

П.І.Бідюк, Навчально-науковий комплекс "Інститут прикладного системного аналізу" Національного технічного університету "Київський політехнічний інститут"

А.В.Кроптя, Навчально-науковий комплекс "Інститут прикладного системного аналізу" Національного технічного університету "Київський політехнічний інститут"

А.С.Гасанов, Навчально-науковий комплекс "Інститут прикладного системного аналізу" Національного технічного університету "Київський політехнічний інститут"

ОПИСАНИЯ НЕВИЗНАЧЕНОСТЕЙ ЭКСПЕРИМЕНТАЛЬНЫХ ДАНИХ В ЗАДАЧАХ АНАЛИЗА РИЗИКОВ

Бідюк П.І. , Кроптя А.В., Гасанов А.С.

Описание неопределенностей экспериментальных данных в задачах анализа рисков

В работе описаны типы неопределенностей моделей, которые используются для принятия решений. Приведена характеристика погрешностей измерений и оценок неизмеримой величины. Определены погрешности оценок неизмеримой величины. Авторами рассмотрено моделирование по методу Монте Карло.

Ключові слова: неопределенности, погрешности, неизмеримая величина, метод Монте Карло.

Бидюк П.И., Кроптя А.В., Гасанов А.С.

Описание неопределенностей экспериментальных данных в задачах анализа рисков

В работе описаны типы неопределенности моделей, которые используются для принятия решений. Приведена характеристика погрешностей измерений и оценок неизмеримой величины. Определены погрешности оценок неизмеримой величины. Авторами рассмотрено моделирование по методу Монте Карло.

Ключевые слова: неопределенности, погрешности, неизмеримая величина, метод Монте Карло.

Поява невизначеностей статистичного характеру означає, що неможливо встановити точне значення величини. Наприклад, можна прогнозувати, що процентні ставки на протязі двох-трьох наступних кварталів будуть знаходитись в інтервалі від 12% до 20%. Неможливість встановлення точних значень процентних ставок вимагає резервування капіталу і пошуку альтернативних джерел для фінансування. Невизначеність такого типу можна віднести до *інтервальної статистичної невизначеності*. Вона розповсюджена надзвичайно широко і зустрічається майже в кожному випадку аналізу процесів довільної природи.

До *параметричних невизначеностей моделей*, які використовуються для прийняття рішень, можна віднести неточності визначення параметрів статистичних і математичних моделей внаслідок таких причин: наближене знання розподілу вибірки, некоректність застосування методу оцінювання та методичні похибки обчислень.

Структурні невизначеності моделей пов'язані з труднощами встановлення точної структури математичної моделі процесу. Тобто, визначення порядку моделі, типу збурень та можливостей їх врахування в моделі, наявності нелінійностей та їх описання, врахування запізнення. Як правило, ми користуємось оцінками вказаних величин, які можуть бути відносно далекими від точних значень.

Іноді можна встановити, які значення є більш ймовірними для конкретного процесу, а які менш ймовірними. Тобто можна з деяким наближенням встановити ймовірність появи тих чи інших значень змінної x з інтервалу можливих значень Δx . Таку складову можна назвати *ймовірнісною невизначеністю*.

Для врахування та описання невизначеностей теорія ймовірностей надає корисний інструментарій. Обов'язковими інструментами є щільність та функція розподілу (ФР) і ймовірнісна міра – функція, яка відображає множини значень у ймовірності того, що конкретна змінна відноситься до даної множини.

При аналізі ризиків головна мета полягає у прийнятті рішень, спрямованих на зменшення цих ризиків та підвищення (максимізацію) позитивного ефекту (функції корисності – ФК) від інвестицій чи інших дій. Ймовірнісний аналіз ризиків дає можливість розрахувати очікувану корисність K для конкретної дії $d \in D$. Тобто метою є визначення очікуваної корисності $K_d(x)$ для кожної допустимої керуючої дії d та для кожного фактичного значення ключової змінної x . Однак, значення x може бути частково або повністю невідомим і його потрібно оцінювати. Очікуване значення корисності $E[K_d(x)]$ можна визначити за допомогою функції розподілу для x . Для обчислення $E[K_d(x)]$ необхідно використати по можливості точніші параметри розподілів. Очевидно також, що самі функції корисності (ФК) можуть мати різну форму для різних практичних задач або навіть декілька різних форм для однієї задачі.

