

Домашнє завдання №2

Зміст

1. Щільність розподілу енергії	2
Експериментальний тест №1	3
Експериментальний тест №2	4
Експериментальний тест №3	5
2. Метод найменших квадратів	6
Модель з глобальною адитивною похибкою	6
Модель з локальною адитивною похибкою	8
Перелік посилань	9

1. Щільність розподілу енергії

Нехай задано таку функціональну залежність значень енергії E та температури T :

$$E(T) = A + B \tanh \left(\frac{T - T_0}{C} \right), \quad (1.1)$$

де значення A , B , C та T_0 в рамках фізичної моделі є константами. Тепер нехай покладемо величину C як неперервно розподілену нормальну випадкову величину з математичним сподіванням μ_c та дисперсією σ_c :

$$C \sim N(\mu_c, \sigma_c^2), \quad (1.2)$$

а отже, щільність розподілу введеної випадкової величини буде такою:

$$f_C(x) = \frac{1}{\sqrt{2\pi\sigma_c^2}} e^{-\frac{(x-\mu_c)^2}{2\sigma_c^2}} \quad (1.3)$$

Відтак, енергія E матиме функціональну залежність (1.1) від випадкової величини C , де значення T , T_0 , A та B є фіксованими:

$$g(C) = A + B \tanh \left(\frac{T - T_0}{C} \right) \quad (1.4)$$

Таким чином, з'ясуємо безпосередній вигляд щільності $f_E(y)$ випадкової величини E як результату перетворень, заданих формулою (1.4):

$$E \sim f_E(y) \quad (1.5)$$

За формулою заміни змінних:

$$f_E(y) = f_C(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right|, \quad (1.6)$$

де обернена залежність $g^{-1}(y)$ задаватиметься виразом

$$g^{-1}(y) = \frac{2(T - T_0)}{\ln(B - A + y) - \ln(B + A - y)}, \quad (1.7)$$

а Якобіан перетворення, відповідно, матиме вид

$$\frac{d}{dy} g^{-1}(y) = -2(T - T_0) \cdot \left(\ln \frac{B - A + y}{B + A - y} \right)^{-2} \cdot \frac{2B}{(B - A + y)(B + A - y)} \quad (1.8)$$

Тоді щільність розподілу $f_E(y)$ записуватиметься так:

$$f_E(y) = \frac{1}{\sqrt{2\pi\sigma_c^2}} \cdot \exp \left[-\frac{1}{2\sigma_c^2} \left(\frac{2(T - T_0)}{\ln(B - A + y) - \ln(B + A - y)} - \mu_c \right)^2 \right] \times \\ \times \left| \left(\ln \frac{B - A + y}{B + A - y} \right)^{-2} \cdot \frac{4B(T - T_0)}{(B - A + y)(B + A - y)} \right| \quad (1.9)$$

Експериментальний тест №1

Для перевірки коректності наведених викладок порівняємо гістограму значень функції $g(C)$ симуляції $N = 10\,000$ значень випадкової величини C та аналітичну криву щільності випадкової величини E (1.9). Першим тестом зафіксуємо значення параметрів, ігноруючи фізичну інтерпретацію (Табл. 1). На Рис. 1 зображено відповідні «результати життєздатності» аналітичного розподілу.

