Моделі статистичного навчання: методи ресемплінгу (передискретизації, повторної вибірки)

Два найпопулярніші методи

1. Перехресна перевірка

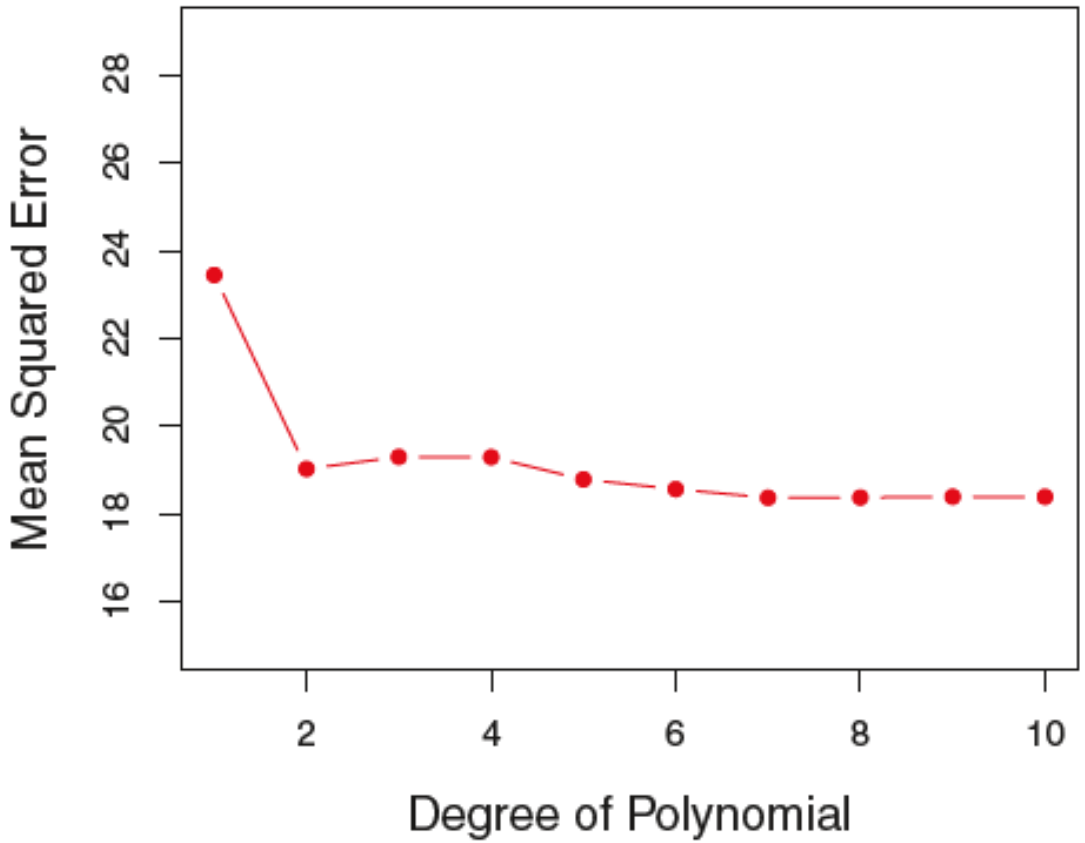
2. Бутстрап

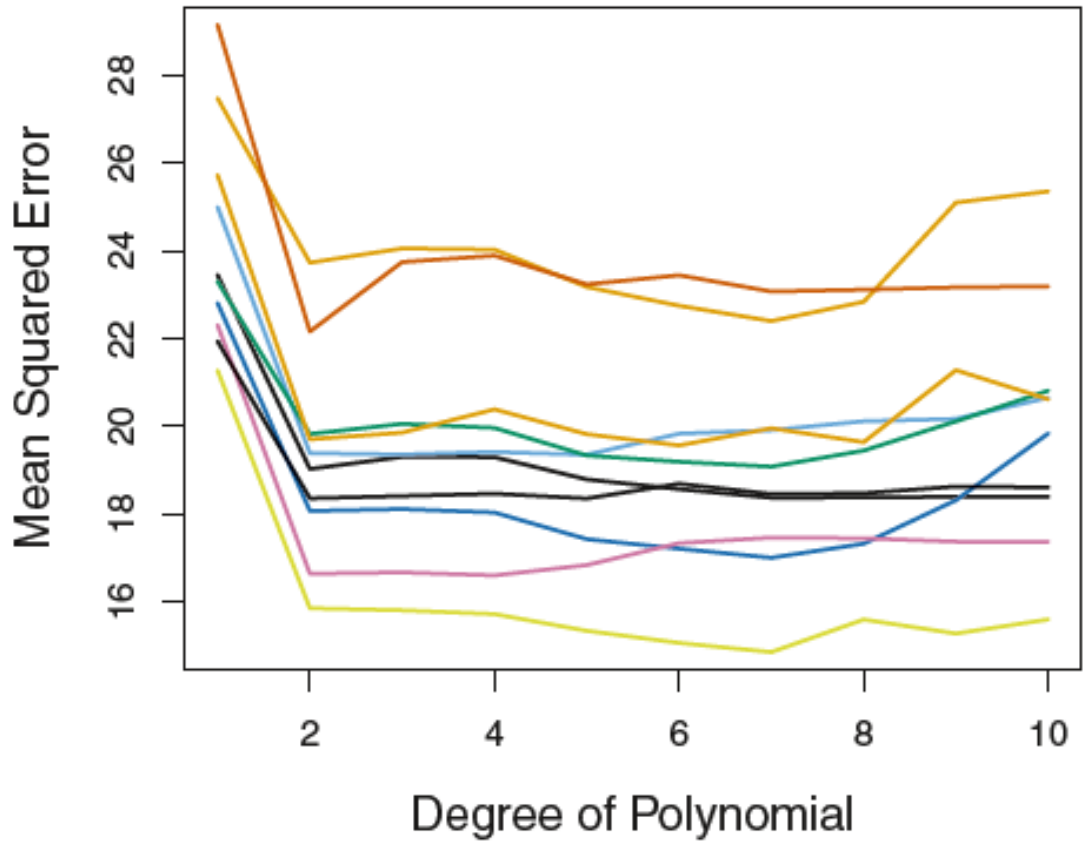
Перехресна перевірка

Перехресна перевірка

Приклад. Дані Auto, mpg, horsepower…

Розбиваємо вибірку на дві (тестову і навчальну) та обчислюємо тестову помилку.


Mean Squared Error
16
18
20
22
24
26
28
2
4
6
8
10
Degree of Polynomial


Mean Squared Error
16
18
20
22
24
26
28
2
4
6
8
10
Degree of Polynomial

Недоліки:

Недоліки:

1. Тестова помилка сильно змінюється, в залежності від того, які спостереження включені до навчального набору, а які спостереження включені до набору перевірок.

Недоліки:

1. Тестова помилка сильно змінюється, в залежності від того, які спостереження включені до навчального набору, а які спостереження включені до набору перевірок.

2. Оскільки лише підмножина спостережень використовується для підгонки моделі, то тестова помилка може бути переоцінена.

