням (default distribution).

Формально такий вибір апіорної ЩРЙ записують так:

$$g(\theta) \propto c = \text{стала величина},$$

для всіх θ : $-\infty < \theta < +\infty$. Очевидно, що такий вибір можна наближено формалізувати так:

$$g(\theta) = \lim_{a \rightarrow \infty} \left(\frac{1}{2a} \right), \quad -a < \theta < a.$$

Тепер застосуємо до статистичних даних (значень скорингу) теорему Байєса у такому вигляді:

$$h(\theta | x_1, \dots, x_n) \propto L(x_1, \dots, x_n | \theta) g(\theta),$$

а після підстановки вибраного неінформативного (плоского) розподілу для θ отримаємо спрощений вигляд ТБ:

$$h(\theta | x_1, \dots, x_n) \propto L(x_1, \dots, x_n | \theta).$$

Тепер запишемо функцію правдоподібності з використанням вибіркового середнього для $n = 50$ (для того, щоб записати дисперсію для групи, індивідуальне значення потрібно поділити на 50). Оскільки для групи маркетологів з 50 чоловік

$$(\bar{x} | \theta) \sim N(\theta, 25/n) = N(\theta; 0,5),$$

то апостеріорна щільність приймає такий вигляд:

$$h(\theta | \bar{x}) \propto \exp \left\{ (-0,5) \left(\frac{\bar{x} - \theta}{0,5} \right)^2 \right\}.$$

Очевидно, що апостеріорна щільність для θ має такий розподіл:

$$(\theta | \bar{x}) \sim N(\bar{x}; 0,5).$$

Зазначимо, що незважаючи на те, що апіорний розподіл був вибраний неналежним, апостеріорний розподіл належний і нормальний. Більше того, оскільки апіорна інформація дуже незначна, то основну роль відіграє експеримент і апостеріорний розподіл центрується навколо вибіркового середнього, а оскільки ця оцінка відповідає методу максимальної правдоподібності, то ми отримали кращий варіант з можливих.



Дисперсія апостеріорного розподілу зменшилась до 0,5 завдяки тому, що розмір вибірки великий ($n = 50$). Цей приклад ілюструє перевагу використання пропорційної форми теореми Байєса. Не було необхідності виконувати формальне інтегрування (у знаменнику) оскільки форму сімейства апостеріорного розподілу можна було розпізнати на основі ядра апостеріорного розподілу.

Якщо припустити, що вибіркове середнє $\bar{x} = 3$; то апостеріорний розподіл приймає такий вигляд:

$$(\theta | \bar{x}) \sim N(3; 0,5).$$

На основі аналізу апостеріорного розподілу встановлено, що в результаті проходження курсів підвищення кваліфікації середнє значення скорингу знаходиться в інтервалі $(0,9; 5,1)$ з ймовірністю 0,997 (у цьому інтервалі крива охоплює 99,7% площі розподілу).

Випадок 2. *Апріорний розподіл параметра нормальний.* Припустимо, що перед проходженням курсів підвищення кваліфікації зміни (прирости) середнього скорингу θ для маркетологів мали розподіл $(\theta) \sim N(5; 2,0)$. Приймаємо цей розподіл за апріорний. У більш загальній формі апріорний розподіл має такий вигляд: $(\theta) \sim N(m, \sigma^2)$, або

$$g(\theta) \propto \exp\left\{-0,5\left(\frac{\theta - m}{\sigma}\right)^2\right\}.$$

Підставимо цей вираз у $h(\theta | x_1, \dots, x_n) \propto L(x_1, \dots, x_n | \theta) g(\theta)$ і отримаємо:

$$h(\theta | \bar{x}) \propto \exp\left\{-0,5\left[\left(\frac{\theta - m}{\sigma}\right)^2 + \left(\frac{\bar{x} - \theta}{0,5}\right)^2\right]\right\},$$

де $\exp\left\{-0,5\left(\frac{\bar{x} - \theta}{0,5}\right)^2\right\}$ взято з попереднього випадку.

Після спрощення виразу під експонентою отримаємо:

$$h(\theta | \bar{x}) \propto \exp\left\{-0,5\left(\frac{\theta - \bar{\theta}}{\gamma}\right)^2\right\},$$

де $\frac{1}{\gamma^2} = \frac{1}{\sigma^2} + \frac{1}{(0,5)^2}$,

а оцінка середнього для θ :

$$\bar{\theta} = \left(\frac{\sigma^{-2}}{\sigma^{-2} + (0,5)^{-2}} \right) m + \left(\frac{(0,5)^{-2}}{\sigma^{-2} + (0,5)^{-2}} \right) \bar{x},$$

тобто $(\theta | \bar{x}) \sim N(\bar{\theta}, \gamma^2)$. Таким чином, у випадку, коли вибіркове середнє \bar{x} має нормальний розподіл із середнім θ і відомою дисперсією, а θ також має нормальний розподіл, то апостеріорний розподіл для θ також нормальний.

Якщо прийняти за точність оцінювання (розподілу) величину обернену дисперсії, то апостеріорна точність γ^{-2} дорівнює сумі точності апіорного розподілу σ^{-2} і точності вибіркового (експериментального) розподілу $(0,5)^{-2}$.

Апостеріорне середнє $\bar{\theta}$ – це опукла комбінація зваженого апіорного середнього m і вибіркового середнього \bar{x} . Вагові коефіцієнти пропорційні точності (оцінювання), яка відповідає апіорному та вибірково му розподілам.

Відповідно до виконаних розрахунків, отримано такі параметри апостеріорного розподілу:

$$(\bar{\theta}) \sim N(3,37; 0,26).$$

Таким чином, найбільш ймовірне значення середнього θ змістилось від значення 5,0 до 3,37. Більше того, дисперсія апостеріорного розподілу зменшилась із значення 2,0 до 0,26, що привело до *суттєвого зменшення невизначеності* цього розподілу завдяки отриманим даних після курсів, тобто до його звуження.

Приклад 3.8. Випадок нормально розподілених даних з відомою дисперсією і невідомим середнім.

Припустимо, що дані деякого виконаного експерименту мають нормальний розподіл, але у загальному випадку їх середнє і дисперсія невідомі. Буде зручно параметризувати розподіл даних у термінах його точності, тобто з використанням величини, оберненої до дисперсії.

Функція правдоподібності

Припустимо, що є незалежні спостереження x_1, \dots, x_n з нормальним розподілом $N(\theta, q^{-1})$, де q – точність (розподілу). Функцію правдоподібності для цих даних можна записати у вигляді:

$$f(x_1, \dots, x_n | \theta, q) \propto q^{0,5n} \exp \left\{ \left(-\frac{q}{2} \right) \sum_{i=1}^n (x_i - \theta)^2 \right\}.$$



Зазначимо, що вираз під сумою можна записати так:

$$\begin{aligned}\sum_{i=1}^n (x_i - \theta)^2 &= \sum_{i=1}^n [(x_i - \bar{x}) + (\bar{x} - \theta)]^2 = \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + n(\theta - \bar{x})^2.\end{aligned}$$

Покладаючи $\sum_{i=1}^n (x_i - \bar{x})^2 = (n-1)s^2$, напишемо функцію

правдоподібності у вигляді:

$$\begin{aligned}f(x_1, \dots, x_n | \theta, q) &\equiv L(\bar{x}, s^2 | \theta, q) \propto \\ &\propto q^{0,5n} \exp\left\{\left(-\frac{q}{2}\right) [(n-1)s^2 + n(\theta - \bar{x})^2]\right\}.\end{aligned}$$

Згідно з теоремою Байєса апостеріорна щільність матиме такий вигляд:

$$h(\theta, q | data) \propto L(\bar{x}, s^2 | \theta, q) g(\theta, q),$$

де $g(\theta, q)$.

Апріорна щільність. Розглянемо можливості використання двох видів апріорних розподілів для $g(\theta, q)$. Спочатку можна скористатись *невизначеним* або *розподілом за замовчуванням* у випадку, коли апріорна інформація незначна. Потім можна вибрати *природний спряжений апріорний розподіл*, який зручний з математичної точки зору.

РОЗДІЛ 4

ОЦІНЮВАННЯ (ВИБІР) АПРІОРНИХ РОЗПОДІЛІВ

4.1 Об'єктивні апріорні розподіли

Безпосередньо у самій теоремі Байєса не формулюється тип апріорного розподілу. Яким же чином можна характеризувати дані (інформацію) стосовно досліджуваного процесу (об'єкта) перед їх практичним використанням? Відповіді на це логічне є частково філософськими, а частково цілком прагматичними. Конкретна відповідь залежить від природи інформації, а також від системи довіри стосовно належного (правильного) компромісу між об'єктивною та суб'єктивною складовими, що містяться у даних, і які будуть надалі внесені у розв'язок поставленої задачі. Тобто апріорний розподіл може бути у більшій або меншій мірі суб'єктивним чи об'єктивним [83, 84].

Сам Томас Байєс застосував свою теорему у випадку, коли дані відповідали біноміальному розподілу з невідомим параметром p – ймовірність успіху експерименту при використанні однієї спроби, а у якості апріорного розподілу він вибрав рівномірний. Іншими словами Байєс вибрав такий апріорний розподіл, згідно з яким всі можливі значення p мали однакову ймовірність. Це і є одним з різновидів об'єктивного апріорного розподілу.

Однак теорема припускає можливість використання даних з широким вибором типів розподілів і забезпечує широкий вибір типів апріорних розподілів. У випадку, коли наявна інформація мінімальна, для оновлення процесу (розв'язання задачі) часто вибирають об'єктивний апріорний розподіл.

Об'єктивний апріорний розподіл

Деякі типи апріорних розподілів були запропоновані з метою відобразити факт відсутності або наявності *незначної* інформації про параметри до отримання даних, при цьому формулювались необхідні обмеження на розподіл. Таку ідею вибору відносять до класу *об'єктивних* апріорних розподілів. Однак, термін «незначна інформація» є невизначеним, а тому його можна трактувати по-різному. Більше того, при



виконанні багатьох досліджень результат наукового експерименту необхідно обґрунтовувати і узагальнювати з мінімальними посиланнями на специфіку конкретної задачі. Надалі така інформація може використовуватись іншими як посилання на це досягнення. Таким чином, факт наявності незначної інформації може трактуватись по-різному у різному контексті.

Апріорний розподіл на основі публічних суджень (думки суспільства)

Нехай ставиться завдання визначити на скільки процентів зміниться кількість добровольців, які побажають вступити у контрактну армію хоча б на один рік, якщо їм запропонувати деякий бонус, наприклад, одноразову початкову виплату 10000 доларів. Військовий аналітик, який має досвід у розв'язанні такого роду завдань, може дати свою оцінку на скільки процентів зміниться кількість добровольців. Можна поставити також експеримент, за допомогою якого можна буде оцінити цей процент. З іншого боку, різні аналітики можуть дати різні варіанти відповіді, при цьому кожний буде обґрунтовувати саме свій варіант.

Тому у багатьох ситуаціях не потрібно шукати апріорний розподіл, який буде узагальнюючим (однаковим для всіх), а логічно запропонувати множину апріорних розподілів, які дадуть можливість отримати множину відповідних апостеріорних висновків. Таким чином можна отримати результат, який буде відображати, у якійсь мірі, «думку суспільства». У таких випадках остаточним висновком (результатом оцінювання) може бути усереднення окремих висновків, обчислених за різними апріорними розподілами [23].

Принцип недостатньої обґрунтованості Лапласа

Наскільки відомо, Лаплас сформулював теорему Байєса (задача про інверсну (обернену, *inverse*) ймовірність) не читаючи його оригінальну працю, яка була опублікована в Англії у 1763 році [98]. Для того щоб вносити найменше суб'єктивної інформації у розв'язання задач він запропонував використовувати *принцип недостатнього обґрунтування* (*the principle of insufficient reason*) – 1812 рік. Згідно з його принципом, у випадках, коли немає причин прийняти альтернативний варіант, всі значення невідомого параметра, що задають появу деяких подій, повинні бути прийняті апріорно як рівноймовірні. Пізніше (у 1812 році) він сформулював *принцип достатнього обґрунтування*, згідно з яким апріорні ймовірності могли бути різними.

Якщо невідомий параметр знаходиться в інтервалі скінченної ширини, то використання рівномірного розподілу відповідає вибору на

основі публічних суджень або об'єктивному вибору. Однак, якщо хоча б одна границя інтервалу визначення параметра залишається нескінченною, то об'єктивний апріорний розподіл буде *неналежним*, тобто інтегрування не дасть в результаті «1». Разом з тим, у таких випадках апостеріорний розподіл залишається, як правило, належно визначеним.

Переваги використання об'єктивних апріорних розподілів

1. Об'єктивні апріорні розподіли іноді використовують як основу (відправну точку), яка не відображає специфічні уподобання аналітика. Вони можуть використовуватись будь-ким незалежно від уподобань та прогнозів конкретного виконавця стосовно остаточного результату.
2. Об'єктивні апріорні розподіли відображають ту ідею, що інформація, наявна про досліджуваний процес, незначна або відсутня.
3. Остаточні результати (апостеріорний байєсівський висновок) стосовно розв'язання конкретної задачі, отримані з використанням об'єктивних апріорних розподілів, іноді бувають математично еквівалентними результатам, отриманим з використанням частотних ймовірностей. Однак, ці результати мають різну філософську інтерпретацію, і ця різниця бути суттєвою.

Недоліки використання об'єктивних апріорних розподілів

Буває нелегко встановити (вибрати) такий *об'єктивний* апріорний розподіл, який відповідав би ситуації з «незначними знаннями» стосовно досліджуваного процесу і задовольняв би обмеженням, які накладаються аналітиком. Тому використання об'єктивних апріорних розподілів має вказані нижче недоліки.

1. Іноді неналежні апріорні розподіли виникають внаслідок повної відсутності апріорного знання досліджуваного процесу, оскільки параметри можуть бути визначені на всій осі дійсних чисел або на половині осі. Це може призвести до неналежного апостеріорного розподілу.
2. Якщо параметрів кілька, то їх часто вважають незалежними між собою у об'єктивному апріорному розподілі. Однак, вони можуть бути корельованими, що призводить до ускладнення задачі.
3. Іноді бажано щоб апостеріорний висновок не залежав від параметризації конкретного типу, що застосовується до задачі. Наприклад, бажано мати інваріантність стосовно деяких типів перетворень даних, таких як афінне або поворотне перетворення.



Однак, такі властивості можуть забезпечити не всі види апіорних розподілів.

4. Іноді вибір, що характеризує відсутність апіорної інформації про параметр, призводить до небажаного (некоректного) висновку стосовно деякої важливої функції цього параметра.
5. Вибір об'єктивного розподілу не завжди відображає фактичні апіорні знання, тобто такий апіорний вибір не має логічної ймовірнісної інтерпретації.
6. Використання неналежного апіорного розподілу призводить до невизначеності при обчисленні фактору Байєса, тому що константа, необхідна для обчислень, залишається невизначеною.
7. Може виникати парадокс при обчисленні апостеріорного байєсівського висновку, тобто обчислення приводять до різних результатів залежно від типу параметризації задачі і від того, яка параметризація була використана для маржиналізації спільного апостеріорного розподілу.

4.2 Суб'єктивні апіорні розподіли

Суб'єктивні апіорні розподіли відображають спроби використати при розв'язанні задачі наявну (наприклад, отриману деяким конкретним дослідником на основі власного досвіду) апіорну інформацію про досліджувані процеси. Їх називають суб'єктивними тому що *вони ґрунтуються не на узагальненнях, законах чи закономірностях, а на деяких окремих результатах*, отриманих окремими дослідниками у процесі розв'язування конкретних задач. Такі апіорні розподіли мають свої переваги та недоліки [102].

Переваги суб'єктивних розподілів

1. Суб'єктивні апіорні розподіли відносяться до належних, тобто вони інтегруються до 1.
2. У загальному випадку суб'єктивні апіорні розподіли характеризуються хорошою аналітичною поведінкою. Вплив суб'єктивного апіорного розподілу на апостеріорний проявляється у тому, що (начебто) має місце додаткове повторення (копіювання) наявних даних про задачу.
3. Суб'єктивні апіорні розподіли дають можливість використовувати при розв'язуванні задачі всі наявні знання, досвід, інтуїцію, експертні і можливі сторонні вказівки та іншу інформацію, якою може володіти дослідник.
4. Досить часто суб'єктивний байєсівський підхід залишається єдиною можливим для розв'язання складних задач, які неможливо

розв'язати іншими методами, наприклад, за допомогою частотного ймовірнісного підходу.

5. Цей підхід відкритий для узагальнення, тобто він може постійно удосконалюватись на основі нових знань та нового досвіду.

Приклад 4.1. Припустимо, що існує невідомий параметр θ деякого явища, оцінку якого необхідно знайти. Нехай завдяки виконаним попереднім дослідженням встановлено, що $\theta = 12$; цю інформацію можна використати як апріорне знання.

Якщо досвід дослідження цього явища свідчить про те, що невизначеність стосовно оцінки параметра має симетричний характер, то логічно вибрати за апріорний нормальний розподіл із середнім 12. Таким чином суб'єктивна апріорна інформація використовується для формулювання апріорного розподілу. Додатковою інформацією може бути така: дуже мало ймовірно, що значення параметра $\theta < 9$ (ймовірність менша 0,01) і що значення параметра $\theta > 15$.

Якщо прийняти ці граничні значення як три стандартних відхилення від середнього з ймовірністю 0,997 ($\mu \pm 3\sigma$), то апріорний розподіл буде мати стандартне відхилення 1 ($(12-9):3 = 1$ і $(15-12):3 = 1$). А тому можна прийняти такий суб'єктивний апріорний розподіл: $(\theta) \sim N(12, 1)$.

Приклад 4.2. Припустимо, що дані мають біноміальний розподіл з невідомим параметром θ для r успішних результатів при виконанні n незалежних дослідів. Ігноруючи константи, запишемо функцію правдоподібності:

$$L(r|\theta, n) = \theta^r (1 - \theta)^{n-r}$$

Прийmemo апріорний розподіл для θ у формі бета розподілу, $dbeta(a, b)$ із щільністю, пропорційною виразу

$$\theta^{a-1}(1-\theta)^{b-1}, \quad a > 0, \quad b > 0,$$

де a, b – відомі параметри. За теоремою Байєса апостеріорний розподіл буде пропорційним виразу:

$$\theta^{(a+r)-1}(1-\theta)^{(b+n-r)-1},$$

тобто ще один бета розподіл з параметрами $(a + r, b + n - r)$. Якщо прийняти за апріорний рівномірний (невизначений) розподіл з параметрами $(a = 0, b = 0)$, то апостеріорна щільність також буде бета розподілом з параметрами $(r, n - r)$.

Таким чином, використання суб'єктивної апріорної інформації у формі бета розподілу привело до того, що апостеріорний розподіл веде себе



так, начебто у досліджуваних даних дійсно існують $(a + r)$ і $(b + n - r)$ ступенів свободи.

У цілому можна стверджувати, що використання суб'єктивної інформації (наявні знання, досвід, інтуїцію, можливі додаткові сторонні вказівки та іншу належну інформацію) надає можливість дослідникам розв'язувати багато задач з високою ефективністю, у тому числі без використання теореми Байєса [119, 120].

Недоліки суб'єктивних розподілів

1. Не завжди можна легко перетворити апріорні знання у адекватний апріорний розподіл. Параметри, які визначають (індексують) суб'єктивний апріорний розподіл, називають *гіперпараметрами* (це параметри для параметрів). Ці гіперпараметри потрібно ідентифікувати (встановити та оцінити), що досить часто потребує розв'язання складних обчислювальних задач, особливо у багатовимірному випадку.
2. Результати байєсівського аналізу, які використовують суб'єктивний апріорний розподіл, мають найбільше значення для того аналітика, який запропонував цей розподіл, у той час як інші дослідники можуть пропонувати свої варіанти.
3. Суб'єктивний апріорний розподіл може бути досить складним для аналітичного дослідження, особливо якщо використовуються складні обчислювальні процедури або цей розподіл не належить до сімейства природних спряжених розподілів. Однак, сучасні обчислювальні засоби надають можливість визначати апостеріорні та прогнознi розподіли з прийнятними витратами часу навіть у тих випадках, коли апріорний розподіл є математично складним.

4.3 Спрощені (невизначені) апріорні розподіли для одного параметра

Розглянемо задачу формування апріорного розподілу у випадку розподілів даних, що індексуються одним параметром. Надалі цей підхід буде узагальнено на багатовимірний випадок.

Апріорний розподіл для одного параметра

Практично для будь-якої невідомої величини можна побудувати апріорний розподіл, який представляє нашу ступінь довіри до неї. Припустимо, що необхідно ідентифікувати апріорний розподіл для щільності кальцію у крові, яку позначимо через θ . Якщо у дослідника немає спеціальних знань з біології, можна просто прийняти неінформативний

рівномірний розподіл. Він передбачає апріорний вибір однакових ймовірностей для всіх значень параметра θ . Однак, такий попередній вибір можна уточнити.

Наприклад, припустимо, що через θ позначено середню густину кальцію – кількість грамів або міліграмів на одиницю об'єму. Очевидно, що з фізичних міркувань $\theta > 0$, а тому апріорну щільність ймовірності для цього параметра можна записати так:

$$g(\theta) = \begin{cases} 1/a, & 0 < \theta < a, \\ 0, & \text{інакше.} \end{cases} \quad (4.1)$$

Але ж яке значення має a ? Якщо необхідно підкреслити, що знання стосовно параметра θ незначні або ж існує невизначеність, то для наближення до рівномірного розподілу краще скористатись $\log(\theta)$ (це згладжує можливі значення). Сам апріорний розподіл матиме такий вигляд:

$$g(\theta) \propto 1/\theta = \theta^{-1}.$$

Для ілюстрації припустимо, що $\{X|\theta\} \sim N(\theta, 1)$, $-\infty < \theta < \infty$ і необхідно знайти оцінку θ . Відсутність апріорної інформації стосовно цієї оцінки може стати причиною вибору неналежного неінформативного розподілу параметра θ вздовж всієї осі дійсних чисел. Якщо поставити вимогу формулювання тільки належного апріорного розподілу, то можна опинитись у важкій ситуації. Надалі буде видно, що у випадку відсутності інформації про θ необхідно знати деяку інформацію стосовно функцій θ і навпаки.

Припустимо, наприклад, ми стараємось представити задачу так, що все рівно яке значення матиме θ . Але з цього випливає таке: якщо ми можемо застосувати до θ просте монотонне перетворення в одиничний інтервал, а потім скористаємось рівномірним апріорним розподілом для цього інтервалу, то проблема зникне. Однак, це не так; на цьому зупинимось трохи нижче.

Наша задача полягає у тому, щоб знайти такий апріорний розподіл, який буде відображати той факт, що все рівно, яке значення прийматиме θ ; нехай

$$\phi \equiv F(\theta), \quad (4.2)$$

де $F(\theta)$ – монотонне не спадаюче диференційоване перетворення θ , таке, що

$$0 \leq F(\theta) \leq 1. \quad (4.3)$$



Наприклад, у якості $F(\cdot)$ може бути використана кумулятивна функція розподілу. Нехай апіорний розподіл для θ є рівномірним, тобто його щільність має вигляд:

$$p(\phi) = \begin{cases} 1, & 0 < \phi < 1, \\ 0, & \text{інакше.} \end{cases} \quad (4.4)$$

Очевидно, що $p(\phi)$ відображає нечутливість до всіх значень ϕ , але обернене перетворення $\theta = F^{-1}(\phi)$ індукує або дає змогу отримати апіорну щільність для θ згідно з основною теоремою перетворення функції розподілу:

$$g(\theta) = \left| \frac{dF(\theta)}{d\theta} \right|. \quad (4.5)$$

Вибрати можна будь-яку функцію $F(\theta)$. Припустимо, що вибрана функція нормального розподілу:

$$\phi \equiv F(\theta) = \int_{-\infty}^{\theta} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} t^2\right) dt. \quad (4.6)$$

Тобто апіорна щільність нормального розподілу для θ має вигляд:

$$g(\theta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \theta^2\right), \quad (4.7)$$

або $(\theta) \sim N(0, 1)$. Апостеріорна щільність приймає такий вигляд:

$$p(\theta | X) \propto \exp\left[-\frac{1}{2} (x - \theta)^2 - \frac{1}{2} \theta^2\right], \quad (4.8)$$

тобто $(\theta | X) \sim N(0, 5x; 0, 5)$. (4.9)

Зазначимо, що у даному випадку спостерігається висока інформативність стосовно θ , незважаючи на те, що всім можливим інтервалам ϕ однакової довжини присвоєні однакові ймовірності. Все, що тут було зроблено, так це те, що ми старались ввести індиферентність (за принципом недостатньої обґрунтованості Лапласа) до всіх значень простого монотонного перетворення θ .

Для того щоб апіорний розподіл θ виражав індиферентність (нечутливість) стосовно конкретних значень θ , необхідно його задати таким чином:

$$g(\theta) \propto \text{константа}, \quad (4.10)$$

а це у свою чергу означає, що $\frac{dF(\theta)}{d\theta}$ – це також константа, тобто $F(\theta)$ – це лінійна функція θ . Однак для θ , визначеного на всій дійсній осі (або навіть на додатній половині осі), неможливо забезпечити виконання умови $0 \leq F(\theta) \leq 1$, і щоб при цьому $F(\theta)$ була прямою лінією для області визначення $-\infty < \theta < \infty$.

Задавання рівномірних розподілів на простих монотонних перетвореннях θ не надають можливості розв'язати задачу ідентифікації апріорних розподілів, те ж саме стосується складних перетворень.

Невизначена апріорна щільність параметра на інтервалі $(-\infty, \infty)$

У подальших викладках, коли буде необхідно підкреслити, що апріорно всі значення параметра θ мають однакову ймовірність появи, буде прийматись невизначена щільність, що пояснюється нижче.

Для того щоб виразити невизначеність інформації (знання) стосовно параметра θ , визначеного на інтервалі $-\infty < \theta < \infty$, використовують неналежний (невизначений) апріорний розподіл у формі $g(\theta) \propto \text{константа}$, яка визначається через граничний перехід:

$$g(\theta) = \lim_{a \rightarrow \infty} g_a(\theta), \quad (4.11)$$

де

$$g_a(\theta) = \begin{cases} 1/2a, & -a < \theta < a, \\ 0, & \text{інакше.} \end{cases} \quad (4.12)$$

Примітка: Граничний перехід, який використовується для отримання невизначеного апріорного розподілу, є неоднозначним.

Невизначена апріорна щільність параметра на інтервалі $(0, \infty)$

Для додатного параметра, наприклад σ , беруть $\log \sigma$ для того, щоб досягнути рівномірного розподілу на всій додатній половині дійсної осі так само, як це було зроблено з параметром θ . Тепер функцію щільності можна задати апріорний розподіл так:

$$g(\sigma) \propto \frac{1}{\sigma},$$

або у еквівалентній формі:



$$g(\sigma^2) \propto \frac{1}{\sigma^2}. \quad (4.13)$$

Такі невизначені апіорні щільності корисні своїми властивостями інваріантності обчислюваних висновків при використанні моделей з простими структурами. Тобто при використанні простих перетворень параметрів θ або σ апостеріорні висновки, що ґрунтуються на нових параметрах, будуть узгоджуватись (будуть консистентними) з висновками, які ґрунтуються на старих параметрах. Тобто, якщо $g(\theta)$ – це апіорна щільність для θ , і якщо виконується перетворення θ до ϕ , наприклад, $\phi = F(\theta)$, і якщо $h(\phi)$ – це апіорна щільність для ϕ , то повинна виконуватись рівність $g(\theta) d\theta = h(\phi) d\phi$. При цьому апостеріорні висновки повинні відображати цю консистентність (узгодження).

4.4 Природні (натуральні) спряжені апіорні розподіли

Припустимо, що існує деяка суб'єктивна апіорна інформація стосовно параметра θ , яку бажано формалізувати і використати у моделюванні. Досить часто достатньо представити таку інформацію деяким гладким розподілом, який відноситься до сімейства суб'єктивних апіорних розподілів, що включає такі розподіли:

- природні (натуральні) спряжені розподіли;
- експоненціальні (еліптичні) розподіли;
- змішані апіорні розподіли.

Одним із таких сімейств є сімейство *природних (натуральних) спряжених розподілів* [52-56, 100, 127]. Це сімейство вибрано завдяки зручності його використання з математичної точки зору. Розподіли, що входять у це сімейство, називають ще *зручними апіорними* розподілами (convenience prior). Це сімейство надає можливість формувати замкнені форми розподілів при переході від апіорного до апостеріорного. Тобто у даному випадку апостеріорний розподіл належить до того ж сімейства, що і апіорний.

Ще одним поширеним сімейством суб'єктивних апіорних розподілів є *сімейство експоненціальних (еліптичних) розподілів*, яке має симетричне представлення після центрування. Спочатку воно призначалось для опису даних, але його також успішно використовують як апіорні розподіли. Воно характеризується наявністю гіперпараметрів, які дають можливість досягти великої гнучкості представлення апіорної інформації у межах цього класу.

Ще один клас суб'єктивних апіорних розподілів утворюють *змішані апіорні розподіли*. Розглянемо коротко характеристики трьох згаданих сімейств розподілів.

Сімейство природних (натуральних) апріорних розподілів

Для створення належного сімейства природних апріорних розподілів необхідно спочатку необхідно виконати такі дії:

- сформулювати функцію правдоподібності;
- поміняти у цій функції ролі випадкової змінної і параметра;
- «збагатити» параметри отриманого ядра щільності розподілу, тобто узагальнити їх значення так чином, щоб вони не залежали від даних;
- ідентифікувати розподіл, який відповідає результуючому ядру і додати до нього необхідну нормуючу константу для того щоб інтеграл щільності давав одиницю.

Приклад 4.3 Природний спряжений розподіл для біноміальних даних.

Припустимо, що функція правдоподібності біноміальна і можна записати:

$$L(\theta) = \theta^m (1 - \theta)^{n-m}. \quad (4.14)$$

Нагадаємо, що функція правдоподібності визначена з точністю до мультиплікативної константи. Нехай у рівнянні (4.14) θ – випадкова змінна, а m і n – невідомі параметри. Тепер $L(\theta)$ можна вважати ядром *бета розподілу*. Замінімо параметри, що характеризують вибірку m і n , на довільні параметри α і β , а також введемо нормування щільності, щоб вона була належною (інтегрування буде давати одиничний результат). Таким чином отримаємо сімейство апріорних бета розподілів:

$$g(\theta) = \begin{cases} \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}, & 0 < \theta < 1, \\ 0, & \text{інакше.} \end{cases} \quad (4.15)$$

для $\alpha > 0$, $\beta > 0$. Тепер можна використати всю наявну апріорну інформацію для оцінювання гіперпараметрів α і β , які індексують сімейство бета розподілів.

Ще один фактор зручності при використанні природних спряжених апріорних розподілів стосується обчислення апостеріорного розподілу. Оскільки за теоремою Байєса апостеріорна щільність пропорційна правдоподібності, помноженій на апріорну щільність (у даному прикладі апріорний бета розподіл), то апостеріорна щільність задається виразами:

$$h(\theta | y) \propto [\theta^m (1 - \theta)^{n-m}] [\theta^{\alpha-1} (1 - \theta)^{\beta-1}], \quad (4.16)$$



$$h(\theta|y) \propto \theta^{m+\alpha-1} (1-\theta)^{n-m+\beta-1},$$

тобто ядром апостеріорної щільності є бета розподіл.

Перевагою використання *природних спряжених апіорних* розподілів є те, що *апостеріорні розподіли мають таку ж форму*, тобто відносяться до того ж сімейства, що і апіорні розподіли. Ця властивість «замкнутості» сімейства природних спряжених розподілів залишається дійсною для всіх членів експоненціального сімейства, тобто для сімейства розподілів з достатньою статистикою (що відповідає деяким простим умовам регулярності).

У таблиці 4.1 подано деякі вибіркові розподіли експоненціального сімейства разом з відповідними їм природним спряженим апіорним розподілом. Рекомендуємо читачу встановити ці взаємовідношення після розгляду прикладу біноміального розподілу, поданого нижче [114, 116, 122].

Таблиця 4.1

Деякі типи природних спряжених апіорних розподілів

№ п/п	Дискретний розподіл даних	Природний спряжений апіорний розподіл
1	Біноміальний	Ймовірність успішної події відповідає бета розподілу (<i>success probability follows a beta distribution</i>)
2	Негативний біноміальний	Ймовірність успішної події відповідає бета розподілу (<i>success probability follows a beta distribution</i>)
3	Пуассона	Середнє відповідає гамма розподілу
4	Експоненціальний із середнім λ	λ відповідає гамма розподілу
5	Нормальний з відомою дисперсією, але невідомим середнім	Середнє відповідає нормальному розподілу
6	Нормальний з відомим середнім, але невідомою дисперсією	Дисперсія відповідає інверсному гамма розподілу

Деякі типи відомих розподілів:

– **Біноміальний**: випадкова змінна $X \in \{0, 1, \dots, n\}$ має біноміальний (n, θ) розподіл, якщо $\theta \in [0, 1]$ і

$$p(X=x|\theta, n) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \quad \text{для } x \in \{0, 1, \dots, n\}.$$

– **Бета розподіл**: випадкова змінна $X \in [0, 1]$ має $\text{beta}(a, b)$ розподіл, якщо $a > 0$, $b > 0$ і

$$p(x|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} \quad \text{для } 0 \leq x \leq 1;$$

$$\Gamma(a) = \int_0^{\infty} t^{a-1} \exp(-t) dt, \quad a > 0;$$

$$p(x|a, b) = \text{dbeta}(x, a, b).$$

Значення гамма функції можна знайти у відповідних таблицях.

– **Розподіл Пуассона:** випадкова змінна $X \in \{0, 1, 2, \dots\}$ має розподіл Пуассона для θ , якщо $\theta > 0$ і ймовірність $p(X=x|\theta)$ обчислюється за виразом:

$$p(X=x|\theta) = \theta^x \frac{e^{-\theta}}{x!} \quad \text{для } x \in \{0, 1, 2, \dots\};$$

$$p(x|\theta) = \text{dpois}(x, \theta).$$

– **Гамма і обернений (інверсний) гамма розподіл:** випадкова змінна $X \in (0, \infty)$ має гамма (a, b) розподіл, якщо $a > 0, b > 0$ і

$$p(x|a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} \quad \text{для } x > 0;$$

$$p(x|a, b) = \text{dgamma}(x, a, b).$$

Випадкова змінна $X \in (0, \infty)$ має обернений (інверсний) гамма (a, b) розподіл, якщо $1/X$ має гамма (a, b) розподіл. Іншими словами, якщо $Y \sim \text{gamma}(a, b)$ і $X = 1/Y$, то $X \sim \text{inverse-gamma}(a, b)$, а функція щільності для X має вигляд:

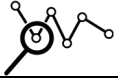
$$p(x|a, b) = \frac{b^a}{\Gamma(a)} x^{-a-1} e^{-bx} \quad \text{для } x > 0.$$

– **Одновимірний нормальний розподіл:** випадкова змінна $X \in \mathfrak{R}$ має нормальний розподіл ($\text{normal}(\theta, \sigma^2)$), якщо $\sigma^2 > 0$ і

$$p(x|\theta, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x-\theta)^2\right) \quad \text{для } -\infty < x < \infty;$$

$$p(x|\theta, \sigma^2) = \text{dnorm}(x, \theta, \sigma).$$

– **Багатовимірний нормальний розподіл:** випадковий вектор $\mathbf{x} \in \mathfrak{R}^p$ має багатовимірний нормальний розподіл ($\text{normal}(\theta, \Sigma)$), якщо Σ – додатне напіввизначена матриця розмірністю $p \times p$ і функція щільності задається так:



$$p(x|\theta, \Sigma) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2} (\mathbf{x} - \theta)^T \Sigma^{-1} (\mathbf{x} - \theta)\right\} \text{ для } \mathbf{x} \in \mathfrak{R}^p ;$$

– **Рівномірний розподіл:** неперервна змінна X має рівномірний розподіл на інтервалі $[a, b]$, який позначається $X \sim U(a, b)$, якщо його функція щільності задається таким виразом:

$$f_{U(x|a,b)} = \begin{cases} \frac{1}{b-a}, & \text{якщо } a \leq x \leq b, \\ 0, & \text{інакше,} \end{cases}$$

де $-\infty < a < b < \infty$.

Експоненціальне сімейство апіорних розподілів

Клас експоненціальних розподілів (степеневих розподілів) для параметра θ має таку щільність:

$$g(\theta) = \frac{k}{\sigma} \exp\left[-0,5 \left|\frac{\theta - \theta_0}{\sigma}\right|^{(2/(1+\beta))}\right], \quad (4.17)$$

де $-\infty < \theta < \infty$; $-1 < \beta \leq 1$; $\sigma > 0$; $k^{-1} = \Gamma(1 + 0,5(1+\beta)) 2^{1+0,5(1+\beta)}$. Можна показати, що

$$E(\theta) = \theta_0,$$

і

$$Var(\theta) = 2^{1+\beta} \sigma^2 \frac{\Gamma[(1+\beta)(3/2)]}{\Gamma[(1+\beta)(1/2)]}, \quad (4.18)$$

Параметр β – це міра ексцесу (kurtosis), який відображає ступінь відхилення розподілу від нормального. Якщо $\beta = 0$, то розподіл нормальний. Коли $\beta = 1$, то розподіл називають подвійним експоненціальним із щільністю:

$$g(\theta) = \frac{1}{\sigma\sqrt{2}} \exp\left\{-\sqrt{2} \left|\frac{\theta - \theta_0}{\sigma}\right|\right\}, \quad -\infty < \theta < \infty, \quad (4.19)$$

Якщо $\beta \rightarrow -1$, то розподіл наближається до рівномірного розподілу на інтервалі: $(\theta_0 - \sigma\sqrt{3}, \theta_0 + \sigma\sqrt{3})$. Таким чином, якщо знайти границю для виразу (4.17), то отримаємо:

$$\lim_{\beta \rightarrow -1} g(\theta) = \frac{1}{2\sigma\sqrt{3}}, \quad (\theta_0 - \sigma\sqrt{3}, \theta_0 + \sigma\sqrt{3}). \quad (4.20)$$

Експоненціальний розподіл для даних

Сімейство експоненціальних степеневих (power) розподілів можна використовувати також для опису *розподілів даних*. Так, для неперервної спостережуваної змінної X щільність розподілу має вигляд:

$$f(x|\theta, \tau, \alpha) = \frac{K}{\tau} \exp \left[-0,5 \left| \frac{x - \theta}{\tau} \right|^{2/(1+\alpha)} \right], \quad (4.21)$$

де $K^{-1} = 2^{1+0,5(1+\alpha)} \Gamma(1+0,5(1+\alpha))$; $-\infty < x < \infty$; $-1 < \alpha \leq 1$; $0 < \tau < \infty$; $-\infty < \theta < \infty$. Припустимо, що α відоме, θ і τ – апріорно незалежні і приймається неінформативна (рівномірна) апріорна щільність для (θ, τ) у вигляді $g(\theta, \tau) \propto \frac{1}{\tau}$. Далі можна отримати апостеріорну щільність для (θ, τ) на основі статистичних даних.

Сімейство змішаних апріорних розподілів

Існує можливість згенерувати узагальнені класи апріорних розподілів, що містять множину гіперпараметрів, які надають можливість адаптувати апріорний розподіл до наявної апріорної інформації використовуючи суміші природних спряжених апріорних розподілів. Якщо наявна апріорна інформація свідчить про те, що дані можуть кластеризуватись навколо більше однієї моди, то для трансформування цієї інформації у математичну форму можна скористатись сімейством змішаних розподілів.

4.5 Апріорні розподіли на основі даних

Апріорні розподіли на основі даних можуть бути таких типів:

1. розподіли на основі історичних даних; – вони використовуються у тих випадках, коли існують вибірки історичних даних у достатньому об'ємі для встановлення типу апріорного розподілу невідомого параметра;
2. якщо історичних даних недостатньо для встановлення типу апріорного розподілу невідомого параметра, то наявні дані навіть невеликого об'єму (при цьому можливе застосування методів розмноження вибірки) використовують для наближеного оцінювання апріорного розподілу і для формування функції правдоподібності; тобто наявні дані ділять на дві частини – одна для оцінювання апріорного розподілу, а друга – для формування функції правдоподібності.



4.6 Приклади розподілів на основі моделей з одним параметром

Моделі з одним параметром – це клас розподілів, що індексуються (визначаються) одним параметром. Розглянемо формування байєсівського висновку на основі моделей з одним параметром: біноміальна і модель Пуассона. Моделі такого типу – це зручний і корисний статистичний інструмент, який дає можливість досить просто пояснити основи байєсівського аналізу даних, а саме:

- використання спряжених апіорних розподілів;
- формування прогнозних розподілів;
- оцінювання довірчих інтервалів (*confidence regions estimation*).

Приклад 4.4. Біноміальна модель.

Розглянемо аналіз даних, що стосуються стану «фінансово щасливих» чоловіків віком 65 років або більше. Позначимо через $y_i = 1$ – відповідь загалом щасливого респондента, а через $y_i = 0$ – протилежний стан. Всього опитано 129 чоловіків, тобто $n = 129$. Якщо немає ніякої додаткової інформації стосовно того, ким являються ці індивідууми, то такі дані можна вважати такими, що характеризують довільних 129 чоловіків (*exchangeable*). Оскільки $n = 129$ – це набагато менше, ніж загальна кількість чоловіків, які проживають у вибраному регіоні, то можна сформулювати деяку спільну інформацію (*joint beliefs*) стосовно цієї групи:

- оцінка (довіра, наближення до) параметра $\theta = \frac{1}{n} \sum_{i=1}^n y_i$;
- значення $\{y_i\} \sim i.i.d.$ – незалежні та однаково розподілені двійкові випадкові числа з математичним сподіванням θ .