Природно вибрати гладку форму ФК $K_d(x)$, яка змінюється в інтервалі Δx можливих значень x . Гладка функція може бути розкладена в ряд Тейлора і при вузькому інтервалі Δx можна скористатись тільки лінійними та квадратичними членами розкладу, зберігаючи прийнятну точність апроксимації:

$$K_a(x) \approx a_0 + a_1(x - x_0) + a_2(x - x_0)^2, \quad (1)$$

де x_0 – точка в інтервалі Δx . Тепер очікуване значення ФК можна обчислити за виразом:

$$\begin{aligned} E[K_a(x)] &\approx E[a_0 + a_1(x - x_0) + a_2(x - x_0)^2] = \\ &= E[K_a(x)] \approx a_0 + a_1 E[(x - x_0)] + a_2 E[(x - x_0)^2]. \end{aligned} \quad (2)$$

Якщо за точку x_0 використати математичне сподівання $E[x]$, то за другий момент можна взяти дисперсію $Var[x]$ розподілу ймовірностей.

У випадку існування порогу x_0 очікуване значення гладко змінюється до досягнення цього значення, а потім стрибкоподібно зменшується і знову неперервно (гладко) змінюється при $x > x_0$. Як правило, з аналізу досліджуваного процесу можна визначити (відносно вузький) інтервал, що містить допустимі значення x . Можна припустити, що $K_d(x)$ наближено дорівнює деякій константі K^+ при $x < x_0$ і константі $K^- < K^+$ при $x > x_0$. У простому випадку можна нормувати ці значення так: $K^+ = 1$ і $K^- = 0$. В такому випадку очікуване значення $E[K_d^{(0)}(x)]$ співпадає з ймовірністю того, що $x < x_0$, тобто з відповідним значенням функції розподілу $F(x_0)$.

При наявності точної інформації щодо функції розподілу необхідно використовувати точні значення $F(x)$. Однак при відсутності точної інформації щодо ФР для кожного значення x матимемо інтервал $[F^-(x), F^+(x)]$ можливих значень ймовірностей $F(x)$. Таку пару ФР: $F^-(x)$ і $F^+(x)$, які обмежують значення (невідомої) фактичної ФР, називають імовірнісним обмеженням або p – обмеженням.

Використання p – обмеження пов'язане із деякими видами відповідних йому невизначеностей. Наприклад, невизначеність відсутня, якщо відоме точне значення x_0 і воно використовується як p – обмеження:

$$F^-(x) = F^+(x) = \begin{cases} 0, & \text{якщо } x \leq x_0, \\ 1, & \text{інакше.} \end{cases} \quad (3)$$

Якщо $x \in [x^-, x^+]$, то граничні значення для ФР можна представити у вигляді:

$$F^-(x) = \begin{cases} 0, & \text{якщо } x \leq x^-, \\ 1, & \text{інакше;} \end{cases} \quad (4)$$

$$F^+(x) = \begin{cases} 0, & \text{якщо } x \leq x^-, \\ 1, & \text{інакше.} \end{cases}$$

Інформацію про моменти також можна представити у термінах p – обмеження. Якщо основна змінна x визначена на інтервалі $[x^-, x^+]$ із середнім \bar{x} , то $F(x) \in [F^-(x), F^+(x)]$, де

$$F^+(x) = \min \left(1, \frac{x^+ - \bar{x}}{x^+ - x} \right). \quad (5)$$

В загальному випадку $F^-(x) \leq F^+(x)$. Окремими випадками p – обмеження є дійсне число, інтервал та функція розподілу.

Можливості використання p – обмеження. Розглянемо випадок, коли необхідно оцінити величину $y(k)$, $k=1, \dots, N$, що не може бути вимірною. Це може висока температура, для вимірювання якої не існує приладів; запаси нафти чи корисних копалин в деякому районі; відстань до зірки і т. ін. Для того обчислити оцінки цих величин, необхідно визначити (виміряти) іншу величину $x(k)$, $k=1, \dots, N$, яка зв'язана з $y(k)$. Наприклад, для того щоб оцінити запаси нафти в деякому регіоні вимірюють відбиті ультразвукові сигнали $x(k)$, які посиляють у дві паралельно розміщені свердловини, потім розв'язують систему диференціальних рівнянь в частинних похідних і обчислюють оцінку запасів y . Це приводить до необхідності розв'язання некоректно поставлених зворотних задач. На практиці моделі, що зв'язують виміри $x(1), \dots, x(N)$ з оцінкою невимірюваної величини $y(k)$, відомі з деяким наближенням. Крім цієї невизначеності моделі існують невизначеності (похибки) вимірів $\Delta x(k) = \tilde{x}(k) - x(k)$, де $\tilde{x}(k)$ – вимір; $x(k)$ – точне значення. Похибки вимірів (реєстрації даних) також призводять до похибок (невизначеностей) щодо значень змінної $y(k)$, тобто $\tilde{y}(k) = f(\tilde{x}(1), \dots, \tilde{x}(N))$. Це приводить до необхідності оцінювання похибок оцінок невимірюваної змінної: $\Delta y(k) = \tilde{y}(k) - y(k)$. В ідеальному випадку відома інформація щодо розподілу похибок вимірів $\Delta x(k) = \tilde{x}(k) - x(k)$, яка надається виробником вимірювальних приладів. Однак, у більшості випадків відома тільки верхня границя похибок вимірів Δ_m , тобто $|\Delta x(k)| \leq \Delta_m$. Фактично кожний вимір належить інтервалу: $[\tilde{x}(k) - \Delta_m; \tilde{x}(k) + \Delta_m]$. В такому випадку оцінювання невимірюваної змінної пов'язане з інтервальними обчисленнями.