Leave-one-out cross-validation (LOOCV)

Leave-one-out cross-validation (LOOCV)

Ідея: у тренувальну вибірку включати одне спостереження.

Leave-one-out cross-validation (LOOCV)

Ідея: у тренувальну вибірку включати одне спостереження.

Оцінюємо моделі на спостереженнях $\{(x_2, y_2),\ldots,(x_n, y_n)\}$, а тестуємо на $\{(x_1, y_1)\}$

Leave-one-out cross-validation (LOOCV)

Ідея: у тренувальну вибірку включати одне спостереження.

Оцінюємо моделі на спостереженнях $\{(x_2, y_2),\ldots,(x_n, y_n)\}$, а тестуємо на $\{(x_1, y_1)\}$

Отримуємо $MSE_1 = (y_1 - \hat{y}_1)^2$ незміщена оцінка тестової помилки.

Leave-one-out cross-validation (LOOCV)

Ідея: у тренувальну вибірку включати одне спостереження.

Оцінюємо моделі на спостереженнях $\{(x_2, y_2),\ldots,(x_n, y_n)\}$, а тестуємо на $\{(x_1, y_1)\}$

Отримуємо $MSE_1 = (y_1 - \hat{y}_1)^2$ незміщена оцінка тестової помилки.

Аналогічно обчислюємо $MSE_2,\ldots, MSE_n$.

Leave-one-out cross-validation (LOOCV)

Ідея: у тренувальну вибірку включати одне спостереження.