Використовуючи біноміальний розподіл

$$p(X=x|\theta, n) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \quad \text{для } x \in \{0, 1, \dots, n\},$$

напишемо ймовірність будь-якого потенційно можливого результату $\{y_1, \dots, y_{129}\}$ з параметром (середнє) θ визначається за виразом:

$$p(y_1, \dots, y_{129} | \theta) = \theta^x (1-\theta)^{n-x},$$

де $x = \sum_{i=1}^{129} y_i$, $n = 129$, тобто x показує скільки було всього позитивних відповідей у процесі опитування. Тепер необхідно визначити *апіорний* розподіл.

Рівномірний апріорний розподіл

У даному випадку параметр θ – це невідоме число, яке знаходиться у діапазоні між 0 і 1. Припустимо, що наявна апріорна інформація свідчить про те, що всі дискретні підінтервали в діапазоні $[0, 1]$ мають однакову ширину і однакову ймовірність попадання туди значень параметра. У символічному вигляді ця інформація може бути подана так:

$$p(a \leq \theta \leq b) = p(a+c \leq \theta \leq b+c) \text{ для } 0 \leq a < b < b+c \leq 1.$$

Ця умова передбачає, що параметр θ повинен мати рівномірну щільність:

$$p(\theta) = 1, \quad \forall \theta \in [0, 1].$$

Для цієї апріорної щільності і поданої вище моделі для бінарних результатів опитування (бінарних подій) правило Байєса приймає вигляд:

$$\begin{aligned} p(\theta | y_1, \dots, y_{129}) &= \frac{p(y_1, \dots, y_{129} | \theta) p(\theta)}{p(y_1, \dots, y_{129})} = \\ &= p(y_1, \dots, y_{129} | \theta) \frac{1}{p(y_1, \dots, y_{129})} \propto \\ &\propto p(y_1, \dots, y_{129} | \theta). \end{aligned}$$

Останній рядок свідчить про те, що у даному конкретному випадку функції $p(\theta | y_1, \dots, y_{129})$ і $p(y_1, \dots, y_{129} | \theta)$ пропорційні між собою як функції θ . Це має місце тому, що апостеріорний розподіл дорівнює значенню $p(y_1, \dots, y_{129} | \theta)$, поділеному на величину, яка не залежить від θ . А це означає, що ці дві наведені функції від θ мають однакову форми, але, можливо, різні масштаби.

Дані та апостеріорний розподіл

Нехай отриманні дані опитування дають такі результати:

- всього опитано 129 чоловіків;
- 118 індивідуумів (91%) дали позитивну відповідь: загалом вони задоволені фінансовим станом;
- 11 індивідуумів (9%) – незадоволені своїм фінансовим станом.

Повна ймовірність цих даних для деякого значення θ складає

$$p(y_1, \dots, y_{129} | \theta) = \theta^{118} (1 - \theta)^{11}.$$



Отриманий вище результат стосовно пропорційності свідчить, що апостеріорний розподіл $p(\theta|y_1, \dots, y_{129})$ бути мати таку ж форму (як подано вище), а тому істинне значення параметра θ буде з високою ймовірністю близьким до 0,91, і майже напевно перевищуватиме значення 0,8.

Однак, відповідь стосовно форми і числового значення $p(\theta|y_1, \dots, y_n)$ може бути точнішою з урахуванням значення константи масштабу. За правилом Байєса отримаємо:

$$\begin{aligned} p(\theta|y_1, \dots, y_{129}) &= \theta^{118} (1-\theta)^{11} \frac{p(\theta)}{p(y_1, \dots, y_{129})} = \\ &= \theta^{118} (1-\theta)^{11} \frac{1}{p(y_1, \dots, y_{129})} \end{aligned}$$

Значення константи масштабу $1/p(y_1, \dots, y_{129})$ за допомогою такого інтегралу:

$$\int_0^1 \theta^{a-1} (1-\theta)^{b-1} d\theta = \frac{\Gamma(a) \Gamma(b)}{\Gamma(a+b)},$$

де значення гамма функції $\Gamma(x)$ для будь-якого $x > 0$ можна визначити за відповідними таблицями або обчислити за допомогою системи **R**, використовуючи такий оператор: *gamma()*.

Як можна обчислити значення (розподіл) $p(\theta|y_1, \dots, y_{129})$? Резюмуємо отримані знання про $p(\theta|y_1, \dots, y_{129})$:

$$\text{а) } \int_0^1 p(\theta|y_1, \dots, y_{129}) d\theta = 1,$$

оскільки всі розподіли інтегруються (або визначається сума) до 1;

$$\text{б) } p(\theta|y_1, \dots, y_{129}) = \theta^{118} (1-\theta)^{11} \frac{1}{p(y_1, \dots, y_{129})}$$

за правилом Байєса.

А тому маємо: $\int_0^1 p(\theta|y_1, \dots, y_{129}) d\theta = 1$ – за допомогою (а); і за допомогою (б):

$$\int_0^1 \theta^{118} (1-\theta)^{11} \frac{1}{p(y_1, \dots, y_{129})} d\theta = 1,$$

або

$$\frac{1}{p(y_1, \dots, y_{129})} \int_0^1 \theta^{118} (1-\theta)^{11} d\theta = 1 = \frac{1}{p(y_1, \dots, y_{129})} \frac{\Gamma(119) \Gamma(12)}{\Gamma(131)},$$

тобто

$$p(y_1, \dots, y_{129}) = \frac{\Gamma(131)}{\Gamma(119) \Gamma(12)}.$$

Цей результат справедливий для будь-якої послідовності $\{y_1, \dots, y_{129}\}$, яка містить 118 одиниць і 11 нулів.

Об'єднуючи отримані результати обчислень, отримаємо:

$$\begin{aligned} p(\theta | y_1, \dots, y_{129}) &= \frac{\Gamma(131)}{\Gamma(119) \Gamma(12)} \theta^{118} (1-\theta)^{11} = \\ &= \frac{\Gamma(131)}{\Gamma(119) \Gamma(12)} \theta^{119-1} (1-\theta)^{12-1}. \end{aligned}$$

Таким чином, отримана щільність параметра θ – це бета розподіл з параметрами $a = 119$ і $b = 12$; його можна обчислити, представити графічно і згенерувати з нього псевдовипадкові значення за допомогою таких функцій системи **R**: *dbeta()* і *rbeta()*.

Використання розподілу бета

Невідома значення параметра θ , яке знаходиться у діапазоні між 0 і 1, має бета розподіл *dbeta(a, b)*, то воно описується так

$$p(\theta) = \text{dbeta}(\theta, a, b) = \frac{\Gamma(a+b)}{\Gamma(a) \Gamma(b)} \theta^{a-1} (1-\theta)^{b-1}, \quad \text{для } 0 \leq \theta \leq 1.$$

Ця випадкова змінна має такі статистичні характеристики:

$$\text{mode}[\theta] = \frac{a-1}{(a-1)+(b-1)}, \quad \text{якщо } a > 1 \text{ і } b > 1;$$

$$E[\theta] = \frac{a}{a+b};$$

$$\text{Var}[\theta] = \frac{ab}{(a+b+1)(a+b)^2} = E[\theta] \frac{E[1-\theta]}{a+b+1}.$$

Для розглянутих даних стосовно «фінансово щасливих» чоловіків $\{y_1, \dots, y_{129}\}$, для яких $\sum_{i=1}^2 y_i = 118$, отримано такі результати оцінювання:



$$\text{mode} [\theta | y_1, \dots, y_{129}] = 0,915;$$

$$E[\theta | y_1, \dots, y_{129}] = 0,908;$$

$$\text{sd} [\theta | y_1, \dots, y_{129}] = 0,025.$$

Поняття взаємного заміщення (*exchangeability*) даних

Повертаючись до наведеного вище прикладу аналізу даних стосовно «фінансово щасливих» чоловіків, зазначимо, що у даному випадку випадкова змінна приймає такі значення:

$$y_i = \begin{cases} 1, & \text{якщо учасник опитування почуває себе фінансово щасливим,} \\ 0, & \text{інакше.} \end{cases}$$

Розглянемо структуру нашої спільної довіри до даних для десяти випадково вибраних учасників опитування: $\{y_1, \dots, y_{10}\}$. Нехай необхідно визначити $p(y_1, \dots, y_{10})$, тобто $\Pr(Y_1 = y_1, \dots, Y_{10} = y_{10})$, де y_i приймає значення 1 або 0.

Припустимо, що необхідно визначити ймовірності трьох різних варіантів відповідей:

$$p(1, 0, 0, 1, 0, 1, 1, 0, 1, 1) = ?$$

$$p(1, 0, 1, 0, 1, 1, 0, 1, 1, 0) = ?$$

$$p(1, 1, 0, 0, 1, 1, 0, 0, 1, 1) = ?$$

Кожна послідовність відповідей містить 6 одиниць і чотири 0. Чи можуть ці три ймовірності бути однаковими?

Означення 4.1. Нехай $p(y_1, \dots, y_n)$ – спільна щільність розподілу ймовірностей для змінних (Y_1, \dots, Y_n) . Якщо $p(y_1, \dots, y_n) = p(y_{\pi 1}, \dots, y_{\pi n})$ для всіх перестановок π з множини $\{1, \dots, n\}$, то змінні характеризуються взаємним заміщенням.

Іншими словами, змінні (Y_1, \dots, Y_n) характеризуються взаємним заміщенням, якщо індекси цих змінних не несуть інформації про результати (тобто їх значення), тобто перестановки значень не впливають на розподіл.

Твердження 4.1. Якщо параметр $\theta \sim p(\theta)$ і змінні Y_1, \dots, Y_n умовно незалежні (за умови існування невідомого параметра θ) і однаково розподілені, тобто

$$\{Y_1, \dots, Y_n\} \sim i.i.d.,$$

то маргінально (тобто безумовно по θ) змінні Y_1, \dots, Y_n характеризуються взаємним заміщенням (*exchangeability*).

Доведення. Припустимо, що змінні Y_1, \dots, Y_n умовно незалежні (за умови існування невідомого параметра θ) і однаково розподілені. Тоді для будь-якої перестановки π з позицій $\{1, \dots, n\}$ і будь-якої множини значень $(y_1, \dots, y_n) \in \mathbf{Y}^n$ справедлива рівність за означенням маргінальної ймовірності:

$$p(y_1, \dots, y_n) = \int p(y_1, \dots, y_n | \theta) d\theta = \\ = \int \left\{ \prod_{i=1}^n p(y_i | \theta) \right\} p(\theta) d\theta = \text{(змінні } Y_i \text{ умовно незалежні та однаково}$$

розподілені);

$$= \int \left\{ \prod_{i=1}^n p(y_{\pi i} | \theta) \right\} p(\theta) d\theta = \text{(добуток не залежить від порядку}$$

використання змінних);

$$= p(y_{\pi 1}, \dots, y_{\pi n}) \text{ (за означенням маргінальної ймовірності).}$$

Теорема де Фінетті

Вище було показано, якщо

$$\left. \begin{array}{l} Y_1, \dots, Y_n | \theta \text{ i.i.d.} \\ \theta \sim p(\theta) \end{array} \right\} \Rightarrow Y_1, \dots, Y_n$$

то такі змінні характеризуються взаємним заміщенням.

Чи можна повернути стрілку у протилежний бік, тобто чи справедливий протилежний результат? Нехай $\{Y_1, Y_2, \dots\}$ – потенційно нескінченна послідовність випадкових змінних, які належать до спільного вибіркового простору \mathbf{Y} .

Теорема 4.1 (де Фінетті). Нехай $Y_i \in \mathbf{Y}$, для всіх $i \in \{1, 2, \dots\}$. Припустимо, що для будь-якого n , модель довіри (belief model) для змінних Y_1, \dots, Y_n характеризується взаємним заміщенням, тобто:

$$p(y_1, \dots, y_n) = p(y_{\pi 1}, \dots, y_{\pi n})$$

для всіх перестановок π з $\{1, \dots, n\}$. Тоді модель довіри можна записати таким чином:

$$p(y_1, \dots, y_n) = \int \left\{ \prod_{i=1}^n p(y_i | \theta) \right\} p(\theta) d\theta$$



для деякого параметра θ , деякого апіорного розподілу для θ і деякої генеруючої моделі $p(y|\theta)$. Типи апіорного розподілу і генеруючої моделі залежать від форми моделі довіри $p(y_1, \dots, y_n)$.

Розподіл ймовірностей $p(\theta)$ представляє (відображає) нашу довіру до результатів експерименту $\{Y_1, Y_2, \dots\}$, що відповідають нашій моделі довіри $p(y_1, y_2, \dots)$. Точніше кажучи,

$p(\theta)$ відображає нашу довіру (адекватність) до $\lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{Y_i}{n}$ у випадку

двійкових даних;

$p(\theta)$ відображає нашу довіру (адекватність) до $\lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{(Y_i \leq c)}{n}$ для

кожного $p(\theta)$ у загальному випадку.

Таким чином, основну ідею перестановки (заміщення) змінних можна сформулювати таким чином: якщо

$$\left. \begin{array}{l} Y_1, \dots, Y_n | \theta \text{ i.i.d.} \\ \theta \sim p(\theta) \end{array} \right\} \Leftrightarrow Y_1, \dots, Y_n,$$

то такі змінні характеризуються взаємним заміщенням для всіх n .

У яких випадках умова « Y_1, \dots, Y_n характеризуються взаємним заміщенням для всіх n » виконується? Для того щоб виконувалась ця умова, необхідно щоб була повторюваність результатів. Взаємне заміщення результатів (exchangeability) має місце тоді, коли індекси вимірів (результатів) не несуть інформації.

Ситуації, у яких повторюваність має місце, включають такі:

- Y_1, \dots, Y_n – це результати повторюваних експериментів;
- Y_1, \dots, Y_n – генеруються із скінченної популяції (множини) із заміщенням;
- Y_1, \dots, Y_n – генеруються із нескінченної популяції (множини) без заміщення.

Якщо Y_1, \dots, Y_n характеризуються заміщенням (переміщенням) і генеруються із скінченної популяції розміром $N \gg n$ без заміщення, то їх можна наближено моделювати як умовно незалежні та однаково розподілені.

4.7 Ймовірнісний висновок для даних, що характеризуються заміщенням

Апостеріорний висновок при використанні рівномірного апріорного розподілу

Якщо $\{Y_1, Y_2, \dots, Y_n | \theta\} \sim i.i.d.$ – це двійкові дані з параметром θ , то вище було показано, що розподіл цього параметра при використанні рівномірного апріорного розподілу має такий вигляд:

$$\begin{aligned} p(\theta | y_1, \dots, y_n) &= \theta^{\sum y_i} (1 - \theta)^{n - \sum y_i} \frac{p(\theta)}{p(y_1, \dots, y_n)} = \\ &= \theta^{\sum y_i} (1 - \theta)^{n - \sum y_i} \times p(\theta) / p(y_1, \dots, y_n). \end{aligned}$$

Порівнюючи відносні ймовірності для будь-яких значень θ , наприклад, θ_a і θ_b , отримаємо такий вираз:

$$\begin{aligned} \frac{p(\theta_a | y_1, \dots, y_n)}{p(\theta_b | y_1, \dots, y_n)} &= \frac{\theta_a^{\sum y_i} (1 - \theta_a)^{n - \sum y_i} \times p(\theta_a) / p(y_1, \dots, y_n)}{\theta_b^{\sum y_i} (1 - \theta_b)^{n - \sum y_i} \times p(\theta_b) / p(y_1, \dots, y_n)} = \\ &= \left(\frac{\theta_a}{\theta_b} \right)^{\sum y_i} \left(\frac{1 - \theta_a}{1 - \theta_b} \right)^{n - \sum y_i} \frac{p(\theta_a)}{p(\theta_b)}. \end{aligned}$$

Таким чином, відношення щільності для θ_a і θ_b залежить від (y_1, \dots, y_n) тільки через $\sum_{i=1}^n y_i$. На основі цього можна показати, що

$$\Pr(\theta \in A | Y_1 = y_1, \dots, Y_n = y_n) = \Pr(\theta \in A | \sum_{i=1}^n Y_i = \sum_{i=1}^n y_i).$$

Це можна інтерпретувати так: сума $\sum_{i=1}^n Y_i$ містить всю інформацію про θ , яка знаходиться в даних, і що $\sum_{i=1}^n Y_i$ – це достатня статистика для θ і $p(y_1, \dots, y_n | \theta)$.

Слово «достатня» означає, що достатньо знати $\sum_{i=1}^n Y_i$ для того щоб сформулювати (зробити) висновок стосовно θ . У даному випадку, коли $\{Y_1, Y_2, \dots, Y_n | \theta\}$ – це незалежні однаково розподілені двійкові випадкові величини з параметром θ , достатня статистика $Y = \sum_{i=1}^n Y_i$ має біноміальний розподіл з параметрами (n, θ) .



Біноміальний розподіл

Випадкова змінна $Y \in \{0, 1, \dots, n\}$ має біноміальний розподіл з параметрами (n, θ) якщо

$$\Pr(Y = y | \theta) = \text{dbinom}(y, n, \theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y},$$
$$y \in \{0, \dots, n\}.$$

Математичне сподівання і дисперсія випадкової змінної з біноміальним розподілом обчислюються за такими виразами:

$$E[Y | \theta] = n \theta;$$

$$\text{Var}[Y | \theta] = n \theta(1 - \theta).$$

Апостеріорний висновок при використанні рівномірного апріорного розподілу

Після отримання спостережень $Y = y$ необхідно визначити апостеріорний розподіл для параметра θ :

$$p(\theta | y) = \frac{p(y | \theta) p(\theta)}{p(y)} = \frac{\binom{n}{y} \theta^y (1 - \theta)^{n-y} p(\theta)}{p(y)} =$$
$$= c(y) \theta^y (1 - \theta)^{n-y} p(\theta),$$

де $c(y)$ – функція від y , але не від θ . Для рівномірного розподілу, тобто $p(\theta) = 1$, можна визначити $c(y)$ таким чином:

$$1 = \int_0^1 c(y) \theta^y (1 - \theta)^{n-y} d\theta,$$
$$1 = c(y) \int_0^1 \theta^y (1 - \theta)^{n-y} d\theta,$$
$$1 = c(y) \frac{\Gamma(y + 1) \Gamma(n - y + 1)}{\Gamma(n + 2)}.$$

Звідси знаходимо значення нормуючої константи:

$$c(y) = \frac{\Gamma(n + 2)}{\Gamma(y + 1) \Gamma(n - y + 1)},$$

а тому

$$\begin{aligned} p(\theta | y) &= \frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)} \theta^y (1-\theta)^{n-y} = \\ &= \frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)} \theta^{(y+1)-1} (1-\theta)^{(n-y+1)-1} = \\ &= \text{beta}(y+1, n-y+1). \end{aligned}$$

Повертаючись до прикладу з опитуванням стосовно задоволеності життям, для якого $Y \equiv \sum y_i = 118$, запишемо тип розподілу:

$$n = 129; Y \equiv \sum y_i = 118 \Rightarrow \{\theta | Y = 118\} \sim \text{beta}(119, 12).$$

Це підтверджує результат достатності для цієї моделі і апріорного розподілу, тобто $\sum y_i = y = 118$ і апостеріорний розподіл параметра має вигляд:

$$p(\theta | y_1, \dots, y_n) = p(\theta | y) = \text{beta}(119, 12).$$

Іншими словами, інформація, що міститься у даних $\{Y_1 = y_1, \dots, Y_n = y_n\}$ така ж само, як інформація, яка міститься в $\{Y = y\}$, де $Y = \sum Y_i$, а $y = \sum y_i$.

Формування апостеріорних розподілів при апріорному beta розподілі

Вище використано рівномірний апріорний розподіл $p(\theta) = 1$ для всіх значень $\theta \in [0, 1]$. Цей розподіл можна розглядати як *beta* розподіл з параметрами $a = 1$, $b = 1$:

$$p(\theta) = \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} \theta^{1-1} (1-\theta)^{1-1} = \frac{1}{1 \times 1} 1 \times 1 = 1.$$

Зазначимо, що $\Gamma(x+1) = x! = x(x-1)\dots 1$, якщо x – додатне ціле число, і $\Gamma(1) = 1$ за домовленістю. У попередньому параграфі показано:

$$\text{якщо } \left\{ \begin{array}{l} \theta \sim \text{beta}(1, 1) \text{ (рівномірний)} \\ Y \sim \text{binomial}(n, \theta) \end{array} \right\}, \text{ то } \{\theta | Y = y\} \sim \text{beta}(1+y, 1+n-y).$$

А тому для того щоб знайти апостеріорний розподіл у випадку, коли апріорним є *beta*($a=1, b=1$), то необхідно просто додати число «1» до параметра a і «0» до параметра b .



Чи справедливий цей результат для довільного апіорного $beta$ розподілу? Це можна довести так: нехай $\theta \sim beta(a, b)$, а $\{Y|\theta\} \sim binomial(n, \theta)$; тоді при $Y = y$ маємо:

$$\begin{aligned} p(\theta|y) &= \frac{p(y|\theta) p(\theta)}{p(y)} = \\ &= \frac{1}{p(y)} \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1} \times \binom{n}{y} \theta^y (1-\theta)^{n-1} = \\ &= c(n, y, a, b) \times \theta^{a+y-1} (1-\theta)^{b+n-y-1} = \\ &= dbeta(\theta, a+y, b+n-y). \end{aligned}$$

Важливо розуміти суть двох останніх рядків, поданих вище. Передостанній рядок означає, що $p(\theta|y)$ є функцією θ , пропорційною добутку $\theta^{a+y-1} (1-\theta)^{b+n-y-1}$. А це означає, що $p(\theta|y)$ має таку ж форму, як і бета розподіл: $dbeta(\theta, a+y, b+n-y)$.

Разом з тим відомо, що інтеграли від $p(\theta|y)$ і щільності бета обидва повинні давати в результаті 1, а тому повинні мати однаковий *масштаб*. Таким чином, все це означає, що знайдена щільність $p(\theta|y)$ і щільність бета розподілу – це фактично одна й та ж функція. В результаті ми встановили, що знайдена апостеріорна щільність пропорційна відомій щільності розподілу ймовірностей, а тому повинна дорівнювати цій щільності [101].

Спряженість розподілів

Вище показано, що апіорний бета розподіл для параметра і біноміальний розподіл у якості моделі даних (*sampling model*) привели до апостеріорного розподілу бета. Для того щоб коректно відобразити цей факт, підкреслимо, що клас апіорних бета розподілів є *спряженим* для біноміальної моделі даних (*sampling model*).

Означення 4.2 (спряженість розподілів). Клас P апіорних розподілів для θ називають спряженим для моделі даних (*sampling model*) $p(y|\theta)$, якщо

$$p(\theta) \in P \Rightarrow p(\theta|y) \in P,$$

тобто апіорний і апостеріорний розподіли належать до одного класу розподілів.

Спряжені розподіли дають можливість спростити обчислення апостеріорних розподілів, але вони можуть не відображати наявну апіорну інформацію. Однак, суміші спряжених апіорних розподілів дуже гнучкі, а

пов'язані з ними обчислювальні процедури відносно просто відслідковуються.

Об'єднання (комбінування) апріорної інформації і даних

Якщо $\theta|\{Y = y\} \sim \text{beta}(a + y, b + n - y)$, то

$$E[\theta|y] = \frac{a + y}{a + b + n}; \quad \text{mode}[\theta|y] = \frac{a + y - 1}{a + b + n - 2};$$

$$\text{Var}[\theta|y] = \frac{E[\theta|y] E[1 - \theta|y]}{a + b + n + 1}.$$

Апостеріорне математичне сподівання $E(\theta|y)$ легко знайти в результаті комбінування апріорної інформації та даних:

$$E[\theta|y] = \frac{a + y}{a + b + n} =$$

(після розділення виразу на дві частини)

$$\begin{aligned} &= \frac{a + y}{a + b + n} \frac{a}{a + b} + \frac{n}{a + b + n} \frac{y}{n} = \\ &= \frac{a + y}{a + b + n} \times \text{prior expectation} + \frac{n}{a + b + n} \times \text{data average}, \end{aligned}$$

де *prior expectation* = математичне сподівання апріорного розподілу; *data average* = усереднення даних.

Для цієї моделі і апріорного розподілу апріорне математичне сподівання (також відоме як апостеріорне середнє подане вище) є зваженим середнім апріорного сподівання та вибіркового середнього з ваговими коефіцієнтами, пропорційними $(a + b)$ і n , відповідно. Це приводить до інтерпретації a і b як «апріорні дані»:

$$\begin{aligned} a &\approx \text{"апріорна кількість одиничок"}, \\ b &\approx \text{"апріорна кількість нулів"}, \\ a + b &\approx \text{"апріорна довжина вибірки"} \end{aligned}$$

Якщо довжина нашої вибірки n перевищує апріорну довжину вибірки $a + b$, то логічно стверджувати, що більша частина інформації про θ повинна надходити від даних, а не від апріорного розподілу. Це дійсно має місце: наприклад, якщо $n \gg (a + b)$, то



$$\frac{a+b}{a+b+n} \approx 0; \quad E[\theta|y] \approx \frac{y}{n}; \quad Var[\theta|y] \approx \frac{1}{n} \frac{y}{n} \left(1 - \frac{y}{n}\right).$$

Прогнозний розподіл для бінарних даних

Важливою відмінною рисою байєсівського висновку є існування прогнозуючого розподілу для нових спостережень. Нехай y_1, \dots, y_n – виміри n бінарних випадкових змінних, і нехай $\tilde{Y} \in [0, 1]$ – додатковий вимір для тих же змінних, який ще отримаємо у майбутньому.

Прогнозний розподіл для \tilde{Y} – це умовний розподіл для \tilde{Y} за умови $\{Y_1 = y_1, \dots, Y_n = y_n\}$. Для умовно незалежних та однаково розподілених бінарних змінних цей розподіл можна отримати за допомогою розподілу \tilde{Y} при заданому θ та апостеріорного розподілу для θ :

$$\begin{aligned} \Pr(\tilde{Y} = 1 | y_1, \dots, y_n) &= \int_{\theta} \Pr(\tilde{Y} = 1, \theta | y_1, \dots, y_n) d\theta = \\ &= \int_{\theta} \Pr(\tilde{Y} = 1 | \theta, y_1, \dots, y_n) p(\theta | y_1, \dots, y_n) d\theta = \\ &= \int_{\theta} \theta p(\theta | y_1, \dots, y_n) d\theta = \\ &= E[\theta | y_1, \dots, y_n] = \frac{a + \sum_{i=1}^n y_i}{a + b + n}; \end{aligned}$$

$$\Pr(\tilde{Y} = 0 | y_1, \dots, y_n) = 1 - E[\theta | y_1, \dots, y_n] = \frac{b + \sum_{i=1}^n (1 - y_i)}{a + b + n}.$$

Стосовно цього прогнозного розподілу можна зробити такі зауваження:

1. Він не залежить від значень невідомих величин. Якби це було так, то його не можна було б використовувати для обчислення оцінок прогнозів.
2. Прогнозний розподіл залежить від наявних (минулих) спостережень. У цьому розподілі \tilde{Y} залежить від Y_1, \dots, Y_n . Це має місце тому, що спостереження Y_1, \dots, Y_n дають інформацію про параметр θ , який, у свою чергу, дає інформацію про \tilde{Y} . Якби було навпаки, тобто \tilde{Y} не залежав від Y_1, \dots, Y_n , то це означало б неможливість отримати інформацію про майбутні (прогнозовані) значення.

Приклад оцінювання прогнозного розподілу

Рівномірний апріорний розподіл, або $beta(1, 1)$ можна розглядати як еквівалент інформації, що міститься у апріорній множині даних, яка складається із однієї «1» і одного «0». При такому апріорному розподілі

$$\Pr(\tilde{Y} = 1 | Y = y) = E[\theta | Y = y] = \frac{2}{2+n} \frac{1}{2} + \frac{n}{2+n} \frac{y}{n},$$

$$\text{mode}(\theta | Y = y) = \frac{y}{n},$$

де $Y = \sum_{i=1}^n Y_i$. Чи має значення різниця між цими двома апостеріорними величинами, отриманими на основі наявної інформації? Розглянемо випадок коли $Y = 0$, а при цьому $\text{mode}(\theta | Y = 0) = 0$, але $\Pr(\tilde{Y} = 1 | Y = 0) = 1/(2+n)$.

4.8 Приклад застосування моделі Пуассона

Для аналізу цілочислових вимірів $Y = \{0, 1, 2, \dots\}$ можна скористатись моделлю Пуассона.

Розподіл Пуассона

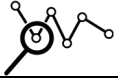
Змінна Y із середнім θ має розподіл Пуассона якщо

$$\Pr(Y = y | \theta) = \text{dpois}(y, \theta) = \frac{\theta^y e^{-\theta}}{y!}, \quad y \in \{0, 1, 2, \dots\}.$$

Математичне сподівання і дисперсія цієї випадкової змінної мають однакові значення:

$$\begin{aligned} E[Y | \theta] &= \theta, \\ \text{Var}[Y | \theta] &= \theta. \end{aligned}$$

Іноді говорять, що сімейство розподілів Пуассона характеризується (споріднене) взаємозв'язком середне-дисперсія, оскільки розподіл Пуассона з більшим середнім буде мати і більшу дисперсію.

**Апостеріорний висновок**

Якщо дані $\{Y_1, \dots, Y_n\} \sim i.i.d.$, тобто незалежні і мають однаковий розподіл із середнім θ , то спільна функція розподілу ймовірностей для цих змінних має вигляд:

$$\begin{aligned} \Pr(Y_1=y_1, \dots, Y_n=y_n | \theta) &= \prod_{i=1}^n p(y_i | \theta) = \prod_{i=1}^n \frac{1}{y_i!} \theta^{y_i} e^{-\theta} = \\ &= c(y_1, \dots, y_n) \theta^{\sum y_i} e^{-n\theta} \end{aligned}$$

де $c(y_1, \dots, y_n)$ – константа, яка залежить від значень y_1, \dots, y_n .

Тепер можна порівняти, наприклад, апостеріорні результати, які можна отримати для двох значень параметра θ , тобто

$$\begin{aligned} \frac{p(\theta_a | y_1, \dots, y_n)}{p(\theta_b | y_1, \dots, y_n)} &= \frac{c(y_1, \dots, y_n) e^{-n\theta_a} \theta_a^{\sum y_i} p(\theta_a)}{c(y_1, \dots, y_n) e^{-n\theta_b} \theta_b^{\sum y_i} p(\theta_b)} = \\ &= \frac{e^{-n\theta_a} \theta_a^{\sum y_i} p(\theta_a)}{e^{-n\theta_b} \theta_b^{\sum y_i} p(\theta_b)} \end{aligned}$$

Так само, як і у випадку бінарних даних (незалежних та однаково розподілених), сума $\sum_{i=1}^n Y_i$ утримує всю інформацію про θ , що міститься у даних. Тобто можна сказати, що $\sum_{i=1}^n Y_i$ – це достатня статистика для опису параметра, більше того, $\left\{ \sum_{i=1}^n Y_i \right\} \sim poisson(n, \theta)$.

Спряжений апіорний розподіл

Скористаємось таким класом спряжених апіорних розподілів, який дасть можливість спростити обчислення апостеріорного розподілу. Згадаємо, що клас апіорних щільностей буде спряженим для розподілу даних $p(y_1, \dots, y_n | \theta)$, якщо апостеріорний розподіл буде належати до цього ж класу. Якщо дані відповідають моделі Пуассона, то апостеріорний розподіл для θ має таку форму:

$$\begin{aligned} p(\theta | y_1, \dots, y_n) &\propto p(\theta) p(y_1, \dots, y_n | \theta) = \\ &\propto p(\theta) \theta^{\sum y_i} e^{-n\theta} \end{aligned}$$

Це означає, що незалежно від того, який розподіл вибрано за спряжений, результат буде містити члени $\theta^{c_1} e^{-c_2\theta}$ з константами c_1 і c_2 . Самий простий клас таких щільностей включає тільки ці члени, а їх відповідні розподіли ймовірностей відомі як сімейства гамма розподілів.

Гамма розподіл для параметра

Додатна величина θ має розподіл $gamma(a, b)$, якщо

$$p(\theta) = dgamma(\theta, a, b) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}; \quad \theta, a, b > 0$$

Описові статистики цієї випадкової змінної описуються так:

$$\begin{aligned} E[\theta] &= a/b; \\ Var[\theta] &= a/b^2; \\ mode[\theta] &= \begin{cases} (a-1)/b, & \text{якщо } a > 1; \\ 0, & \text{якщо } a \leq 1. \end{cases} \end{aligned}$$

Апостеріорний розподіл для θ

Нехай дані $\{Y_1, \dots, Y_n\} \sim i.i.d.$, тобто незалежні і мають однаковий розподіл Пуассона, а параметр має апріорний гамма розподіл: $p(\theta) = dgamma(\theta, a, b)$. Тепер апостеріорний розподіл для параметра буде такий:

$$\begin{aligned} p(\theta | y_1, \dots, y_n) &= \frac{p(\theta) p(y_1, \dots, y_n | \theta)}{p(y_1, \dots, y_n)} = \\ &= \left\{ \theta^{a-1} e^{-b\theta} \right\} \cdot \left\{ \theta^{\sum y_i} e^{-n\theta} \right\} \cdot c(y_1, \dots, y_n, a, b) = \\ &= \left\{ \theta^{a+\sum y_i-1} e^{-(b+n)\theta} \right\} \cdot c(y_1, \dots, y_n, a, b) \end{aligned}$$

Очевидно, що апостеріорним розподілом параметра θ є гамма розподіл. Тобто ми показали, що сімейство гамма розподілів є спряженим для розподілу Пуассона для даних:

$$\left. \begin{array}{l} \theta \sim gamma(a, b) \\ Y_1, \dots, Y_n | \theta \sim Poisson(\theta) \end{array} \right\} \Rightarrow \{\theta | Y_1, \dots, Y_n\} \sim gamma\left(a + \sum_{i=1}^n Y_i, b+n\right)$$

Оцінювання і прогнозування виконується аналогічно до біноміальної моделі. Апостеріорне сподівання для θ є опуклою комбінацією апріорного сподівання і вибіркового середнього:

$$E[\theta | y_1, \dots, y_n] = \frac{a + \sum y_i}{b + n} = \frac{b}{b+n} \frac{a}{b} + \frac{n}{b+n} \frac{\sum y_i}{n}$$



де b – інтерпретується як кількість апіорних спостережень; a – інтерпретується як сума значень (counts) для b апіорних спостережень.

При великих значеннях n інформація, що надходить від даних, домінує над апіорною інформацією:

$$n \gg b \Rightarrow E[\theta | y_1, \dots, y_n] \approx \bar{y}; \quad \text{Var}[\theta | y_1, \dots, y_n] \approx \bar{y}/n$$

Оцінки прогнозів (майбутніх значень) можна обчислити за допомогою апостеріорного прогнозного розподілу:

$$\begin{aligned} p(\hat{y} | y_1, \dots, y_n) &= \int_0^{\infty} p(\hat{y} | \theta, y_1, \dots, y_n) p(\theta | y_1, \dots, y_n) d\theta = \\ &= \int_0^{\infty} p(\hat{y} | \theta) p(\theta | y_1, \dots, y_n) d\theta = \\ &= \int d\text{pois}(\hat{y}, \theta) d\text{gamma}(\theta, a + \sum y_i, b+n) d\theta = \\ &= \int \left\{ \frac{1}{\hat{y}!} \theta^{\hat{y}} e^{-\theta} \right\} \left\{ \frac{(b+n)^{a+\sum y_i}}{\Gamma(a+\sum y_i)} \theta^{a+\sum y_i-1} e^{-(b+n)\theta} \right\} d\theta = \\ &= \frac{(b+n)^{a+\sum y_i}}{\Gamma(\hat{y}+1) \Gamma(a+\sum y_i)} \int_0^{\infty} \theta^{a+\sum y_i+\hat{y}-1} e^{-(b+n+1)\theta} d\theta. \end{aligned}$$

Для обчислення поданого вище інтегралу скористаємось наявною інформацією стосовно щільності гамма розподілу:

$$1 = \int_0^{\infty} \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} d\theta \quad \text{для будь-яких } a, b > 0$$

Звідси знайдемо, що

$$\int_0^{\infty} \theta^{a-1} e^{-b\theta} d\theta = \frac{\Gamma(a)}{b^a} \quad \text{для будь-яких } a, b > 0$$

Тепер підставимо $a + \sum y_i + \hat{y}$ замість a , і $b + n + 1$ замість b і отримаємо:

$$\int_0^{\infty} \theta^{a + \sum y_i + \hat{y} - 1} e^{-(b+n+1)\theta} d\theta = \frac{\Gamma(a + \sum y_i + \hat{y})}{(b+n+1)^{a + \sum y_i + \hat{y}}}$$

Після деяких спрощень отримаємо:

$$p(\hat{y} | y_1, \dots, y_n) = \frac{\Gamma(a + \sum y_i + \tilde{y})}{\Gamma(\hat{y}+1) \Gamma(a + \sum y_i)} \left(\frac{b+n}{b+n+1} \right)^{a + \sum y_i} \left(\frac{1}{b+n+1} \right)^{\hat{y}},$$

для $\hat{y} \in \{0, 1, 2, \dots\}$.

Це негативний біноміальний розподіл з параметрами $(a + \sum y_i, b + n)$. Математичне сподівання і дисперсія для цього розподілу обчислюються за виразами:

$$E[\hat{Y} | y_1, \dots, y_n] = \frac{a + \sum y_i}{b + n} = E[\theta | y_1, \dots, y_n];$$

$$\begin{aligned} \text{Var}[\hat{Y} | y_1, \dots, y_n] &= \frac{a + \sum y_i}{b + n} \frac{b + n + 1}{b + n} = \text{Var}[\theta | y_1, \dots, y_n] \cdot (b + n + 1) = \\ &= E[\theta | y_1, \dots, y_n] \cdot \frac{b + n + 1}{b + n}. \end{aligned}$$

Розглянемо глибше вираз для обчислення прогнозової дисперсії. Прогнозна дисперсія є мірою апостеріорної невизначеності стосовно нової вибірки \hat{Y} із генеральної сукупності. Невизначеність стосовно \hat{Y} зумовлена наявністю невизначеності стосовно генеральної сукупності, а також варіабельністю випадкової вибірки з цієї сукупності.

При великих значеннях n невизначеність стосовно θ незначна $((b + n + 1)/(b + n) \approx 1)$, а невизначеність стосовно \hat{Y} зумовлена, в основному, варіабельністю випадкової вибірки з генеральної сукупності, яка дорівнює θ – параметру розподілу Пуассона. А при невеликих значеннях n загальна невизначеність \hat{Y} включає варіабельність випадкової вибірки і невизначеність, зумовлену короткою вибіркою, а тому у цілому невизначеність буде вищою від варіабельності $((b + n + 1)/(b + n) > 1)$.

Приклад 4.5. Застосування розподілу Пуассона і гамма розподілу. Протягом 1990-х років були зібрані дані стосовно рівня освіти і кількості дітей у жінок США віком 40 років (на момент збору даних). У цьому прикладі буде зроблено порівняння стосовно кількості дітей у жінок, які мали ступінь бакалавра, і у жінок без ступеня. Позначимо через $Y_{1,1}, \dots, Y_{n_1,1}$ – кількість дітей у жінок n_1 без ступеня бакалавра, $Y_{1,2}, \dots, Y_{n_2,2}$ – кількість дітей у жінок n_2 у жінок із ступенем бакалавра або вище. Для цього прикладу скористаємось такими моделями (розподілами) для даних:

$$Y_{1,1}, \dots, Y_{n_1,1} | \theta_1 \sim i.i.d. \text{Poisson}(\theta_1);$$

$$Y_{1,2}, \dots, Y_{n_2,2} | \theta_2 \sim i.i.d. \text{Poisson}(\theta_2).$$

Можливість використання цього розподілу для даних підтверджена експериментально.

Групові суми і середні обчислених емпіричних подані нижче:



Нижче ступеня бакалавра: $n_1 = 111$; $\sum_{i=1}^{n_1} Y_{i,1} = 217$; $\bar{Y}_1 = 1,95$;

Ступінь бакалавра або вище: $n_2 = 44$; $\sum_{i=1}^{n_2} Y_{i,2} = 66$; $\bar{Y}_2 = 1,5$.

Для випадку, коли $\{\theta_1, \theta_2\} \sim i.i.d. \text{gamma}(a=2, b=1)$ (апостеріорний розподіл), отримано такі апостеріорні розподіли:

$$\theta_1 | \{n_1 = 111; \sum Y_{i,1} = 217\} \sim \text{gamma}(2+217, 1+111) = \text{gamma}(219, 112);$$

$$\theta_2 | \{n_2 = 44; \sum Y_{i,2} = 66\} \sim \text{gamma}(2+66, 1+44) = \text{gamma}(68, 45).$$

Апостеріорні середні, моди і довірчі інтервали на рівні довіри 95% для параметрів θ_1 і θ_2 можна отримати з їх апостеріорних гамма розподілів.

За допомогою системи R отримано такі результати: $\theta_1 > \theta_2$. Наприклад, $\Pr(\theta_1 > \theta_2 | \sum Y_{i,1} = 217; \sum Y_{i,2} = 66) = 0,97$.