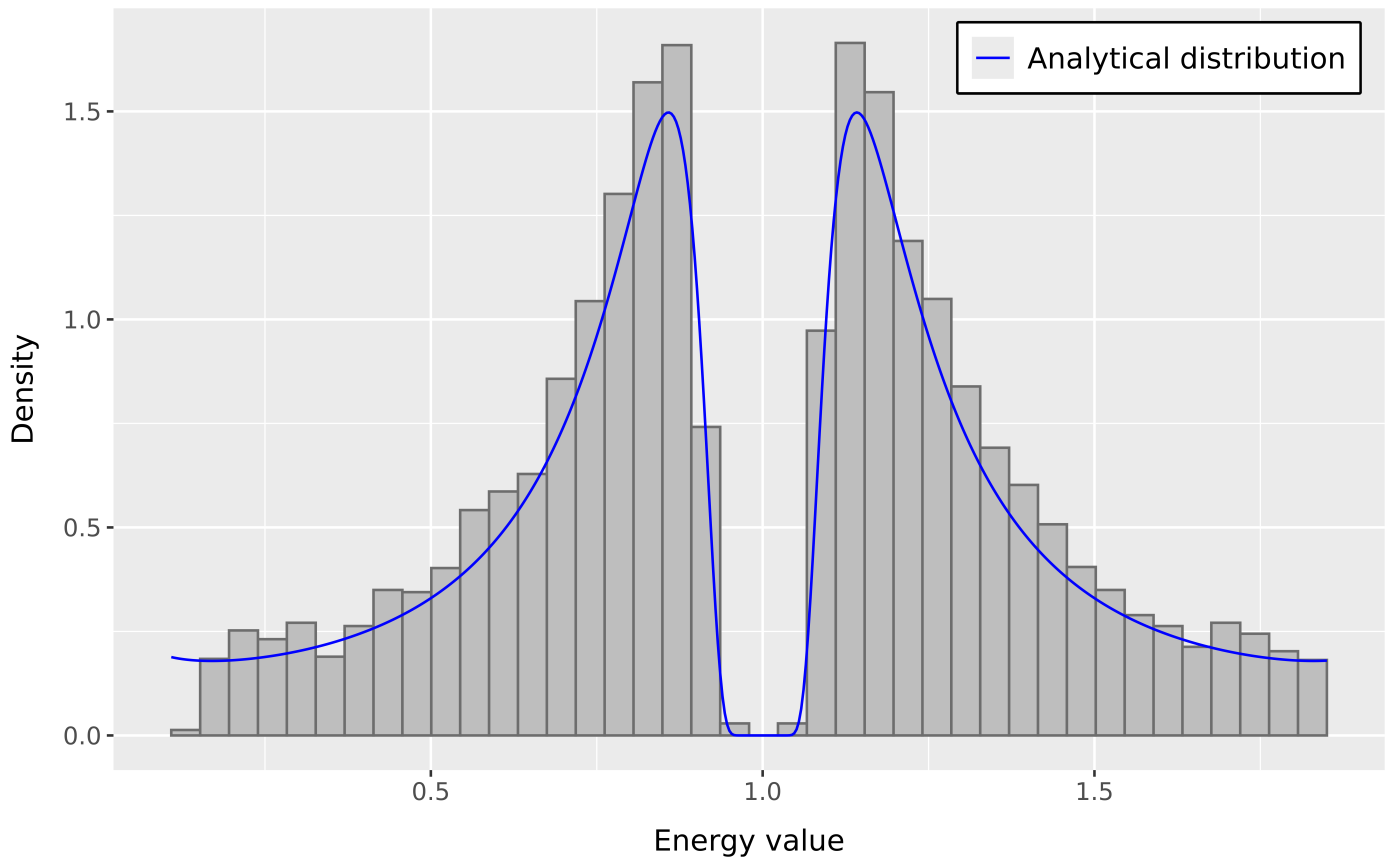


Рис. 1: Порівняльний графік симуляції та аналітичного розподілу (Табл. 1)

μ_c	σ_c	A	B	T	T_0
0	5	1	1	2	1

Таблиця 1: Значення параметрів (без фізичного контексту)

Експериментальний тест №2

Наступним тестом спробуємо зафіксувати значення параметрів з огляду на фізичний зміст. До прикладу [1, розділ 3, Табл. 4], використовуючи значення $A = 46.4 \pm 6.7$, $B = 51.5 \pm 8.0$, $C = 18.9 \pm 5.9$ та $T_0 = -64.5 \pm 3.5$ було отримано довірчий інтервал $T_{85} = -46 \pm 20$ для відповідного значення енергії $E = 85$. Спробуємо інтерпретувати цей приклад в рамках нашої задачі (Табл. 2). Отримані результати зображені на Рис. 2.