Характеристики похибок вимірів та оцінок невимірюваної величини.

Вираз для похибок оцінок невимірюваної величини

$\Delta y(k) = \tilde{y}(k) - y(k) = f(\tilde{x}(1), \dots, \tilde{x}(N)) - f(x(1), \dots, x(N))$ можна спростити завдяки використанню розкладу f в ряд Тейлора та обмеженням лінійними величинами:

$$\Delta y(k) \stackrel{def}{=} c_1 \Delta x(1) + \dots + c_n \Delta x(N), \quad (6)$$

де def – означає „за визначенням”; c_i – часткові похідні

$$c_i = \frac{\partial f}{\partial x(k)|_{(\tilde{x}(1), \dots, \tilde{x}(N))}} \quad (7)$$

в точці $(\tilde{x}(1), \dots, \tilde{x}(N))$.

Припустимо, що похибка $\Delta y(k)$ оцінок $\tilde{y}(k)$ – це лінійна комбінація незалежних випадкових гаусових змінних $\Delta x(k)$, тобто $\Delta y(k)$ – нормально розподілена випадкова величина з нульовим середнім та стандартним відхиленням:

$$\sigma = \sqrt{c_1^2 \sigma_1^2 + \dots + c_N^2 \sigma_N^2}. \quad (8)$$

При інтервальному оцінюванні ймовірності похибок $\Delta x(k)$ невідомі, а відомо тільки, що $|\Delta x(k)| \leq \Delta_m$. При цьому сума (6) приймає максимально можливе значення, якщо кожний член цієї суми також приймає максимально можливе значення. Максимально можливе значення цього члена складає $|c_i| \Delta_m$, а тому сума (6) може приймати максимальне значення

$$\Delta_{\max} = |c_1| \Delta(1) + \dots + |c_N| \Delta(N) = \Delta_m \sum_{k=1}^N |c_k|. \quad (9)$$

По аналогії, мінімально можливим значенням для $\Delta y \in -\Delta_{\max}$. Таким чином, $\Delta y(k) \in [-\Delta_{\max}, \Delta_{\max}]$.

Визначення похибок оцінок невимірюваної величини. Можна формалізувати постановку задачі оцінювання невимірюваної величини: для заданої функції $f(x(1), \dots, x(N))$, N вимірів $\tilde{x}(1), \dots, \tilde{x}(N)$ та N додатніх чисел $\sigma_1, \dots, \sigma_N$ (або $\Delta_1, \dots, \Delta_N$) обчислити вираз (8) або (9). Для виконання обчислень необхідно знайти частинні похідні. Прямолінійним методом розв’язання задачі є *чисельне диференціювання*. Для того щоб обчислити i -у часткову похідну, необхідно дати приріст i -й вхідній величині $x(i)$, тобто покласти $\tilde{x}(i) + h(i)$, а всі інші виміри залишити без змін: $\delta(i) = h(i)$ для i -го виміру і $\delta(j) = 0, \forall j \neq i$. Оцінка коефіцієнта c_i визначається за виразом:

$$c_i = \frac{1}{h(i)} [f(\tilde{x}(1), \dots, \tilde{x}(i-1), \tilde{x}(i)+h(i), \tilde{x}(i+1), \dots, \tilde{x}(N)) - y]. \quad (10)$$

Цей метод називають *чисельним диференціюванням*.