Тепер розглянемо два випадково вибраних суб'єкти – по одному із кожної групи жінок. У якій мірі можна очікувати, що представниця групи без бакалаврського ступеня може мати більше дітей ніж інша? Ймовірність цієї події можна розрахувати точно. Обидва апостеріорні прогнозні розподіли для \hat{Y}_1 і \hat{Y}_2 відносяться до негативних біноміальних розподілів.

Зазначимо, що між цими двома розподілами набагато більше спільного ніж між апостеріорними розподілами для параметрів θ_1 і θ_2 . Наприклад,

$$\Pr(\hat{Y}_1 > \hat{Y}_2 | \sum Y_{i,1} = 217; \sum Y_{i,2} = 66) = 0,48,$$

а

$$\Pr(\hat{Y}_1 = \hat{Y}_2 | \sum Y_{i,1} = 217; \sum Y_{i,2} = 66) = 0,22.$$

Дуже важливо розуміти різницю між подіями $\{\theta_1 > \theta_2\}$ і $\{\hat{Y}_1 > \hat{Y}_2\}$. Наявність значного свідчення відмінності між двома різними групами (популяціями) ще не означає, що ця відмінність справді велика.

РОЗДІЛ 5

АЛГОРИТМИ МОНТЕ-КАРЛО ДЛЯ МАРКОВСЬКИХ ЛАНЦЮГІВ

5.1 Метод Монте-Карло

У попередньому розділі наведено приклади, у яких спряжений апіорний розподіл для параметра θ приводив до апостеріорного розподілу. Отриманий таким чином розподіл давав можливість просто обчислювати апостеріорне середнє та дисперсію. Однак, часто необхідно зробити інші висновки на основі апостеріорного розподілу. Наприклад, необхідно обчислити ймовірність $\Pr(\theta \in A | y_1, \dots, y_n)$, де A – деяка довільна множина можливих значень параметра. Також може виникнути задача визначення середніх та стандартних відхилень для деяких функцій від θ , або прогнозного розподілу для пропущених або не спостережуваних даних.

При порівнянні двох або більше вибірок даних (популяцій) може виникнути задача знаходження апостеріорного розподілу для $|\theta_1 - \theta_2|$, θ_1 / θ_2 або $\max\{\theta_1, \dots, \theta_m\}$. Всі ці вирази є функціями більше одного параметра. Обчислення точних значень цих апостеріорних величин може бути пов'язано із значними труднощами. Але якщо можна згенерувати випадкові значення цих параметрів з їх апостеріорних розподілів, то всі ці апостеріорні величини можна апроксимувати із бажаною точністю за допомогою методу Монте-Карло [62, 63].

У останньому розділі було отримано два апостеріорних розподіли для кількості народжених дітей у жінок без ступеня бакалавра і жінок, які мають цей ступінь (або вищий), відповідно:

$$p(\theta_1 | \sum_{i=1}^{111} Y_{i,1} = 217) \sim \text{gamma}(\theta_1, 219, 112);$$
$$p(\theta_2 | \sum_{i=1}^{44} Y_{i,2} = 66) \sim \text{gamma}(\theta_2, 68, 45).$$

Параметри θ_1 і θ_2 моделювались як умовно незалежні для наявних даних. Було показано, що

$$\Pr(\theta_1 > \theta_2 | \sum Y_{i,1} = 217, \sum Y_{i,2} = 66) = 0,97.$$



Яким чином була розрахована ця ймовірність?

$$\begin{aligned} \Pr(\theta_1 > \theta_2 \mid y_{1,1}, \dots, y_{n_2,2}) &= \\ &= \int_0^{\infty} \int_0^{\theta_1} p(\theta_1, \theta_2 \mid y_{1,1}, \dots, y_{n_2,2}) d\theta_2 d\theta_1 = \\ &= \int_0^{\infty} \int_0^{\theta_1} \text{dgamma}(\theta_1, 219, 112) \cdot \text{dgamma}(\theta_2, 68, 45) d\theta_2 d\theta_1 = \\ &= \frac{112^{219} 45^{68}}{\Gamma(219) \Gamma(68)} \int_0^{\infty} \int_0^{\theta_1} \theta_1^{218} \theta_2^{67} \exp(-112 \theta_1 - 45 \theta_2) d\theta_2 d\theta_1. \end{aligned}$$

Цей інтеграл можна обчислити різними способами. Його можна обчислити вручну, використовуючи знання з математичного аналізу, або за допомогою чисельних методів. Однак, можливість використання того чи іншого методу залежить від особливостей моделі, типу апіорного розподілу, а також постановки задачі стосовно обчислення конкретної ймовірності [51].

Альтернативним методом обчислення цього та багатьох інших інтегралів є апроксимація за методом Монте-Карло, яка ґрунтується на повторному генеруванні псевдовипадкових чисел з наступним усередненням результатів.

Нехай θ – параметр, оцінку якого необхідно знайти, і нехай y_1, \dots, y_n – вибірка даних з розподілу $p(y_1, \dots, y_n \mid \theta)$. Припустимо, що існує можливість генерування M незалежних значень параметра θ з апостеріорного розподілу $p(\theta \mid y_1, \dots, y_n)$, тобто

$$\{\theta^{(1)}, \dots, \theta^{(M)}\} \sim i.i.d. p(\theta \mid y_1, \dots, y_n).$$

Тоді емпіричний розподіл значень $\{\theta^{(1)}, \dots, \theta^{(M)}\}$ буде апроксимувати розподіл $p(\theta \mid y_1, \dots, y_n)$ з точністю, яка залежить від кількості M згенерованих псевдовипадкових чисел (ПВЧ). Емпіричний розподіл $\{\theta^{(1)}, \dots, \theta^{(M)}\}$ називають *апроксимацією за методом Монте-Карло* розподілу $p(\theta \mid y_1, \dots, y_n)$.

Відомо, що якість апроксимації зростає із збільшенням кількості згенерованих чисел M . Для узагальнення припустимо додатково, що $g(\theta)$ – деяка довільна функція. Згідно із законом великих чисел, якщо $\{\theta^{(1)}, \dots, \theta^{(M)}\}$ – незалежні та однаково розподілені числа з розподілу $p(\theta \mid y_1, \dots, y_n)$, то

$$\frac{1}{M} \sum_{k=1}^M g(\theta^{(k)}) \rightarrow E[g(\theta) | y_1, \dots, y_n] = \int_{\theta} g(\theta) p(\theta | y_1, \dots, y_n) d\theta.$$

$$M \rightarrow \infty.$$

А звідси випливає, що при $M \rightarrow \infty$ виконуються такі умови:

$$\bar{\theta} = \frac{\sum_{k=1}^M \theta^{(k)}}{M} \rightarrow E[\theta | y_1, \dots, y_n];$$

$$\frac{\sum_{k=1}^M (\theta^{(k)} - \bar{\theta})^2}{M-1} \rightarrow Var[\theta | y_1, \dots, y_n];$$

$$\frac{\#(\theta^{(m)} \leq c)}{M} \rightarrow \Pr(\theta \leq c | y_1, \dots, y_n),$$

де символ «#» означає «кількість»;

емпіричний розподіл $\{\theta^{(1)}, \dots, \theta^{(M)}\} \rightarrow p(\theta | y_1, \dots, y_n)$;

медіана $\{\theta^{(1)}, \dots, \theta^{(M)}\} \rightarrow \theta_{1/2}$;

α – процентиль емпіричного розподілу $\{\theta^{(1)}, \dots, \theta^{(M)}\} \rightarrow \theta_{\alpha}$.

Таким чином, стосовно будь-якого аспекту (характеристики) апостеріорного розподілу можна стверджувати, що його можна апроксимувати із бажаною точністю, якщо вибірка згенерованих псевдовипадкових чисел достатньо велика [48].

Розрахунок чисельних значень для прикладу з народжуваністю

Порівняємо значення деяких апостеріорних величин, обчислених за методом Монте-Карло, з іншими процедурами. Припустимо, що дані задовольняють властивості

$$\{Y_1, \dots, Y_n | \theta\} \sim i.i.d. \text{Poisson}(\theta),$$

а параметр має апріорний гамма розподіл:

$$\theta \sim \text{gamma}(a, b).$$

Після отримання спостережень $\{Y_1 = y_1, \dots, Y_n = y_n\}$ апостеріорний розподіл буде гамма розподілом, тобто $\theta \sim \text{gamma}(a + \sum y_i, b + n)$.

У прикладі з аналізом народжуваності дітей для жінок із ступенем бакалавра використано такі значення параметрів:

$$(a = 2, b = 1) \text{ і } (\sum y_i = 66, n = 44).$$



Математичне сподівання, тобто апостеріорне середнє

$$\bar{\theta} = \frac{a + \sum y_i}{b + n} = \frac{68}{45} = 1,51.$$

Згідно з апроксимацією за методом Монте-Карло для різних значень кількості ітерацій $S \in \{10, 100, 1000\}$ (у системі **R** [37]) отримано такі значення:

$$\begin{aligned}\theta_{MC} &= 1,53279, & S &= 10, \\ \theta_{MC} &= 1,51395, & S &= 100, \\ \theta_{MC} &= 1,50101, & S &= 1000.\end{aligned}$$

Результати обчислень можуть відрізнятись залежно від початкових умов генератора псевдовипадкових чисел.

Ймовірності: апостеріорну ймовірність того, що $\{\theta < 1,75\}$ можна знайти з високою точністю у системі **R** за допомогою команди `pgamma(1.75, a + s y, b + n)`. В результаті отримано значення 0,8998.

Квантили: 95%-й довірчий інтервал на основі квантилів можна обчислити за командою у системі **R** таким чином: `qgamma(c(0.025, 0.975), a + s y, b + n)`; результат такий: (1.173, 1.891).

Додатково можна розрахувати стандартні похибки методу Монте-Карло для встановлення точності апроксимації апостеріорних середніх [26, 48]. Так, вибіркове середнє для значень, згенерованих за методом Монте-Карло, складає

$$\bar{\theta} = \frac{1}{M} \sum_{k=1}^M \theta_k.$$

Центральна гранична теорема говорить, що $\bar{\theta}$ має наближено нормальний розподіл з математичним сподіванням $E[\theta | y_1, \dots, y_n]$ і стандартним відхиленням $\sqrt{Var[\theta | y_1, \dots, y_n] / M}$. Стандартна похибка методу Монте-Карло є апроксимацією для цього стандартного відхилення. Якщо за оцінку дисперсії (тобто апроксимацію для $Var[\theta | y_1, \dots, y_n]$) взяти величину

$$\hat{\sigma}_{\theta}^2 = \frac{1}{M-1} \sum_{k=1}^M (\theta_k - \bar{\theta})^2,$$

то стандартна похибка методу Монте-Карло визначиться за виразом:

$$SE_{MC} = \sqrt{\hat{\sigma}^2 / M}.$$

Довірчий інтервал для оцінки параметра за методом Монте-Карло (апостеріорного середнього) на рівні 95% наближено визначається величиною:

$$\Delta \hat{\theta}_{95} = \hat{\theta} \pm 2\sqrt{\hat{\sigma}^2 / M}.$$

Вибір кількості ітерацій M ґрунтується на тому, що стандартна похибка методу Монте-Карло має бути меншою ніж похибка визначення математичного сподівання $E[\theta | y_1, \dots, y_n]$. Наприклад, припустимо, що при використанні значення $M = 100$ отримано оцінку дисперсії $Var[\theta | y_1, \dots, y_n] = 0,024$. При цьому наближена стандартна похибка методу Монте-Карло складе

$$SE_{MC} = \sqrt{0,024/100} = 0,015.$$

Якщо ж задача ставилась таким чином, щоб різниця між математичним сподіванням розподілу і його оцінкою за методом Монте-Карло склала менше 0,01 з високою ймовірністю, то розмір згенерованої вибірки $\{\hat{\theta}_1, \dots, \hat{\theta}_M\}$ необхідно збільшити так, щоб виконувалась нерівність:

$$2\sqrt{0,024/M} < 0,01, \text{ тобто } M > 960.$$

5.2 Апостеріорний висновок для довільних функцій

Припустимо, що необхідно знайти апостеріорний розподіл для деякої функції $g(\theta)$, властивості якої можуть бути проаналізовані аналітично або чисельними методами. Наприклад, при використанні біноміальної моделі зустрічається функція такого вигляду:

$$\gamma = \log \frac{\theta}{1 - \theta}.$$

Згідно із законом великих чисел, якщо генерується послідовність значень $\{\theta_1, \theta_2, \dots, \theta_k, \dots\}$ з апостеріорного розподілу для θ , то середнє значення $\log \frac{\theta_k}{1 - \theta_k}$ збігається до



$$E \left[\log \frac{\theta_k}{1 - \theta_k} \mid y_1, \dots, y_n \right].$$

Однак, часто виникає необхідність дослідити інші аспекти апостеріорного розподілу величини $\gamma = \log \frac{\theta}{1 - \theta}$. Все це можна зробити у

межах застосування методу Монте-Карло:

- згенерувати $\theta^{(1)} \sim p(\theta \mid y_1, \dots, y_n)$, обчислити $\gamma^{(1)} = g(\theta^{(1)})$,
- згенерувати $\theta^{(2)} \sim p(\theta \mid y_1, \dots, y_n)$, обчислити $\gamma^{(2)} = g(\theta^{(2)})$,
- ⋮
- згенерувати $\theta^{(M)} \sim p(\theta \mid y_1, \dots, y_n)$, обчислити $\gamma^{(M)} = g(\theta^{(M)})$.

Обчислення у кожному рядку виконуються незалежно. Послідовність $\{\gamma^{(1)}, \dots, \gamma^{(M)}\}$ – утворює вибірку з M незалежних значень з розподілу $p(\gamma \mid y_1, \dots, y_n)$ і при $M \rightarrow \infty$ виконуються такі умови:

$$\bar{\gamma} = \frac{1}{M} \sum_{k=1}^M \gamma^{(k)} \rightarrow E[\gamma \mid y_1, \dots, y_n],$$
$$\frac{1}{M-1} \sum_{k=1}^M (\gamma^{(k)} - \bar{\gamma})^2 \rightarrow \text{Var}[\gamma \mid y_1, \dots, y_n],$$

а емпіричний розподіл прямує до істинного:

$$\{\gamma^{(1)}, \dots, \gamma^{(M)}\} \rightarrow p(\gamma \mid y_1, \dots, y_n).$$

5.3 Генерування значень з прогнозного розподілу

Як було показано у попередньому розділі, прогнозний розподіл випадкової змінної \hat{Y} – це такий розподіл ймовірностей для \hat{Y} , який має такі характеристики:

- відомі величини обчислюються за вибраних (відомих) умов;
- залежність від невідомих величин ліквідується шляхом інтегрування (або обчисленням відповідної суми у дискретному випадку).

Наприклад, нехай \hat{Y} – це кількість дітей особи, яка є членом популяції жінок віком 40 років із ступенем бакалавра. Якби була відома істинна середня народжуваність θ цієї популяції, то невизначеність стосовно \hat{Y} можна описати розподілом Пуассона (прогнозна модель даних):

$$\Pr(\hat{Y} = \hat{y} | \theta) = p(\hat{y} | \theta) = \frac{1}{\hat{y}!} \theta^{\hat{y}} e^{-\theta}.$$

Однак, за цією моделлю не можна обчислювати прогнози, оскільки параметр θ в дійсності (з самого початку) невідомий. Якби не існувала вибірка даних з цієї популяції (сукупності осіб), то прогнозний розподіл можна було б отримати шляхом інтегрування

$$\Pr(\hat{Y} = \hat{y}) = \int p(\hat{y} | \theta) p(\theta) d\theta.$$

У випадку, коли $\theta \sim \text{gamma}(a, b)$, прогнозним буде (як було показано у попередньому розділі) негативний біноміальний розподіл з параметрами (a, b) : $\text{nbinoomial}(a, b)$.

Прогнозний розподіл, який інтегрується по невідомому параметру, але не концентрується на спостереженнях, називають *апріорним прогнозним розподілом*. Такий розподіл може бути корисним у випадку, коли апріорний розподіл параметра θ дійсно узгоджується із спостереженнями Y .

Після отримання вибірки спостережень Y_1, \dots, Y_n з популяції відповідний (апостеріорний) прогнозний розподіл для нового спостереження приймає вигляд:

$$\begin{aligned} \Pr(\hat{Y} = \hat{y} | Y_1 = y_1, \dots, Y_n = y_n) &= \int p(\hat{y} | \theta, y_1, \dots, y_n) p(\theta | y_1, \dots, y_n) d\theta = \\ &= \int p(\hat{y} | \theta) p(\theta | y_1, \dots, y_n) d\theta. \end{aligned}$$

Такий розподіл називають *апостеріорним прогнозним розподілом*, оскільки він концентрується на отриманих спостереженнях. У попередньому розділі показано, що при використанні моделі Пуассона для даних і апріорного гамма розподілу для параметра апостеріорним прогнозним буде негативний біноміальний розподіл з параметрами $(a + \sum y_i, b + n)$.

У багатьох випадках моделювання існує можливість генерувати числа з розподілу $p(\theta | y_1, \dots, y_n)$ і $p(y | \theta)$, але генерувати безпосередньо з розподілу $p(\hat{y} | y_1, \dots, y_n)$ може бути занадто складно. У такому випадку можна генерувати з апостеріорного прогнозного розподілу непрямо за допомогою процедури Монте-Карло [81, 82].

Оскільки

$$p(\hat{y} | y_1, \dots, y_n) = \int p(\hat{y} | \theta) p(\theta | y_1, \dots, y_n) d\theta,$$

то $p(\hat{y} | y_1, \dots, y_n)$ – це апостеріорне сподівання для $p(\hat{y} | \theta)$. Для того щоб знайти апостеріорну прогнозну ймовірність того, що $\hat{Y} = \hat{y}$ (тобто \hat{Y}



дорівнює деякому конкретному значенню \hat{y}), можна скористатись методом Монте-Карло таким чином:

- згенерувати послідовність $\{\theta^{(1)}, \dots, \theta^{(M)}\} \sim i.i.d. p(\theta | y_1, \dots, y_n)$, тобто за апостеріорним розподілом;
- апроксимувати $p(\hat{y} | y_1, \dots, y_n) = \sum_{k=1}^M p(\hat{y} | \theta^{(k)}) / M$.

Ця процедура буде працювати задовільно, якщо $p(y | \theta)$ – дискретний розподіл і нас цікавлять величини, які можна легко обчислити з $p(y | \theta)$.

Однак, у загальному випадку буде корисно отримати (згенерувати) множину значень \hat{Y} з його апостеріорного прогнозного розподілу. Це можна зробити таким чином:

- згенерувати $\theta^{(1)} \sim p(\theta | y_1, \dots, y_n)$, згенерувати $\hat{y}^{(1)} \sim p(\hat{y} | \theta^{(1)})$,
 згенерувати $\theta^{(2)} \sim p(\theta | y_1, \dots, y_n)$, згенерувати $\hat{y}^{(2)} \sim p(\hat{y} | \theta^{(2)})$,
 \vdots
 згенерувати $\theta^{(M)} \sim p(\theta | y_1, \dots, y_n)$, згенерувати $\hat{y}^{(M)} \sim p(\hat{y} | \theta^{(M)})$.

Послідовність $\{(\theta, \hat{y})^{(1)}, \dots, (\theta, \hat{y})^{(M)}\}$ – це M незалежних значень із спільного апостеріорного розподілу (θ, \hat{Y}) , а послідовність $\{\hat{y}^{(1)}, \dots, \hat{y}^{(M)}\}$ – це M незалежних значень із маргінального апостеріорного розподілу для \hat{Y} , який являє собою апостеріорний прогнозний розподіл [77, 78].

Приклад 5.1. Генерування значень з розподілу Пуассона

Можна показати, що 40-літні жінки без ступеня бакалавра будуть мати (у середньому) більше дітей ніж 40-літні жінки із ступенем з ймовірністю 0,48. Це потребує обчислення нескінченної подвійної суми такого виду:

$$\begin{aligned} & \Pr(\hat{Y}_1 > \hat{Y}_2 | \sum Y_{i,1} = 217, \sum Y_{i,2} = 66) = \\ & = \sum_{\hat{y}_2=0}^{\infty} \sum_{\hat{y}_1=\hat{y}_2+1}^{\infty} dnbinom(\hat{y}_1, 219, 112) \cdot dnbinom(\hat{y}_2, 68, 45). \end{aligned}$$

Альтернативним методом обчислення цієї ймовірності є апроксимація за методом Монте-Карло. Оскільки величини \hat{Y}_1 і \hat{Y}_2 є апостеріорно незалежними, то генерування псевдовипадкових величин можна виконувати з кожного індивідуального розподілу незалежно.

Апостеріорні прогнози значення із спряженої моделі Пуассона можна згенерувати таким чином:

- згенерувати $\theta^{(1)} \sim \text{gamma}(a + \sum y_i, b+n)$ і $\hat{y}^{(1)} \sim \text{poisson}(\theta^{(1)})$,
- згенерувати $\theta^{(2)} \sim \text{gamma}(a + \sum y_i, b+n)$ і $\hat{y}^{(2)} \sim \text{poisson}(\theta^{(2)})$,
- ⋮
- згенерувати $\theta^{(M)} \sim \text{gamma}(a + \sum y_i, b+n)$ і $\hat{y}^{(M)} \sim \text{poisson}(\theta^{(M)})$.

Операція генерування досить просто виконується за допомогою кількох команд у системі **R** [37]:

```
> a<-2 ; b<-1
> sy1<-217 ; n1<-111
> sy2<-66 ; n2<-44

• theta1.mc<-rgamma(10000, a+sy1, b+n1)
• theta2.mc<-rgamma(10000, a+sy2, b+n2)
• y1.mc<-rpois(10000, theta1.mc)
• y2.mc<-rpois(10000, theta2.mc)

• mean(y1.mc>y2.mc)
[1] 0.4823
```

Після генерування цих величин з апостеріорних прогнозних розподілів їх можна використати для обчислення інших величин, які необхідні для виконання аналізу. Наприклад, можна знайти апроксимацію розподілу різниці кількості дітей у двох індивідуумів, взятих по одному з кожної групи: $D = (\hat{Y}_1 - \hat{Y}_2)$.

5.4 Генерування неперервних випадкових величин

Розглянемо задачу генерування випадкових (псевдовипадкових) величин з розподілу заданого типу. Така задача постійно виникає при використанні методів імітаційного моделювання, прогнозування та керування. Усереднення коректно згенерованих значень дає можливість знайти невідомі параметри апостеріорного розподілу та використати їх для підтримки прийняття рішень [79].

5.4.1 Рівномірний розподіл

Для функції розподілу випадкової величини $U(a, b)$ можна легко знайти обернену шляхом розв'язання рівняння $u = F(x)$ з метою обчислення x для рівномірно розподіленого $0 \leq u \leq 1$:

$$x = F^{-1}(u) = a + (b-a)u.$$



Таким чином, процедура генерування рівномірно розподіленого випадкового числа складається з двох таких кроків [29]:

Крок 1: Згенерувати $\{u\} \sim U(0, 1)$ – нормально розподілені числа в інтервалі $[0, 1]$.

Крок 2: Обчислити рівномірно розподілені числа $U(a, b)$ в інтервалі $[a, b]$: $X = a + (b - a)U$.

5.4.2 Експоненціальний розподіл

Експоненційно розподілена випадкова величина із середнім $\beta > 0$ генерується за таким алгоритмом зворотного перетворення:

Крок 1: Згенерувати $\{u\} \sim U(0, 1)$ – нормально розподілене число в інтервалі $[0, 1]$.

Крок 2: Обчислити експоненційно розподілені числа за виразом: $X = -\beta \ln(U)$.

5.4.3 Гамма розподіл

При заданому $X \sim \text{gamma}(\alpha, \beta)$ для будь-якого $\beta > 0$ можна обчислити випадкову величину X' за виразом: $X' = \beta X$. При цьому обмежується генерування величин з розподілу $\text{gamma}(\alpha, 1)$.

Більш загальний алгоритм генерування складається з трьох кроків, поданих нижче (при цьому величину $b = (e + \alpha)/e$ необхідно визначити наперед).

Крок 1: Згенерувати $U_1 \sim U(0, 1)$ і обчислити $P = bU_1$. Якщо $P > 1$, то переходимо на крок 3. Інакше, переходимо на крок 2.

Крок 2: Обчислюємо $Y = P^{1/\alpha}$ і генеруємо $U_2 \sim U(0, 1)$. Якщо $U_2 \leq e^{-Y}$, то покладаємо $X = Y$. Інакше повертаємось на крок 1.

Крок 3: Обчислюємо $Y = -\ln[(b-P)/\alpha]$ і генеруємо $U_2 \sim U(0, 1)$. Якщо $U_2 \leq Y^{\alpha-1}$, то покладаємо $X = Y$. Інакше повертаємось на крок 1.

5.5 Модель у формі нормального розподілу

Випадкова змінна Y має нормальний розподіл з параметром θ і дисперсією $0 < \sigma^2 \leq c$ (обмежена сталою c) якщо її функція щільності має такий вигляд:

$$p(y|\theta, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y-\theta}{\sigma}\right)^2\right], \quad -\infty < y < \infty.$$

Деякі важливі властивості нормального розподілу:

- розподіл симетричний стосовно моди, медіани і середнього – всі ці величини дорівнюють $\bar{\theta}$;
- біля 97,5% значень розподілу знаходяться в інтервалі $\pm 1,96 \sigma$; а 99,9% значень нормальної вибірки знаходяться в інтервалі $\pm 3,09 \sigma$; таблиця квантилів стандартного нормального розподілу така:

Ймовірність	0,500	0,8413	0,900	0,950	97,50	0,9772	0,9900	0,9990	0,9999
Значення	0,000	1,000	1,282	1,645	1,960	2,000	2,3260	3,0900	3,7150

- якщо $X \sim normal(\mu, \tau^2)$, $Y \sim normal(\theta, \sigma^2)$ і X та Y – незалежні, то $aX + bY \sim normal(a\mu + b\theta, a^2\tau^2 + b^2\sigma^2)$;
- команди в **R**: *dnorm*, *rnorm* і *qnorm* використовують як аргумент σ , а не σ^2 .

Важливість нормального розподілу впливає, перш за все, із центральної граничної теореми (ЦГТ), яка стверджує, що при дуже загальних умовах сума (або їх середнє) множини випадкових змінних наближено має нормальний розподіл. На практиці це означає, що модель для генерування нормально розподілених чисел буде прийнятною для отримання даних, які формуються під впливом множини (великої кількості) випадкових факторів [86-88].

Приклад 5.2. Дослідження зросту жінок

У ході дослідження 1100 сімей в Англії за період часу з 1893-го по 1898-й рік зібрано дані стосовно зросту $n = 1375$ жінок віком вище 18 років. Вибіркове середнє цих даних склало $\bar{y} = 63,75$ дюйма або $\bar{y} = 63,75 \cdot 2,54 \approx 162,0$ см; а вибіркове стандартне відхилення $s = 2,62$ дюйма, або $s = 2,62 \cdot 2,54 \approx 6,65$ см. Одним з пояснень варіабельності зросту жінок є те, що вони гетерогенні в смислі множини факторів, які впливають на зріст людини: *генетичні фактори*, *дієта* (тип продуктів харчування), *хвороби*, *стреси* і т. ін. Варіабельність цих факторів спричиняє



варіабельність величини зросту жінок. Так, зріст жінки y_i можна формально описати адитивною лінійною моделлю, яка враховує згадані фактори:

$$\begin{aligned}y_1 &= a + b \cdot gene_1 + c \cdot diet_1 + d \cdot disease_1 + \dots, \\y_2 &= a + b \cdot gene_2 + c \cdot diet_2 + d \cdot disease_2 + \dots, \\&\vdots \\y_n &= a + b \cdot gene_n + c \cdot diet_n + d \cdot disease_n + \dots,\end{aligned}$$

де $gene_i$ – може означати наявність спеціального гена, який сприяє збільшенню зросту; $diet_i$ – може служити виміром деякого аспекту харчування жінок; $disease_i$ – може вказувати на те, чи хворіла жінка конкретною хворобою.

Загалом може бути багато факторів, які впливають на зріст людини. Якщо ці фактори мають наближено адитивний вплив, то кожний вимір зросту y_i дорівнює приблизно лінійній комбінації досить великої кількості членів. Для таких ситуацій центральна гранична теорема формулюється таким чином: емпіричний розподіл значень зросту $Y = \{y_1, y_2, \dots, y_n\}$ буде наближатись до нормального із збільшенням кількості вимірів. Таким чином, нормальний розподіл можна вибрати як належну модель (апостеріорний розподіл) для генерування даних.

5.6 Висновок щодо середнього і дисперсії нормального розподілу

Нехай модель даних має вигляд:

$$Y = \{y_1, y_2, \dots, y_n \mid \theta, \sigma^2\} \sim i.i.d. normal(\theta, \sigma^2).$$

Тепер спільна щільність (правдоподібність) для цих даних буде такою:

$$\begin{aligned}p(y_1, y_2, \dots, y_n \mid \theta, \sigma^2) &= \prod_{i=1}^n p(y_i \mid \theta, \sigma^2) = \\&= \prod_{i=1}^n \frac{1}{\sigma \sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{y_i - \theta}{\sigma}\right)^2\right] = \\&= (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2} \sum_{i=1}^n \left(\frac{y_i - \theta}{\sigma}\right)^2\right].\end{aligned}$$

Розкладаючи квадратичний член під експонентою, знайдемо, що $p(y_1, y_2, \dots, y_n \mid \theta, \sigma^2)$ залежить від даних $Y = \{y_1, y_2, \dots, y_n\}$ через

$$\sum_{i=1}^n \left(\frac{y_i - \theta}{\sigma} \right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^n y_i^2 - 2 \frac{\theta}{\sigma^2} \sum_{i=1}^n y_i + n \frac{\theta^2}{\sigma^2}.$$

На основі цього виразу можна стверджувати, що члени $\{\sum y_i^2, \sum y_i\}$ утворюють двовимірну достатню статистику. Знання цих двох величин еквівалентно знанню таких значень:

$$\bar{y} = \frac{1}{n} \sum y_i, \quad S^2 = \frac{1}{n-1} \sum (y_i - \bar{y})^2,$$

а тому $\{\bar{y}, S^2\}$ також утворюють достатню статистику.

Формування висновку для цієї моделі з двома параметрами можна розділити на дві частини, формування висновку для кожного параметра окремо.

Почнемо з формування висновку стосовно θ за умови відомої дисперсії σ^2 і скористаємось спряженим апіорним розподілом для θ . Для будь-якого (умовного) апіорного розподілу $p(\theta | \sigma^2)$ апостеріорний розподіл буде задовольняти виразу:

$$\begin{aligned} p(\theta | y_1, y_2, \dots, y_n, \sigma^2) &\propto p(\theta | \sigma^2) \exp\left(-\frac{1}{2\sigma^2} \sum (y_i - \theta)^2\right) \propto \\ &\propto p(\theta | \sigma^2) \exp(c_1 (\theta - c_2)^2). \end{aligned}$$

Згадаємо, що клас апіорних розподілів буде спряженим стосовно моделі для вибіркового даного якщо результуючий апостеріорний розподіл знаходиться у тому ж класі.

З наведених вище формулювань видно таке: для того щоб розподіл $p(\theta | \sigma^2)$ був спряженим, він повинен містити квадратичні члени типу $\exp(c_1 (\theta - c_2)^2)$.

Самим простим класом розподілів ймовірностей у просторі дійсних чисел \mathfrak{X} є сімейство нормальних розподілів. Якщо апіорний розподіл $p(\theta | \sigma^2)$ нормальний і дані

$$\{y_1, y_2, \dots, y_n | \theta, \sigma^2\} \sim i.i.d. \text{ normal}(\theta, \sigma^2),$$

то апостеріорний розподіл також буде нормальним. Тобто якщо $\theta \sim \text{normal}(\mu_0, \tau_0^2)$, то



$$\begin{aligned}
 p(\theta | y_1, y_2, \dots, y_n, \sigma^2) &= p(\theta | \sigma^2) \frac{p(y_1, y_2, \dots, y_n | \theta, \sigma^2)}{p(y_1, y_2, \dots, y_n | \sigma^2)} \propto \\
 &\propto p(\theta | \sigma^2) p(y_1, y_2, \dots, y_n | \theta, \sigma^2) \propto \\
 &\propto \exp\left\{-\frac{1}{2\tau_0^2}(\theta - \mu_0)^2\right\} \cdot \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \theta)^2\right\}.
 \end{aligned}$$

Якщо додати члени під експонентою і тимчасово не брати до уваги коефіцієнт $(-1/2)$, то отримаємо такий вираз для апостеріорного розподілу:

$$\frac{1}{\tau_0^2} (\theta^2 - 2\theta\mu_0 + \mu_0^2) + \frac{1}{\sigma^2} (\sum y_i^2 - 2\theta \sum y_i + n\theta^2) = a\theta^2 - 2b\theta + c,$$

де $a = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}$, $b = \frac{\mu_0}{\tau_0^2} + \frac{\sum y_i}{\sigma^2}$, $c = c(\mu_0, \tau_0^2, \sigma^2, y_1, \dots, y_n)$.

Тепер визначимо форму розподілу $p(\theta | y_1, y_2, \dots, y_n, \sigma^2)$:

$$\begin{aligned}
 p(\theta | y_1, y_2, \dots, y_n, \sigma^2) &\propto \exp\left\{-\frac{1}{2}(a\theta^2 - 2b\theta)\right\} = \\
 &\propto \exp\left\{-\frac{1}{2}a\left(\theta^2 - \frac{2b\theta}{a} + \frac{b^2}{a^2}\right) + \frac{b^2}{2a}\right\} \propto \\
 &\propto \exp\left\{-\frac{1}{2}a\left(\theta - \frac{b}{a}\right)^2\right\} = \\
 &= \exp\left\{-\frac{1}{2}\left(\frac{\theta - b/a}{1/\sqrt{a}}\right)^2\right\}.
 \end{aligned}$$

Ця функція має таку ж форму як і нормальний розподіл, у якому $1/\sqrt{a}$ грає роль стандартного відхилення, а b/a – середнє. Оскільки розподіли ймовірностей визначаються за своєю формою, це означає, що $p(\theta | y_1, y_2, \dots, y_n, \sigma^2)$ дійсно відноситься до нормальних розподілів.

Позначимо середнє і дисперсію цього розподілу через μ_n і τ_n^2 , де

$$\mu_n = \frac{b}{a} = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}, \quad \tau_n^2 = \frac{1}{a} = \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}.$$

Комбінування інформації

Апостеріорні (умовні) параметри μ_n і τ_n^2 визначаються комбінацією апіорних параметрів μ_0 і τ_0^2 з членами (розподілу), які характеризують дані.

Апостеріорна дисперсія і точність: вираз для $1/\tau_n^2$ має вигляд:

$$\frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2},$$

Тобто апіорна *обернена* дисперсія комбінується з оберненою дисперсією даних. Обернену дисперсію ще називають *точністю*. Для моделі у формі нормального розподілу введемо позначення:

$\tilde{\sigma}^2 = 1/\sigma^2$ – точність генерування даних з розподілу, тобто наскільки близькими є значення y_i до параметра розподілу θ ;

$\tilde{\tau}_0^2 = 1/\tau_0^2$ – точність апіорної дисперсії параметра;

$\tilde{\tau}_n^2 = 1/\tau_n^2$ – точність апостеріорної дисперсії параметра.

Зручно представляти точність як кількість інформації за адитивною шкалою. Для нормальної моделі $\frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}$ загальна точність має

вигляд:

$$\tilde{\tau}_n^2 = \tilde{\tau}_0^2 + n \tilde{\sigma}^2,$$

тобто

апостеріорна інформація = апіорна інформація + інформація від даних.

Апостеріорне середнє: зазначимо, що

$$\mu_n = \frac{\tilde{\tau}_0^2}{\tilde{\tau}_0^2 + n \tilde{\sigma}^2} \mu_0 + \frac{n \tilde{\sigma}^2}{\tilde{\tau}_0^2 + n \tilde{\sigma}^2} \bar{y},$$

тобто апостеріорне середнє – це зважене середнє апіорного середнього і вибіркового середнього. Ваговим коефіцієнтом для вибіркового середнього є n/σ^2 , тобто вибіркова точність вибіркового середнього. Ваговим



коефіцієнтом для апіорного середнього є $1/\tau_0^2$, тобто точність для апіорного значення.

Якби апіорне середнє ґрунтувалось на κ_0 апіорних спостереженнях з тієї ж (або подібної) вибірки даних Y_1, \dots, Y_n , тоді можна було б покласти $\tau_0^2 = \sigma^2 / \kappa_0$, для дисперсії *середнього* апіорних спостережень. У такому випадку формула для апостеріорного середнього спрощується і приймає такий вигляд:

$$\mu_n = \frac{\kappa_0}{\kappa_0 + n} \mu_0 + \frac{n}{\kappa_0 + n} \bar{y}.$$

Прогнозування при використанні нормального розподілу

Розглянемо задачу прогнозування подальшого розвитку процесу \hat{Y} на основі наявної вибірки даних (y_1, \dots, y_n) . Для того щоб знайти прогнозний розподіл скористаємось такою наявною інформацією:

$$\{\hat{Y} | \theta, \sigma\} \sim normal(\theta, \sigma^2) \Leftrightarrow \hat{Y} = \theta + \hat{\epsilon}, \quad \{\hat{\epsilon} | \theta, \sigma\} \sim normal(0, \sigma^2).$$

Іншими словами, той факт, що \hat{Y} має нормальний розподіл із середнім θ – це те ж саме, що \hat{Y} дорівнює θ плюс нормально розподілена випадкова шумова складова з нульовим середнім і дисперсією σ^2 . Ґрунтуючись на такому поданні \hat{Y} , знайдемо спочатку апостеріорне середнє і дисперсію \hat{Y} :

$$\begin{aligned} E[\hat{Y} | y_1, \dots, y_n, \sigma^2] &= E[\theta + \hat{\epsilon} | y_1, \dots, y_n, \sigma^2] = \\ &= E[\theta | y_1, \dots, y_n, \sigma^2] + E[\hat{\epsilon} | y_1, \dots, y_n, \sigma^2] = \\ &= \mu_n + 0 = \mu_n. \end{aligned}$$

$$\begin{aligned} Var[\hat{Y} | y_1, \dots, y_n, \sigma^2] &= Var[\theta + \hat{\epsilon} | y_1, \dots, y_n, \sigma^2] = \\ &= Var[\theta | y_1, \dots, y_n, \sigma^2] + Var[\hat{\epsilon} | y_1, \dots, y_n, \sigma^2] = \\ &= \tau_n + \sigma^2. \end{aligned}$$

Згідно з ЦГТ сума незалежних нормально розподілених змінних також має нормальний розподіл. Оскільки обидві величини, θ і $\hat{\epsilon}$, визначаються на основі вибірки даних (y_1, \dots, y_n) і σ^2 , і мають нормальний розподіл, то $\hat{Y} = \theta + \hat{\epsilon}$ також має нормальний розподіл. Таким чином, прогнозний розподіл визначається так:

$$(\hat{Y} | \sigma^2, y_1, \dots, y_n) \sim \text{normal}(\mu_n, \tau_n^2 + \sigma^2).$$

Щодо форми дисперсії \hat{Y} необхідно зазначити: загалом невизначеність щодо нового (наступного) значення \hat{Y} є функцією невизначеності стосовно центру вибірки (τ_n^2), а також наскільки варіабельною є вибірка даних (σ^2).

Якщо $n \rightarrow \infty$, то невизначеність стосовно місцезнаходження центру популяції θ зменшується і апостеріорна дисперсія τ_n^2 параметра θ прямує до нуля. Однак, визначеність щодо θ не приводить до зменшення варіабельності вибірки σ^2 , а тому невизначеність щодо \hat{Y} ніколи не зменшується нижче σ^2 .

Приклад 5.3. Аналіз довжини крил мушки (midge)

Розглянемо дев'ять вимірів довжини крила мошки (комара-дергуна). На основі 9-ти вимірів розмаху крила необхідно зробити висновок щодо середнього значення розмаху крила θ цієї популяції мошок. Вивчення інших типів популяцій мошок свідчить, що довжина крила складає у типовому випадку 1,9 мм, а тому покладемо $\mu_0 = 1,9$. Також очевидно, що довжина крила – додатне число, тобто $\theta > 0$.

Таким чином, в ідеалі апіорний розподіл для θ передбачає, що $\theta > 0$ (тільки за такої умови з'являється поняття маси). Це обмеження можна врахувати за допомогою нормального апіорного розподілу таким чином: оскільки для будь-якого нормального розподілу максимальна площа під кривою знаходиться в інтервалі, що визначається двома стандартними відхиленнями, то виберемо τ_0^2 таким, що $\mu_0 - 2\tau_0 > 0$ або (що еквівалентно) $2\tau_0 < \mu_0$ або $\tau_0 < 1,9/2 = 0,95$. Прийнемо $\tau_0 = 0,95$, що у деякій мірі перевищує апіорну невизначеність щодо θ .