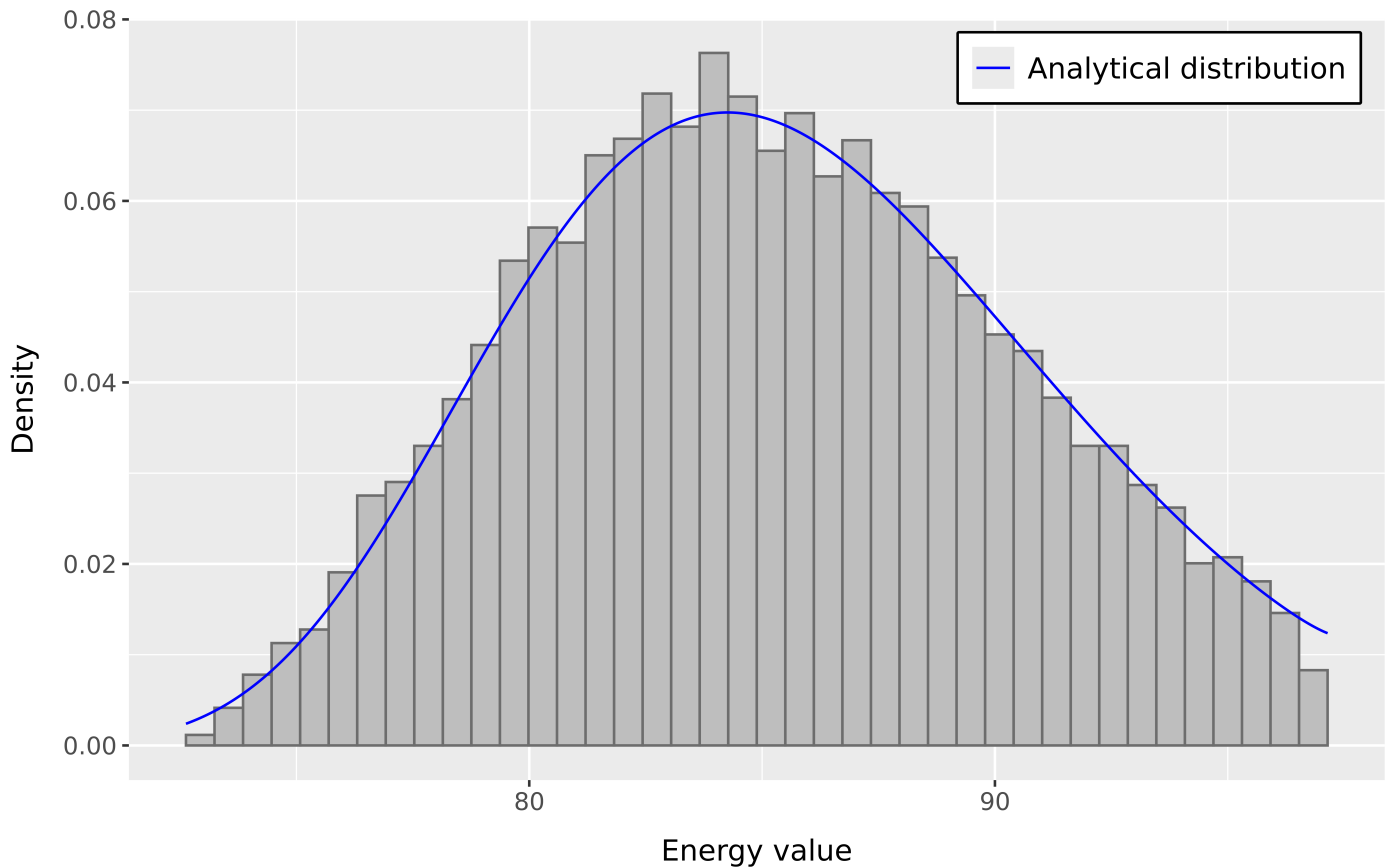


Рис. 2: Порівняльний графік симуляції та аналітичного розподілу (Табл. 2)

μ_c	σ_c	A	B	T	T_0
18.9	5.0	46.4	51.5	-46.0	-64.5

Таблиця 2: Значення параметрів (з огляду на приклад [1, розділ 3, Табл. 4])

Експериментальний тест №3

Ще один тест проведемо для дещо інших значень параметрів (Табл. 3), при цьому коефіцієнти A та B визначимо через U та L таким чином:

$$A = \frac{U + L}{2}, \quad B = \frac{U - L}{2} \quad (1.10)$$

Крім того, зобразимо розподіл енергії для 12 різних точок температури T , рівновіддалено розподілених на проміжку від -90 до 0 (Рис. 3).