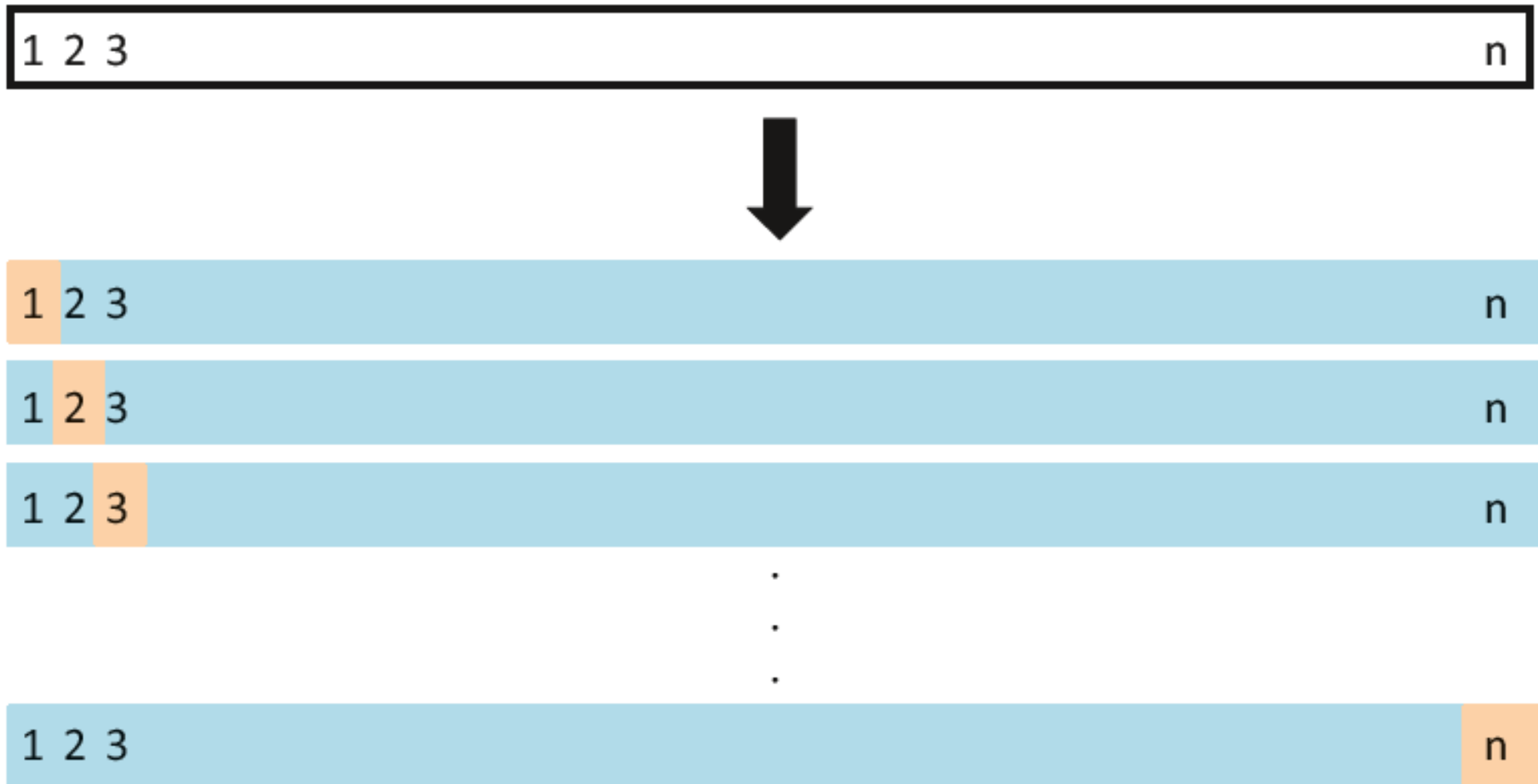
Оцінюємо моделі на спостереженнях $\{(x_2, y_2),\ldots,(x_n, y_n)\}$, а тестуємо на $\{(x_1, y_1)\}$