Отримано такі спостереження за розміром крил:

1,64; 1,70; 1,72; 1,74; 1,82; 1,82; 1,82; 1,90; 2,08

із вибірковою середнім $\bar{y} = 1,804$. Використовуючи наведені вище формули для μ_n і τ_n^2 , отримаємо

$$\{\theta | y_1, \dots, y_9, \sigma^2\} \sim \text{normal}(\mu_n, \tau_n^2),$$

де



$$\mu_n = \frac{\frac{1}{\tau_0^2} \mu_0 + \frac{n}{\sigma^2} \bar{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} = \frac{1,11 \cdot 1,9 + \frac{9}{\sigma^2} 1,804}{1,11 + \frac{9}{\sigma^2}};$$
$$\tau_n^2 = \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} = \frac{1}{1,11 + \frac{9}{\sigma^2}}.$$

Якщо $\sigma^2 = S^2 = 0,017$ (де S^2 – вибіркова дисперсія, яка розраховується за 9-ма спостереженнями), то

$$\{\theta | y_1, \dots, y_9, \sigma^2 = 0,017\} \sim normal(1,805; 0,002).$$

Обчислений 95%-й довірчий інтервал для θ , який ґрунтується на цьому розподілі, має такі границі: (1,72; 1,89). Однак, цей довірчий інтервал передбачає, що $\sigma^2 = S^2$, де S^2 – наближена оцінка вибіркової дисперсії, яка розраховується за 9-ма спостереженнями. Тобто точніше представлення наявної інформації вимагає врахування того факту, що фактична дисперсія σ^2 невідома.

5.7 Спільний висновок щодо середнього і дисперсії нормального розподілу

Концептуально байєсівський висновок для двох або більше невідомих параметрів не відрізняється від випадку, коли аналізується один параметр [76]. Для будь-якого спільного апіорного розподілу $p(\theta, \sigma^2)$, для параметрів θ і σ^2 , апостеріорний висновок можна знайти за правилом Байєса таким чином:

$$p(\theta, \sigma^2 | y_1, \dots, y_n) = \frac{p(y_1, \dots, y_n | \theta, \sigma^2) p(\theta, \sigma^2)}{p(y_1, \dots, y_n)}.$$

Так само як і раніше, почнемо аналіз задачі з вибору простого спряженого класу апіорних розподілів, який забезпечить нескладне отримання апостеріорного розподілу.

Скористаємось аксіомою теорії ймовірностей, яка дає можливість виразити спільний розподіл двох величин у вигляді добутку умовної та маргінальної ймовірностей:

$$p(\theta, \sigma^2) = p(\theta | \sigma^2) p(\sigma^2).$$

У попередньому параграфі показано: якщо σ^2 відома, то спряженим апіорним розподілом для θ буде нормальний з параметрами (μ_0, τ_0^2) . Розглянемо окремий випадок, коли $\tau_0^2 = \sigma^2 / \kappa_0$:

$$p(\theta, \sigma^2) = p(\theta | \sigma^2) p(\sigma^2) = \text{dnorm}(\theta, \mu_0, \tau_0 = \sigma / \sqrt{\kappa_0}) p(\sigma^2).$$

У такому випадку параметри μ_0 і κ_0 можна інтерпретувати як середнє і розмір вибірки з множини апіорних спостережень.

Для опису σ^2 необхідно вибрати сімейство апіорних розподілів, опорною множиною значень якого є інтервал $(0, \infty)$. Одним із таких сімейств є сімейство гамма розподілів, яким можна скористатись так, як це вже було продемонстровано на прикладі моделі Пуассона для даних. Однак, це сімейство не є спряженим для нормально розподіленої дисперсії. Але, разом з тим, сімейство гамма розподілів виявляється спряженим класом для значень $1/\sigma^2$ (точність).

Коли використовується такий апіорний розподіл, то говорять, що σ^2 має *обернений гамма* розподіл:

$$\begin{aligned} \text{точність} &= 1/\sigma^2 \sim \text{gamma}(a, b), \\ \text{дисперсія} &= \sigma^2 \sim \text{inverse-gamma}(a, b). \end{aligned}$$

Тепер замість параметрів a і b скористаємось іншою параметризацією (розглянемо інтерпретацію на пізніше) цього апіорного розподілу такого вигляду:

$$1/\sigma^2 \sim \text{gamma}\left(\frac{v_0}{2}, \frac{v_0}{2} \sigma_0^2\right).$$

При використанні цієї параметризації описові статистики мають такий вигляд:

- $E[\sigma^2] = \sigma_0^2 \frac{v_0/2}{v_0/2 - 1}$;
- $\text{mode}[\sigma^2] = \sigma_0^2 \frac{v_0/2}{v_0/2 + 1}$, тобто $\text{mode}[\sigma^2] < \sigma_0^2 < E[\sigma^2]$;
- $\text{Var}[\sigma^2]$ зменшується при збільшенні значень v_0 .

Апіорні параметри (σ_0^2, v_0) можна інтерпретувати (як буде показано нижче) як вибірку дисперсію і величину вибірки апіорних спостережень.



Апостеріорний висновок

Припустимо, що апіорні розподіли і модель даних мають такий вигляд:

$$\begin{aligned}1/\sigma^2 &\sim \text{gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0}{2} \sigma_0^2\right), \\ \theta/\sigma^2 &\sim \text{normal}(\mu_0, \sigma^2/\kappa_0), \\ \{Y_1, \dots, Y_n | \theta, \sigma^2\} &\sim \text{i.i.d. normal}(\theta, \sigma^2).\end{aligned}$$

Оскільки апіорний розподіл для θ і σ^2 можна записати (розкласти) таким чином

$$p(\theta, \sigma^2) = p(\theta | \sigma^2) p(\sigma^2),$$

то апостеріорний розподіл також можна подати у вигляді розкладу:

$$p(\theta, \sigma^2 | y_1, \dots, y_n) = p(\theta | y_1, \dots, y_n, \sigma^2) \cdot p(\sigma^2 | y_1, \dots, y_n).$$

Умовний розподіл для θ (за умови наявних даних і параметра σ^2) можна отримати на основі результатів, отриманих у попередньому підрозділі:

$$\{\theta | y_1, \dots, y_n, \sigma^2\} \sim \text{normal}(\mu_n, \sigma^2/\kappa_n),$$

$$\text{де } \kappa_n = \kappa_0 + n; \quad \mu_n = \frac{(\kappa_0/\sigma^2) \mu_0 + (n/\sigma^2) \bar{y}}{\kappa_0/\sigma^2 + n/\sigma^2} = \frac{\kappa_0 \mu_0 + n \bar{y}}{\kappa_n}.$$

А тому, якщо μ_0 – це середнє для κ_0 апіорних спостережень, то $E[\theta | y_1, \dots, y_n, \sigma^2]$ – це вибіркове середнє для поточних та апіорних спостережень, а $Var[\theta | y_1, \dots, y_n, \sigma^2]$ – це σ^2 , поділене на загальну кількість спостережень (апіорних та поточних).

Апостеріорний розподіл для σ^2 можна отримати шляхом інтегрування по невідомому значенні θ :

$$\begin{aligned}p(\sigma^2 | y_1, \dots, y_n) &= p(\sigma^2) \cdot p(y_1, \dots, y_n | \sigma^2) = \\ &= p(\sigma^2) \int p(y_1, \dots, y_n | \theta, \sigma^2) p(\theta | \sigma^2) d\theta.\end{aligned}$$

Цей інтеграл можна взяти за допомогою методу з математичного аналізу [117]. В результаті отримаємо такий вираз:

$$\{1/\sigma^2 | y_1, \dots, y_n, \sigma^2\} \sim \text{gamma}(\nu_n/2, \nu_n \sigma_n^2/2),$$

де $v_n = v_0 + n$;

$$\sigma_n^2 = \frac{1}{v_n} \left[v_0 \sigma_0^2 + (n-1) S^2 + \frac{\kappa_0 n}{\kappa_n} (\bar{y} - \mu_0)^2 \right].$$

Тут v_0 інтерпретується як розмір апіорної вибірки, яка має апіорну дисперсію σ_0^2 , а вибіркова дисперсія $S^2 = \sum_{k=1}^n \frac{[y(k) - \bar{y}]^2}{n-1}$; $(n-1) S^2$ – сума квадратів відхилень від середнього (або просто сума квадратів). За аналогією $v_0 \sigma_0^2$ і $v_n \sigma_n^2$ – апіорна та апостеріорна суми квадратів відповідно.

Якщо помножити останнє рівняння зліва і справа на v_n , то отримаємо майже таке: «апостеріорна сума квадратів відхилень дорівнює апіорній сумі квадратів відхилень плюс сума квадратів для даних». Однак третій член у цьому рівнянні, тобто $\frac{\kappa_0 n}{\kappa_n} (\bar{y} - \mu_0)^2$, інтерпретується інакше: він говорить про те, що велике значення $(\bar{y} - \mu_0)^2$ збільшує апостеріорну ймовірність появи великих значень σ^2 .

Це має смисл для конкретного спільного апіорного розподілу для θ і σ^2 : якщо μ_0 – вибіркове середнє для κ_0 апіорних спостережень з дисперсією σ^2 , то $\frac{\kappa_0 n}{\kappa_0 + n} (\bar{y} - \mu_0)^2$ – це оцінка для σ^2 і ми використовуємо інформацію, яку дає цей член. У тих випадках, коли μ_0 не може розглядатись як середнє апіорних спостережень, буде запропоновано альтернативний апіорний розподіл (у наступному розділі).

Приклад 5.4. Обчислення параметрів розподілу

Повернемось до прикладу з мошками: згідно з наявними даними вимірів істинне середнє і стандартне відхилення не можуть суттєво відрізнятись від значень 1,9 і 0,1 мм, відповідно. Таким чином, припускаємо, що $\mu_0 = 1,9$, а $\sigma_0^2 = 0,01$. Однак, наявні дані можуть відрізнятись від інших можливих вимірів стосовно довжини крила, а тому покладемо: $\kappa_0 = v_0 = 1$. Це свідчить про те, що вибрані апіорні розподіли відносно слабо центруються (групуються) навколо цих оцінок.



Вибіркове середнє і дисперсія наявних даних складають $\bar{y} = 1,804$ і $S^2 = 0,0169$ ($S = 0,130$). Використовуючи ці значення і апіорні параметри, обчислимо μ_n і σ_n^2 :

$$\mu_n = \frac{\kappa_0 \mu_0 + n \bar{y}}{\kappa_n} = \frac{1,9 + 9 \cdot 1,804}{1 + 9} = 1,814,$$
$$\sigma_n^2 = \frac{1}{v_n} \left[v_0 \sigma_0^2 + (n-1) S^2 + \frac{\kappa_0 n}{\kappa_n} (\bar{y} - \mu_0)^2 \right] =$$
$$= \frac{0,010 + 0,135 + 0,008}{10} = 0,015.$$

Спільний апостеріорний розподіл параметрів повністю визначається такими параметрами: $\mu_n = 1,814$; $\kappa_n = 10$; $\sigma_n^2 = 0,015$; $v_n = 10$, і може буде заданий виразами:

$$\{\theta | y_1, \dots, y_n, \sigma^2\} \sim \text{normal}(1,814, \sigma^2/10),$$
$$\{1/\sigma^2 | y_1, \dots, y_n\} \sim \text{gamma}(10/2; 10 \cdot 0.015/2)$$

Використання згенерованих значень за методом Монте-Карло

У багатьох випадках аналізу фактичних даних ставиться задача оцінювання середнього популяції θ , а тому необхідно обчислювати математичне сподівання $E[\theta | y_1, \dots, y_n]$, стандартне відхилення $SD[\theta | y_1, \dots, y_n]$, ймовірність $\Pr(\theta_1 < \theta_2 | y_{1,1}, \dots, y_{n,2})$ і т. ін. Ці величини визначаються за *маргінальним* апостеріорним розподілом параметра θ за умови наявних даних. Однак, все що відомо на даний момент – це те, що умовний розподіл для θ (за умови наявних даних і σ^2) є нормальним, а розподілом для σ^2 (за умови наявних даних) є обернений (інверсний) розподіл гамма. Якби можна було генерувати *маргінальні* значення θ з розподілу $p(\theta | y_1, \dots, y_n)$, то для оцінювання згаданих параметрів можна було б скористатись методом Монте-Карло [51].

Це робиться досить просто шляхом генерування величин θ і σ^2 з їх спільного апостеріорного розподілу:

$$\begin{aligned} \sigma^{2(1)} &\sim \text{inverse gamma} (v_n/2, \sigma_n^2 v_n/2), \quad \theta^{(1)} \sim \text{normal} (\mu_n, \sigma^{2(1)} / \kappa_n), \\ &\vdots \\ \sigma^{2(S)} &\sim \text{inverse gamma} (v_n/2, \sigma_n^2 v_n/2), \quad \theta^{(S)} \sim \text{normal} (\mu_n, \sigma^{2(S)} / \kappa_n). \end{aligned}$$

Зазначимо, що кожне значення $\theta^{(S)}$ генерується із свого умовного розподілу за умови наявних даних і $\sigma^2 = \sigma^{2(S)}$. Ця процедура Монте-Карло легко програмується на мові **R** [37].

5.8 Алгоритм Метрополіса-Хастінгса (Metropolis-Hastings)

Розглянемо узагальнений підхід на основі імітаційного моделювання до обчислення величин (моментів і квантилів апостеріорного та прогнозного розподілів, які виникають у байєсівському аналізі даних. Припустимо, що необхідно знайти середнє апостеріорного розподілу

$$E(\theta | \mathbf{y}) = \int \theta h(\theta | \mathbf{y}) d\theta,$$

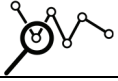
де θ – вектор параметрів; \mathbf{y} – вектор спостережень. Іноді такий інтеграл можна знайти аналітично, але у більшості випадків це неможливо. Тому частіше для обчислення таких інтегралів використовують метод Монте-Карло, який ґрунтується на генеруванні псевдовипадкових послідовностей із заданими параметрами та усередненні результатів [26, 48, 106, 107].

Загальна ідея оцінювання полягає у тому щоб правильно згенерувати псевдовипадкові числа із апостеріорної щільності $h(\theta | \mathbf{y})$. Якщо існує можливість згенерувати послідовність значень параметрів

$$\{\theta^{(1)}, \dots, \theta^{(M)}\} \sim h(\theta | \mathbf{y})$$

з апостеріорної щільності, то за умови досягнення достатньо великої вибірки можна обчислити не тільки згаданий вище інтеграл, але й інші характеристики апостеріорного розподілу шляхом формування відповідних оцінок. Наприклад, середнє згенерованої вибірки – це буде апостеріорне середнє, обчислене шляхом імітації процесу, а квантілі згенерованого розподілу – це оцінки квантилів апостеріорного розподілу. При дотриманні законів великих чисел, тобто коли потужності вибірок будуть достатньо великими, обчислені оцінки будуть збігатись до апостеріорних величин.

Якщо модель, що будується, надає можливість використовувати природну спряжену апіорну щільність, то апостеріорна щільність також буде належати до того ж сімейства, що і апіорна. Тобто буде відомо, за яким виразом генерувати псевдовипадкову послідовність. Якщо ж такий



підхід (з використанням спряженої щільності) неможливий, то генерування з апостеріорної щільності буде складнішою задачею.

Генерування складної апостеріорної щільності високої розмірності можна виконувати за допомогою методів Монте-Карло для марковських ланцюгів (МКМЛ). Цей клас методів ґрунтується на імітаційному моделюванні належним чином сформованих марковських ланцюгів, які збігаються до щільності бажаного типу (як правило, це апостеріорна щільність).

Основна властивість марковського ланцюга полягає у тому, що умовна щільність параметра $\theta^{(j)}$ (це j -й елемент послідовності) узгоджується із заданим розподілом, але значення цього елемента залежить тільки від попереднього значення $\theta^{(j-1)}$. Позначимо цю умовну перехідну щільність так: $T(\theta^{(j-1)}, \cdot | \mathbf{y})$.

Метод моделювання МКМЛ повинен забезпечити генерування такої перехідної щільності, яка буде збігатись до апостеріорної щільності, починаючи з початкової умови θ_0 . Збіжність виконується у тому смислі, що для будь-якої зліченної множини A з області h , розподіл $P(\theta^{(j)} \in A | \mathbf{y}, \theta_0)$ збігається до $\int_A h(\theta | \mathbf{y}) d\theta$ при $j \rightarrow \infty$.

Таким чином, при використанні такого підходу можна отримати псевдовипадкові числа, що відносяться до апостеріорної щільності, за такою рекурсивною схемою:

$$\begin{aligned}\theta^{(1)} &\sim T(\theta^{(0)}, \cdot | \mathbf{y}) \\ \theta^{(2)} &\sim T(\theta^{(1)}, \cdot | \mathbf{y}) \\ &\vdots \\ \theta^{(j)} &\sim T(\theta^{(j-1)}, \cdot | \mathbf{y}) \\ &\vdots\end{aligned}$$

Оскільки згенерований таким чином ланцюг повинен збігатись до апостеріорної щільності, то після деякої кількості n_0 ітерацій послідовність

$$\theta^{(n_0+1)}, \theta^{(n_0+2)}, \dots, \theta^{(n_0+M)}$$

можна розглядати як значення, що належать апостеріорному розподілу $h(\theta | \mathbf{y})$. Початкові n_0 значень ланцюга, як правило не використовуються, оскільки це перехідний процес з досить грубими величинами, неприйнятними для подальшого використання. Цей перехідний процес ще називають «періодом входу» («burn-in» period). Необхідно зазначити, що псевдовипадкові числа, згенеровані за процедурою марковського ланцюга,

– це корельована вибірка з апостеріорної щільності. Однак, ці числа можна використовувати для обчислення вибірових середніх та квантилів.

За допомогою закону великих чисел для марковських послідовностей можна показати, що обчислені таким чином оцінки будуть збігатись до апостеріорних величин при великій кількості згенерованих чисел. Тобто для будь-якої інтегрованої функції від θ , наприклад $g(\theta)$, вибірове середнє значень $\{g(\theta^{(j)})\}$ збігається, при виконанні слабких умов регулярності, до його математичного сподівання:

$$M^{-1} \sum_{j=1}^M g(\theta^{(j)}) \rightarrow \int_{\theta} g(\theta) h(\theta|\mathbf{y}) d\theta \quad \text{при } M \uparrow \infty. \quad (5.1)$$

Корисність генерування псевдовипадкових послідовностей типу марковських ланцюгів підтверджується тим фактом, що, як правило, можливо побудувати перехідну щільність $T(\theta^{(j-1)}, \cdot | \mathbf{y})$, яка збігається до цільової щільності [112].

Алгоритм Метрополіса-Хастінгса (Metropolis-Hastings)

Нехай необхідно генерувати послідовність з неперервної цільової щільності $h(\theta|\mathbf{y})$, де θ – вектор параметрів; $h(\theta|\mathbf{y})$ – неперервна щільність. Припустимо, що значення θ генеруються у формі одного блоку, а марковський ланцюг визначається перехідною щільністю $q(\theta', \theta|\mathbf{y})$, де (θ', θ) – два будь-яких значення із вибірки. Нехай вказана перехідна щільність задається без зв'язку із цільовою щільністю, а тому вона не збігається до цільової щільності [106, 107].

Наприклад, за перехідну щільність $q(\theta', \theta|\mathbf{y})$ можна прийняти багатовимірну нормальну щільність з вектором середніх θ' і коваріаційною (дисперсійною) матрицею \mathbf{V} , тобто $E[\theta'_i (\theta'_i)^T] = \mathbf{V}$. Таким чином, маючи на вході $q(\theta', \theta|\mathbf{y})$, необхідно побудувати марковський ланцюг, який збігається до $h(\theta|\mathbf{y})$. Це можна зробити за допомогою методу та алгоритму Метрополіса-Хастінгса (алгоритм М-Х). Ідея методу полягає у тому, щоб модифікувати розподіл $q(\theta', \theta|\mathbf{y})$ таким чином, щоб перехідна щільність модифікованого ланцюга «збігалась» до цільової щільності.

Для того щоб визначити послідовність операцій при виконанні алгоритму М-Н, задамо початкове значення ланцюга $\theta^{(0)}$ з подальшою метою знайти (згенерувати) послідовність значень: $\theta^{(0)}, \theta^{(1)}, \dots, \theta^{(j-1)}$.



Наступне значення ланцюга $\theta^{(j)}$ обчислюється за допомогою процедури, що складається з двох кроків, поданих нижче [103, 106].

Крок пропозиції (*proposal step*), тобто генерування кандидата. Кандидат на чергове значення ланцюга θ генерується з розподілу $q(\theta^{(j-1)}, \theta | \mathbf{y})$ (щільність пропозиції) і обчислюється така величина:

$$\alpha(\theta^{(j-1)}, \theta | \mathbf{y}) = \min \left\{ 1, \frac{h(\theta | \mathbf{y})}{h(\theta^{(j-1)} | \mathbf{y})} \frac{q(\theta, \theta^{(j-1)} | \mathbf{y})}{q(\theta^{(j-1)}, \theta | \mathbf{y})} \right\}. \quad (5.2)$$

Крок заповнення ланцюга (*move step*) наступним значенням. Прийняти наступне значення ланцюга на основі аналізу такої умови:

$$\theta^{(j)} = \begin{cases} \theta, & \text{з ймовірністю } \alpha(\theta^{(j-1)}, \theta | \mathbf{y}), \\ \theta^{(j-1)} & \text{з ймовірністю } 1 - \alpha(\theta^{(j-1)}, \theta | \mathbf{y}). \end{cases}$$

Необхідно зазначити, що $q(\theta', \theta | \mathbf{y})$ – це щільність, з якої генеруються дані з метою імітації марковського ланцюга, її називають щільністю для генерування кандидатів або щільністю, що генерує пропозиції для ланцюга (*proposal density*). Загальні способи його задавання будуть розглянуті нижче. Також зазначимо, що функцію $\alpha(\theta^{(j-1)}, \theta | \mathbf{y})$ у (5.2) можна обчислити без знання нормуючої константи апостеріорної щільності $h(\theta | \mathbf{y})$. Величину $\alpha(\theta^{(j-1)}, \theta | \mathbf{y})$ називають *ймовірністю прийняття* (*acceptance probability*) або *ймовірністю заповнення ланцюга* (*probability of move*).

Теоретичні властивості наведеного вище алгоритму залежать у значній мірі від природи (способу визначення) щільності, з якої генерується псевдовипадкова величина. Типовою вимогою до цієї генеруючої щільності є її додатна визначеність на опорній множині апостеріорної щільності. Звідси випливає, що ланцюг, згенерований за алгоритмом М-Х, може перейти у будь-яку точку (прийняти будь-яке значення) опорної множини за один крок.

На додаток до цього зазначимо, що внаслідок процедури рандомізації на кроці заповнення ланцюга (*move step*) перехідна щільність $T(\theta^{(j-1)}, \cdot | \mathbf{y})$ ланцюга, що генерується, має цікаву форму, оскільки вона є сумішшю неперервної щільності (для віддалення від $\theta^{(j-1)}$) і дискретної складової (для забезпечення деякого (ненульового) значення ймовірності перебування в околі $\theta^{(j-1)}$). Перехідна щільність має таку форму:

$$T(\theta^{(j-1)}, \cdot | \mathbf{y}) = q(\theta^{(j-1)}, \theta | \mathbf{y}) \alpha(\theta^{(j-1)}, \theta | \mathbf{y}) + r(\theta^{(j-1)} | \mathbf{y}) \delta_{\theta^{(j-1)}}, \quad (5.3)$$

де $r(\theta^{(j-1)} | \mathbf{y}) = \int_{\theta} q(\theta^{(j-1)}, \theta | \mathbf{y}) \alpha(\theta^{(j-1)}, \theta | \mathbf{y}) d\theta$;

$\delta_{\theta^{(j-1)}}$ – функція Дірака для $\theta^{(j-1)}$, яка визначається таким чином:

$$\delta_{\theta}(\theta') = \begin{cases} 0, & \text{якщо } \theta' \neq \theta, \\ 1, & \text{якщо } \int \delta_{\theta}(\theta') d\theta' = 1, \end{cases}$$

або ж наближено так:

$$\delta_{\theta}(\theta') = \begin{cases} 0, & \text{якщо } \theta' \neq \theta, \\ 1, & \text{якщо } \theta' = \theta. \end{cases}$$

Легко перевірити, що інтеграл перехідної щільності по всіх можливих значеннях θ дорівнює одиниці, як це і має бути. Зазначимо, що функції $T(\theta^{(j-1)}, \theta | \mathbf{y})$ і $r(\theta^{(j-1)} | \mathbf{y})$ не обчислюються у процесі реалізації алгоритму М-Х, поданого вище [26].

Внаслідок вибраного способу визначення оцінки $\theta^{(j)}$ та її подання в явному вигляді у формі перехідної щільності, за алгоритмом М-Х, значення $\theta^{(j-1)}$ може бути повторене у марковському ланцюгу, що генерується. Очевидно, що для забезпечення ефективного переміщення згенерованих значень по опорній множині цільової щільності ланцюг не повинен часто затримуватись в одній точці на багатьох ітераціях. Такої поведінки алгоритму можна уникнути завдяки належному вибору пропонууючої щільності $q(\theta^{(j-1)}, \theta | \mathbf{y})$.

Контроль збіжності марковського ланцюга

Для поточного контролю процесу оцінювання параметрів за алгоритмом М-Х можна скористатись критерієм, який отримав назву *період (час) автокореляції* або *фактор неефективності* (inefficiency factor), який розраховується для кожного скалярного параметра:

$$IF = 1 + 2 \sum_{s=1}^M \left(1 - \frac{s}{M}\right) \rho(s),$$

де $\rho(s)$ – вибіркова автокореляційна функція з лагами s , яка розраховується для M згенерованих елементів марковського ланцюга, що представляють собою еволюцію оцінки вибраного параметра:

$$\theta^{(n_0+1)}, \theta^{(n_0+2)}, \dots, \theta^{(n_0+M)}.$$



Критерій IF можна інтерпретувати як ефективний розмір вибірки (*effective sample size (ESS)*), що обчислюється таким чином:

$$ESS = \frac{M}{IF}. \quad (5.4)$$

Якщо згенеровані значення марковського ланцюга незалежні між собою, то період автокореляції теоретично дорівнює 1, а $ESS = M$. Коли фактор неефективності приймає великі значення, то ефективний розмір вибірки суттєво зменшується.

На практиці ефективний розмір вибірки означає скільки приблизно ітерацій алгоритму М-Х необхідно зробити (після відкидання початкових ітерацій (“burn-in” period), які ми не враховуємо у процесі оцінювання) для того щоб оцінка параметра перестала змінюватись. Тобто значення M залежить від розмірності вектора параметрів та коректності функціонування алгоритму. Кількість початкових ітерацій, які не враховуються у процесі оцінювання параметра, може складати від кількох тисяч до кількох десятків тисяч [107, 117].

5.9 Приклад застосування алгоритму М-Х (бінарні дані)

Розглянемо приклад даних, що характеризують інфікування (або відсутність інфікування) пацієнтів після складної операції (табл. 5.1).

Числа у першому стовпчику, наприклад 11/87, означають, що було виконано 98 операцій із значеннями незалежних змінних ($x_1 = 1, x_2 = 1, x_3 = 1$); після цих 98-ми операцій було 11 випадків інфікування пацієнта і 87 випадків без інфікування. Всього розглянуто 251 випадок.

Таблиця 5.1

Дані стосовно інфікування пацієнтів після операції

y (1/10)	x_1	x_2	x_3
11/87	1	1	1
1/17	0	1	1
0/2	0	0	1
23/3	1	1	0
28/30	0	1	0
0/9	1	0	0
8/32	0	0	0

Таким чином, основна (залежна) змінна $y(\cdot)$ є двійковою і для її формального опису скористаємось моделлю *пробіт*, тобто ймовірність інфікування i – го пацієнта буде розраховуватись за виразом:

$$\Pr(y_i = 1 | \mathbf{x}_i, \beta) = \Phi(\mathbf{x}_i^T \beta), \quad (5.5)$$

де $\mathbf{x}_i = (1, x_{i1}, x_{i2}, x_{i3})^T$ – вектор незалежних змінних; $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^T$ – вектор невідомих параметрів (коефіцієнтів) моделі; $\Phi(\cdot)$ – кумулятивна функція стандартного нормального розподілу випадкової змінної (або просто функція розподілу). Якщо прийняти, що результати інфікування, $\mathbf{y} = (y_1, y_2, \dots, y_{251})$, – це умовно незалежні події, то функцію правдоподібності можна записати у вигляді:

$$\begin{aligned} L(\mathbf{y} | \mathbf{X}, \beta) &= \prod_{i=1}^{251} \Phi(\mathbf{x}_i^T \beta)^{y_i} [1 - \Phi(\mathbf{x}_i^T \beta)]^{(1-y_i)} = \\ &= \prod_{i=1}^{251} \begin{cases} \Phi(\mathbf{x}_i^T \beta), & \text{якщо } y_i = 1, \\ 1 - \Phi(\mathbf{x}_i^T \beta), & \text{якщо } y_i = 0. \end{cases} \end{aligned}$$

Ця функція правдоподібності не передбачає прийняття природної спряженої апіорної щільності. Припустимо, що апіорна інформація стосовно вектора параметрів β може бути подача багатовимірною нормальною щільністю з нульовим середнім для кожного параметра, а дисперсія задається величиною $5\mathbf{I}_4$, де \mathbf{I}_4 – одинична матриця розмірністю 4×4 . Тобто апіорна щільність вектора параметрів β описується виразом:

$$g(\beta) \propto \exp[-0,5 \beta^T (5\mathbf{I}_4)^{-1} \beta],$$

при цьому апостеріорна щільність приймає вигляд:

$$\begin{aligned} h(\beta | \mathbf{y}, \mathbf{X}) &\propto g(\beta) L(\mathbf{y} | \mathbf{X}, \beta) \propto \\ &\propto \exp[-0,5 \beta^T (5\mathbf{I}_4)^{-1} \beta] \prod_{i=1}^{251} \Phi(\mathbf{x}_i^T \beta)^{y_i} [1 - \Phi(\mathbf{x}_i^T \beta)]^{(1-y_i)}. \end{aligned}$$

Очевидно, що ця апостеріорна щільність не належить до відомого сімейства розподілів.

Якщо ми будемо оцінювати вектор параметрів β як середнє відповідних, згенерованих за цим розподілом, марковських ланцюгів, то потрібно буде обчислити вибіркові середні і, можливо, їх коваріаційну матрицю для контролю якості оцінок. Простим способом розв'язання цієї задачі є генерування псевдовипадкових послідовностей з цього апостеріорного розподілу, а потім використання цих послідовностей для



обчислення шуканих апостеріорних оцінок параметрів. Надалі покажемо, що згенерувати дані з цієї апостеріорної щільності можна досить просто за процедурою МКМЛ [42], яка ґрунтується на модифікації цільового розподілу шляхом додавання прихованих (*latent*) даних. А зараз розглянемо алгоритм М-Х без прихованих даних.

Модель випадкового кроку як пропонуючий розподіл (proposal density)

Для того щоб вибрати щільність вихідного пропонуючого розподілу, розглянемо спочатку функцію правдоподібності [48]. Нехай за методом максимальної правдоподібності (ММП), з використанням процедури Ньютона-Рафсона, знайдено такі оцінки вектора невідомих параметрів:

$$\hat{\beta} = \arg \max \ln[L(\mathbf{y} | \mathbf{x}, \beta)] = (-1,093022; 0,607643; 1,197543, -1,904730)^T,$$

і нехай симетрична матриця, отримана в результаті обернення гессіана з протилежним знаком (матриця других похідних) логарифмованої функції правдоподібності, оціненої в точці $\hat{\beta}$, має вигляд:

$$\mathbf{V} = \begin{bmatrix} 0,040745 & -0,007038 & -0,039399 & 0,004829 \\ & 0,073101 & -0,006940 & -0,050162 \\ & & 0,062292 & -0,016803 \\ & & & 0,080788 \end{bmatrix}.$$

Тепер припустимо, що початкові *пропоновані значення* оцінок параметрів генеруються за моделлю випадкового кроку (random walk) таким чином:

$$\beta = \beta^{(j-1)} + \varepsilon^{(j)}, \quad \varepsilon^{(j)} \sim N_4(0, \tau \mathbf{V}), \quad (5.6)$$

де τ – параметр масштабу, який використовується для налаштування алгоритму М-Х так, щоб генерувати конкуруючі псевдовипадкові послідовності. У цьому випадку пропонуюча щільність має такий вигляд:

$$q(\beta^{(j-1)}, \beta | \mathbf{y}) \propto \exp \left[-0,5 (\beta - \beta^{(j-1)})^T (\tau \mathbf{V})^{-1} (\beta - \beta^{(j-1)}) \right], \quad (5.7)$$

де немає нормуючої константи у знаменнику.

Для того щоб генерувати псевдовипадкові числа з багатовимірної нормальної щільності скористаємось тим фактом з матричної алгебри, що будь-яку додатно визначену матрицю \mathbf{A} можна однозначно розкласти на

$\mathbf{L}\mathbf{L}^T$, де \mathbf{L} – нижня трикутна матриця з додатними елементами на головній діагоналі. Припустимо, що така матриця знайдена, тобто $\tau\mathbf{V} = \mathbf{L}\mathbf{L}^T$. Значення вектора $\boldsymbol{\varepsilon}^{(j)}$ можна обчислити з розподілу $N_4(\mathbf{0}, \tau\mathbf{V})$ таким чином:

$$\boldsymbol{\varepsilon}^{(j)} = \mathbf{L} \mathbf{z}^{(j)},$$

де $\mathbf{z}^{(j)} \sim N_4(\mathbf{0}, \mathbf{I}_4)$.

Припустимо, що виконано $(j-1)$ -й крок алгоритму і необхідно генерувати числа на наступному кроці. Спочатку генерується пропоноване значення β за рівнянням $\beta = \beta^{(j-1)} + \boldsymbol{\varepsilon}^{(j)}$, $\boldsymbol{\varepsilon}^{(j)} \sim N_4(\mathbf{0}, \tau\mathbf{V})$, а потім обчислюється ймовірність заповнення ланцюга за виразом:

$$\alpha(\beta^{(j-1)}, \beta | \mathbf{y}) = \min \left\{ 1, \frac{h(\beta | \mathbf{y})}{h(\beta^{(j-1)} | \mathbf{y})} \frac{q(\beta, \beta^{(j-1)} | \mathbf{y})}{q(\beta^{(j-1)}, \beta | \mathbf{y})} \right\}, \quad (5.8)$$

де $\alpha(\beta^{(j-1)}, \beta | \mathbf{y})$ означає умовну ймовірність вибору значення $\beta^{(j-1)}$ або β для заповнення ланцюга.

Зазначимо, що члени, які містять пропонууючу щільність, скорочуються (тобто члени $q(\beta, \beta^{(j-1)} | \mathbf{y})$ і $q(\beta^{(j-1)}, \beta | \mathbf{y})$). Це наслідок того, що пропонууюча щільність у вигляді рівняння випадкового кроку є симетричною стосовно своїх аргументів. Якщо скористатись рівномірним розподілом, тобто $U \sim Uniform(0, 1)$, то на наступній ітерації буде виконуватись таке присвоєння:

$$\beta^{(j)} = \begin{cases} \beta, & \text{якщо } U < \alpha(\beta^{(j-1)}, \beta | \mathbf{y}), \\ \beta^{(j-1)}, & \text{інакше.} \end{cases}$$

Далі цей процес повторюється з метою отримання послідовності $\{\beta^{(1)}, \beta^{(2)}, \dots, \beta^{(n_0+M)}\}$.

Важливо розуміти роль параметра налаштування алгоритму τ . З виразу для ймовірності заповнення ланцюга видно, що при великих значеннях τ значення, що пропонуються для заповнення ланцюга, будуть далекими від поточного значення і ймовірно будуть відхилені. Якщо ж значення τ занадто мале, то пропоноване значення скоріше всього буде прийняте, але наступні значення будуть близькими між собою і таким чином ланцюг буде повільно (слабко) використовувати апостеріорний розподіл. У обох випадках значення β , згенеровані з апостеріорного розподілу, будуть мати значну послідовну кореляцію.



Для того щоб уникнути цієї проблеми, доцільно зробити пробні запуски алгоритму з різними значеннями τ . Існує рекомендація [62, 63] використовувати таке значення τ , яке забезпечить прийняття запропонованих значень на рівні (30 – 50)%. Виходячи з таких рекомендацій, для даних, що відносяться до цього прикладу, використано значення параметра налагодження $\tau = 1$. Кількість початкових ітерацій алгоритму М-Х, що не використовувались для обчислення оцінок параметрів, взято 100, а для обчислення оцінок використано 5000 наступних ітерацій.

Для того щоб ілюструвати результати обчислень (імітаційного моделювання) існують різні способи: графіки, таблиці та окремі числа. У поданій нижче таблиці 5.2 наведені значення апіорних та апостеріорних перших двох моментів, 2,5 (нижній) і 97,5 (верхній) процентилі маргінальної щільності для β . Кожна з цих величин обчислюється за апостеріорним розподілом з використанням відомих формул; апостеріорне стандартне відхилення обчислюється – це стандартне відхилення для згенерованих значень, а апостеріорні процентилі також розраховані на основі згенерованих послідовностей [58].

Таблиця 5.2

Результати застосування алгоритму М-Х до оцінювання параметрів з використання рівняння випадкового кроку для пропонованих значень

β_i	Апіорні параметри		Апостеріорні параметри			
	середнє	станд. відхил.	середнє	станд. відхил.	нижній процентиль	верхній процентиль
β_0	0,000	3,162	-1,110	0,224	-1,553	-0,677
β_1	0,000	3,162	0,612	0,254	0,116	1,127
β_2	0,000	3,162	1,198	0,263	0,689	1,725
β_3	0,000	3,162	-1,901	0,275	-2,477	-1,354

Як і очікувалось, перший і другий регресори збільшують ймовірність інфікування ($\beta_1 > 0$; $\beta_2 > 0$), а третій (профілактика з використанням антибіотика) зменшує ймовірність інфікування.

Форма маргінальної апостеріорної щільності для чотирьох оцінених параметрів подана на рис. 4.1. Графіки отримані шляхом згладжування гістограм згенерованих значень гаусівським ядром. На рисунку також

подано автокореляційні функції для оцінок (згенерованих ланцюгів) кожного параметра. Значення автокореляційних функцій спадають майже до нуля при значенні лагу 20.

Скорочений варіант пропонууючої щільності

Псевдовипадкові числа, які є кандидатами на нові значення ланцюга, можна згенерувати за дещо іншою схемою. Якщо вибрати нормальний розподіл, то форма пропонууючої щільності буде такою:

$$q(\beta | \mathbf{y}) \propto \exp \left\{ -0,5 (\beta - \hat{\beta})^T (\tau \mathbf{V})^{-1} (\beta - \hat{\beta}) \right\}.$$

Ця пропонууюча щільність подібна до тієї, що ґрунтується на рівнянні випадкового кроку, за винятком того, що цей розподіл центрується навколо фіксованої точки (значенні) $\hat{\beta}$, а не попередньому значенні ланцюга. Цю щільність називають скороченою пропонууючою щільністю. Пропоновані значення генеруються за виразом:

$$\begin{aligned} \beta &= \hat{\beta} + \varepsilon^{(j)}, \\ \varepsilon^{(j)} &\sim N_4(\mathbf{0}, \tau \mathbf{V}), \end{aligned}$$

де $\hat{\beta}$ – значення, отримані в результаті застосування ММП до логарифмованої функції правдоподібності.

Ймовірність заповнення ланцюга новим значенням визначається так:

$$\alpha(\beta^{(j-1)}, \beta | \mathbf{y}) = \min \left\{ 1, \frac{w(\beta | \mathbf{y})}{w(\beta^{(j-1)} | \mathbf{y})} \right\}, \quad (5.9)$$

де $w(\beta | \mathbf{y}) = \frac{h(\beta | \mathbf{y}, \mathbf{X})}{q(\beta | \mathbf{y})}$; $w(\beta^{(j-1)} | \mathbf{y}) = \frac{h(\beta^{(j-1)} | \mathbf{y}, \mathbf{X})}{q(\beta^{(j-1)} | \mathbf{y})}$.

Для того щоб цей скорочений варіант пропонууючої щільності функціонував коректно, необхідно щоб функція $w(\cdot | \mathbf{y}, \mathbf{X})$ була обмежена. Для унімодальної, абсолютно неперервної функції $h(\cdot | \mathbf{y}, \mathbf{X})$ ця умова виконується у випадку, коли хвости пропонууючої щільності товстіші ніж хвости цільової щільності. Цю умову необхідно перевіряти аналітично, але вона, як правило, виконується коли $\tau > 1$.



5.10 Алгоритм Гіббса для генерування марковських ланцюгів

Одним із простих алгоритмів генерування марковських ланцюгів за методом Монте-Карло є алгоритм Гіббса. Він запропонований у роботі [79] де розглядалась задача розпізнавання образів; у роботі [122] його застосували до розв'язання задачі заповнення пропусків даних; а у [78] його використали до розв'язання ряду задач байєсівського аналізу даних.

Побудова марковського ланцюга за алгоритмом Гіббса реалізується шляхом генерування множини умовних розподілів такого типу:

$$\begin{aligned} &h(\theta_1 | \mathbf{y}, \theta_2, \theta_3, \dots, \theta_p), \\ &h(\theta_2 | \mathbf{y}, \theta_1, \theta_3, \dots, \theta_p), \\ &\vdots \\ &h(\theta_p | \mathbf{y}, \theta_1, \theta_2, \dots, \theta_{p-1}). \end{aligned} \quad (5.10)$$

Кількість згенерованих блоків повинна бути такою, щоб гарантувати належне трактування (тобто виконувати необхідний аналіз) кожної повної умовної щільності. При розв'язанні багатьох задач байєсівського аналізу даних процедура генерування за Гіббсом впливає із самої структури моделі.