μ_c	σ_c	L	U	T_0
15	5	2	210	-50

Таблиця 3: Значення параметрів (моделювання Монте-Карло)

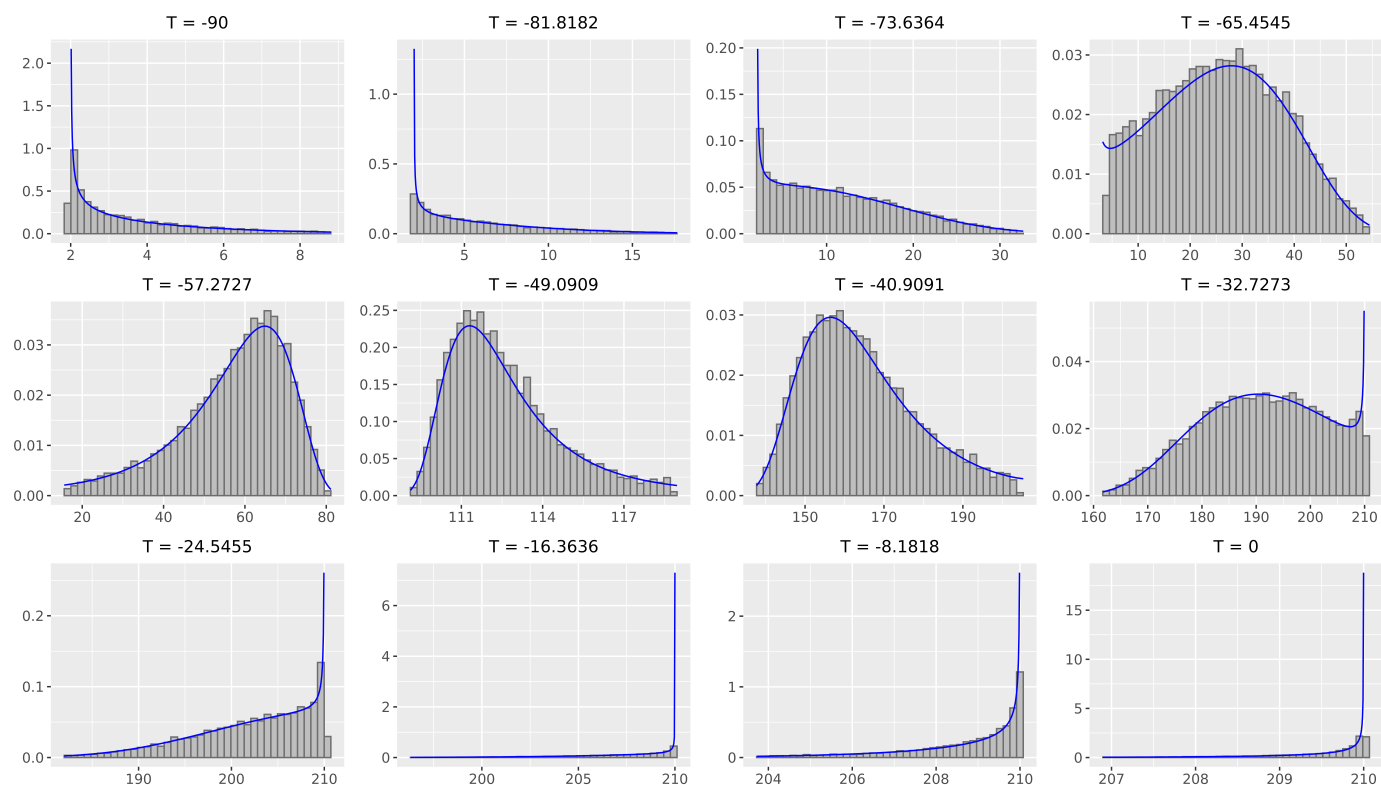


Рис. 3: Порівняльний графік симуляції та аналітичного розподілу (Табл. 3)

2. Метод найменших квадратів

Модель з глобальною адитивною похибкою

Нехай задана така нелінійна регресійна модель:

$$E_i = A + B \tanh\left(\frac{T_i - T_0}{C}\right) + \varepsilon_i, \quad i = \overline{1, n}, \quad (2.1)$$