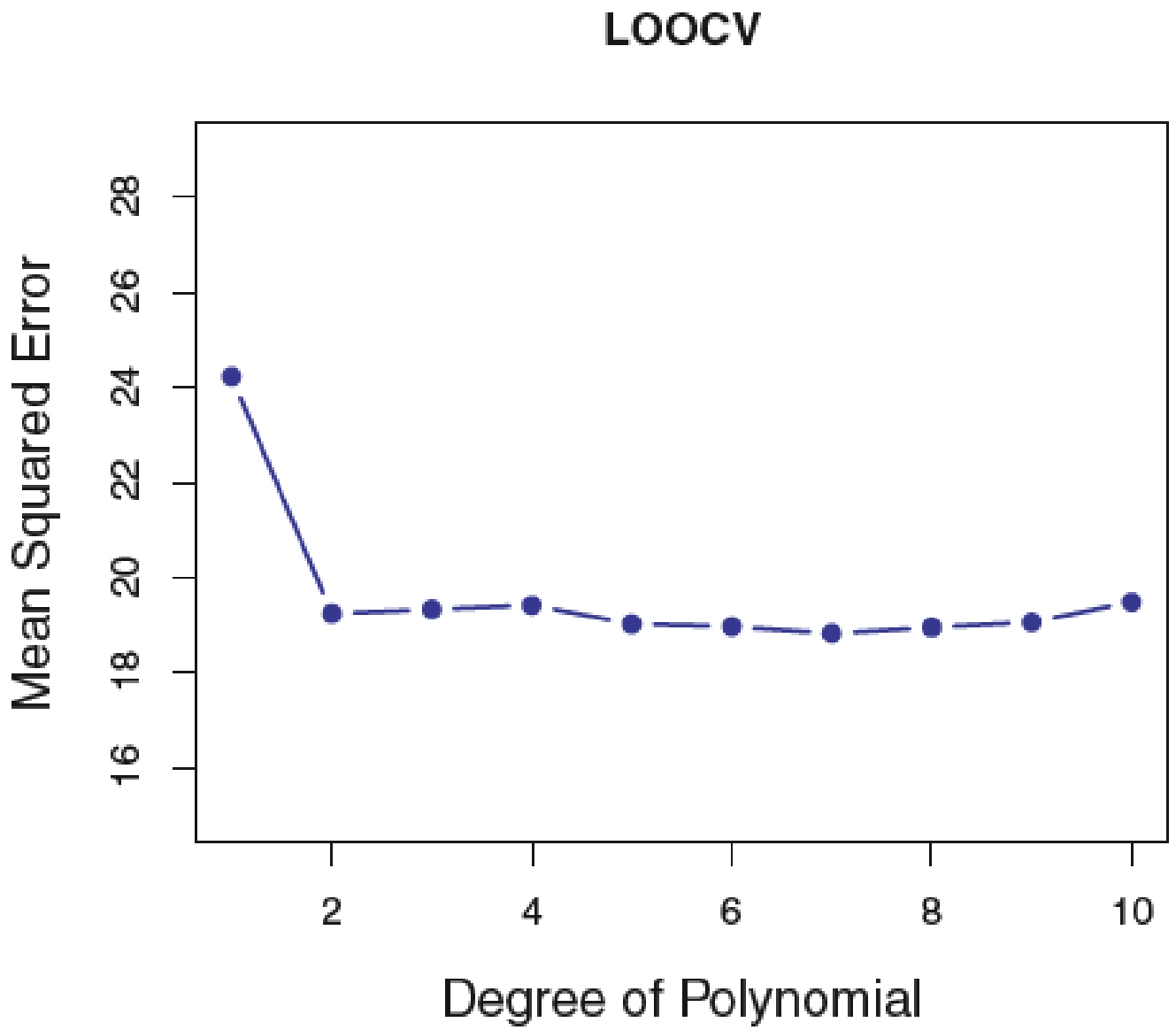
Отримуємо $MSE_1 = (y_1 - \hat{y}_1)^2$ незміщена оцінка тестової помилки.

Аналогічно обчислюємо $MSE_2,\ldots, MSE_n$.

Оцінка для тестової помилки

$$\mathrm{CV}_{(n)} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{MSE}_i \qquad\qquad \mathrm{CV}_{(n)} = \frac{1}{n}\sum_{i=1}^{n} \left(\frac{y_i - \hat{y}_i}{1 - h_i}\right)^2$$


1 2 3
n
1 2 3
n
1 2 3
n
1 2 3
n
1 2 3
n


LOOCV
Mean Squared Error
16
18
20
22
24
26
28
2
4
6
8
10
Degree of Polynomial

k-Fold Cross-Validation

k-Fold Cross-Validation

Ідея: розбиваємо загальну вибірку на k груп приблизно однакового розміру.

k-Fold Cross-Validation

Ідея: розбиваємо загальну вибірку на k груп приблизно однакового розміру.