Як простий приклад розглянемо значення $\mathbf{y} = (y_1, \dots, y_n)$, які генеруються за допомогою моделі лінійної регресії:

$$\begin{aligned} \{y_i | \beta, \sigma^2\} &\sim N(\mathbf{x}_i^T \beta, \sigma^2), \quad i = 1, 2, \dots, n; \\ \beta &\sim N_k(\beta_0, \mathbf{B}_0), \\ \sigma^2 &\sim IG(0,5 \nu_0, 0,5 \delta_0), \end{aligned}$$

де через IG – позначено обернений гамма розподіл із щільністю

$$p(\sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{0,5\nu_0+1} \exp\left(-\frac{\delta_0}{2\sigma^2}\right), \quad \sigma^2 > 0.$$

такому випадку, коли розподіли β і σ^2 розглядаються двома окремими блоками, повна умовна щільність для β визначається таким чином:

$$\begin{aligned} h(\beta | \mathbf{y}, \sigma^2) &\propto f(\mathbf{y}, \beta, \sigma^2) \propto \\ &\propto g(\beta) L(\mathbf{y} | \beta, \sigma^2) \propto \\ &\propto \exp[-0,5 (\beta - \hat{\beta})^T \mathbf{B}_n^{-1} (\beta - \hat{\beta})], \end{aligned}$$

де

$$\hat{\beta} = \mathbf{B}_n \left(\mathbf{B}_0^{-1} \beta_0 + \sigma^{-2} \sum_{i=1}^n \mathbf{x}_i y_i \right);$$

і

$$\mathbf{B}_n = \left(\mathbf{B}_0^{-1} + \sigma^{-2} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T \right)^{-1}.$$

Вектор β має багатовимірну нормальну повну умовну щільність. Можна показати, що повна умовна щільність параметра σ^2 , тобто $h(\sigma^2 | \mathbf{y}, \beta)$ має оновлений обернений гамма розподіл:

$$(\sigma^2 | \mathbf{y}, \beta) \sim IG \left\{ \frac{\nu_0 + n}{2}, \frac{\delta_0 + \sum_{i=1}^n (y_i - \mathbf{x}_i^T \beta)^2}{2} \right\}.$$

Таким чином, є можливість виконувати аналіз кожної з розглянутих повної умовної щільності. Зазначимо, що процедура розширення даних (data augmentation), яка буде розглянута нижче, часто надає допомогу в отриманні множини повних умовних розподілів, що піддаються аналізу.

Один цикл виконання алгоритму Гіббса закінчується генеруванням оцінок параметрів $\{\theta_k\}_{k=1}^p$ з кожного повного умовного розподілу, що дає можливість рекурсивно оновлювати оцінки невідомих параметрів. Алгоритм Гіббса, у якому кожний блок реалізується у фіксованому порядку, подано нижче [79].

Алгоритм генерування псевдовипадкових величин за Гіббсом

Крок 1: Задати початкові значення вектора параметрів:

$$\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_1^{(p)}).$$

Крок 2: Повторити для $j = 1, 2, \dots, n_0 + M$ такі обчислення:

- згенерувати $\theta_1^{(j)}$ з $h(\theta_1 | \mathbf{y}, \theta_2^{(j-1)}, \theta_3^{(j-1)}, \dots, \theta_p^{(j-1)})$,
- згенерувати $\theta_2^{(j)}$ з $h(\theta_2 | \mathbf{y}, \theta_1^{(j)}, \theta_3^{(j-1)}, \dots, \theta_p^{(j-1)})$,
- ⋮
- згенерувати $\theta_p^{(j)}$ з $h(\theta_p | \mathbf{y}, \theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_{p-1}^{(j)})$.



Крок 3: Обчислити і зберегти значення оцінок векторів

$$\{\theta^{(n_0+1)}, \theta^{(n_0+2)}, \dots, \theta^{(n_0+M)}\}.$$

У цьому алгоритмі блок θ_k генерується з повного умовного розподілу

$$h(\theta_k | \mathbf{y}, \theta_1^{(j)}, \dots, \theta_{k-1}^{(j)}, \theta_{k+1}^{(j-1)}, \dots, \theta_p^{(j-1)}).$$

У якому елементи, які задають умову, відображають той факт, що коли досягається k -й блок, то попередні $(k-1)$ блоків вже оновлені. Перехідна щільність ланцюга (знову за припущення, що h – абсолютно неперервна) задається добутком перехідної щільності для кожного блока:

$$T(\theta^{(j-1)}, \theta^{(j)} | \mathbf{y}) = \prod_{k=1}^p h(\theta_k | \mathbf{y}, \theta_1^{(j)}, \dots, \theta_{k-1}^{(j)}, \theta_{k+1}^{(j-1)}, \dots, \theta_{k+1}^{(j-1)}) \quad (5.11)$$

Для ілюстрації способу, за яким блоки оновлюються, розглянемо структуру з двох блоків; у кожному блоці міститься одна компонента (θ_1 або θ_2). Зазначимо, що протягом однієї ітерації алгоритму оновлюються обидві компоненти алгоритму. При цьому кожна компонента оновлюється вздовж своєї осі координат. Ця властивість алгоритму може спричиняти проблеми, якщо компоненти вектора параметрів високо корельовані між собою: внаслідок цього контури стискаються і рухи вздовж осей координат стають дуже незначними (тобто природи значень параметрів стають малопомітними) [49, 51, 79].

Використання алгоритму Гіббса ґрунтується на припущенні, що повні умовні розподіли, які використовуються у процесі генерування ланцюгів, мають чітке зрозуміле (tractable) представлення. Якщо один або більше повних умовних розподілів приймають форму, яку неможливо розпізнати, тоді генерування ланцюга для цього блока параметрів можна виконати з використанням належної пропонууючої щільності і мультиблокового алгоритму Метрополіса-Хастінгса. Такий спосіб генерування ланцюга іноді називають ще «Метрополісом всередині Гіббса».

РОЗДІЛ 6

ПРОГНОЗУВАННЯ ВОЛАТИЛЬНОСТІ ЗА СТОХАСТИЧНОЮ МОДЕЛЛЮ

6.1 Вступ

Окрім моделей АРУГ були запропоновані і розвинені моделі стохастичної волатильності (МСВ), в яких на мінливу структуру дисперсії впливає стохастичний процес. В одній із версій такої моделі доходи від цінних паперів були описані добутком двох незалежних стохастичних процесів [46, 49, 56, 115, 121]. Стохастична модель волатильності може бути представлена як

$$x(k) = \psi_0 + \psi_1 x(k-1) + y(k), \quad (6.1)$$

$$y(k) = \sqrt{h(k)}u(k), \quad u(k) \approx N(0,1), \quad (6.2)$$

$$\log h(k+1) = \alpha + \phi \log h(k) + \eta(k), \quad \eta(k) \approx N(0, \sigma^2), \quad (6.3)$$

де $x(k)$ – часовий ряд статистичних даних, на основі яких будується модель; $u(k)$ і $\eta(k)$ – це стохастичні незалежні процеси білого шуму. У деяких формулюваннях припускається, що ψ_0 та ψ_1 дорівнюють нулю. Процес AR(1) з інновацією $y(k)$ в часовому ряду, що визначається рівнянням (6.1), пояснює існування можливої авторегресії в процесі $x(k)$.

Запропоновані також різні методи оцінювання таких моделей. В роботах [49, 50, 124] запропоновано Байєсів підхід, в якому поєднані метод Монте-Карло та ланцюги Маркова (МКМЛ). Вибірки об'єднаних апіорних розподілів невідомих параметрів та прихованих змінних отримані із застосуванням методу генерації вибірки за Гіббсом. Традиційні труднощі оцінювання параметрів моделі стохастичної волатильності розв'язані завдяки розвитку комп'ютерних технологій, які фактично надали можливість реалізувати метод МКМЛ. Таким чином, на сьогодні практичне застосування МКМЛ можливе так само, як і моделювання рівнянь АРУГ.

Асиметрія у доході від акцій є предметом багатьох досліджень. Можлива асиметрична поведінка доходів від акцій під час масових торгів може бути описана пороговою моделлю з умовною гетероскедастичністю.



Наявні результати показали, що структура умовного середнього може істотно залежати від росту та падіння продаж на протязі минулого дня. Крім того, в багатьох дослідженнях стверджувалось, що дисперсія реагує асиметрично на минулі доходи від паперів. Асиметрія має місце внаслідок того, що дисперсія має тенденцію швидше збільшуватись під впливом поганих новин ніж під впливом хороших. Цьому ефекту дали назву “ефект левериджу”. Зниження цін акцій приводить до збільшення впливу фінансових важелів, що, в свою чергу, призводить до збільшення дисперсії цін акцій [4, 6-8].

В даному розділі розглянемо порогову стохастичну модель волатильності (ПМСВ), яка описує одночасно асиметрію середнього та дисперсії. Лінійні структури, представлені в (6.1) та (6.3) узагальнюються на порогові нелінійні структури, в яких авторегресійна динаміка середнього та дисперсії керуються попередніми значеннями реалізації процесу. В такій специфікації асиметрія середнього та дисперсії можуть бути описані через нелінійні структури для $r(k)$ та $\log h(k)$, відповідно. В моделі також враховується асиметрія швидкості зміни дисперсії. Нова модель отримала назву *порогової моделі стохастичної волатильності* (ПМСВ).

В наступному параграфі наведено визначення моделі ПМСВ та наведено формулювання порогової структури. В підрозділі 6.3 описується процедура розширення масивів даних і метод МКМЛ для генерації вибірок із спільного апостеріорного розподілу невідомих параметрів та прихованих змінних. Тут також розглянуті методи прогнозування та згладжування. В підрозділі 6.4 представлено результати комп'ютерного моделювання, які демонструють ефективність підходу Монте Карло для марковських ланцюгів. Результати моделювання свідчать про те, що асиметрія дисперсії, що розглядається в багатьох роботах, може виникати внаслідок нехтування асиметрією середнього.

В підрозділі 6.5 наведено результати дослідження чутливості, які свідчать, що наведені алгоритми не дуже чутливі до припущень щодо моделей. Описана методологія застосовувалась до даних біржі Standard & Poor 500 (S&P 500) та індексу цін на Гонконгській біржі Hong Kong Hang Seng (HSI); відповідні результати наведено в підрозділі 6.6. При цьому отримано переконливі свідчення щодо наявності асиметрії доходів та дисперсії. Асиметрія швидкості зміни дисперсії виявлена в цінах на ринку акцій Гонконга.

Виконані дослідження ПМСВ свідчать про її суттєві переваги над відомими моделями гетероскедастичних процесів.

6.2 Порогова модель стохастичної волатильності

Дисперсія доходів має тенденцію збільшуватись після того, як ціна акцій знижується. Таку поведінку дисперсії можна описати за допомогою постійного коефіцієнта кореляції ρ між $u(k)$ і $\eta(k)$, зберігаючи всі інші припущення незмінними. В початковій моделі, визначеній рівняннями (6.1) – (6.3), ρ дорівнює нулю. Численні емпіричні дослідження показали, що коефіцієнти ρ мають від'ємні значення в припущенні, що від'ємні доходи пов'язані з додатними дисперсіями нев'язок [17].

Нижче пропонується новий підхід до фіксації асиметрії дисперсії часового ряду. Оскільки було встановлено, що дисперсія має тенденцію зростати під впливом поганих (невтішних) новин, то цілком імовірно, що динаміка авторегресії в рівнянні (6.3) визначається знаком доходів на попередньому відрізку часу. Існує гіпотеза, що величина доходів залежить від знаку доходів. Цей вид асиметрії доходів також можна взяти до уваги, узагальнюючи рівняння (6.1) до кусочно-лінійної структури. Отже природніше буде розглянути порогові нелінійні структури ніж лінійні авторегресійні процеси, представлені рівняннями (6.1) – (6.3).

Визначимо набір випадкових змінних Бернуллі $s(k)$:

$$s(k) = \begin{cases} 0, & \text{якщо } x(k-1) < 0, \\ 1, & \text{якщо } x(k-2) \geq 0. \end{cases} \quad (6.4)$$

Тепер порогова модель стохастичної волатильності приймає такий вигляд:

$$x(k) = \psi_{0s(k)} + \psi_{1s(k)}x(k-1) + y(k), \quad (6.5)$$

$$y(k) = \sqrt{h(k)}u(k), \quad u(k) \sim N(0,1), \quad (6.6)$$

$$\log h(k+1) = \alpha_{s_k+1} + \phi_{s_k+1} \log h(k) + \eta(k), \quad \eta(k) \sim N(0, \sigma^2). \quad (6.7)$$

Як і в початковому формулюванні $u(k)$ та $\eta(k)$ є стохастично незалежними. В момент часу $k-1$, коли має місце несподіване пониження ціни внаслідок надходження невітшних новин, $x(k-1) < 0$ та $s(k) = 0$. Навпаки, якщо є хороші новини в момент часу $k-1$, то $x(k-1) > 0$ і $s(k) = 1$. Тому значення $s(k)$ визначається знаком величини $x(k-1)$. В пороговій моделі стохастичної волатильності значення параметрів ψ_0 , ψ_1 , α та ϕ переключаються між цими двома режимами, які відповідають підвищенню та падінню цін активів.

В симетричному випадку обидва набори параметрів ідентичні між собою. Зокрема, якщо $\phi_0 = \phi_1$, то із того факту, що $\alpha_0 \geq \alpha_1$ випливає, що дисперсія буде вищою, коли минулий дохід від'ємний, ніж коли він



додатний. В узагальненій моделі ϕ_0 може відрізнятися від ϕ_1 . Дійсно коефіцієнт, ϕ_{s_t} характеризує величину впливу минулого доходу на поточну дисперсію. Якщо ϕ_0 більше ніж ϕ_1 , то дисперсія в минулі періоди буде мати більший вплив на поточну дисперсію після падіння цін, ніж після підвищення цін. Очікується, що в подібній гіпотетичній ситуації ринку знадобиться більше часу для “обробки” невітійної інформації, яка міститься в дисперсії минулих даних. Цей вид асиметрії ще недостатньо розглядався в літературі з аналізу стохастичної волатильності.

6.3 Байєсівський аналіз порогової моделі стохастичної волатильності

Генерування вибірки за Гіббсом

В стандартному байєсівському виводі використовуються маржинальні апостеріорні розподіли невідомих параметрів. Проте, в багатьох випадках спільний апостеріорний розподіл або навіть маржинальний апостеріорний розподіл не мають замкнених форм. Також досить непросто отримати модельні значення від бажаних апостеріорних розподілів. Підходи на основі методу Монте Карло для марковських ланцюгів представляють собою марковські алгоритми оновлення, що мають за мету отримання вибірки із спільного апостеріорного розподілу. Окремим випадком є генерація вибірки за Гіббсом. Цей метод на основі процедури Монте Карло є близьким до підходу на основі розширення вибірок даних.

Нехай необхідно зробити вибірку з об'єднаного розподілу $F(\omega_1, \dots, \omega_m)$ де $\omega = (\omega_1, \dots, \omega_m)^T$ – вектор невідомих параметрів або прихованих змінних. При відомих початкових значеннях $[\omega_1^{(0)}, \dots, \omega_m^{(0)}]$ алгоритм надає оцінку значення $\omega_1^{(i)}$ з $F(\omega_1 | \omega_2^{(i-1)}, \dots, \omega_m^{(i-1)})$, значення $\omega_2^{(i)}$ з $F(\omega_2 | \omega_1^{(i)}, \omega_3^{(i-1)}, \dots, \omega_m^{(i-1)})$,; значення $\omega_m^{(i)}$ з $F(\omega_m | \omega_1^{(i-1)}, \dots, \omega_{m-1}^{(i-1)})$, при $i = 1, \dots, M + N$. При нежорстких умовах вектор параметрів $[\omega_1^{(i)}, \dots, \omega_m^{(i)}]$ збігається за розподілом до спільного розподілу $F(\omega_1, \dots, \omega_m)$ при $i \rightarrow \infty$. Зазвичай перші M перехідних ітерацій пропускаються, а останні N ітерацій приймаються за наближену вибірку, залежну від $F(\omega_1, \dots, \omega_m)$. Оцінки густини ймовірностей можуть бути отримані двома різними способами. Один підхід полягає у традиційному оцінюванні густини ймовірностей ядра. Друий підхід полягає у використанні виразу:

$$f(\omega_j) = \frac{1}{N} \sum_{i=M+1}^{M+N} f(\omega_j | \omega_{-j}^{(i)}),$$

де ω_{-j} означає всі параметри, окрім ω_j . Точкові оцінки будь-якої функції ω , скажемо $g(\omega)$, також можна знайти, використовуючи генерацію вибірки за Гіббсом. Один із загальноприйнятих підходів полягає у використанні апостеріорного середнього, тобто

$$\frac{1}{N} \sum_{i=M+1}^{M+N} g(\omega^{(i)}).$$

Алгоритм дискретизації (генерації вибірки) Гіббса можна використати у випадку, якщо є можливість моделювати за допомогою усіх повних умовних розподілів $F(\omega_j | \omega_{j-1})$. Дискретизацію за Гіббсом та інші методи на основі МКМЛ можна знайти в роботі [80].

Реалізація методу МКМЛ

Перед тим як перейти до докладного описання алгоритму реалізації методу Монте-Карло для марковських ланцюгів, параметризуємо (тобто, запишемо рівняння для обчислення параметрів) $\psi_{0_{s(k)}}$, $\psi_{1_{s(k)}}$ та $\phi_{s(k)}$ таким чином [77]:

$$\psi_{0_{s(k)}} = \psi_0 + \delta s(k),$$

$$\psi_{1_{s(k)}} = \psi_0 + cs(k),$$

$$\alpha_{s(k)} = \alpha + \gamma s(k),$$

$$\phi_{s(k)} = \phi + d_{s(k)}.$$

Окремий випадок, $\delta = c = 0$, відповідає наявності асиметрії лише у рівнянні для дисперсії. Аналогічно, у випадку $\gamma = d = 0$ припускається, що немає асиметрії в рівнянні для дисперсії, а у випадок $\delta = c = \gamma = d = 0$ відповідає симетричній моделі. У двох різних режимах рівнянь для середнього та дисперсії, коефіцієнти перетину (константи рівнянь), відрізняються величинами δ і γ відповідно. Параметри c і d представляють збільшення коефіцієнта авторегресії для компонент середнього та дисперсії, відповідно.

Нехай

$$H(k) = [h(1), \dots, h(k)]^T, \quad S(k) = [s(1), \dots, s(k)]^T, \quad X(k) = [x(1), \dots, x(k)]^T \quad \text{та} \\ Y(k) = [y(1), \dots, y(k)]^T \quad \text{при } k \geq 1 - \text{вектори зазначених часових рядів, і нехай}$$



$\theta = [\psi_0, \delta, \psi_1, c, \alpha, \gamma, \sigma^2, \phi, d]^T$ – вектор невідомих параметрів. Зазначимо, що x_0 – одна із прихованих змінних. Якщо відоме значення $X(k-1)$, то визначається значення $S(k)$. Використовуючи ідею розширення даних, сформуємо гіпервектор параметрів $\omega = [x_0, H_n^T, \theta^T]^T$, який необхідно оцінити.

Для того щоб скористатись методом МКМЛ, необхідно моделювати придатні умовні розподіли. З теореми Байєса і наведених вище припущень щодо моделі, маємо (для зручності аргумент k напишемо як нижній індекс):

$$\begin{aligned}
 & f(\omega_i | \omega_{-i}, X_n) \propto f(\omega, R_n) = f(x_0, X_n, H_n | \theta) f(\theta) = \\
 & = \prod_{k=2}^n f(x_k, h_k | x_0, X_{k-1}, H_{k-1}, \theta) f(x_1 | h_1, x_0, \theta) f(h_1 | \theta, x_0) = \\
 & = \prod_{k=2}^n [f(x_k | h_k, x_0, X_{k-1}, H_{k-1}, \theta) f(h_k | x_0, X_{k-1}, H_{k-1}, \theta)] f(x_1 | h_1, x_0, \theta) f(h_1 | \theta, x_0) \\
 & \times f(\theta) f(x_0) = \\
 & = \prod_{k=1}^n f(x_k | h_1, x_{k-1}, \theta) \prod_{k=2}^n f(h_k | X_{k-1}, H_{k-1}, \theta) f(h_1 | \theta, x_0) f(\theta) f(x_0) = \\
 & = \prod_{k=1}^n f(x_k | h_1, x_{k-1}, \theta) \prod_{k=2}^n f(h_k | s_k, h_{k-1}, \theta) f(h_1 | \theta, x_0) f(\theta) f(x_0) = \\
 & = \alpha \prod_{k=1}^n h_k^{-1/2} \exp\left(-\frac{1}{2} \frac{y_k^2}{h_k}\right) \prod_{k=2}^n (\sigma h_k)^{-1} \exp\left\{-\frac{1}{2\sigma^2} (\log h_k - \alpha_{s_k} - \phi_{s_k} \log h_{k-1})^2\right\} \times \\
 & \quad \times f(h_1 | \theta, x_0) f(\theta) f(x_0),
 \end{aligned} \tag{6.8}$$

де

$$\begin{aligned}
 y(k) &= x(k) - \psi_{0_{s(k)}} - \psi_{1_{s(k)}} x(k-1), \\
 s(k) &= \begin{cases} 0 & \text{якщо } x(k) < 0 \\ 1 & \text{якщо } x(k) \geq 0 \end{cases}
 \end{aligned}$$

Генерування вибірки θ і x_0 за допомогою повного умовного розподілу кожного параметра – це не дуже складна процедура, якщо використовуються спряжені апіорні змінні. Наприклад, якщо для $\psi_0, \delta, \psi_1, c$ і x_0 вибрані нормально розподілені апіорні значення, то можна показати, що їх повні умовні розподіли будуть нормальними. Так само, якщо за апіорний розподіл для σ^2 використаємо інверсний гамма-розподіл, то його повний умовний розподіл також буде інверсним гамма-

розподілом. Таким чином, використовуючи стандартні спряжені апіорні дані, можна згенерувати вибірку за Гіббсом шляхом дискретизації параметрів індивідуально на основі стандартних розподілів. Однак, отримання оцінок параметрів індивідуально може призвести до високих рівнів кореляції між ітераціями, що спричиняє повільну збіжність алгоритму [96].

Тому пропонується згрупувати високо корельовані параметри разом і генерувати вибірки одночасно. Сформуємо три групи для θ ; перша група – параметри середніх для рівнянь: $\psi_0, \delta, \psi_1, c$; друга група – параметри α, γ, ϕ та d для рівняння дисперсії, і остання група – дисперсія σ^2 . Таке групування є необхідним, оскільки очікуються високі значення кореляції в межах кожної групи ітерацій методу МКМЛ.

Оскільки для першої групи параметрів прийнято багатовимірний нормальний розподіл, то можна показати, що її повний умовний розподіл також буде багатовимірним нормальним. Докладніше така процедура розглядається нижче в даному розділі.

По аналогії, для другої групи α, γ, ϕ та d скористаємось багатовимірним нормальним розподілом, задаючи спряжені нормально розподілені апіорні значення. Для того щоб прискорити збіжність, пропонується оцінювати значення σ^2 з $f(\sigma^2 | \theta_{-\sigma^2}, x_0, X_n)$, а не з $f(\sigma^2 | \theta_{-\sigma^2}, x_0, H_n, X_n)$. Іншими словами, ми отримуємо σ^2 з розподілу, не накладаючи умову щодо H_n . Відповідно до результатів, отриманих в роботі [89], виключення умови по H_n може прискорювати збіжність алгоритму генерації вибірок. Це особливо стосується моделей часових рядів, що включають приховані змінні. Головна проблема полягає в тому, що $f(\sigma^2 | \theta_{-\sigma^2}, x_0, X_n)$ – нестандартний розподіл, а тому з нього не просто робити вибірку. Щоб полегшити визначення σ^2 , було використано симетричну модель випадкового кроку Метрополіса, згідно з якою на ітерація $i + 1$ формується значення $\sigma^{2(i+1)}$ за виразом:

$$\sigma^{2(i+1)} = \sigma^{2(i)} + a, \quad a \sim N(0, \tau^2);$$

а нове значення приймається з ймовірністю

$$\min \left(1, \frac{f(\sigma^{2(i+1)} | \theta_{-\sigma^2}, x_0, X_n)}{f(\sigma^{2(i)} | \theta_{-\sigma^2}, x_0, X_n)} \right).$$

Інакше, покладаємо, що $\sigma^{2(i+1)} = \sigma^{2(i)}$. Оцінювання ймовірності прийняття нового значення виконується досить просто, оскільки величину



$$f(\sigma^2 | \theta_{-\sigma^2}, x_0, X_n) \Rightarrow f(X_n | \theta, x_0) \Rightarrow f(Y_n | \theta)$$

можна обчислити за одну реалізацію фільтра Калмана з точністю до коефіцієнта пропорційності, використовуючи гаусову модель в просторі станів, наведену нижче. Додатний скаляр τ^2 визначає величину кроку в моделі Метрополіса. На практиці, можна підбирати величину кроку τ^2 , настраюючи прийнятну величину ймовірності прийняття нового значення таким чином, щоб досягти швидкої збіжності. Досвід показує, що “оптимальне” значення ймовірності прийняття зазвичай знаходиться між 0,2 і 0,5.

Для того щоб завершити схему генерації вибірки за методом МКМЛ, необхідно обчислити $h(k)$, $k=1, \dots, n$ з належних повних умовних щільностей. В даній схемі вибрано багатокрокову вибірку, що одночасно обчислює $h(k)$. Ми сформулювали порогову модель стохастичної волатильності у вигляді часткової негаусівської форми простору станів. Ідея цього підходу полягає у тому, щоб апроксимувати розподіл $\log \chi_1^2$ для $\log[u(k)]^2$ сумішшю нормального розподілу, тобто

$$f(\log[u(k)]^2) \approx \sum_{i=1}^p q_i f(\log[u(k)]^2 | \rho(k) = i),$$

де

$$\begin{aligned} \log([u(k)]^2 | \rho(k) = i) &\sim N(\mu_i - 1.2704, \tau_i^2), \\ q_i &= \Pr(\rho(k) = i), \end{aligned}$$

$\rho(k)$ – дискретна змішана змінна змішування, що приймає значення від 1 до P . В роботах [80, 88] показано, що при $p=7$ можна отримати задовільне наближення істинного розподілу $\log \chi_1^2$. Значення параметрів змішаного нормального розподілу наведено нижче в даному розділі.

При відомих значеннях ρ_i і при використанні нормального змішаного наближення, порогову модель стохастичної волатильності можна записати у вигляді звичайної лінійної моделі в просторі станів з параметрами, залежними від часу, тобто:

$$\begin{aligned} \log[y(k)]^2 &= \log h(k) + \varepsilon(k), \\ \log h(k+1) &= \alpha_{s_{k+1}} + \phi_{s_{k+1}} \log h(k) + \eta(k), \end{aligned} \quad (6.9)$$

де

$$\begin{aligned} \varepsilon(k) | \rho(k) &\sim N(\mu^T(k), \tau^2(k)), \\ \mu^2(k) &= \mu_{\rho_k} - 1.2704, \end{aligned}$$

$$\tau^2(k) = \tau_{\rho_k}^2.$$

Для декомпозиції умовної об'єднаної щільності $\log h(1), \dots, \log h(n)$,
 $f(\log h(1), \dots, \log h(n) | \rho(1), \dots, \rho(n), \theta, x_0, X_n)$,

в однофакторні щільності використано результати, отримані в роботі [80]:

$$\begin{aligned} & f(\log h_1, \dots, \log h_n | \rho_1, \dots, \rho_n, \theta, x_0, X_n) = \\ & = f(\log h_1, \dots, \log h_n | \rho_1, \dots, \rho_n, \theta, x_0, Y_n) = \\ & = f(\log h_n | \rho_1, \dots, \rho_n, \theta, x_0, Y_n) \prod_{k=1}^{n-1} f(\log h_k | \log h_{k+1}, \dots, \log h_n, \rho_1, \dots, \rho_n, \theta, Y_n) = \\ & = f(\log h_n | \rho_1, \dots, \rho_n, \theta, x_0, Y_n) \prod_{k=1}^{n-1} f(\log h_k | \log h_{k+1}, \rho_1, \dots, \rho_n, \theta, Y_k). \end{aligned}$$

Обчислення виконується послідовно від $\log h_n$ до $\log h_1$. Усі варіанти імітаційного моделювання включають генерацію вибірки з нормальних розподілів із середніми та різницями, оціненими стандартним фільтром Калмана. Алгоритм генерування даних та обчислення параметрів складається з наступних кроків:

Крок 1: Використовуючи дифузійний (рівномірний) апріорний розподіл для початкового стану $\log h_0$, покладемо

$$\begin{aligned} a(1) &= \log y(1)^2 - \mu_1^T, \\ p(1) &= \tau(1)^2. \end{aligned}$$

Крок 2: Обчислити рекурсивне $a(k)$ і $p(k)$ за виразами:

$$\begin{aligned} a(k) &= \phi_{s_k} a(k-1) + \alpha_{s_k} + \frac{\rho(k)}{\tau(1)^2} [\log y(k)^2 - \phi_{s_k} a(k-1) - \alpha_{s_k} - \mu^T(k)], \\ p(k) &= \frac{[\phi_{s_k}^2 p(k-1) + \sigma^2] \tau(k)^2}{\phi_{s_k}^2 p(k-1) + \sigma^2 + \tau(k)^2}. \end{aligned}$$

Крок 3: Обчислити $\log h(n)$ з $N[a(n), p(n)]$.

Крок 4: Обчислити $\log h(k)$ з $k = n-1, \dots, 1$ по розподілу:

$$N\left(a(k) + \frac{\phi_{s_{k+1}} p(k) [\log h(k+1) - \phi_{s_{k+1}} a(k) - \alpha_{s_{k+1}}]}{\phi_{s_{k+1}}^2 p(k) + \sigma^2}, \frac{\sigma^2 p(k)}{\phi_{s_{k+1}}^2 p(k) + \sigma^2} \right).$$

Крок 5: Обчислити $\rho(k)$ з умови, що повна умовна щільність $\rho(k)$ є пропорційною до



$$f(\log y(k)^2 | \rho(k), h(k)) f(\rho(k)) \propto \frac{1}{\tau_{p_k}} \exp \left[-\frac{1}{2\tau_{p_k}^2} [\log y(k)^2 - \log h(k) - \mu_{p_k} + 1.2704]^2 \right] f(p(k)).$$

Підсумовуючи, сформулюємо алгоритм формування вибірки за методом МКМЛ та обчислення параметрів розподілу:

- 1) Обчислити $\rho(k)$, $k = 1, \dots, n$.
- 2) Обчислити σ^2 за моделлю випадкового кроку Метрополіса.
- 3) Обчислити $h(k)$, $k = 1, \dots, n$, використовуючи багатокрокову дискретизацію.
- 4) Обчислити $(\psi_0, \delta, \psi_1, c)^T$ з його багатовимірного нормального повного умовного розподілу.
- 5) Обчислити x_0 з його нормального повного умовного розподілу.
- 6) Обчислити $(\alpha, \gamma, \phi, d)^T$ з його багатовимірного нормального повного умовного розподілу.

Закінчення ітерації алгоритму за методом МКМЛ.

Згладжування і прогноз дисперсії

Тепер розглянемо згладжування і прогноз дисперсії, використовуючи вибірку за Гіббсом [117]. Після виконання ітерації, необхідних для дискретизації за Гіббсом, отримуємо наближення (апроксимацію) вибірки із спільного апостеріорного розподілу $f(x_0, H_n, \theta | X_n)$, позначеного як $(x_0^{(i)}, h_1^{(i)}, \dots, h_n^{(i)}, \theta^{(i)})$, $i = M + 1, \dots, M + N$. Згладжені оцінки $h(k)$, $(k = 1, \dots, n)$ є оцінками h_t , отриманими з маржинального апостеріорного розподілу $f(h(k) | X_n)$. Природним вибором є маржинальне апостеріорне очікування, $E(h(k) | X_n)$, яке може бути оцінене як вибіркоче середнє:

$$\frac{1}{N} \sum_{i=M+1}^{M+N} h(k)^{(i)}.$$

Щоб спрогнозувати майбутню дисперсію, використовуючи наявну поточну доступну інформацію, необхідно згенерувати вибірки з $f(h_{n+j} | X_n)$ при $j > 0$. Це може бути ефективно виконано методом композиції. Так, при $j = 1$ можна записати:

$$\begin{aligned} f(h_{n+1} | X_n) &= \int f(h_{n+1} | h_n, \theta, X_n) f(h_n, \theta | X_n) d(h_n, \theta) = \\ &= \int f(h_{n+1} | h_n, \theta, s_{n+1}) f(h_n, \theta | X_n) d(h_n, \theta). \end{aligned}$$

Тому $h_{n+1}^{(2)}, i = M + 1, \dots, M + N$, як наближена вибірка з $f(h_{n+1} | X_n)$, моделюється за допомогою логнормальних щільностей розподілу $f(h_{n+1} | h_n^{(i)}, \theta^{(i)}, s_{n+1})$. Використовуючи цю вибірку, можна сформулювати оцінки параметрів густини розподілу як точкові оцінки h_{n+1} . Даний підхід узагальнено на обчислення багатокрокового прогнозу. Наприклад, оскільки

$$f(h_{n+2} | X_n) = \int f(h_{n+2} | h_{n+1}, h_n, \theta, X_{n+1}) f(x_{n+1} | h_{n+1}, h_n, \theta, X_n) f(h_{n+1} | h_n, \theta, X_n) \cdot f(h_n, \theta | X_n) d(x_{n+1}, h_{n+1}, h_n, \theta) =$$

$$= \int f(h_{n+2} | h_{n+1}, h_n, \theta, s_{n+2}) f(x_{n+1} | h_{n+1}, \theta, x_n, s_{n+1}) f(h_{n+1} | h_n, \theta, x_n) \cdot f(h_n, \theta | X_n) d(x_{n+1}, h_{n+1}, h_n, \theta)$$

то для прогнозування h_{n+2} необхідно додати ще два кроки:

Крок 1: Обчислити $x_{n+1}^{(i)}$ з $f(x_{n+1} | h_{n+1}^{(i)}, \theta^{(i)}, x_n, s_{n+1})$.

Крок 2: Обчислити $h_{n+2}^{(i)}$ з $f(h_{n+2} | h_{n+1}^{(i)}, \theta^{(i)}, s_{n+2})$

для $i = M + 1, \dots, M + N$,

де $s(n+2)^{(i)} = \begin{cases} 0 & \text{якщо } x(n+1)^{(i)} < 0 \\ 1 & \text{якщо } x(n+1)^{(i)} \geq 0 \end{cases}$.

Перший і другий кроки включають генерування виборок з нормального та логнормального розподілів, відповідно. Прогнозовані точкові значення h_{n+2} сформовані таким же чином, використовуючи вибіркове середнє чи інші локальні міри.

Важливо підкреслити, що $x(n+1)^{(i)}, i = M + 1, \dots, M + N$, обчислені на кроці 1, представляють собою модельні значення з $f(x(n+1) | X_n)$. Якщо необхідно мати екстремальний p -й процентильний прогноз, скажімо при $p=1$ для оцінювання значення VaR (*Value-at-Risk*) на фінансових ринках, то вибірка $x(n+1)^{(i)}$ забезпечить вибір з p -го емпіричного процентилю. Очевидно, що як тільки ми маємо $h_{n+j}^{(i)}$ з розподілу $f(h_{n+j} | X_n)$, то можна знайти також значення багатокрокового прогнозу VaR.

6.4 Дослідження моделі стохастичної волатильності

Продемонструємо байєсівську методологію, описану в попередньому підрозділі на трьох модельних прикладах [14, 17, 27, 40, 49, 115]. Як сказано вище, повна ПМСВ охоплює лише модель з асиметрією в рівнянні для середнього ($\gamma = d = 0$), а модель з асиметрією в рівнянні для дисперсії ($\delta = c = 0$) тільки як окремі випадки. Першу модель назвемо моделлю M , а



останню моделлю V . В експериментах з моделювання, було модельовано три ряди (A , B і C) за допомогою моделі M , моделі V і моделі F (повна модель), відповідно:

Ряд A :

$$x(k) = \begin{cases} -0.002 - 0.3x(k-1) + y(k), & \text{якщо } x(k-1) < 0, \\ 0.003 + 0.3x(k-1) + y(k), & \text{якщо } x(k-1) \geq 0. \end{cases}$$

$$y(k) = \sqrt{h(k)} u(k), \quad u(k) \sim N(0,1),$$

$$\log h(k) = -0.5 + 0.9 \log h(k-1) + \eta(k),$$

$$\eta(k) \sim N(0,1).$$

Ряд B :

$$x(k) = 0.002 + 0.2x(k-1) + y(k-1),$$

$$y(k) = \sqrt{h(k)} u(k), \quad u(k) \sim N(0,1),$$

$$\log h(k) = \begin{cases} -0.5 + 0.6 \log h(k-1) + \eta(k), & \text{якщо } x(k-1) < 0, \\ -1 + 0.9 \log h(k-1) + \eta(k), & \text{якщо } x(k-1) \geq 0, \end{cases}$$

$$\eta(k) \sim N(0,1).$$

Ряд C :

$$x(k) = \begin{cases} -0.002 - 0.3x(k-1) + y(k), & \text{якщо } x(k-1) < 0, \\ 0.003 + 0.3x(k-1) + y(k), & \text{якщо } x(k-1) \geq 0, \end{cases}$$

$$y(k) = \sqrt{h(k)} u(k), \quad u(k) \sim N(0,1),$$

$$\log h(k) = \begin{cases} -0.5 + 0.6 \log h(k-1) + \eta(k), & \text{якщо } x(k-1) < 0, \\ -1 + 0.9 \log h(k-1) + \eta(k), & \text{якщо } x(k-1) \geq 0, \end{cases}$$

$$\eta(k) \sim N(0,1).$$

В моделі M , що використовується для генерування ряду A , вибрано від'ємний параметр ψ_1 і додатний $\psi_1 + c$. Вибір двох коефіцієнтів авторегресії у рівнянні для середнього виконано свідчить про те, що коефіцієнт авторегресії переважно додатний при $x(k-1) \geq 0$, а зворотнє твердження є істинним для тих періодів часу, коли $x(k-1) < 0$.

Оскільки висока стійкість дисперсії була виявлена в більшості літературних джерел із стохастичної волатильності, то покладаємо, що $\phi_{s_k} = 0,9$ в моделі M . Ряд B , моделювався з моделі V . Для врахування існування авторегресії в ряді доходів від цінних паперів, було прийнято мале значення ψ_1 .

Підсумкова статистика з маржинальних апостеріорних розподілів усіх параметрів наведена в таблицях 6.1 – 6.3. Апостеріорне середнє, стандартне відхилення (у квадратних дужках) і 90 % Байєсівський інтервал, у круглих дужках.

Таблиця 6.1

Підсумкова статистика з маржинальних апостеріорних розподілів усіх параметрів на основі вибраних моделей для ряду A , істинною є модель M

Істинне значення		S	M	V	F
ψ_0	-0.002	0.0043[0.0009] (0.0028,0.0058)	-0.0016[0.0011] (-0.0035,0.0002)	0.0040[0.0009] (0.0025,0.0055)	-0.0016[0.0011] (-0.0035,0.0002)
δ	0.005	0	0.0051[0.0015] (0.0027,0.0077)	0	0.0052[0.0015] (0.0027,0.0076)
ψ_1	-0.3	0.1508[0.0327] (0.098,0.2054)	-0.2506[0.0585] (-0.3470,-0.1533)	0.1592[0.0335] (0.1050,0.2146)	(0.0027,0.0076)
c	0.6	0	0.5580[0.0716] (0.4412,0.6791)	0	0.5648[0.0727] (0.4446,0.6829)
α	-0.5	-0.5249[0.1007] (-0.6965,-0.359)	-0.5730[0.1089] (-0.761,-0.4019)	-0.2185[0.1929] (-0.5512,0.0800)	-0.5964[0.2082] (-0.9404,-0.2581)
γ	0	0	0	-0.5145[0.2760] (-0.9563,-0.0407)	0.0174[0.2938] (-0.4613,0.4978)
σ^2	1	1.0006[1.1299] (0.8086,1.2344)	1.1039[0.1506] (0.8687,1.3670)	1.0086[0.1244] (0.8155,1.2263)	1.1411[0.1505] (0.9101,1.4059)
ϕ	0.9	0.9051[0.0172] (0.8760,0.9325)	0.8980[0.0183] (0.8673,0.9264)	0.9381[0.0319] (0.8820,0.9881)	0.8996[0.0336] (0.8439,0.9538)
d	0	0	0	-0.0561[0.0443] (-0.1260,0.0193)	-0.0060[0.0460] (-0.0823,0.0697)

У всіх експериментах дискретизацію виконано за методом МКМЛ $M+N=10000$ разів. Перші $M = 2000$ ітерацій були пропущені, а останні $N = 8000$ ітерацій використані для формування апостеріорної вибірки, яка прийнята за наближену вибірку із спільного апостеріорного розподілу всіх невідомих параметрів і прихованих змінних.

Апостеріорні середні і апостеріорні стандартні відхилення розраховані відповідно як вибіркові середні та вибіркові стандартні відхилення. Нижні і верхні границі 90 % апостеріорних інтервалів обчислені як 5-й вибірковий центиль і 95-й вибірковий центиль з апостеріорної вибірки.

Від'ємне значення коефіцієнта γ прийнято завдяки загальновизнаному факту – дисперсія завжди вища при надходженні невістивих новин, ніж хороших. Параметр d був обраний від'ємним в припущенні, що вплив невістивих новин є менш стійким, ніж ефект хороших новин. Остаточна модель формується шляхом комбінування перших двох випадків.