де значення A , B та T_0 є фіксованими (Табл. 4), величина C є невідомим параметром моделі, а випадкова похибка ε_i має нормальний розподіл:

$$\varepsilon_i \sim N(0, \sigma_c^2), \quad i = \overline{1, n} \quad (2.2)$$

Першим кроком згенеруємо вибірку даних $(T_i, E_i)_{i=\overline{1, n}}$, на основі якої наступним кроком оцінимо параметр C за методом найменших квадратів (МНК). Для етапу генерування даних покладемо параметр $C = \mu_c$.

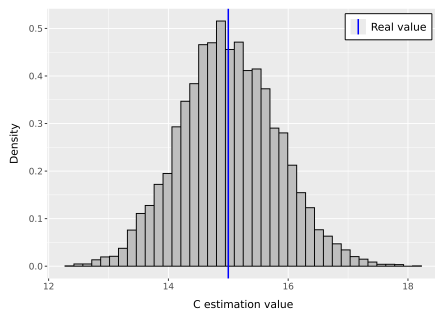
A	B	T_0	μ_c	σ_c
106	104	-50	15	5

Таблиця 4: Значення параметрів регресійної моделі

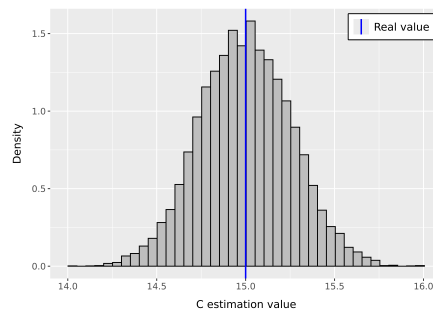
У рамках задачі фокус зосереджений навколо 12 точкових значень температури T , рівновіддалено розподілених на проміжку $[-90, 0]$. Таким чином, організуємо генерування даних так: збиратимемо по k спостережень на кожну з 12 досліджуваних точок, тобто в результаті матимемо вибірку даних розміром $n = 12k$. Такий підхід дозволить спостерігати за результатами МНК при збільшенні розміру вибірки.

Отже, на основі наявної згенерованої вибірки розміром $n = 12k$ маємо змогу обчислити оцінку параметра C методом МНК. Для більш наочної картини проведемо серію з $N = 10\,000$ повторних запусків МНК для щоразу інших згенерованих даних згідно моделі (2.1).

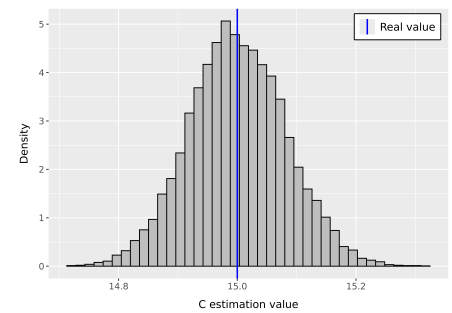
Виконаємо аналогічні обчислення при розмірі вибірки в межах кожного запуску МНК як $n = 12$, $n = 120$ та $n = 1200$, тобто при генеруванні $k = 1$, $k = 10$ та $k = 100$ спостережень на кожну фокус-точку, відповідно. Графіки гістограм оцінок параметра C продемонстровано на Рис. 4. З Табл. 5 бачимо, що зі збільшенням розміру вибірки n МНК демонструє збіжні властивості.



а) Вибірка $n = 12$



б) Вибірка $n = 120$



в) Вибірка $n = 1200$

Рис. 4: Результати серії з $N = 10\,000$ повторних запусків МНК (глобальна адитивна помилка)

Розмір вибірки	Істинне значення	Вибіркове середнє	Стандартне відхилення
$n = 12$	15	15.0097	0.8150
$n = 120$		15.0002	0.2572
$n = 1200$		15.0006	0.0814

Таблиця 5: Статистичні характеристики серії з $N = 10\,000$ повторних запусків МНК (глобальна адитивна помилка)

Імплементація МНК засобами мови R за допомогою бібліотеки `gslnls` наведена у Лістингу 1.