Моделювання за методом Монте Карло. Для того щоб змоделювати нормальний розподіл із стандартним відхиленням σ_i , необхідно помножити вихідне значення стандартного генератора нормального розподілу $\{\alpha(k)\} \sim N(0,1)$ на σ_i , тобто $\delta(i) = \sigma(i) \cdot \alpha(i)$. Застосування процедури Монте Карло дає N нормально розподілених значень:

$$c(1) = \mathbf{c} \delta(1), \dots, c(M) = \mathbf{c} \delta(M) \quad (11)$$

із заданим стандартним відхиленням σ , яке можна оцінити за виразом:

$$\sigma = \sqrt{\frac{1}{M-1} \sum_{k=1}^M [c(k)]^2}. \quad (12)$$

Відносна похибка цієї оцінки залежить тільки від M , вона пропорційна $1/\sqrt{M}$ і не залежить від числа вимірів N . Тобто число кроків (циклів) моделювання M_f , яке необхідне для досягнення заданої точності, не залежить від числа вимірів. Похибка розглянутого методу асимптотична нормально розподілена із стандартним відхиленням $\sigma_e \sim \sigma/\sqrt{2M}$. Таким чином, якщо скористатись обмеженням „дві сигми”, то із ймовірністю 95% цей алгоритм дає оцінку для σ , яка відрізняється від свого фактичного значення на $\leq 2\sigma_e = 2\sigma/\sqrt{2M}$. Це похибка, з якою оцінюється похибка оцінки невимірюваної величини. Оскільки в реальних ситуаціях не вимагається висока точність таких оцінок, тобто ми говоримо, що похибка складає $\pm 15\%$ або $\pm 20\%$, то достатньо отримати відносну оцінку в межах, скажімо $\pm 20\%$. Для згаданої вище оцінки в інтервалі „дві сигми” це означає, що необхідно вибрати таке найменше значення M (число ітерацій алгоритму оцінювання), яке забезпечує виконання обмеження:

$$2\sigma_e = \frac{2\sigma}{\sqrt{2M}} \leq 0,2\sigma, \quad (13)$$

$2\sigma = 0,2\sigma\sqrt{2M}$ або $M = 100/2 = 50$, тобто необхідно покласти $M = 50$. Якщо вибрати 20%–е стандартне відхилення (правило двох сигм дасть інтервал 40%), то в такому випадку необхідно виконати тільки 13 ітерацій алгоритму оцінювання:

$$\frac{2\sigma}{\sqrt{2M}} = 0,4\sigma \Rightarrow 10 = 2\sqrt{2M}, \text{ або } M = 100/8 \approx 13.$$

Однак, якщо необхідно гарантувати 20%–у точність в 99,9% випадків, що відповідає правилу „трьох сигм”, то необхідно скористатись таким M , яке задовольняє співвідношенню: $3\sigma_e = \frac{3\sigma}{\sqrt{2M}} \leq 0,2\sigma$, $\frac{3\sigma}{\sqrt{2M}} = 0,2\sigma \Rightarrow 9 = 0,2^2 2M$, або $M = 450/4 \approx 113$. Для $N = 10^3$ всі наведені значення M є набагато меншими ніж $M = N$, що необхідно у випадку чисельного диференціювання. Таким чином, при виборі між (детермінованим) чисельним диференціюванням та рандомізованим алгоритмом Монте Карло необхідно керуватись таким: вибирати детермінований алгоритм у випадку, коли число вимірів N задовольняє нерівності: $N \leq M_0$, де $M_0 \approx 50$; застосовувати рандомізований метод, якщо $N \geq M_0$. Необхідно зазначити, що завдяки незалежності випадкових величин метод Монте Карло можна реалізувати у вигляді паралельно працюючих алгоритмів, що суттєво скорочує час, необхідний для обчислення оцінок невимірюваної змінної.

Література

1. **Айвазян С.А., Енюков И.С., Мешалкин Л.Д.** Прикладная статистика. Основы моделирования и первичная обработка данных. – М. : Финансы и статистика, 1983. – 470 с.
2. **Кобзарь А.И.** Прикладная математическая статистика. – М. : Физматлит, 2006. – 816 с.

Bidyuk P.I., Krptya A.V., Gasanov A.S.

Description of uncertainties of experimental data in tasks of risks analysis

The types of models uncertainties used for decisions making are described in the work. The characteristics of errors of measuring and evaluation of the immeasurable value. The errors of evaluation of the immeasurable value are determined. The authors considered modeling with Monte Carlo method.

Keywords: uncertainties, errors, immeasurable value, Monte Carlo method.