Для кожного ряду було згенеровано $n = 400$ спостережень. Для того щоб підкреслити наявність незначної апріорної інформації про параметри, було використано апріорний розподіл з високим значенням дисперсії.

Таблиця 6.2

Підсумкова статистика з маржинальних апостеріорних розподілів усіх параметрів на основі вибраних моделей для ряду B , істинною є модель V

Істинне значення		S	M	V	F
ψ_0	0.002	0.0051[0.0012] (0.0031,0.0070)	0.0213[0.0041] (0.0143,0.0277)	0.0028[0.0009] (0.0014,0.0042)	0.0072[0.0093] (-0.0086,0.0222)
δ	0	0	-0.0174[0.0042] (-0.0239,-0.0101)	0	-0.0042[0.0094] (-0.0194,0.0116)
ψ_1	0.2	0.2184[0.0256] (0.1757,0.2601)	0.2250[0.0546] (0.1339,0.3138)	0.2165[0.0258] (0.1741,0.2588)	0.2854[0.0572] (0.1926,0.3797)
c	0	0	0.0244[0.0631] (-0.0783,0.1307)	0	-0.0876[0.0652] (-0.1943,0.0208)
α	-0.5	-0.6708[0.1036] (-0.8463,0.5067)	-0.6314[0.1003] (-0.7996,-0.4701)	-0.5999[0.1457] (-0.8426,-0.3647)	-0.6061[0.1339] (-0.8363,-0.3821)
γ	-0.5	0	0	-0.5109[0.2026] (-0.8416,-0.1700)	0.0174[0.2938] (-0.8579,-0.1997)
σ^2	1	2.0983[0.2270] (1.7344,2.4949)	2.0292[0.2225] (1.6761,2.4050)	1.1497[0.1457] (0.9285,1.3903)	1.1677[0.1451] (0.9447,1.4232)
ϕ	0.6	0.8217[0.0240] (0.7811,0.8595)	0.8321[0.0232] (0.7929,0.8698)	0.5772[0.0410] (0.5088,0.6432)	0.5733[0.0387] (0.5086,0.6351)
d	0.3	0	0	0.3131[0.0488] (0.2336,0.3941)	0.3134[0.0467] (0.2358,0.3899)

Як точкові оцінки параметрів, більшість апостеріорних середніх є близькими до істинних значень, при умові що застосована (оцінена) коректна модель. Дійсно, всі істинні значення параметрів лежать в межах 90%-х апостеріорних інтервалів для істинної моделі. У випадку відсутності асиметрії в рівнянні для середнього, обидві моделі S (симетрична модель) і V мають додатне апостеріорне середнє для ψ_1 , коли виконувалась підгонка до ряду A , хоча істинне значення є від'ємним.

І середній і Байєсів інтервал у моделі V припускають, що $\gamma < 0$, але фактично γ дорівнює нулю. Таким чином, у випадку ігнорування асиметрії середнього можна отримати невірний результат. Що стосується повної моделі, то можна стверджувати, що $\gamma = d = 0$, відзначаючи при цьому велике апостеріорне стандартне відхилення і той факт, що обидва байєсівські інтервали включають нуль. В ряді B , дохід є процесом AR(1), дисперсія якого реагує асиметрично до знака минулих значень доходу. Як показано в таблиці 6.2, апостеріорне середнє параметра ψ_1 у всіх підгонках моделі є близьким до відповідних фактичних значень.

Таблиця 6.3

Підсумкова статистика з маржинальних апостеріорних розподілів усіх параметрів на основі вибраних моделей для ряду C , істинною є повна модель F

Істинне значення		S	M	V	F
ψ_0	-0.002	0.0033[0.0004] (0.0027,0.0041)	0.0160[0.0044] (0.0079,0.0227)	0.0030[0.0004] (0.0024,0.0036)	0.0002[0.0078] (-0.0129,0.0130)
δ	0.005	0	-0.0133[0.0044] (-0.0201,-0.0052)	0	0.0025[0.0078] (-0.0102,0.0156)
ψ_1	-0.3	0.2418[0.0245] (0.2012,0.2823)	-0.3439[0.0726] (-0.4643,-0.2295)	0.2329[0.0246] (0.1926,0.2734)	-0.3161[0.0881] (-0.4637,-0.1727)
c	0.6	0	0.6684[0.0777] (0.5447,0.8011)	0	0.5995[0.0916] (0.4500,0.7514)
α	-0.5	-0.6346[0.1058] (-0.8094,0.4661)	-0.6379[0.1064] (-0.8157,-0.4675)	-0.1828[0.1774] (-0.4744,0.1097)	-0.4810[0.1842] (-0.7889,-0.1793)
γ	-0.5	0	0	-0.9720[0.2229] (-1.3407,-0.6056)	-0.5563[0.2309] (-0.9347,-0.1785)
σ^2	1	2.2160[0.2401] (1.8394,2.6282)	2.1570[0.2290] (1.8058,2.5408)	1.0438[0.1205] (0.8616,1.2484)	1.0972[0.1358] (0.8947,1.3360)
ϕ	0.6	0.8629[0.0203] (0.8284,0.8950)	0.8653[0.0201] (0.8315,0.8973)	0.6515[0.0395] (0.5862,0.7155)	0.6122[0.0415] (0.5446,0.6811)
d	0.3	0	0	0.2424[0.0445] (0.1695,0.3148)	0.2922[0.0462] (0.2160,0.3688)

Крім того, значення c є нульовим при налаштуванні моделей M і F . Тому, пояснюється чи не пояснюється асиметрія дисперсії підгонкою моделі, – це не впливає суттєво на результати щодо рівняння для середнього. З таблиці 6.2 також видно, що оцінка σ^2 може бути надзвичайно завищеною, якщо ігнорується асиметрія дисперсії. Нарешті, підсумкова статистика для ряду C представлена в таблиці 6.3. Як і в ряді B , оцінка σ^2 надзвичайно завищена у моделях M та S .

Крім того, видно, що апостеріорні середні забезпечують хороші оцінки для ψ_1 і c в моделях M і F , які пояснюють асиметрію середнього. Це сумісно з попереднім аргументом, що невдача пояснення асиметрії дисперсії, здається суттєво не стосується підгонки середнього. Однак, апостеріорні середні α та γ у моделі V відрізняються істотно від тих, що у моделі F , вказуючи, що ігнорування асиметрії в рівнянні середнього може призвести до неправильної інтерпретації дисперсії. Це суперечить припущенню, використаному в літературі щодо стохастичної волатильності, яке говорить, що інформація про середнє (в деякому наближенні) не залежить від інформації щодо дисперсії. Крім того, результати, наведені в таблиці 6.1 показують, що очевидна асиметрія



дисперсії, наведена в літературі, може виникати внаслідок нехтування асиметрією середнього.

6.5 Аналіз чутливості моделі до припущень

У цьому параграфі виконується аналіз чутливості з метою дослідження, чи є МКМЛ алгоритми високо чутливими до припущень щодо моделі. Перша частина аналізу стосується варіювання значення σ^2 у моделі. Це необхідно для того щоб побачити як впливає дисперсія сигналу на характеристики використаних алгоритмів. В другій частині аналізу розглядається випадок, коли інноваційний сигнал $u(k)$ має розподіл з товстим хвостом в процесі імітаційного моделювання даних. Використані алгоритми були застосовані до модельованих даних з метою визначення чи будуть отримані результати високо чутливими до розподілу інновацій $u(k)$.

Для побудови ряду C було використано всі параметри, крім σ^2 , і було згенеровано три набори даних при $\sigma^2 = 0.5, 0.1$ і 0.01 . Як і в попередньому підрозділі, алгоритми проганялись 10,000 разів; при цьому останні 8000 ітерацій було використано як наближення апостеріорної вибірки. Автокореляційні функції для 8000 ітерацій алгоритму МКМЛ наведені на рис. 6.1 – 6.3.

Що стосується середніх значень параметрів ψ_0, δ, ψ_1 та c , автокореляції цих ітерацій спадають дуже швидко. При $\sigma^2 = 0.01$ поведінка ітерацій відповідає випадковим послідовностям. Навпаки, для параметрів дисперсії $\alpha, \gamma, \sigma^2, \phi$ та d автокореляції ітерацій збільшуються, коли σ^2 зменшується. Імовірно, що причиною такої поведінки є зменшення σ^2 приводить до зменшення відношення сигнал-шум.

Однак, навіть в екстремальному випадку при $\sigma^2 = 0.01$ автокореляції параметрів дисперсії усе ще швидко спадають, що свідчить про швидку збіжність до бажаного спільного апостеріорного розподілу. Крім того, з таблиці 6.4 видно, що всі середні є близькими до відповідних істинних значень параметрів і всі апостеріорні інтервали перекривають істинні значення. Таким чином, є впевненість, що використані алгоритми не будуть давати результати, які можуть призвести до невірної інтерпретації навіть при малому значенні σ^2 .

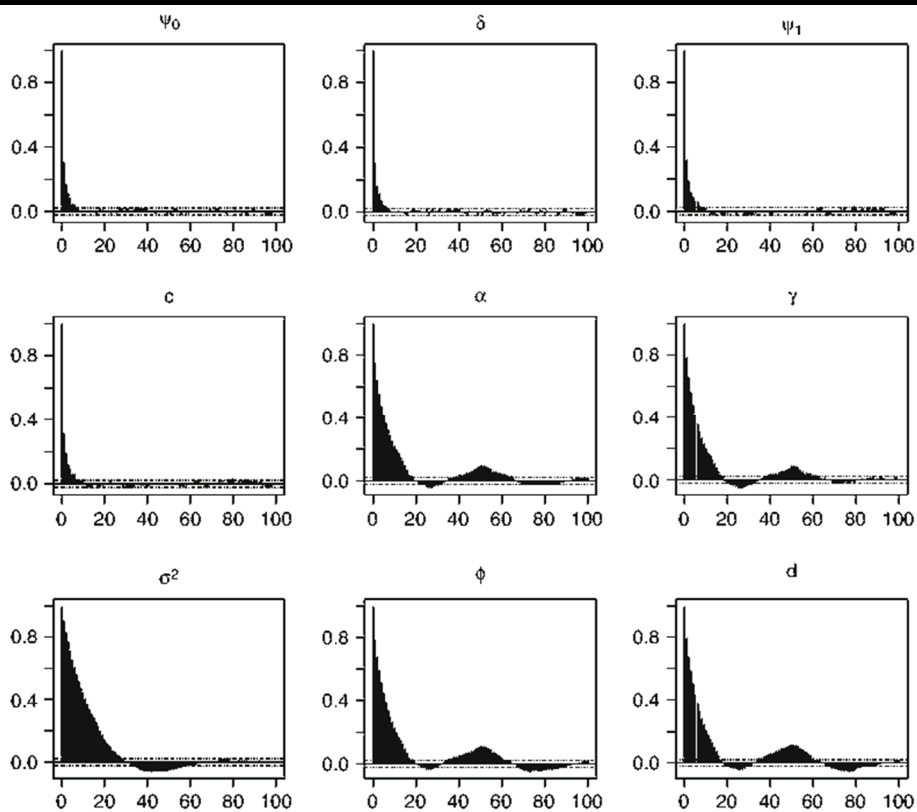


Рисунок 6.1– Автокореляційні функції ітерацій алгоритму МКМЛ, отримані при оцінюванні моделі ПМСВ; $\sigma^2 = 0.5$

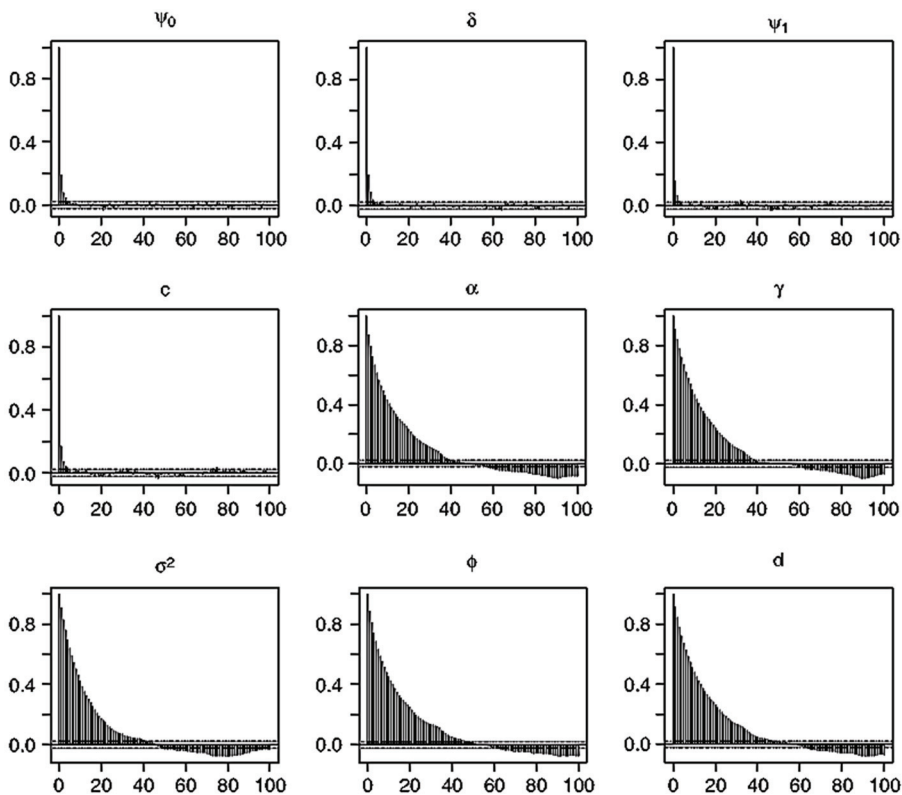


Рисунок 6.2–Автокореляційні функції ітерацій алгоритму МКМЛ, отримані при оцінюванні моделі ПМСВ; $\sigma^2 = 0.1$

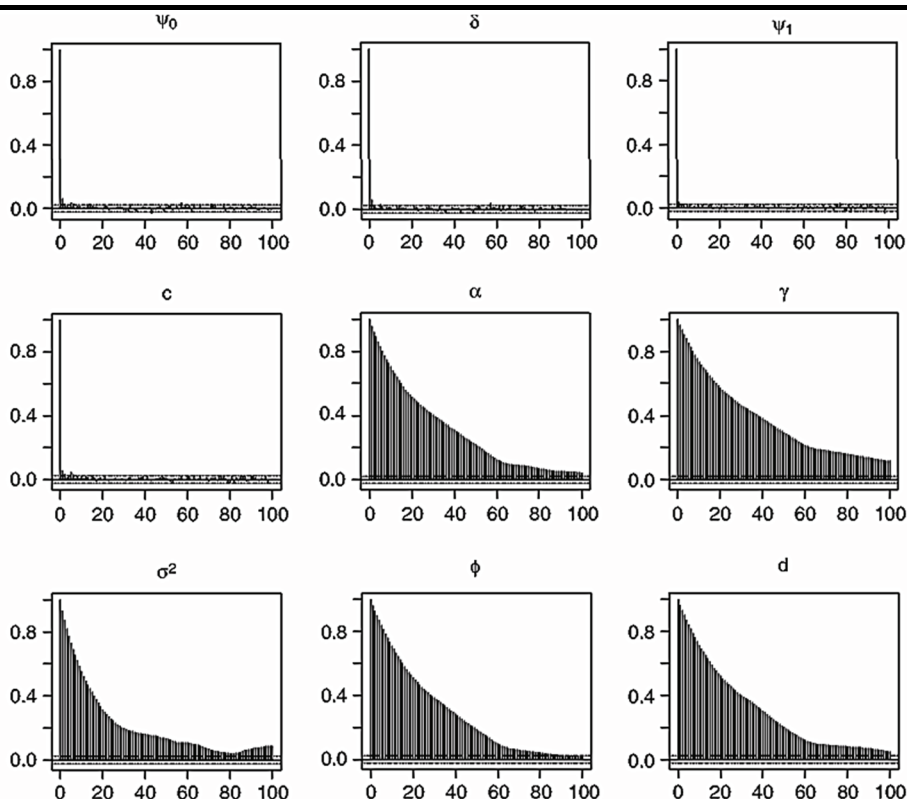


Рисунок.6.3 – Атокореляції МКМЛ ітерацій, отримані при оцінюванні моделі ПМСВ; $\sigma^2 = 0.01$

Таблиця 6.4

Апостеріорне середнє, стандартне відхилення (у квадратних дужках) і 90 % байесівський інтервал (в круглих дужках), отриманий з трьох варіантів ПМСВ моделей при $\sigma^2 = 0.5, 0.1$ та 0.01 .

Істинне значення		$\sigma^2 = 0.5$	$\sigma^2 = 0.1$	$\sigma^2 = 0.01$
ψ_0	-0.002	0,0096 [0.0108] (-0.0076, 0.0276)	0.0042 [0.0116] (-0.0147, 0.0236)	-0.0171 [0.0130] (-0.0385, 0.0041)
δ	0.005	-0.0083 [0.0109] (-0.0265, 0.0091)	-0.0024 [0.0121] (-0.0224, 0.0174)	0.0190 [0.0133] (-0.0027, 0.0412)
ψ_1	-0.3	-0.2733 [0.1001] (-0.4383, -0.1075)	-0.3373 [0.1071] (-0.5130, -0.1597)	-0.3539 [0.0936] (-0.5086, -0.2031)
c	0.6	0.5472 [0.1045] (0.3766, 0.7197)	0.6292 [0.1125] (0.4424, 0.8154)	0.6714 [0.0996] (0.5073, 0.8341)
α	-0.5	-0.5224 [0.1570] (-0.7878, -0.2677)	-0.4877 [0.1488] (-0.7344, -0.2392)	-0.6922 [0.1286] (-0.9085, -0.4879)
γ	-0.5	-0.4329 [0.2163] (-0.7896, -0.0757)	-0.4851 [0.2293] (-0.8608, -0.1057)	-0.3448 [0.1816] (-0.6457, -0.0548)
σ^2	0.5;0.1;0.001	0.4855 [0.0826] (0.3564, 0.6256)	0.1428 [0.0344] (0.0906, 0.2042)	0.0303 [0.0194] (0.0049, 0.0685)
ϕ	0.6	0.5835 [0.0412] (0.5163, 0.6505)	0.6209 [0.0374] (0.5607, 0.6833)	0.5694 [0.0333] (0.5120, 0.6217)
d	0.3	0.3293 [0.0523] (0.2439, 0.4155)	0.2868 [0.0541] (0.1948, 0.3750)	0.3237 [0.0417] (0.2548, 0.3918)

Для того щоб продемонструвати, що використані алгоритми є робастними до розподілу $u(k)$, було згенеровано три набори даних з асиметрією в моделі середнього і дисперсії для ряду C , за винятком того, що $u(k)$ мало стандартизований t -розподіл із степенями свободи $\nu = 8, 10$ і 12 . Стандартизований t -розподіл означає, що дисперсія розподілу дорівнює одиниці, тобто $\text{var}[u(k)]=1$. Було обрано три значення для ступенів свободи $8, 10$ і 12 , тому цей діапазон значень ν рекомендується в літературі із стохастичної волатильності для випадку, коли базисна модель стохастичної волатильності підганяється до реальних даних.

Алгоритми було застосовано до змодельованих даних в припущенні, що процес відповідає параметрам $N(0,1)$. В таблиці 6.5 наведено підсумкові результати виконання алгоритму МКМЛ. Хоча розподіл $u(k)$ було специфіковано неправильно, як нормальний, у виводі Байеса, встановлено, що більшість апостеріорних середніх є досить близькими до відповідних істинних значень. Дійсно, всі наступні інтервали включають істинні значення, крім σ^2 при $\nu = 10$.

Таблиця 6.5

Апостеріорне середнє, стандартне відхилення (у квадратних дужках) і 90 % байесівський інтервал (в круглих дужках), отриманий з трьох варіантів ПМСВ моделей, для випадку, коли розподіл u_t отримано за допомогою стандартизованого t -розподілу з степенями свободи $\nu = 8, 10$ і 12

Істинне значення		$\nu = 8$	$\nu = 10$	$\nu = 12$
ψ_0	-0.002	-0,0044 [0.0090] (-0.0191, 0.0102)	0.0098 [0.0085] (-0.0040, 0.0238)	-0.068 [0.0106] (-0.0244, 0.0102)
δ	0.005	0.0055 [0.090] (-0.0095, 0.0202)	-0.0065 [0.0085] (-0.0207, 0.0072)	0.0098 [0.0106] (-0.0073, 0.0274)
ψ_1	-0.3	-0.2346 [0.0695] (-0.3507, -0.1218)	-0.2902 [0.0809] (-0.4217, -0.1568)	-0.3942 [0.0845] (-0.5316, -0.2558)
c	0.6	0.5390 [0.0761] (0.4164, 0.6646)	0.5814 [0.0844] (0.4420, 0.7191)	0.6939 [0.0885] (0.5462, 0.8380)
α	-0.5	-0.5291 [0.1806] (-0.8319, -0.2377)	-0.4923 [0.2009] (-0.8256, -0.1720)	-0.7229 [0.1928] (-1.0431, -0.4138)
γ	-0.5	-0.5339 [0.2546] (-0.9566, -0.1174)	-0.8640 [0.2571] (-1.2917, -0.4403)	-0.3332 [0.2313] (-0.7071, 0.0460)
σ^2	1	1.2070 [0.1542] (0.9736, 1.4854)	1.2388 [0.1480] (1.0057, 1.5050)	1.1291 [0.1324] (0.9254, 1.3685)
ϕ	0.6	0.6087 [0.0431] (0.5365, 0.6795)	0.5803 [0.0450] (0.5049, 0.6523)	0.5272 [0.0467] (0.4486, 0.6030)
d	0.3	0.2761 [0.0566] (0.1826, 0.3697)	0.2667 [0.0515] (0.1820, 0.3516)	0.3794 [0.0493] (0.2988, 0.4629)

Алгоритми було застосовано до змодельованих даних в припущенні, що процес відповідає параметрам $N(0,1)$. В таблиці 6.5 наведено підсумкові



результати виконання алгоритму МКМЛ. Хоча розподіл $u(k)$ було специфіковано неправильно, як нормальний, у виводі Байеса, встановлено, що більшість апостеріорних середніх є досить близькими до відповідних істинних значень. Дійсно, всі наступні інтервали включають істинні значення, крім σ^2 при $\nu = 10$.

Очевидним ефектом того, що $u(k)$ має розподіл з товстим хвостом, є перевизначення σ^2 . Однак, цей результат залишається прийнятним в тому смислі, що апостеріорні середні є близькими до істинних значень. З апостеріорної підсумкової статистики, наведеної в таблиці 6.5, очевидно, що і асиметрія середнього, і асиметрія дисперсії є дуже істотними факторами. Таким чином результат залишається прийнятним навіть якщо ігнорується наявність товстого хвоста розподілу значень ряду $x(k)$. В цілому було встановлено, що використані алгоритми не дуже чутливі до припущень щодо моделі і генерують точні результати навіть при малому значення σ^2 або у випадку, коли розподіл $x(k)$ має помірні товсті хвости, що відповідає реальностям, наведеним в літературі із стохастичної волатильності.

6.6 Застосування методу до реальних даних

Модель ПМСВ було застосовано до двох реальних рядів даних, які взяті з мережі Інтернет [125, 126]. Це щоденний індекс цін на біржі Hang Seng (HSI) Гонконгу і щоденний індекс біржі США Standard&Poor 500 (SPI) – на часовому інтервалі з 1981 по 1986 роки. Для кожного побудовано повну модель з асиметрією в рівняннях для середнього та дисперсії. Так само як і в процесі імітаційного моделювання число ітерацій алгоритму вибрано 10000. Останні 8000 ітерацій було використано як апостеріорну вибірку. В таблиці 6.6 наведено апостеріорне середнє, апостеріорне стандартне відхилення та 90% апостеріорний інтервал для всіх невідомих параметрів плюс деякі інші цікаві параметри. На рис. 6.4 – 6.9 показано результати ітерацій алгоритму МКМЛ, автокореляції функції ітерацій і маржинальні апостеріорні щільності для всіх параметрів. Оцінки апостеріорної щільності отримані за допомогою гаусового алгоритму (*Splus*) оцінювання густини ядра з належним вибором ширини вікна.

З графіків для сліду та коефіцієнтів автокореляції для обох індексів спостерігається швидка збіжність до апостеріорного розподілу. Автокореляції ітерацій спадають швидко для всіх параметрів. З порівняння рис. 6.5 та 6.8 випливає, що для індексу HSI збіжність є кращою ніж для SPI. Це можна пояснити тим, що HSI має більше значення σ^2 , тобто більше значення відношення сигнал-шум. Крім того, ітерації алгоритму МКМЛ для параметрів у рівнянні для середнього поведуться подібно некорельованим послідовностям. Це вказує на те, що генерування вибірки за алгоритмом МКМЛ дає можливість створити наближені випадкові вибірки для параметрів ψ_0, δ, ψ_1 та c .

Таблиця 6.6

Апостеріорне середнє, стандартне відхилення (у квадратних дужках) і 90 % байєсівський інтервал (у круглих дужках) з повної моделі для індексу Hong Kong Seng і S&P 500 індексів на протязі 1981-1986 років

Параметри	Індекс HSI	Індекс SPI
ψ_0	-0.006 [0.0008] (-0.0019, 0.0007)	0.0004 [0.0005] (-0.0003, 0.0012)
δ	0.0006 [0.0011] (-0.0011, 0.0023)	0.0002 [0.0006] (-0.0008, 0.013)
$\psi_0 + \delta$	-0.0000 [0.0008] (-0.0013, 0.0012)	0.0006 [0.0004] (-0.0001, 0.0014)
ψ_1	-0.0388 [0.0587] (-0.1353, 0.0565)	0.1280 [0.0608] (0.0278, 0.2283)
c	0.2241 [0.0821] (0.0909, 0.3589)	-0.0990 [0.0816] (-0.2330, 0.0356)
$\psi_1 + c$	0.1853 [0.0556] (0.0935, 0.2765)	0.0290 [0.0541] (-0.0611, 0.1181)
α	-0.0300 [0.0987] (-0.2213, 0.0836)	-0.3409 [0.3193] (-0.9654, -0.0054)
γ	-0.8903 [0.2276] (-1.3217, -0.4695)	-0.1231 [0.4565] (-0.8017, 0.6918)
$\alpha + \gamma$	-0.9203 [0.2161] (-1.2976, -0.5851)	-0.4640 [0.2754] (-0.9731, -0.0782)
σ^2	0.0571 [0.0157] (0.0356, 0.0858)	0.0195 [0.0106] (0.0070, 0.0391)
ϕ	0.9869 [0.0117] (0.9640, 0.9992)	0.9628 [0.0329] (0.8984, 0.9975)
d	-0.0871 [0.0297] (-0.1354, -0.0384)	-0.0094 [0.0470] (-0.0793, 0.0736)
$\phi + d$	0.8997 [0.0251] (0.8565, 0.9385)	0.9534 [0.0286] (0.9004, 0.9932)
$-\frac{\gamma}{d}$	-10.5475 [8.0815] (-12.5146, -9.3336)	-9.8646 [32.8908] (-16.5332, -3.6758)

Що стосується асиметрії середнього в HSI, то з апостеріорного середнього ψ_1 та $\psi_1 + c$ для індексу HSI видно, що коефіцієнти авторегресії в рівнянні для доходів мають протилежні знаки. Можна стверджувати, що ψ_1 , власне кажучи, дорівнює нулю, тому що 90% апостеріорний інтервал включає нуль. Однак, апостеріорний характер маржинального розподілу для HSI строго від'ємний, хоча і дуже близький до нуля. Це свідчить про від'ємність ψ_1 для HSI. Навпаки, $\psi_1 + c$ додатній для випадку, коли $r_{t-1} \geq 0$. Протилежність знаків коефіцієнтів авторегресії була встановлена в декількох роботах, що пояснювало слабку автокорельованість ряду доходу. Цей результат є сумісним з іншими їхніми результатами, що ґрунтуються на побудові порогової моделі АРУГ.

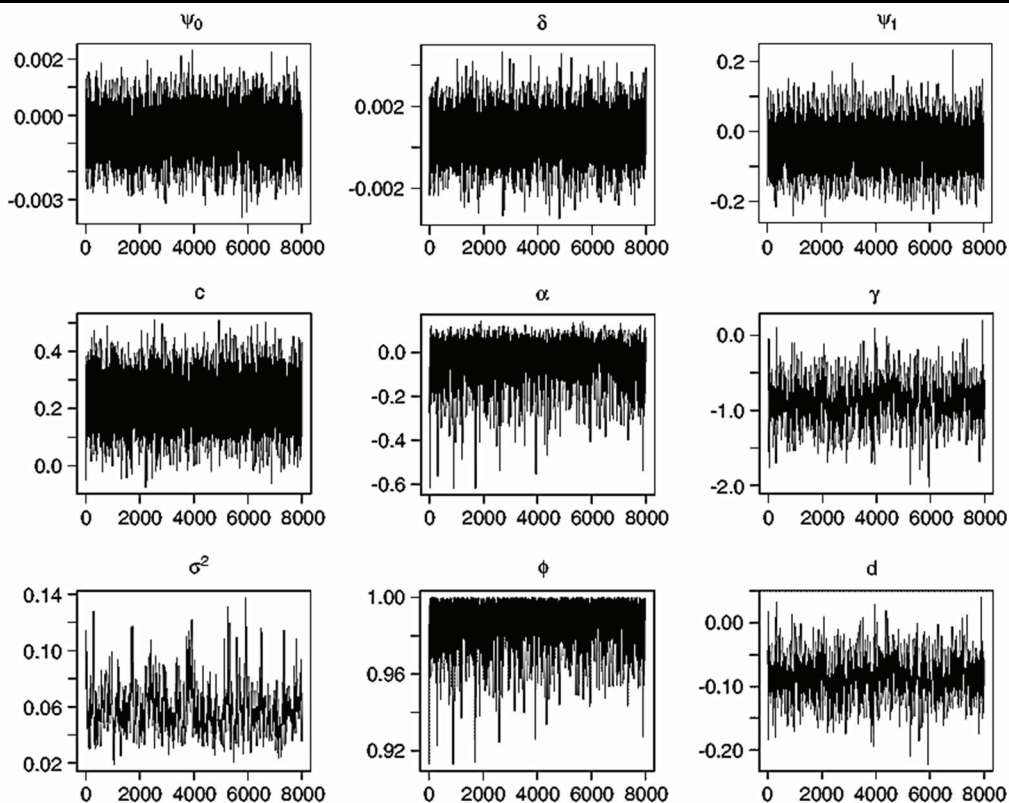


Рисунок 6.4 – Зображення 8000 МКМЛ ітерацій, отримані при оцінюванні моделі ПМСВ для HSI

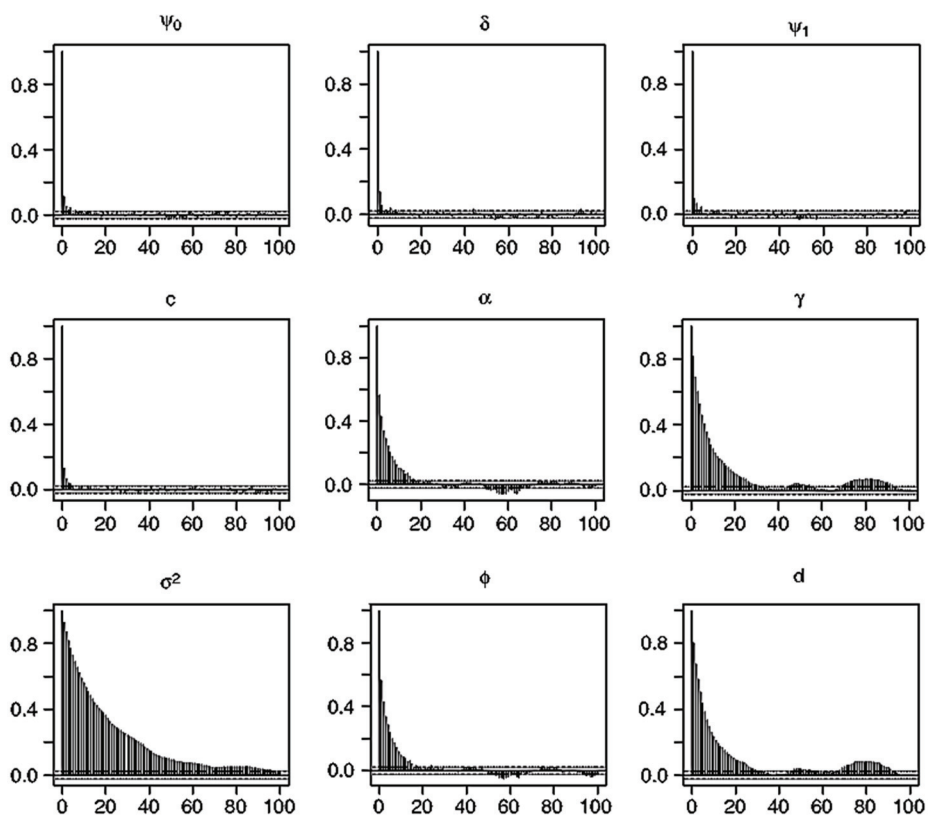


Рисунок 6.5 – Автокореляція ітерацій алгоритму МКМЛ, отримана при оцінюванні моделі ПМСВ для ряду HSI

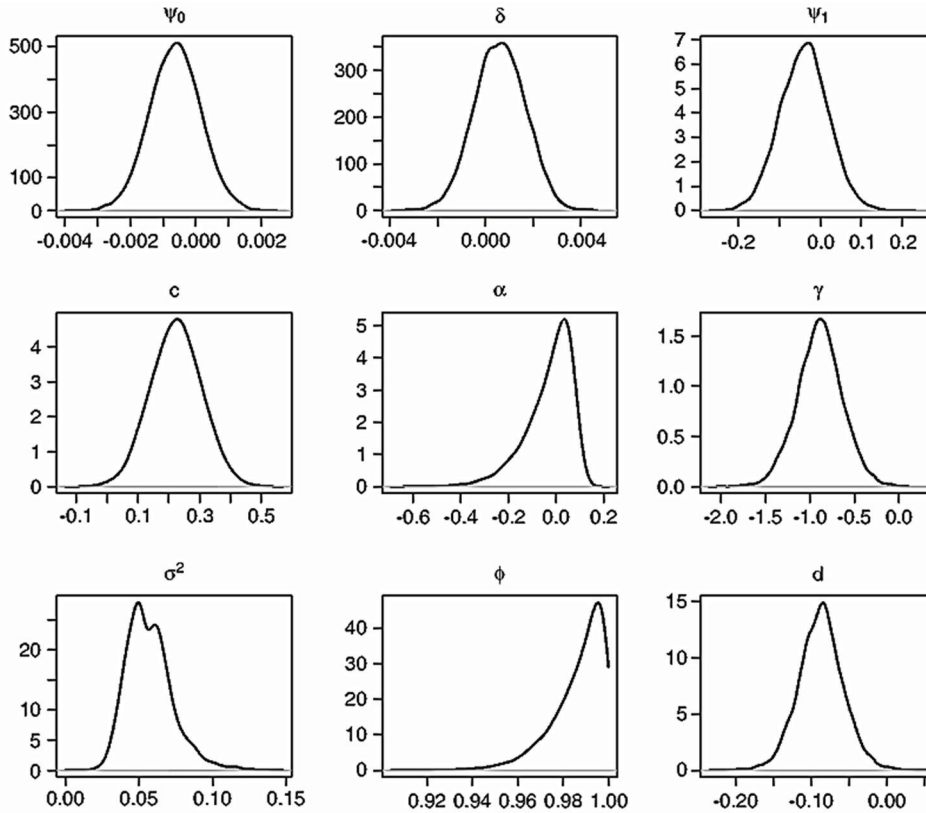


Рисунок 6.6 – Апостеріорні середні для невідомих параметрів моделі ПМСВ при оцінюванні моделі для індексу HSI

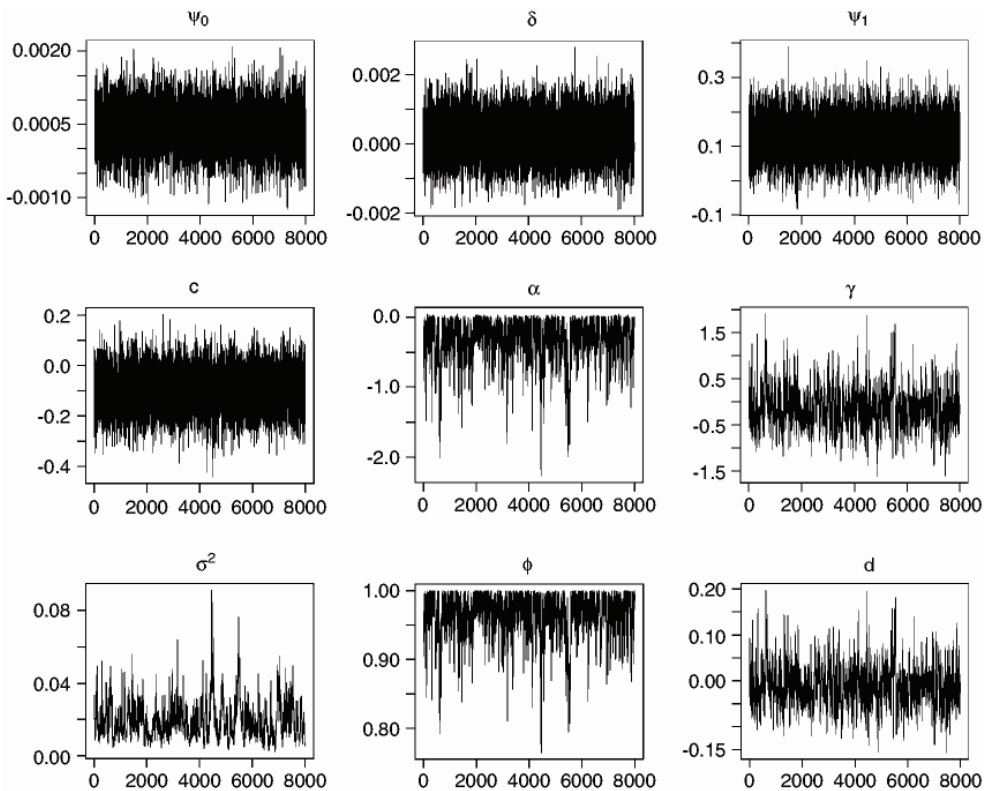


Рисунок 6.7 – Характеристики 8000 ітерацій алгоритму МКМЛ, отриманих при побудові моделі ПМСВ для ряду SPI

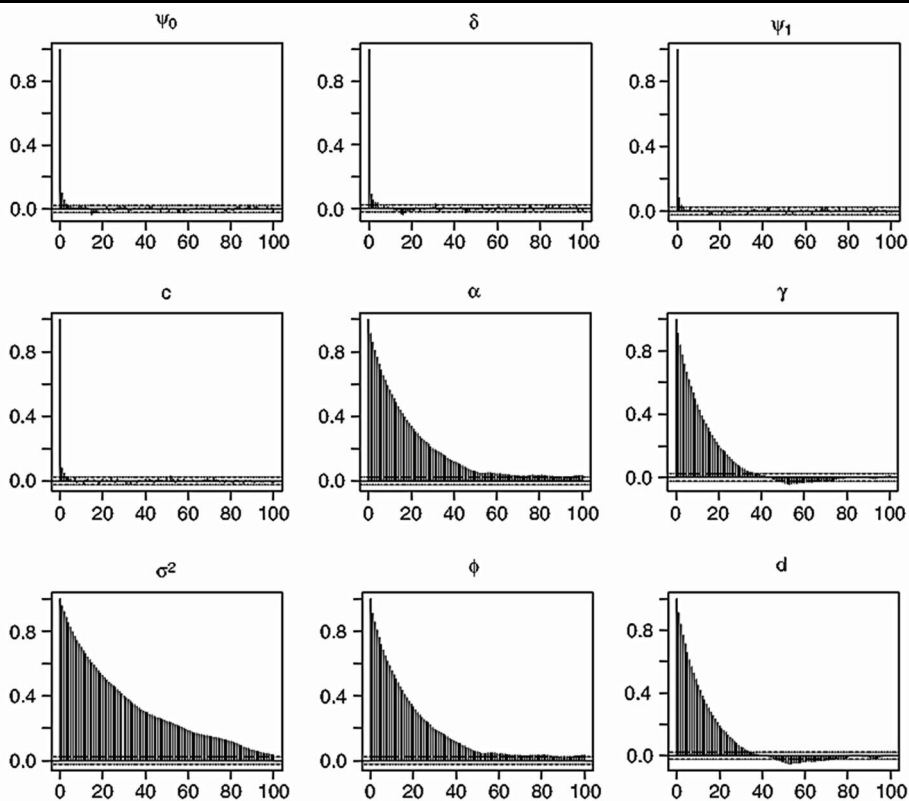


Рисунок 6.8 – Автокореляційні функції ітерацій алгоритму МКМЛ, отримані при оцінюванні моделі ПМСВ для ряду SPI

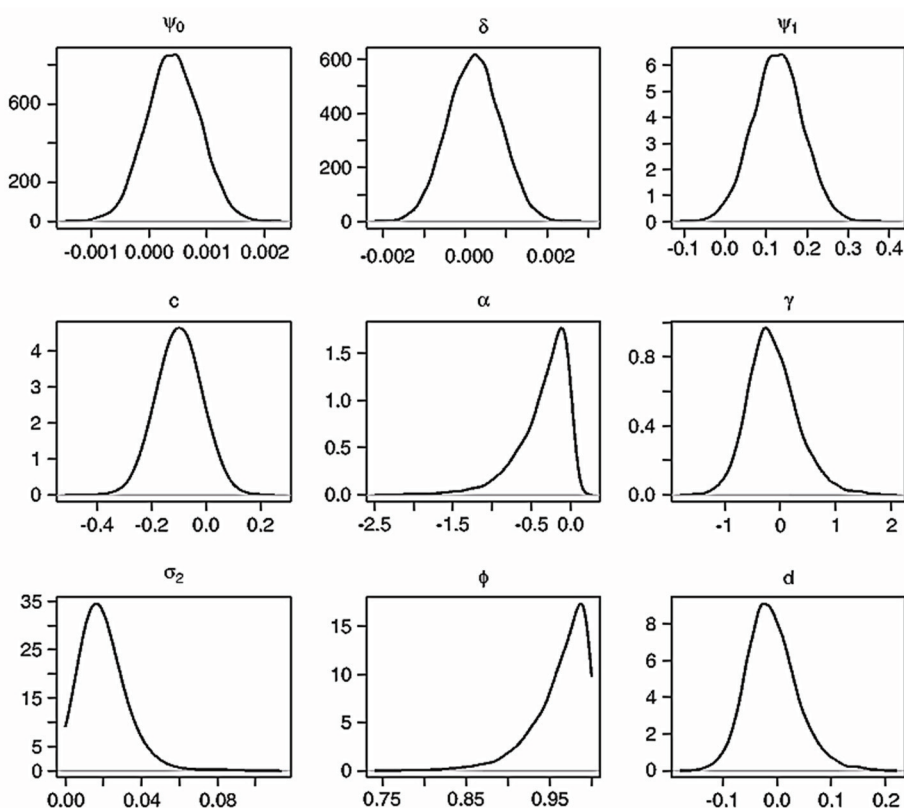


Рисунок 6.9 – Апостеріорні середні для невідомих параметрів, отриманих при оцінюванні моделі ПМСВ для ряду SPI

В ряді даних, отриманих з біржі S&P 500, явище асиметрії середнього не спостерігається. Апостеріорні середні для ψ_1 і $\psi_1 + c$ виявились в даному випадку додатними. Враховуючи асиметрію дисперсії, було встановлено, що апостеріорні середні змінної γ є від'ємними як для HSI, так і для ряду S&P 500. Обидва апостеріорні режими функціонування процесів розташовані у від'ємній області дійсної осі. Цей факт підтримує припущення, що дисперсія є вищою коли новини розцінюються як невтїшні. Оскільки розходження $\log h(k)$ між цими двома режимами визначається через $-\gamma - d \log h(k-1)$, то згадане припущення має силу тоді і тільки тоді, коли $-\gamma - d \log h(k-1)$ завжди додатне, тобто

$$\log h(k-1) > -\frac{\gamma}{d} \text{ для всіх } t, \quad (6.10)$$

якщо коефіцієнт d від'ємний.