Лістинг 1: Запуск МНК для нелінійної регресійної моделі (2.1)

```

1 energy_dependency <- function(T, T0, A, B, C) {
2   return(A + B * tanh((T - T0) / C))
3 }
4
5 data <- data.frame(
6   T = as.vector(x_generated_points),
7   y = as.vector(y_generated_points)
8 )
9
10 # Run nonlinear least-squares model
11 model_gslnls <- gsl_nls(
12   fn = y ~ energy_dependency(
13     T,
14     T0 = input_parameters$T0,
15     A = input_parameters$A,
16     B = input_parameters$B,
17     C
18   ),
19   data = data,
20   algorithm = "lmaccel",
21   start = c(C = 1)
22 )

```


Модель з локальною адитивною похибкою

Нехай тепер задана ось така модель:

$$E_i = A + B \tanh \left(\frac{T_i - T_0}{C + \varepsilon_i} \right), \quad i = \overline{1, n}, \quad (2.3)$$

де значення A , B та T_0 є фіксованими (Табл. 4), величина C є невідомим параметром моделі, а випадкова похибка ε_i має нормальний розподіл:

$$\varepsilon_i \sim N(0, \sigma_c^2), \quad i = \overline{1, n} \quad (2.4)$$

Організація генерування даних у контексті моделі (2.3) аналогічна попередньому прикладу з глобальною адитивною похибкою. Однак, отримані результати свідчать про неспроможність МНК (Рис. 5).

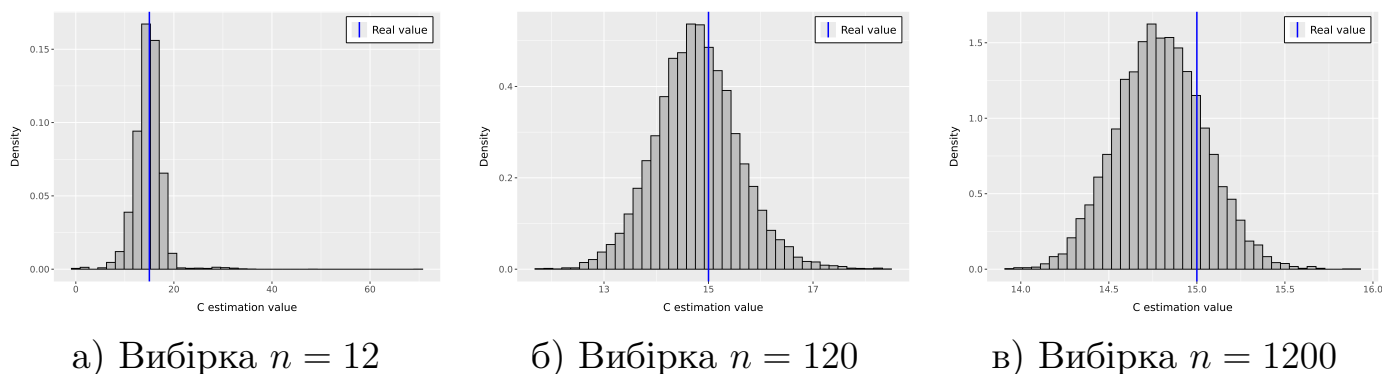


Рис. 5: Результати серії з $N = 10\,000$ повторних запусків МНК (локальна адитивна помилка)

Розмір вибірки	Істинне значення	Вибіркове середнє	Стандартне відхилення
$n = 12$	15	14.7290	2.8841
$n = 120$		14.7791	0.7993
$n = 1200$		14.7945	0.2491

Таблиця 6: Статистичні характеристики серії з $N = 10\,000$ повторних запусків МНК (локальна адитивна помилка)

Перелік посилань

1. Algorithms to Estimate the Ductile to Brittle Transition Temperature, Upper Shelf Energy, and Their Uncertainties for Steel Using Charpy V-Notch Shear Area and Absorbed Energy Data / N. T. Switzner [та ін.] // Metals. — 2023. — Трав. — Т. 13, № 5. — С. 877. — ISSN 2075-4701. — DOI: 10.3390/met13050877. — URL: <http://dx.doi.org/10.3390/met13050877>.