З таблиці 6.6 видно, що апостеріорні середні для $-\gamma/d$ мають значення -10.55 і -9.86 для HSI і SPI, відповідно. Якщо прийняти за апостеріорні середні точкові оцінки, то нерівність (6.10) виконується коли волатильність $\sqrt{h(k-1)}$ перевищує значення 0.0051 і 0.0072 для HSI і SPI, відповідно. При таких обставинах дисперсія буде вищою якщо новини невтїшні, ніж при хороших новинах. На додаток до вищезгаданої асиметрії дисперсії, загальним спостереженням, зазначеним в літературі з стохастичної волатильності та побудови моделей АРУГ є те, що дисперсія є суттєво наявною в процесі. Це пояснює результати отримані щодо змінних $\phi_0 = \phi$ і $\phi_1 = \phi + d$, які стосуються побудови моделі ПМСВ. Апостеріорні середні, апостеріорні байєсівські інтервали і оцінки щільності ядра для ϕ_0 і ϕ_1 показують, що ϕ_0 і ϕ_1 є близькими до одиниці. Крім того, з таблиці 6.6 і рис.6.9 видно, що d має, власне кажучи, нульове значення для SPI. Однак, d можна розглядати як від'ємне для HSI із значенням -0.1 , про що свідчать апостеріорне середнє і апостеріорний режим процесу.

Інший вид асиметрії дисперсії було виявлено для даних, отриманих з HSI. З від'ємності d випливає, що дисперсія, яка є результатом невтїшних новин, є стійкішою ніж та, що є результатом хороших новин. Ймовірно це пояснює те, що ефект від хороших новин може бути “сприйнятий і опрацьований” Гонконгським ринком акцій швидше ніж ефект невтїшних новин. Необхідно підкреслити, що такий вид асиметрії параметра стійкості ще недостатньо розглядався в літературі із стохастичної мінливості. Отримані результати свідчать, що модель ПМСВ може допомогти при розкритті спеціальних структур дисперсії, наприклад, асиметрії стійкості, що має місце в індексі HSI.



Для того щоб прогнозувати майбутню мінливість і доходи, тобто $\sqrt{h(n+1)}$ і $x(n+1)$, можна скористатись методами, описаними вище. На основі 8000 ітерацій алгоритму МКМЛ можна оцінити прогноз густини $f(\sqrt{h(n+1)} | X_n)$ та $f(x(n+1) | X_n)$ стандартними методами прогнозування ядра. На рис.6.10 показано прогнозовані значення густини HSI і SPI.

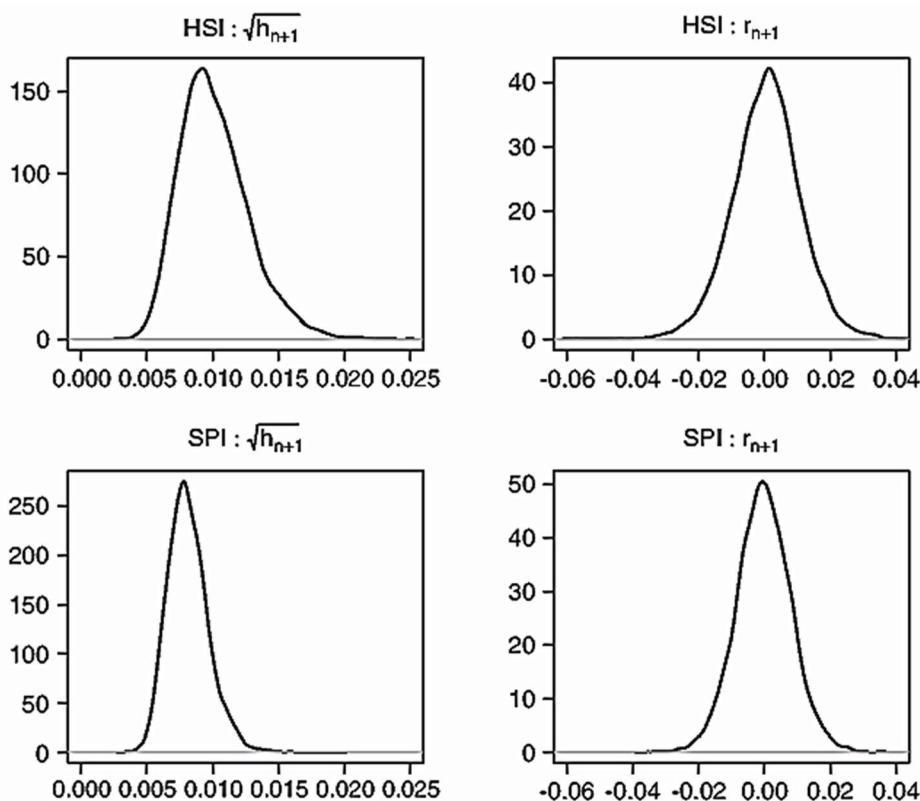


Рисунок 6.10. – Однокроковий прогнозуючий розподіл volatility,

$f(\sqrt{h_{n+1}} | X_n)$ і дохід $f(x_{n+1} | X_n)$ для HSI і SPI зроблені з використанням 8000 МКМЛ ітерацій, де $k = n + 1$ – перший день торгівлі.

Точкові оцінки прогнозу для $\sqrt{h(n+1)}$ також можуть бути побудовані на основі апостеріорних середніх. Це значення 0.0101 і 0.0081 для HSI та SPI. З двох графіків волатильності густини, волатильність HSI більш розсіяна і має більшу площу. Це вказує на те, що HSI є більш мінливим ніж SPI в перший день торгівлі в 1987 році ($k = n + 1$). Обидва розподіли мають деякий перекош (асиметричні). Що стосується розподілів доходів, то обидві прогнозуючі густини $x(n+1)$ є грубо симетричними і мають близькі до нуля моди. Як і варто було очікувати, з розподілів волатильності видно, що значення $x(n+1)$ для HSI є більш розсіяними. З модельної вибірки $x(n+1)^{(i)}$ можна також спрогнозувати, що 1-й перцентиль буде мати значення – 0.0257 і – 0.0199 для HSI і SPI, відповідно. Їх можна інтерпретувати як значення VaR в тому смислі, що HSI (SPI) зменшується більше, ніж на 2.57% (1.99 %) з імовірністю 0.01.

Аналіз результатів

Таким чином, вище було наведено новий підхід до одночасного моделювання асиметрії середнього і асиметрії дисперсії часових рядів. Основний результат полягає у розширенні звичайної симетричної моделі стохастичної волатильності до порогової моделі стохастичної волатильності. Встановлення порогового значення в моделі ПМСВ дозволяє середньому та дисперсії реагувати по-різному на попередні значення часового ряду. Властивості моделі дають можливість відтінити асиметричну поведінку біржових акцій, а також описати відомий ефект левериджу в рамках структури стохастичної волатильності. Розроблена схема генерації послідовностей алгоритмом МКМЛ із спільної апостеріорної густини невідомих параметрів та прихованих змінних. Генерування вибірки здійснюється відносно просто завдяки тому що ми використовували, головним чином, стандартні типи розподілів. Інформацію щодо будь-якої функції параметрів можна отримати шляхом маржиналізації.

Для імітаційного моделювання було використано чотири моделі. Це модель симетричної стохастичної волатильності, модель асиметрії тільки в середньому, модель з асиметрією тільки в дисперсії і повна порогова модель стохастичної волатильності. В цілому метод на основі використання МКМЛ дозволяє успішно вирішити проблему оцінювання. За точкові оцінки невідомих параметрів прийнято апостеріорні середні, які є близькими до істинних значень. Відсутність можливості пояснити асиметрію дисперсії суттєво не впливає на оцінку середнього.

Однак, ігнорування асиметрії в рівнянні для середнього може призвести до невірної інтерпретації поведінки дисперсії. Таким чином, очевидна асиметрія дисперсії, відзначена у фаховій літературі, може бути наслідком неврахування асиметрії середнього. Крім того, дослідження чутливості свідчить про те, що використані алгоритми є в незначній мірі чутливими до припущень щодо моделі. Можна отримати надійні результати оцінювання параметрів навіть у випадку, коли σ^2 приймає малі значення або розподіл вибірки $u(k)$ має товстий хвіст.

Модель ПМСВ застосована до щоденних значень індексів доходів HSI і S&P 500 за період з 1981 по 1986 роки. На основі маржинальних апостеріорних щільностей ψ_1 та $\psi_1 + c$ встановлено, що параметр ψ_1 , імовірноше всього, буде від'ємним, а $\psi_1 + c$ – додатним для індексу HSI. Цей факт необхідний для пояснення того, чому важко відхилити гіпотезу ефективності ринку. Крім того, було встановлено, що дисперсії обох рядів ставали більшими після падіння на ринку, ніж після його піднесення. Є також деяке свідчення того, що дисперсія ряду доходу HSI є менш стійкою після хороших новин ніж після невтішних. Вважаємо за доцільне дослідити



– чому це явище існує виключно на Гонконгській біржі, але воно не спостерігається на біржах США. Також виявилось, що значення індексу HSI, прогнозоване на перший торговий день 1987 року, з використанням щільностей $f(\sqrt{h(n+1)} | X_n)$ та $f(x(n+1) | X_n)$, виявилось більш мінливим ніж значення SPI. При цьому прогноз індексу HSI має, також, менше 1%-е значення VaR.

Очевидно, що в цілому модель ПМСВ дозволяє ідентифікувати складні асиметричні структури моделей рядів доходу в припущенні існування нелінійних порогових процесів в рівняннях середнього і дисперсії. Очікується, що така асиметрична модель стохастичної буде корисною для прогнозування дисперсії та при дослідженні інших властивостей фінансових процесів. Очевидно, що розглянуту методологію можна узагальнити на випадки, коли середнє і дисперсія описуються пороговими авторегресійними процесами вищих порядків. Це потребує окремого дослідження.

5.7 Умови імітаційного моделювання

В проведеному дослідженні було використано незалежні апіорні розподіли, хоча можна використовувати також залежні. Для x_0 і θ були вибрані стандартні спряжені апостеріорні розподіли. Для x_0 і $\theta_{-\sigma^2}$ вибрано нормальні апіорні розподіли, а для σ^2 – обернений гамма-розподіл $IG(\nu_1, \nu_2)$. Область визначення для параметрів ψ_1, c, ϕ була обмежена таким чином, щоб і ψ_{1s_k} , і ϕ_{s_k} мали величини менші одиниці.

Спільна умовна апостеріорна густина для параметрів ψ_0, c, ψ_1 та x_0 отримана з виразу (6.8) таким чином:

$$f(\psi_0, \delta, \psi_1, c, x_0 | \theta_{-(\psi_0, \delta, \psi_1, c)}, H_n, X_n) \propto \exp \left\{ - \sum_{k=1}^n \frac{1}{2h_k} [x_k - \psi_0 - \delta s_k - (\psi_1 + c s_k) x_{k-1}]^2 \right\} f(\psi_0, \delta, \psi_1, c, x_0). \quad (6.10)$$

Апіорний розподіл для x_0 взято з параметрами $N(\kappa_1, \kappa_2)$, де κ_1 і κ_2 можуть бути замінені на вибіркове середнє і вибірккову дисперсію, відповідно. Повна умовна щільність розподілу для x_0 має вигляд:

$$x_0 | \omega_{-x_0} \sim N \left(\frac{\kappa_2 \frac{1}{h_1} (r_1 - \psi_0 - \delta s_k) (\psi_1 + c s_k) + \kappa_1}{\kappa_2 \frac{(\psi_1 + c s_k)^2}{h_1} + 1}, \frac{\kappa_2}{\kappa_2 \frac{(\psi_1 + c s_k)^2}{h_1} + 1} \right).$$

Повний умовний розподіл для середніх значень вектора параметрів $\beta = [\psi_0, \delta, \psi_1, c]^T$ можна отримати з (6.10) як

$$f(\beta | \theta_{-\beta}, x_0, H_n, X_n) \propto \exp \left\{ -\frac{1}{2} (\beta - \mu)^T \Sigma^{-1} (\beta - \mu) - \frac{1}{2} (\beta - \mu_0)^T \Lambda_0^{-1} (\beta - \mu_0) \right\} \times I(|\psi_1 + c| < 1), \quad I(|\psi_1| < 1),$$

де $N(\mu_0, \Lambda_0)$ $I(|\psi_1 + c| < 1)I(|\psi_1| < 1)$ – апіорний розподіл; $I(\cdot)$ – індикаторна функція;

$$\mu = (\mathbf{X}^T, \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}, \quad \Sigma = (\mathbf{X}^T \mathbf{X})^{-1}.$$

$$\mathbf{X} = \begin{pmatrix} 1/\sqrt{h_1} & s_1/\sqrt{h_1} & x_0/\sqrt{h_1} & s_1 x_0/\sqrt{h_1} \\ \vdots & \vdots & \vdots & \vdots \\ 1/\sqrt{h_k} & s_k/\sqrt{h_k} & x_{k-1}/\sqrt{h_k} & s_k x_{k-1}/\sqrt{h_k} \\ \vdots & \vdots & \vdots & \vdots \\ 1/\sqrt{h_n} & s_n/\sqrt{h_n} & x_{n-1}/\sqrt{h_n} & s_n x_{n-1}/\sqrt{h_n} \end{pmatrix}$$

Байєсова оцінка параметрів регресії має вигляд:

$$\beta | \theta_{-\beta}, x_0, H_n, X_n \sim N(\mu_p, \Sigma_p) I(|\psi_1 + c| < 1) I(|\psi_1| < 1),$$

де

$$\Sigma_p^{-1} = \Sigma^{-1} + \Lambda_0^{-1} \quad \text{та} \quad \mu_p = \Sigma_p (\Sigma^{-1} \mu + \Lambda_0^{-1} \mu_0).$$

Таким чином, середні значення параметрів можна обчислити на основі наведеного вище редукованого багатовимірного нормального розподілу, використовуючи генерацію вибірки за методом резекції.

Зазначимо, що член $f(h_1 | \theta, x_0)$ у (6.8) мало впливає на спільний умовний апостеріорний розподіл $f(\omega_i | \omega_{-i}, X_n)$. Нехтуючи цим членом, отримаємо наступний вираз для розподілу:

$$f(\alpha, \gamma, \phi, d | \theta_{-(\alpha, \gamma, \phi, d)}, H_n, X_n) \propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{k=2}^n [\log h_k - \alpha - \gamma s_k - (\phi + d s_k) \log h_{k-1}]^2 \right\} f(\alpha, \gamma, \phi, d). \quad (6.11)$$

Аналогічно, розглянемо вектор параметрів рівняння для дисперсії $\beta = [\alpha, \gamma, \phi, d]^T$. Використовуючи (6.11) з апіорним розподілом



$N(\mu_0, \Lambda_0) I(|\psi_1 + c| < 1) I(|\psi_1| < 1)$, отримаємо наступний вираз для розподілу параметрів

$$f(\beta | \theta_{-\beta}, x_0, H_n, X_n) \propto \exp \left\{ -\frac{1}{2} (\beta - \mu)' (\sigma^2 \Sigma)^{-1} (\beta - \mu) - \frac{1}{2} (\beta - \mu_0)' \Lambda_0^{-1} (\beta - \mu_0) \right\} \times I(|\phi + d| < 1) I(|\phi| < 1)$$

де

$$\mu = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}, \quad \Sigma = (\mathbf{X}^T \mathbf{X})^{-1},$$

де $N(\mu_0, \Lambda_0) I(|\psi_1 + c| < 1) I(|\psi_1| < 1)$ – це апіорні розподіли, $I(\cdot)$ – індикаторна функція;

$$\mathbf{X} = \begin{pmatrix} 1 & s_2 & \log h_1 & s_2 \log h_1 \\ & & \vdots & \\ 1 & s_n & \log h_{t-1} & s_t \log h_{t-1} \\ & & \vdots & \\ 1 & s_n & \log h_{n-1} & s_n \log h_{n-1} \end{pmatrix} \quad \text{та} \quad \mathbf{y} = \begin{pmatrix} \log h_2 \\ \vdots \\ \log h_t \\ \vdots \\ \log h_n \end{pmatrix}.$$

І насамкінець, розподіл вектора параметрів має вигляд:

$$\beta | \theta_{-\beta}, x_0, H_n, X_n \sim N(\mu_p, \Sigma_p) I(|\phi + d| < 1) I(|\phi| < 1),$$

де

$$\Sigma_p^{-1} = \sigma^{-2} \Sigma^{-1} + \Lambda_0^{-1} \quad \text{та} \quad \mu_p = \Sigma_p (\sigma^{-2} \Sigma^{-1} + \Lambda_0^{-1} \mu_0).$$

Перелік посилань

1. *Бідюк П.І.* Системний підхід до прогнозування на основі моделей часових рядів. ISSN 1681–6048 System Research & Information Technologies, 2003, № 3, с.88-110.
2. *Бідюк П.І., Баклан І.В., Рифа В.Н.* Системный подход к построению регрессионной модели по временным рядам // Системні дослідження та інформаційні технології. — 2002. — № 3. — С. 114–131.
3. *Бідюк П.І., Гожий О.П.* Ймовірісно-статистичні методи моделювання і прогнозування: монографія. Миколаїв: Вид-во ЧДУ ім. Петра Могили, 2014. 440 с.
4. *Бідюк П.І., О. Гожий О.П., Коновалюк М.М.* Прогнозування волатильності валютного ринку за нелінійними моделями. Національний університет «Львівська політехніка» Lviv Polytechnic National University Institutional Repository <http://ena.lp.edu.ua>; 2013, с. 257-265.
5. *Бідюк П.І., Кожухівська О.А.* Моделювання і короткострокове прогнозування гетероскедастичних процесів. Індуктивне моделювання складних систем, випуск 4, 2012, с. 48-63.
6. *Бідюк П.І., Коновалюк М.М.* Визначення величини ризику VAR на основі оцінок параметрів моделі стохастичної волатильності. Системні дослідження та інформаційні технології, 2012, № 3, с. 85-94.
7. *Бідюк П.І., Коновалюк М.М., Кожухівська О. А.* Оцінювання моделей стохастичної волатильності та УАРУГ на Java // Наукові праці [Чорноморського державного університету імені Петра Могили]. Сер.: Комп'ютерні технології. - 2012. - Т. 191, Вип. 179. - С. 14-20. - Режим доступу: http://nbuv.gov.ua/UJRN/Npchduct_2012_191_179_4
8. *Бідюк П.І., Коновалюк М.М.* Прогнозування волатильності валютного ринку за нелінійними моделями // Вісник Національного університету «Львівська політехніка», № 719. – 2011 р., с. 154 – 163.
9. *Бідюк П.І., Меньяйленко О.С., Половцев О.В.* Методи прогнозування. – Луганськ: Альма-Матер, 2008. – 607 с.
10. *Бідюк П.І., Половцев О.В.* Аналіз та моделювання економічних процесів перехідного періоду. — Київ: НТУУ «КПІ», 1999. — 230 с.
11. *Бідюк П.І., Романенко В.Д., Тимошук О.Л.* Аналіз часових рядів. — К.: Політехніка, 2012. — 520 с.
12. *Бідюк П.І., Терентьев А.Н.* Построение и методы обучения байесовских сетей. НТУ «КПИ», Информатика и кибернетика. «Таврический вестник информатики и математики», №2 2004, с. 139-154.
13. *Бідюк П.І., Терентьев А.Н., Гасанов А.С.* Построение и методы обучения Байесовских сетей // Кибернетика и системный анализ. – 2005. – 4. – С. 133-147.



14. Бідюк П.І., Трофимчук О.М., Кожухівська О.А. Прогнозування волатильності фінансових процесів за альтернативними моделями. Наукові вісті НТУУ "КПІ", 2012 / 6, с.36-45.
15. Бретт Л. Машинное обучение на R: экспертные техники для прогностического анализа. – СПб.: Питер, 2020. – 464 с.
16. Брюс П., Брюс Э. Практическая статистика для специалистов Data Science: Пер. с англ. — СПб.: БХВ-Петербург, 2018. — 304 с. ISBN 978-5-9775-3974-6
17. Гожий О.П., Калініна І.О. Підхід до представлення порогової моделі стохастичної волатильності // Могилянські читання – 2018: Досвід та тенденції розвитку суспільства в Україні: глобальний, національний та регіональний аспекти: XXI Всеукр. наук.-метод. конф. : тези доповідей Комп'ютерні науки. Технічні науки, Миколаїв, 12-17 листоп. 2018 р. ЧНУ ім. Петра Могили, 2018. – с. 18-20.
18. Грас Д. Data Science. Наука о данных с нуля: Пер. с англ. - 2-е изд., перераб. и доп. - СПб.: БХВ-Петербург, 2021. - 416 с.
19. Грін В.Г. Економетричний аналіз. — К.: Основи, 2005. — 1198 с.
20. Дауни А.Б. Байесовские модели / пер. с англ. В.А. Яроцкого. – М.: ДМК Пресс, 2018. – 182 с.
21. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. 2. — М.: Финансы и статистика, 1986. — 366 с.
22. Дэвидсон-Пайлон К. Вероятностное программирование на Python: байесовский вывод и алгоритмы. — СПб.; Питер, 2019. — 256 с.
23. Згуровський М. З., Бідюк П. І., Терентьєв О. М., Просянкін-Жарова Т. І. Байєсівські мережі в системах підтримки прийняття рішень. — Київ : ТОВ «Видавниче Підприємство «Едельвейс», 2015. — 300 с.
24. Зельнер А. Байесовские методы в эконометрии. — М.: Статистика, 1980. — 438 с.
25. Ивахненко А.Г. Индуктивный метод самоорганизации моделей сложных систем. — Киев: Наук. думка, 1982. — 296 с.
26. Калініна І.О., Гожий В.О. Особливості використання алгоритму Метраполіса-Хастінгса в процедурах машинного навчання // Ольвійський форум – 2019: стратегії країн Причорноморського регіону в геополітичному просторі: XII Міжнар. наук.-практ. конф. 6-9 червня 2019 р., м. Миколаїв: Вид-во ЧНУ ім. Петра Могили, 2019. – с. 64-66.
27. Коновалюк М.М. Байєсівський аналіз моделі стохастичної волатильності в середовищі OpenBUGS // Наук. вісті НТУУ «КПІ». — 2011. — № 2. — С. 77–84.
28. Курт У. Байесовская статистика: Star Wars, LEGO, резиновые уточки и многое другое. – СПб.: Питер, 2021. – 304 с.
29. Лоу А.М., Кельтон В.Д. Имитационное моделирование. – СПб: ПИТЕР, 2004, 846 с.



30. Лук'яненко І., Краснікова Л. *Економетрика*. — Київ: Знання, 1998. — 494 с.
31. Малюгин В.И. Рынок ценных бумаг. — М.: «Дело», 2003. — 322 с.
32. Мартин О. Байесовский анализ на Python/ пер. с англ. А.В. Снастина. - М.: ДМК Пресс, 2020. — 340 с.
33. Пфеффер А. Вероятностное программирование на практике. / Пер. с англ. Слинкин А. А. - М.: ДМК Пресс, 2017. - 462 с.
34. Рынок ценных бумаг / Под ред. В.А. Голованова, А.И. Басова. — 2-е изд. перераб. И доп. — М.: Финансы и статистика, 2006. — 448 с.
35. Терентьев А.Н., Бидюк П.И. Методы построения Байесовских сетей // Адаптивные системы автоматического управления (межведомственный научно-технический сборник). — 2005. — 8. — С. 130 — 141.
36. Терентьев А.Н., Бидюк П.И. Эвристический метод построения Байесовских сетей // Математические машины и системы. — 2006. — 3. — С. 12-23.
37. Уикем Х., Гроулмунд Г. Язык R в задачах науки о данных: импорт, подготовка, обработка, визуализация и моделирование данных.: Пер. с англ. — СПб.: ООО «Диалектика», 2019. — 592 с.
38. Фишер Р. Трейдинг по Фибоначчи: практические приемы и методы. — М.: ЕВРО, 2002. — 192 с.
39. Шведов А.С. О методах Монте-Карло с цепями Маркова. Экономический журнал ВШЭ 2010, №2, с.227-243.
40. Шуряев А.Н. Основы стохастической финансовой математики // Том 1, Факты. Модели. — М.: ФАЗИС, — 1998. — 512 с.
41. Affet-Sahalia Y. Transition densities for interest rate and other nonlinear diffusions // The Journal of Finance. — 1999. — Vol. 54, No. 4.— Pp. 1361–1395.
42. Albert J., Chib S. Bayesian analysis of binary and polychotomous response data. Journal of the American statistical association 88(422), 1993, pp. 669-679.
43. Alexander C. The handbook of risk management and analysis. — NY: John Wiley & Sons, 1996. — 293 p.
44. Andrieu C., De Freitas N., Doucet A., Jordan M.I. An Introduction to MCMC for machine learning. Machine Learning, 50, 5–43, 2003
45. Astrom K.J. Computer control of a paper machine // IBM Journal of Research and Development, 1967, Vol. 11, pp. 389-405.
46. Bera A., Higgins M. ARCH models: properties, estimation and testing // Journal of Economic Surveys. — 1993. — № 7. — P. 305–366.
47. Besag J. Spatial interaction and the statistical analysis of lattice systems // Journal of the Royal Statistical Society, Series B, 1974, vol. 36, pp. 192-236.
48. Bidyuk P., Beglytsia V., Gozhyj A., Kalinina I. Using the Metropolis-Hastings algorithm in Bayesian data analysis procedures. In: 2019 IEEE 14th International Conference on Computer Sciences and Information



- Technologies (CSIT). Pp. 98-101. IEEE (2019) DOI: 10.1109/STC-CSIT.2019.8929797
49. *Bidyuk P., Gozhyj A., Kalinina I., Szymanski Z., Beglytsia V.* The methods Bayesian analysis of the threshold stochastic volatility model // Proceedings of the 2018 IEEE 2nd International Conference on Data Stream Mining and Processing, DSMP 2018, Lviv, Pp. 70-75. IEEE C.N.:CFP 18J13-CDR
 50. *Bidyuk P.I., Korshevnyuk L.O., Gozhyi A.P., Kalinina I.O., Prosyankina-Zharova T.I., Terentiev O.M.* Modeling and forecasting financial and economic processes with decision support system // Science News KPI. №5-6 (2019) pp. 7-17. DOI: <https://doi.org/10.20535/kpi-sn.2019.5-6.176835>
 51. *Bidyuk P., Matsuki Y., Gozhyj A., Beglytsia V., Kalinina I.* Features of application of Monte Carlo method with Markov chain algorithms in Bayesian data analysis. In: Advances in Intelligent Systems and Computing IV. Springer. Vol.1080. pp.361-376. https://doi.org/10.1007/978-3-030-33695-0_25
 52. *Bollerslev T.* A conditionally heteroskedasticity time series model for speculative prices and rates of return // The Review of Economics and Statistics. – 1987. – Vol. 69. – P. 542–547.
 53. *Bollerslev T., Engle R.F., Nelson D.* ARCH models. Handbook of Econometrics. — Amsterdam: North-Holland, 1993. — 4. — 1078 p
 54. *Bollerslev T.* General autoregressive conditional heteroscedasticity // Journal of Econometrics. — 1986. — 31. — P. 518–537.
 55. *Bollerslev T.* Generalized autoregressive conditional heteroscedasticity, J. of Econometrics Volume, vol. 31, no. 3, pp. 307—327, 1986.
 56. *Bollerslev T., Chow R., Kroner K.* ARCH modeling in finance: a review of the theory and empirical evidence // Journal of Econometrics. — 1992. — 52. — P. 5–59.
 57. *Broemeling L.D.* Bayesian analysis of time series. Taylor & Francis group/ CRC, 2019, 293 p.
 58. *Brooks S.P., Gareth O.* Roberts Assessing Convergence of Markov Chain Monte Carlo Algorithms. – School of Mathematics, University of Bristol, 1997, 27 p.
 59. *Broze L., Scaillet O., Zako J.* Quasi-indirect inference for diffusion processes / Broze L., Scaillet O., Zako J. // Econometric Theory.– 1998. – Vol. 14, № 02. – Pp. 161-186.
 60. *Buntine W.L.* A Guide to the literature on probabilistic network from data. IEEE Transactions on Knowledge and Data Engineering. – 1996. – Vol. 8, issue 2, April. – P.195-210
 61. *Buntine V.L.* Learning with Bayesian models / Report 94-04-13, NASA Ames Research Center. – 1998.–78p.
 62. *Chib S, Greenberg E.* Markov chain monte carlo simulation methods in econometrics. Econometric Theory, 12, 1996.409-431. Printed in the United States of America.



63. *Chib S, Greenberg E.* Understanding the Metropolis-Hastings algorithm. *The American Statistician*, Vol. 49, No. 4. (Nov., 1995), pp. 327-335.
64. *Chou R.Y.* Volatility persistence and stock returns – some empirical evidence using GARCH // *Journal of Applied Econometrics*. – 1987. – No. 3, 279 – 294.
65. *Congdon P.* *Applied Bayesian modelling*. – 2nd ed., John Wiley and Sons, Ltd, 2014, 465 p.
66. *Congdon P.* *Bayesian hierarchical models with application using R*. – 2nd ed., Taylor & Francis group / CRC, 2020, 593 p.
67. *Cox D.R.* Statistical analysis of time series: some recent developments // *Scandinavian Journal of Statistics*. — 1981. — № 8. — P. 93–115.
68. *Darrell Huff* How to Lie with Statistics (illust. I. Geis), Norton, New York, 1954. ISBN 0-393-31072-8 Даррелла Хаффа «Як обманювати за допомогою статистики».
69. *Dechter R.* Bucket elimination: a unifying framework for reasoning // *ACM Press*, Vol. 28, article № 61, December 1996. – P. 1-51.
70. *Druzdzal M.J., Gaag L.C.* Building probabilistic networks: where do the numbers come from. // *IEEE Transactions on Knowledge and Data Engineering*. – 2000. – Vol.12, No.4. – P.481-486
71. *Enders W.* *Applied econometric time series*. — New York: John Wiley & Sons, Inc., 1995. — 434 p.
72. *Engle F.R.* Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, *Econometrica*, vol. 50, no. 4, pp. 987—1007, 1982.
73. *Engle R.F., Bollerslev T.* Modeling the persistence of conditional variance // *Econometric Reviews*. – Vol. 5. – 1986. – P. 1–50.
74. *Engle R.F., Lilien D.M., Robins R.P.* Estimating time-varying risk premia in the term structure: ARCH-M model // *Econometrica*. – 1987. – Vol. 55. – No. 2, pp. 391 – 408.
75. *Ferat S.* A Baesian network approach to the self-organization and learning in intellident agents / Dissertation submitted ti the Faculty of Virginia Politechnic and State University in fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical and Computer Engineering, 2000. – 251 p.
76. *Friedman N.* The Baesian structural EM algorithm. Fourteenth conference on Uncertainty in Artificial Intelligence (UAI). – 1998 – 10 p. <http://www.cs.hiji.ac.il>
77. *Gamerman D., Lopes H.F.* *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. – 2nd ed., Chapman & Hall/ CRC, 2006, 343 p. ISBN 1-58488-587-4
78. *Gelfand A.E., Smith A.F.M.* Sampling based approaches to calculating marginal densities // *Journal of the American Statistical association*, 1990, vol. 85, pp. 398-409.



79. *Geman S., Geman D.* Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images // *IEEE Transaction on pattern analysis and machine intelligence*, 1984, 6, pp. 721-741.
80. *Ghysels E., Harvey A., Renault E.* Stochastic volatility. Statistical methods in finance. — Amsterdam: North-Holland, 1996. — P. 733.
81. *Gilks W.R., Richardson S., Spiegelhalter D.J.* Markov chain Monte Carlo in practice. — New York: Chapman & Hall/CRC, 1996. — 486 p.
82. *Gourieroux C., Monfort A., Renault E.* Indirect inference // *Journal of Applied Econometrics*. — 1993. — Vol. 8. — Pp. S85–S118.
83. Guo H. Learning Bayes networks from data. — 2000. — 17., <http://www.kddresearch.org>
84. *Haipeng G.* Dynamic Bayesian networks. — 2002. — 21 p. <http://www.kddresearch.org>
85. *Hamilton J.D.* A new approach to the economic analysis of nonstationary time series and the business cycle // *Econometrica*. — 1989. — № 57/2. — P. 357–384.
86. *Hansen L.* Large sample properties of generalized method of moments estimators // *Econometrica*. — 1982. — Vol. 50, № 4. — Pp. 1029–1054.
87. *Hansen L., Heaton J., Yaron A.* Finite-sample properties of some alternative GMM estimators. // *Journal of Business & Economic Statistics*. — 1996. — Vol. 14, №. 3. — pp. 262–280.
88. *Hansen L., Scheinkman J.* Back to the future: generating moment implications for continuous-time Markov processes // *Econometrica*. — 1995. — Vol. 63, № 4. pp. 767–804.
89. *Harvey A.C., Ruiz E., Shephard N.* Multivariate stochastic variance models // *Review of Economic Studies*, vol. 61, 1994, pp. 75-93.
90. *Harvey A., Shepard N.* Estimation of an asymmetric stochastic volatility model for asset returns // *Journal of Business and Economic Statistics*. — 1996. — Vol. 14, No.4. — Pp. 429-434.
91. *Hautaniemi S.K.* Target identification with Bayesian networks. Master of science thesis. — 2000. — 99 p., <http://www.cs.tut.fi/samba/Publications>
92. *Hautaniemi S.K., Korpisaari P. T., Saarinen J.P.P.* Target identification with dynamic hybrid Bayesian networks. — 2000. — 11 p.
93. *Heckerman D.* A tutorial on learning with Bayesian networks / Technical report MSR -TR-95-06. — Microsoft Research, Redmond, WA (USA). — 1995.— 57p.
94. *Heckerman D., Breeze J.S.* Causal independence for probability assessment and indifference using Bayesian networks / Technical report MSR-TR-94-08. — Microsoft Research, Redmond, WA (USA). — 1995.—14 p.
95. *Huang C., Darwiche A.* Inference in belief networks: g procedural guide // *International journal of approximate reasoning*. — 1994. — 11. — P. 1-45.
96. *Hull J., White A.* The pricing of options on assets with stochastic volatilities // *Journal of Finance*. — 1987. — № 42. — P. 281–300.



97. International convergence of capital measurement and capital standards: Revised Framework. — Basel Committee on Banking Supervision, 2004. — 62 p.
98. *Jacquier E., Polson N.G., Rossi P.E.* Bayesian analysis of stochastic variance models (with discussion) // *Journal of Business and Economic Statistics*, vol. 12 1994, pp. 87-103.
99. *Johnston J., DiNardo J.* Econometric methods. — New York: McGraw-Hill, Inc., 1997. — 530 p.
100. *Lauritzen S.L., Spiegelhalter D.J.* Local computations with probabilities on graphical structures and their application to expert systems // *Journal Royal Statistics Society B.* — 1988, 50(2). - p.157-194.
101. *Lepar V., Shenoy P.* A comparison of lauritzen-spiegelhalter, hugin, and shenoy-shafer architectures for computing marginals of probability distributions // *Uncertainty in artificial intelligence.* — 1999. — 14. — P. 328-337.
102. *Leray P., Francois O.* BNT structure learning package: documentation and experiments / Technical report, laboratory PSI-INSA Rouen-FRE CNRS 2645, November 2004. — 27 p.
103. *Liang F., Liu Ch., Carroll R.J.* Advanced Markov chain Monte Carlo methods. Learning from past samples. John Wiley and Sons, Ltd. 2010, 380 p.
104. *Lokeswarappa K.G.* Junction trees: motivation / Seminar CSE 714 on advanced topics in machine learning, March 2005. — 57 p.
105. *MacKay D.J.C.* Probable networks and plausible production — a review of practical Bayesian methods for supervised neural networks. *Jornal «Network Computation in Neural Systems».* — 1995. — №6. — P. 469-505.
106. *Metropolis N., Rosenbluth A., Rosenbluth M., Teller A., Teller E.* Equation of state calculations by fast computing machines // *J. Chem. Phys.*, 21, 6, pp. 1087-1092, 1953.
107. *Metropolis N., Ulam S.* The Monte Carlo method // *J. Amer. Statistical Assoc.* 1949, vol. 44, № 247, pp. 335-341.
108. *Murphy K.* A Brief introduction to graphical models and Bayesian networks. — 2001. — 19 p. <http://www.cs.Berkeley.edu>
109. *Murphy K.P.* An introduction to graphical model. — 2001 — 19 p., <http://www.ai.mit.edu>
110. *Murphy K.P.* Dynamic Bayesian networks: representation, inference and learning / A PhD dissertation, University of California, Berkeley, 2002. — 225 p.
111. *Murphy K., Mian S.* Modeling expression data using dynamic Bayesian networks. Technical report, Berkeley, CA. — 1999 — 12 p., <http://citeseer.ist.psu.edu>
112. *Nelson D.B.* Conditional heteroscedasticity in asset returns: a new approach // *Econometrica.* 1991. — Vol. 59. — No. 2, pp. 347 — 370.



113. *Nodelman U., Shelton Ch.R., Koller D.* Learning Continuous Time Bayesian Networks. // Proceedings of the Nineteenth International Conference on Uncertainty in Artificial Intelligence. – 2003. – 451-458p.
114. *Pearl J.* Causality. – Cambridge: Cambridge University, 2000. – 384 p
115. *Poon S.H.* Practical guide to forecasting financial market volatility. New York: John Wiley & Sons, Inc., 2005, 238 p.
116. *Pritsker M.* Nonparametric density estimation and tests of continuous time interest rate models // Review of Financial Studies. – 1998.– Vol.11, №3.– Pp. 449–487.
117. *Roberts G.O., Smith A.F.M.* Simple conditions for the convergence of the Gibbs sampler and metropolis Hastings algorithms // Stochastic Processes and Applications, 1994, vol. 49, pp. 207-216.
118. *Robinson R.W.* Counting unlabeled acyclic digraphs / Proceeding of fifth Australian conference on combinatorial mathematics, Melbourne, Australia, 1976. – P. 28-43.
119. *Sebastiani P., Ramoni M.* Bayesian inference with missing data using bound and collapse. Technical Report KMi-TR-58. – 1997 – 21 p., <http://chip.tch.harvard.edu>
120. *Singleton K.* Estimation of an asset pricing models using the empirical characteristic function / Singleton K. //Journal of Econometrics. – 2001. – Vol.102, № 1.– Pp. 111-141.
121. *Shephard N.* Statistical aspects of ARCH and stochastic volatility. Time series models with econometric, finance and other applications. — London: Chapman & Hall. — 1996. — 677 p.
122. *Tanner M.A., Wong W.H.* The calculation of posterior distributions by data augmentation // Journal of the American Statistical Association, Vol. 82, No. 398 (Jun., 1987), pp. 528-540. URL: <http://www.jstor.org/stable/2289457>
123. *Taylor S.J.* Modelling financial time series. — Chichester: John Wiley, 1986. — 268 p. 13. Clark P.K. A subordinated stochastic process model with fixed variance for speculative prices // Econometrica. — 1973. — № 41. — P. 135–156.
124. *Taylor S.J.* Modeling stochastic volatility: A review and comparative study, Mathematical Finance, vol. 4, no. 2, pp. 183—204, 1994.
125. *Tsay L.* Analysis of financial time series. – New York: Wiley series in probability and statistics, 2002. – 455 p.
126. *Tsay R.S.* Analysis of financial time series. New York: John Wiley & Sons, Inc., 2010, 715 p.
127. *Zweig G.G.* Speech recognition with dynamic Bayesian networks / University of California, Berkeley. – 1994.–169 p., <http://www.cse.msu.edu>

Додаток А

Деякі види розподілу випадкових величин

Базові імовірнісні розподіли

У базовій установці **R** (пакет *stats*) реалізовані такі ймовірнісні розподіли:

1. Дискретні:

- біноміальне;
- пуассонівське;
- геометричне;
- гіпергеометричний;
- від'ємнобіноміальне;
- поліноміальне.

2. Безперервні:

- бета-розподіл;
- розподіл Коші;
- експоненціальне;
- χ^2 -розподіл;
- розподіл Фішера (f-розподіл);
- гамма-розподіл;
- логнормальне;
- логістичне;
- нормальне;
- розподіл Стьюдента (t-розподіл);
- рівномірне;
- розподіл Вейбулла.

3. Рангові розподіли Вілкоксона.

Для кожного з розподілів в **R** є чотири функції:

- *щільність розподілу* (для безперервних випадкових величин) та ймовірність прийняття випадковою величиною конкретного значення (дискретні випадкові величини) – префікс *d* перед назвою розподілу;
- *функція розподілу* (ФР) випадкові величини – префікс *p* перед назвою розподілу;
- *квантилі розподілу* – префікс *q* перед назвою розподілу;
- *випадкова вибірка за заданим розподілом* – префікс *r* перед назвою розподілу.

Примітка: Зауважимо, що ФР в **R** є ймовірність $F_{\xi}(x) = P\{\xi \leq x\}$. Далі розглянемо розподіли докладно та наведемо приклади.

Біноміальний розподіл

Дискретна випадкова величина має біноміальний розподіл з параметрами $n \in \mathbb{Z}^+$, $p: X \sim B(n, p)$, якщо вона приймає цілочисельні значення $k=0, 1, \dots, n$ з ймовірностями, що визначаються формулою Бернуллі

$$p_k = P\{X = k\} = C_n^k p^k q^{n-k},$$

де $C_n^k = \frac{n!}{k!(n-k)!}$; $p \in (0, 1)$; $q = 1 - p$.



Біноміальний розподіл виникає в послідовності з n незалежних випробувань з постійною ймовірністю успіху в кожному випробуванні $p = \text{const}$ та повністю визначається значеннями параметрів n та p :

$$X \sim \left(\begin{matrix} 0 & 1 & \dots & k & \dots & n \\ q^n C_n^1 p^1 q^{n-1} \dots C_n^k p^k q^{n-k} \dots p^n \end{matrix} \right).$$

Функція розподілу випадкової величини, що підпорядковується біноміальному закону $X \sim B(n, p)$, має вигляд:

$$F(x) = \begin{cases} 0, & \text{якщо } x \leq 0; \\ \sum_{k < x} C_n^k p^k q^{n-k}, & \text{якщо } 0 < x \leq n; \\ 1, & \text{якщо } x > n. \end{cases}$$

Приклад А-1. Представити графіки ймовірностей p_k та функції розподілу $F(x)$ біноміально розподіленої випадкової величини $X \sim B(n, p)$ при заданих значеннях параметрів n та p за допомогою R.

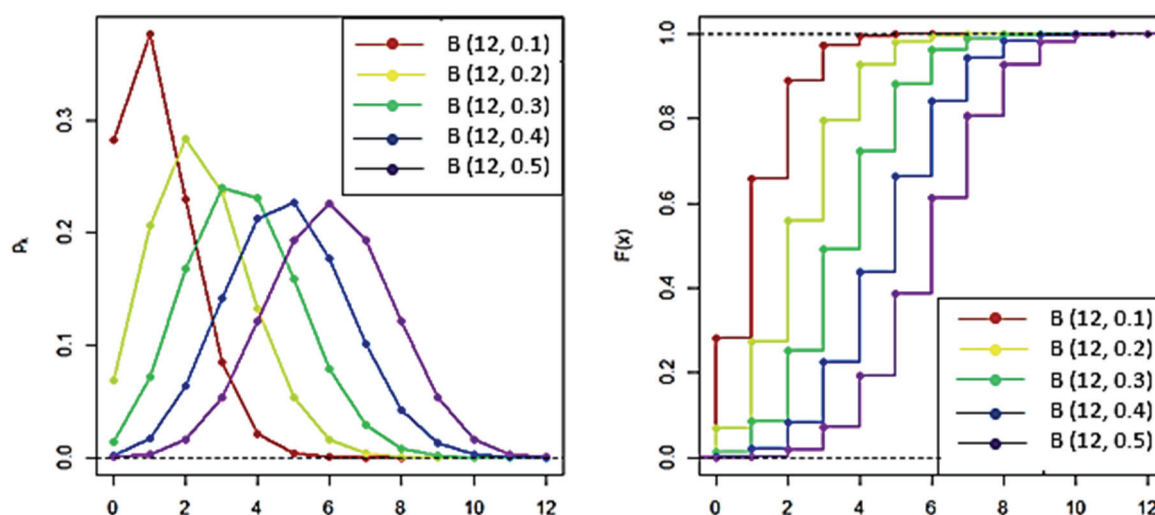


Рисунок А-1 – Біноміальний розподіл ймовірностей p_k та функція біноміального розподілу $F(x)$

На рис. А-1 показано приклади побудови графіків ймовірностей p_k та функції розподілу біноміально розподіленої випадкової величини $X \sim B(n, p)$ при $n=12$ і $p = \frac{1}{10}, \frac{2}{10}, \dots, \frac{5}{10}$.

```
1 source("plot2dist.r")
2 k <- seq(0, n <- 12)
3 p <- seq(5)/10
4 pk <- sapply(p, function(i) dbinom(k, n, i))
5 Fk <- sapply(p, function(i) pbinom(k, n, i))
6 lt <- sapply(p, function(i) sprintf("B(%.0f, %.3g)", n, i))
7 pmfPlot(k, pk, lt)
8 cmfPlot(k, Fk, lt)
```

Команда «source ("plot2dist.r")» у рядку **1** здійснює завантаження із зазначеного файлу вихідного коду, що містить функції для побудови графіків розподілів випадкових величин (див. додаток Б).



Функція `seq()` у рядках 2–3 генерує вектор послідовних значень від першого до другого аргументу; зауважимо, що третій аргумент цієї функції дозволяє вказати приріст у послідовності значень, за умовчанням дорівнює ± 1 .

Функція «`sapply(p,...)`» здійснює підстановку кожної компоненти вектора «`p`» у певну функцію. Таким чином, у рядках 4, 5 за допомогою функцій `dbinom()` і `rbinom()` по вектору абсцис «`k`» обчислюються ординати ймовірності «`pk`» і функції біноміального розподілу «`Fk`» для кожної пари параметрів «`n, p`», а в рядку 6 значення цих параметрів формують написи на графіках.

Функції «`pmfPlot()`» і «`cmfPlot()`» визначені в бібліотеці користувача «`plot2graph.r`» і проводять побудову графіків ймовірностей та функцій розподілу дискретної випадкової величини за переданими векторами абсцис «`k`» і ординат «`pk`» або «`Fk`».

Розподіл Пуассона

Дискретна випадкова величина X має розподіл Пуассона з параметром $\lambda > 0$: $X \sim P(\lambda)$, якщо вона набуває цілих значень $k=0, 1, \dots, \infty$ з ймовірностями, що визначаються формулою Пуассона

$$p_k = P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \text{ де } \lambda > 0.$$

Розподіл Пуассона є граничним випадком біномного розподілу при $n \rightarrow \infty$, а так, що $\lambda=np=\text{const}$.

Воно виникає при розгляді поодиноких незалежних випадкових подій з постійною інтенсивністю λ та повністю визначається її значенням

$$X \sim \begin{pmatrix} 0 & 1 & 2 & \dots \\ \frac{1}{e^\lambda} & \frac{\lambda^1}{e^{\lambda 1!}} & \frac{\lambda^2}{e^{\lambda 2!}} & \dots \end{pmatrix}.$$

Функція розподілу випадкової величини, що підпорядковується закону Пуассона, має вигляд:

$$F(x) = \begin{cases} 0, & \text{якщо } x \leq 0; \\ \sum_{k < x} \frac{\lambda^k}{k!} e^{-\lambda}, & \text{якщо } x > 0. \end{cases}$$

Приклад А-2. Представити графіки ймовірностей p_k і функції розподілу $F(x)$ для випадкової пуассонівської величини $X \sim P(\lambda)$ при заданих значеннях параметра λ .

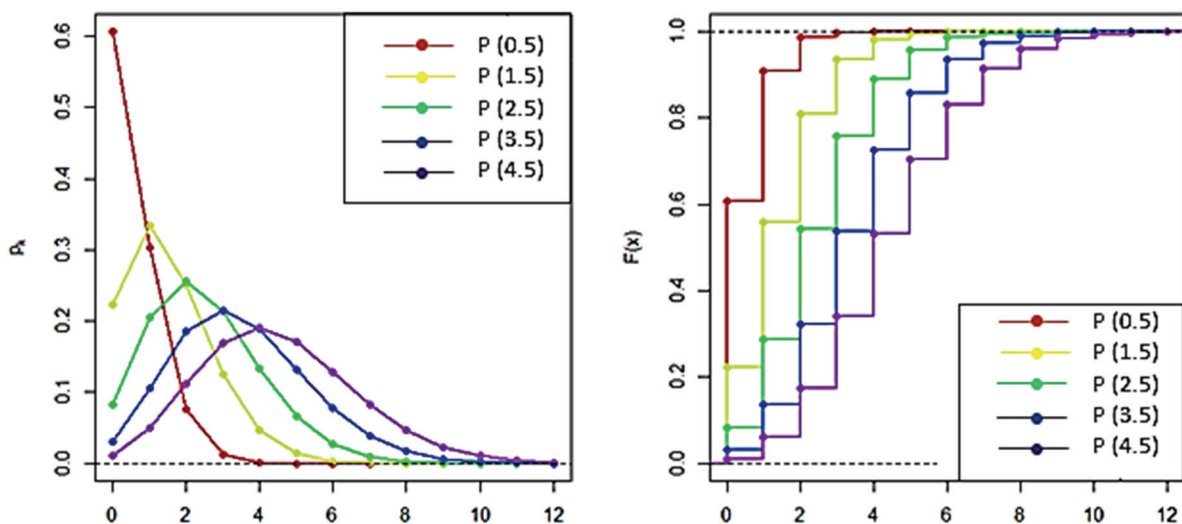


Рисунок А-2 – Пуассонівський розподіл ймовірностей p_k та функція пуассонівського розподілу $F(x)$

На рис. А-2 показано приклади побудови графіків ймовірностей p_k та функції розподілу $F(x)$ пуассонівської випадкової величини при $X \sim P(\lambda)$ при $\lambda = \frac{1}{2}, \frac{3}{2}, \dots, \frac{9}{2}$.

```
lambda <- seq(5, 45, 10)/10
pk <- sapply(lambda, function(i) dpois(k, i))
Fk <- sapply(lambda, function(i) ppois(k, i))
lt <- sapply(lambda, function(i) sprintf("P(%.3g)", i))
par(mfrow=c(1,2))
pmfPlot(k, pk, lt)
cmfPlot(k, Fk, lt)
```

Геометричний розподіл

Дискретна випадкова величина X має *геометричний розподіл* з параметром p : $X \sim G(p)$, якщо вона набуває цілих значень $k=0,1,\dots, \infty$ з ймовірностями, що визначаються формулою

$$p_k = P\{X = k\} = q^k p, \text{ де } p \in (0, 1); q = 1 - p.$$

Геометричне розподіл має випадкова величина X , що дорівнює кількості випробувань у послідовності Бернуллі, що проходять до появи першого успіху. Геометричний розподіл повністю визначається значеннями параметра p :

$$X \sim \begin{pmatrix} 0 & 1 & 2 & \dots \\ p & qp & q^2 p & \dots \end{pmatrix}.$$

Функція розподілу випадкової величини X , що підпорядковується геометричному закону $X \sim G(p)$, має вигляд

$$F(x) = \begin{cases} 0, & \text{якщо } x \leq 0; \\ \sum_{k < x} q^k p, & \text{якщо } x > 0. \end{cases}$$



Приклад А-3. Побудувати графіки ймовірностей p_k та функції розподілу $F(x)$ для геометрично розподіленої випадкової величини $X \sim G(p)$ при заданих значеннях параметра p .

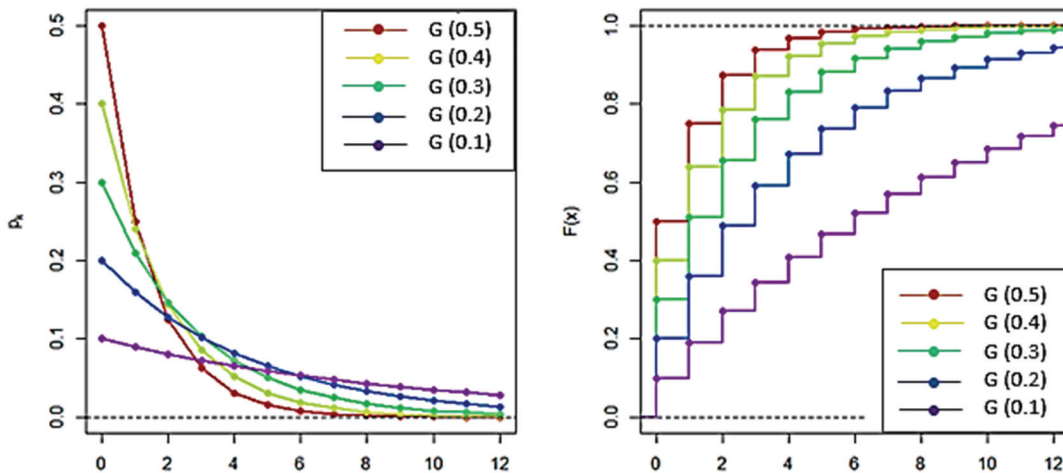


Рисунок А-3 – Геометричний розподіл ймовірностей p_k та функція геометричного розподілу $F(x)$

На рис. А-3 надані приклади побудови графіків ймовірностей p_k та функції розподілу $F(x)$ для геометрично розподіленої випадкової величини $X \sim G(p)$ при $p = \frac{1}{10}, \frac{2}{10}, \dots, \frac{5}{10}$.

```
p <- seq(5, 1)/10
pk <- sapply(p, function(i) dgeom(k, i))
Fk <- sapply(p, function(i) pgeom(k, i))
lt <- sapply(p, function(i) sprintf("G(%.3g)", i))
par(mfrow=c(1,2))
pmfPlot(k, pk, lt)
cmfPlot(k, Fk, lt)
```

Рівномірний розподіл

Найпростішим із безперервних розподілів є рівномірний розподіл, що виникає при узагальненні поняття n рівно ймовірних випадкових подій на випадок $n \rightarrow \infty$. Безперервна випадкова величина X має *рівномірний розподіл* на відрізку $[a, b]$: $X \sim U(a, b)$, якщо її щільність ймовірності постійна і відмінна від нуля тільки на цьому відрізку:

$$f(x) = \begin{cases} 0, & x \notin [a, b]; \\ \frac{1}{b-a}, & x \in [a, b]. \end{cases}$$

Рівномірний розподіл повністю визначається координатами кінців відрізка $[a, b]$. Функція розподілу випадкової величини, що підпорядковується рівномірному закону $X \sim U(a, b)$, має вигляд:

$$F(x) = \begin{cases} 0, & \text{якщо } x \leq a; \\ \frac{x-a}{b-a}, & \text{якщо } a < x \leq b; \\ 1, & \text{якщо } x > b. \end{cases}$$



Приклад А-4. Побудувати графіки щільності ймовірностей $f(x)$ та функції розподілу $F(x)$ для рівномірно розподіленої випадкової величини $X \sim U(a, b)$ при вищевказаних значеннях параметрів a та b .

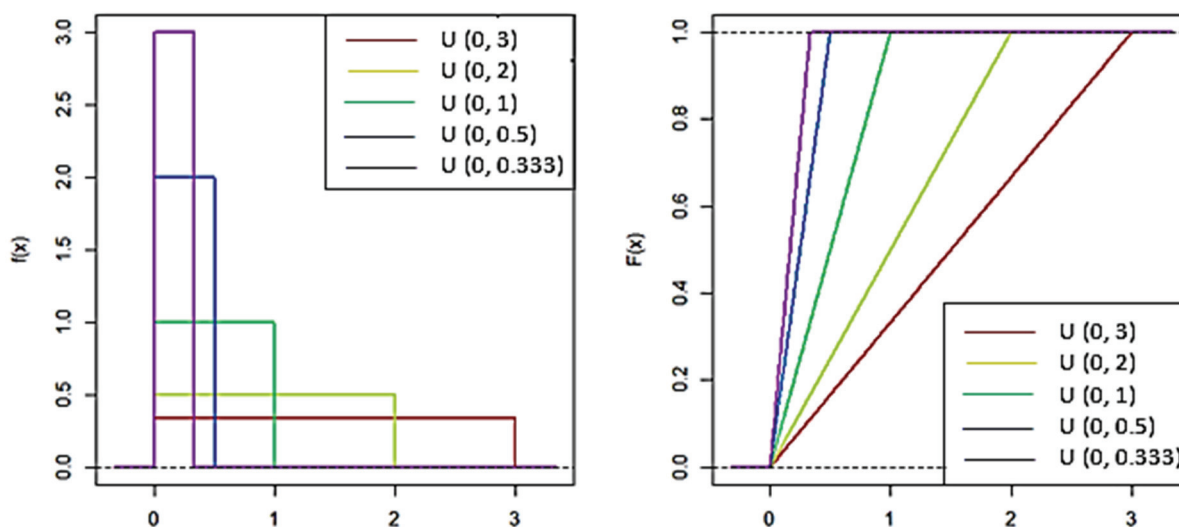


Рисунок А-4 – Щільність рівномірного розподілу $f(x)$ та функція рівномірного розподілу $F(x)$

На рис. А-4 наведено приклади побудови графіків щільності ймовірності $f(x)$ та функції розподілу $F(x)$ рівномірно розподіленої випадкової величини $X \sim U(a, b)$ при значеннях параметрів: $a=0$, $b = \frac{1}{3}, \frac{1}{2}, 1, 2, 3$.

```
a <- 0; b <- c(seq(3,2), 1/seq(3))
x <- seq(a-min(b), max(b)+min(b), len=900)
fx <- sapply(b, function(t) dunif(x, a, t))
Fx <- sapply(b, function(t) punif(x, a, t))
lt <- sapply(b, function(t) sprintf("U(%.3g, %.3g)", a, t))
par(mfrow=c(1,2))
pdfPlot(x, fx, lt)
cdfPlot(x, Fx, lt)
```

Показниковий розподіл

Показниковий розподіл виникає при моделюванні часу між послідовними реалізаціями однієї й тієї ж випадкової події. Безперервна випадкова величина X має показовий розподіл: $X \sim \square(\lambda)$ якщо її щільність ймовірності має вигляд:

$$f(x) = \begin{cases} 0, & \text{якщо } x < 0, \\ \lambda e^{-\lambda x}, & \text{якщо } x \geq 0, \end{cases}$$

де $\lambda > 0$ – параметр, що інтерпретується як середня кількість випадкових подій в одиницю часу.

Функція розподілу показово розподіленої випадкової величини: $X \sim \square(\lambda)$ має вигляд:

$$F(x) = \begin{cases} 0, & \text{якщо } x < 0, \\ 1 - e^{-\lambda x}, & \text{якщо } x \geq 0. \end{cases}$$



Приклад А-5. Побудувати графіки щільності ймовірностей $f(x)$ та функції розподілу $F(x)$ для показово розподіленої випадкової величини $X \sim \square(\lambda)$ при вказаних вище значеннях параметра λ .

На рис. А-5 надані приклади побудови графіків щільності ймовірності $f(x)$ та функції розподілу $F(x)$ показово розподіленої випадкової величини $X \sim \square(\lambda)$ при значеннях параметра $\lambda = \frac{1}{2}, 1, 2, 3, 4$.

```
lambda <- c(1/2, seq(4))
x <- seq(-1/4, 4, len=900)
fx <- sapply(lambda, function(t) dexp(x, t))
Fx <- sapply(lambda, function(t) pexp(x, t))
lt <- sapply(lambda, function(t) sprintf("E(%.3g)", t))
par(mfrow=c(1,2))
pdfPlot(x, fx, lt)
cdfPlot(x, Fx, lt)
```

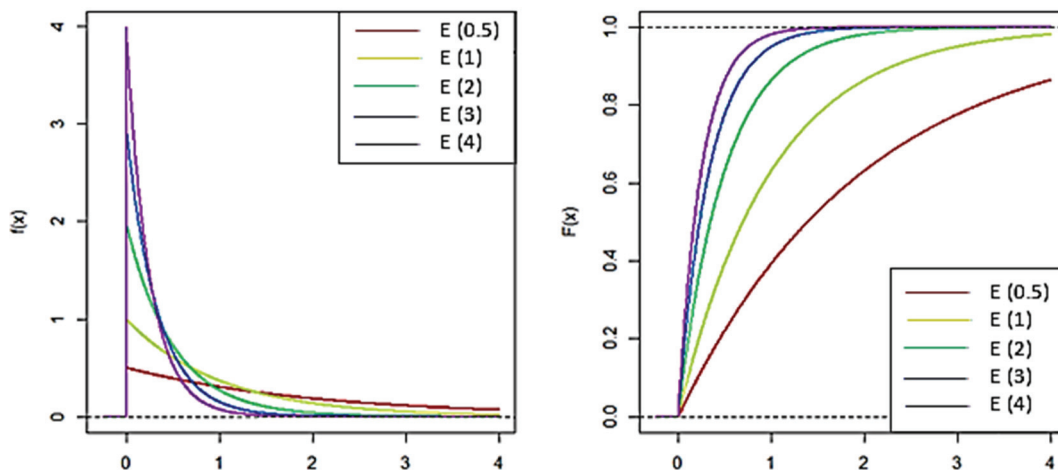


Рисунок А-5 – Щільність показового розподілу $f(x)$ та функція показового розподілу $F(x)$

Нормальний розподіл

Нормальний розподіл зазвичай виникає під час розгляду суми великої кількості незалежно розподілених випадкових величин з кінцевою дисперсією. Безперервна випадкова величина X має *нормальний розподіл*: $X \sim N(a, \sigma)$, якщо її щільність ймовірності має вигляд:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} = \varphi\left(\frac{x-a}{\sigma}\right),$$

де $x, a \in \mathbf{R}$; $\sigma > 0$; $\varphi(z)$ – функція Гауса, яка визначається рівнянням

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}.$$

Нормальний розподіл повністю визначається параметрами a і σ . Функція розподілу випадкової величини, що підпорядковується нормальному закону $X \sim N(a, \sigma)$, має вигляд:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-a)^2}{2\sigma^2}} dt = \frac{1}{2} + \Phi\left(\frac{x-a}{\sigma}\right),$$



де $\Phi(z)$ – функція Лапласа, яка визначається рівнянням

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{y^2}{2}} dy.$$

Приклад А-6. Побудувати графіки щільності ймовірностей $f(x)$ та функції розподілу $F(x)$ для нормально розподіленої випадкової величини $X \sim N(a, \sigma)$ при вказаних вище значеннях параметрів a і σ .

На рис. А-6 надані приклади побудови графіків щільності ймовірності $f(x)$ та функції розподілу $F(x)$ нормально розподіленої випадкової величини $X \sim N(a, \sigma)$ при значеннях параметрів: $a=0, \sigma=1, 2, \dots, 5$.

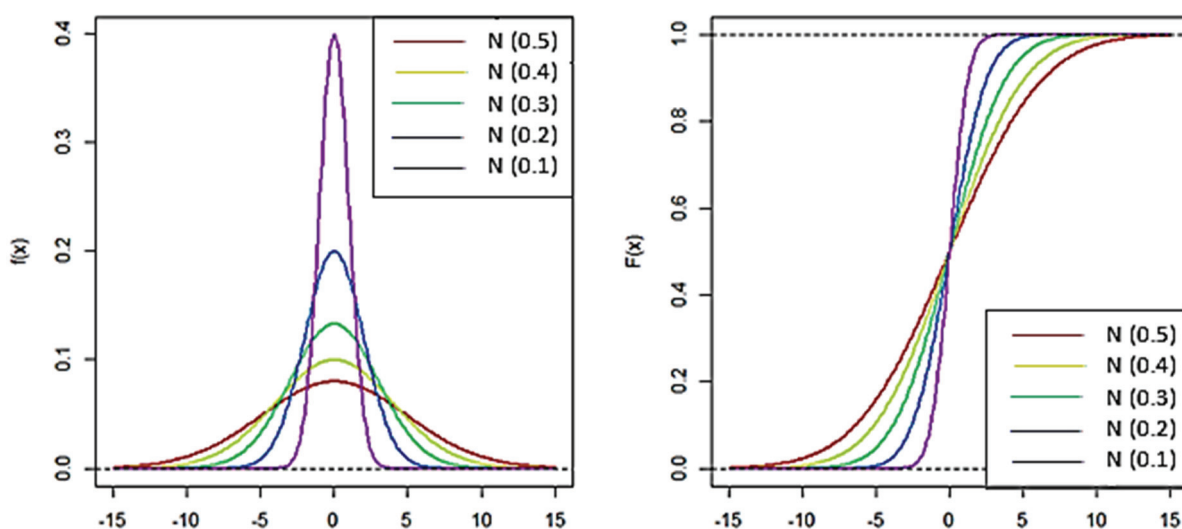


Рисунок А-6 – Щільність нормального розподілу $f(x)$ та функція нормального розподілу $F(x)$

```
a <- 0; sigma <- seq(5, 1)
x <- seq(a-3*max(sigma), a+3*max(sigma), len=300)
fx <- sapply(sigma, function(t) dnorm(x, a, t))
Fx <- sapply(sigma, function(t) pnorm(x, a, t))
lt <- sapply(sigma, function(t) sprintf("N(%.3g, %.3g)", a, t))
par(mfrow=c(1,2))
pdfPlot(x, fx, lt)
cdfPlot(x, Fx, lt)
```

Логнормальний розподіл

Безперервна випадкова величина X має *логарифмічно нормальний* чи *логнормальний розподіл*, якщо її логарифм нормально розподілений. Подібно до нормального розподілу логнормальне виникає при розгляді добутку великої кількості незалежних випадкових величин з кінцевою дисперсією. Щільність ймовірності логарифмічно нормального розподілу має вигляд:



$$f(x) = \begin{cases} 0, & \text{якщо } x \leq 0; \\ \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - a)^2}{2\sigma^2}}, & \text{якщо } x > 0, \end{cases}$$

де $x \in \mathbf{R}; x > 0; \sigma > 0$.

Логарифмічно нормальне розподілення повністю визначається параметрами a і σ . Функція розподілу логарифмічно нормальної випадкової величини $X \sim \ln N(a, \sigma)$ має вигляд:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x t^{-1} e^{-\frac{(\ln t - a)^2}{2\sigma^2}} dt = \Phi\left(\frac{\ln x - a}{\sigma}\right) + \frac{1}{2},$$

де $\Phi(x)$ – функція Лапласа.

Приклад А-7. Побудувати графіки щільності ймовірностей $f(x)$ та функції розподілу $F(x)$ для логарифмічно нормально розподіленої випадкової величини $X \sim \ln N(a, \sigma)$ при вказаних значеннях параметрів a і σ .

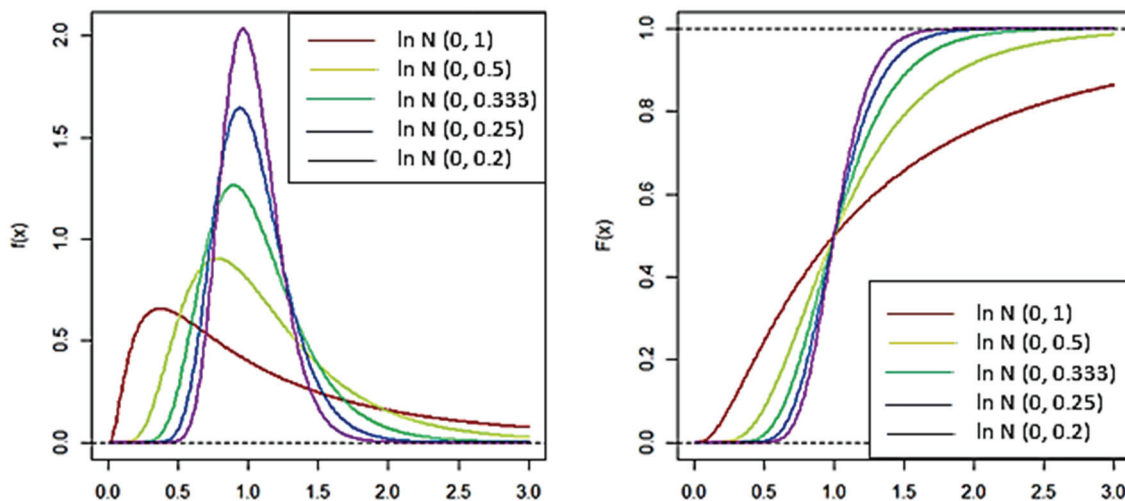


Рисунок А-7 – Щільність логарифмічно нормального розподілу $f(x)$ та функція логарифмічно нормального розподілу $F(x)$

На рис. А-7 наведено приклади побудови графіків щільності ймовірності $f(x)$ та функції розподілу $F(x)$ логарифмічно нормально розподіленої випадкової величини $X \sim \ln N(a, \sigma)$ при значеннях параметрів: $a=0$,

$$\sigma = \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 1.$$

```
a <- 0; sigma <- 1/seq(5)
x <- seq(a, a+3*max(sigma), len=300)
fx <- sapply(sigma, function(t) dlnorm(x, a, t))
Fx <- sapply(sigma, function(t) plnorm(x, a, t))
lt <- sapply(sigma, function(t) sprintf("ln N(%.3g, %.3g)", a, t))
par(mfrow=c(1,2))
pdfPlot(x, fx, lt)
cdfPlot(x, Fx, lt)
```



Розподіл Пірсона

Якщо $X_i \sim N(0,1)$, де $i = 1, 2, \dots, n$ - незалежні стандартні нормальні випадкові величини, то сума n квадратів цих величин має χ^2 -розподіл (Пірсона) з n ступенями свободи:

$$\chi_n^2 = \sum_{i=1}^n X_i^2.$$

Щільність розподілу χ^2 описується формулою:

$$f(x) = \begin{cases} 0, & \text{якщо } x \leq 0; \\ \frac{1}{2^{n/2}\Gamma(n/2)} x^{\frac{n-2}{2}} e^{-x/2}, & \text{якщо } 0 < x; \end{cases} \quad \Gamma(p) = \int_0^{\infty} t^{p-1} e^{-t} dt,$$

де $\Gamma(p)$ – гамма-функція Ейлера. При зростанні числа ступенів свободи $n \rightarrow \infty$ розподіл χ^2 асимптотично нормально.

Приклад А-8. Побудувати графіки щільності ймовірності $f(x)$ і функції розподілу $F(x)$ для випадкової величини $X \sim \chi_n^2$ при значеннях параметра n .

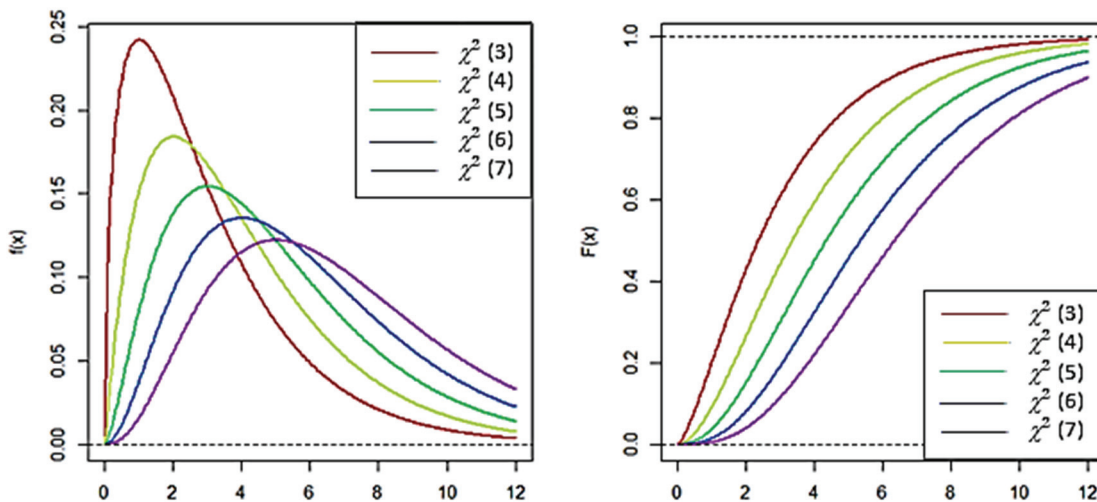


Рисунок А-8 – Щільність χ^2 -розподілу $f(x)$ та функція χ^2 -розподілу $F(x)$

На рис. А-8 наведено приклади побудови графіків щільності ймовірності $f(x)$ та функції розподілу $F(x)$ випадкової величини $X \sim \chi_n^2$ при числі ступенів свободи: $n=3, 4, \dots, 7$.

```
n <- seq(3, 7)
x <- seq(0, 12, len=300)
fx <- sapply(n, function(i) dchisq(x, i))
Fx <- sapply(n, function(i) pchisq(x, i))
par(mfrow=c(1,2))
lt <- sapply(n, function(i) parse(text=sprintf("chi^2*(%.1g)", i)))
pdfPlot(x, fx, lt)
cdfPlot(x, Fx, lt)
```



Розподіл Стюдента

Якщо випадкові величини $Z \sim N(0,1)$ і $U \sim \chi_n^2$ - незалежні, то випадкова величина

$$t_n = \frac{Z}{\sqrt{\frac{U}{n}}}$$

має розподіл Стюдента або t -розподіл з n ступенями свободи.

Щільність t -розподілу має вигляд:

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}},$$

де $x \in \mathbb{R}$; $\Gamma(p)$ – Гамма-функція Ейлера. У разі зростання числа ступенів свободи $n \rightarrow \infty$ розподіл Стюдента асимптотично нормальний.

Приклад А-9. Побудувати графіки щільності ймовірностей $f(x)$ і функції розподілу $F(x)$ випадкової величини $X \sim t_n$ при значеннях параметра n .

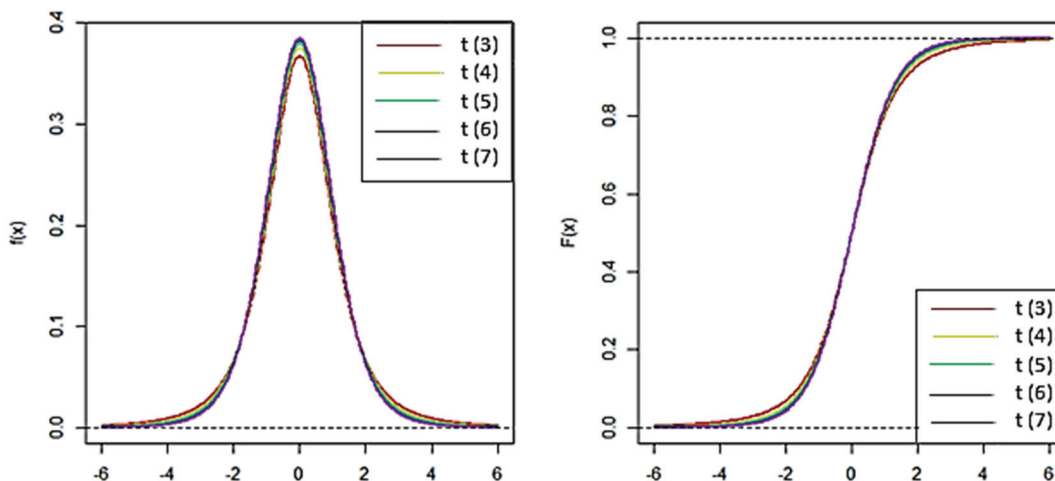


Рисунок А-9 – Щільність t -розподілу та функція t -розподілу

На рис. А-9 наведено приклади побудови графіків щільності ймовірності $f(x)$ та функції розподілу $F(x)$ випадкової величини $X \sim t_n$ при числі ступенів свободи $n=3, 4, \dots, 7$.

```
n <- seq(3, 7)
x <- seq(-6, 6, len=300)
fx <- sapply(n, function(i) dt(x, i))
Fx <- sapply(n, function(i) pt(x, i))
lt <- sapply(n, function(i) sprintf("t(%.0f)", i))
par(mfrow=c(1,2))
pdfPlot(x, fx, lt)
cdfPlot(x, Fx, lt)
```



Розподіл Фішера

Якщо випадкові величини $U \sim \chi_m^2$ і $V \sim \chi_n^2$ - незалежні, то випадкова величина

$$F_{\frac{m}{n}} = \frac{\frac{1}{m}U}{\frac{1}{n}V}$$

має розподіл Фішера або F -розподіл зі ступенями свободи чисельника m та знаменника n . Щільність F -розподілу:

$$f(x) = \begin{cases} 0, & x \leq 0; \\ \frac{\sqrt{\frac{(mx)^m n^n}{(mx+n)^{m+n}}}}{xB\left(\frac{m}{2}, \frac{n}{2}\right)}, & x > 0; \end{cases} \quad B(u, v) = \int_0^1 t^{u-1}(1-t)^{v-1} dt,$$

де $m, n > 0$; $B(u, v)$ – бета-функція Ейлера. У разі зростання числа ступенів свободи $n \rightarrow \infty$ розподіл Фішера асимптотично нормальний.

Приклад А-10. Побудувати графіки щільності ймовірностей $f(x)$ і функції розподілу $F(x)$ випадкової величини $X \sim F_{\frac{m}{n}}$ при заданих значеннях параметрів m і n .

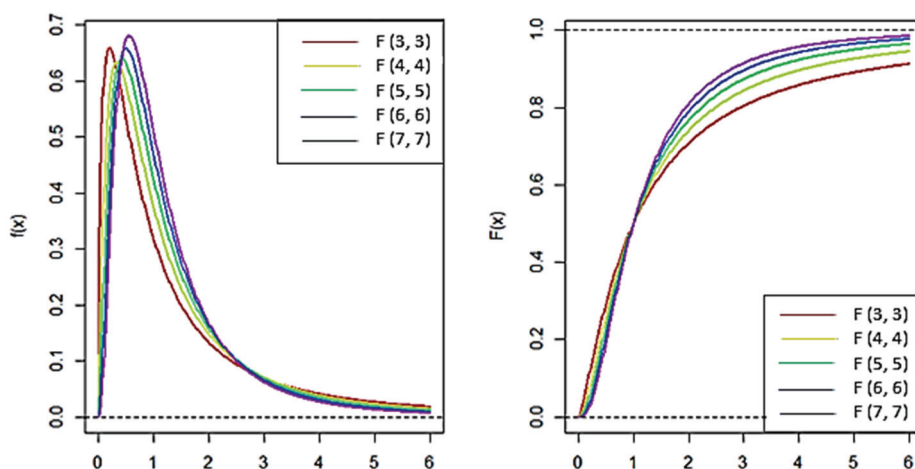


Рисунок А-10 – Щільність F -розподілу та функція F -розподілу

На рис. А-10 зображено приклади побудови графіків щільності ймовірності $f(x)$ та функції розподілу $F(x)$ випадкової величини $X \sim F_{\frac{m}{n}}$ при однакових значеннях ступенів свободи $m=n=3, 4, \dots, 7$.

```
m<-n<-seq(3,7); k<-seq_along(m)
x<-seq(0,6,len=300)
fx<-sapply(k, function(i) df(x,m[i],n[i]))
Fx<-sapply(k, function(i) pf(x,m[i],n[i]))
lt<-sapply(k, function(i) sprintf("F(%.0f,%.0f)", m[i],n[i]))
par(mfrow=c(1,2))
pdfPlot(x,fx,lt)
cdfPlot(x,Fx,lt)
```


Наукове видання

**Бідюк Петро Іванович
Калініна Ірина Олександрівна
Гожий Олександр Петрович**

Байєсівський аналіз даних

Монографія

ISBN 978-617-7941-52-0

Підписано до друку 26.11.2021. Формат 60х 84/16. Папір офсетний
Наклад 300 примірників. Гарнітура Times.
Друк цифровий. Ум. друк. арк. 14,29. Обл.-вид. арк. 15,37.
Замовлення № 2096.

Надруковано з готового оригінал-макету у
книжковому видавництві ФОП Вишемирський В.С.
Свідоцтво про внесення до державного реєстру суб'єктів видавничої
справи: серія ХС № 48 від 14.04.2005
видано Управлінням у справах преси та інформації
73000, Україна, м. Херсон, вул. Соборна, 2.
Тел. (050) 133-10-13, (050) 514-67-88
e-mail: printvvs@gmail.com