Оцінюємо модель на спостереженнях з k – 1 груп, а тестуємо на іншій групі

k-Fold Cross-Validation

Ідея: розбиваємо загальну вибірку на k груп приблизно однакового розміру.

Оцінюємо модель на спостереженнях з k – 1 груп, а тестуємо на іншій групі

Отримуємо MSE_1,…, MSE_k .

k-Fold Cross-Validation

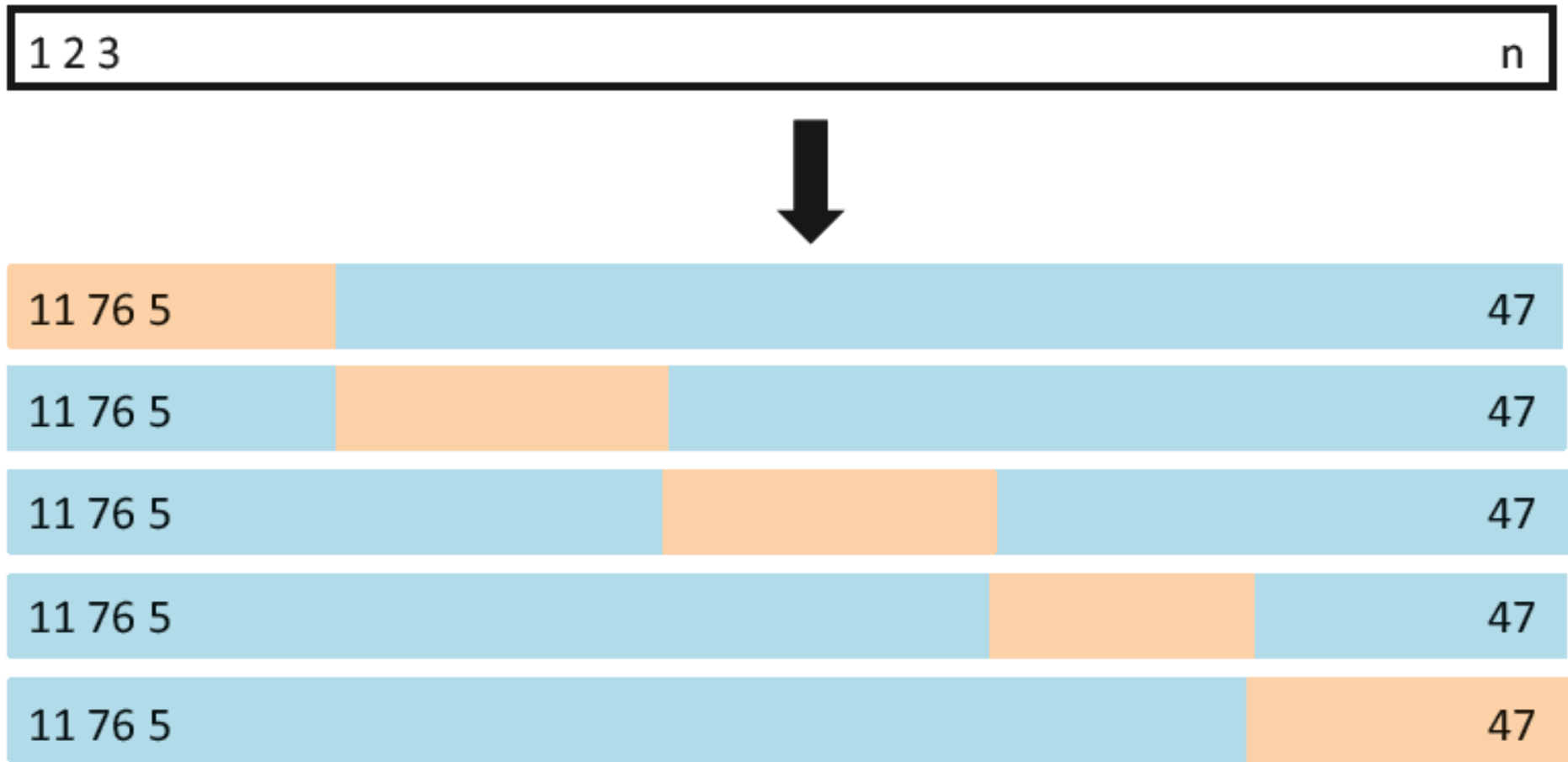
Ідея: розбиваємо загальну вибірку на k груп приблизно однакового розміру.

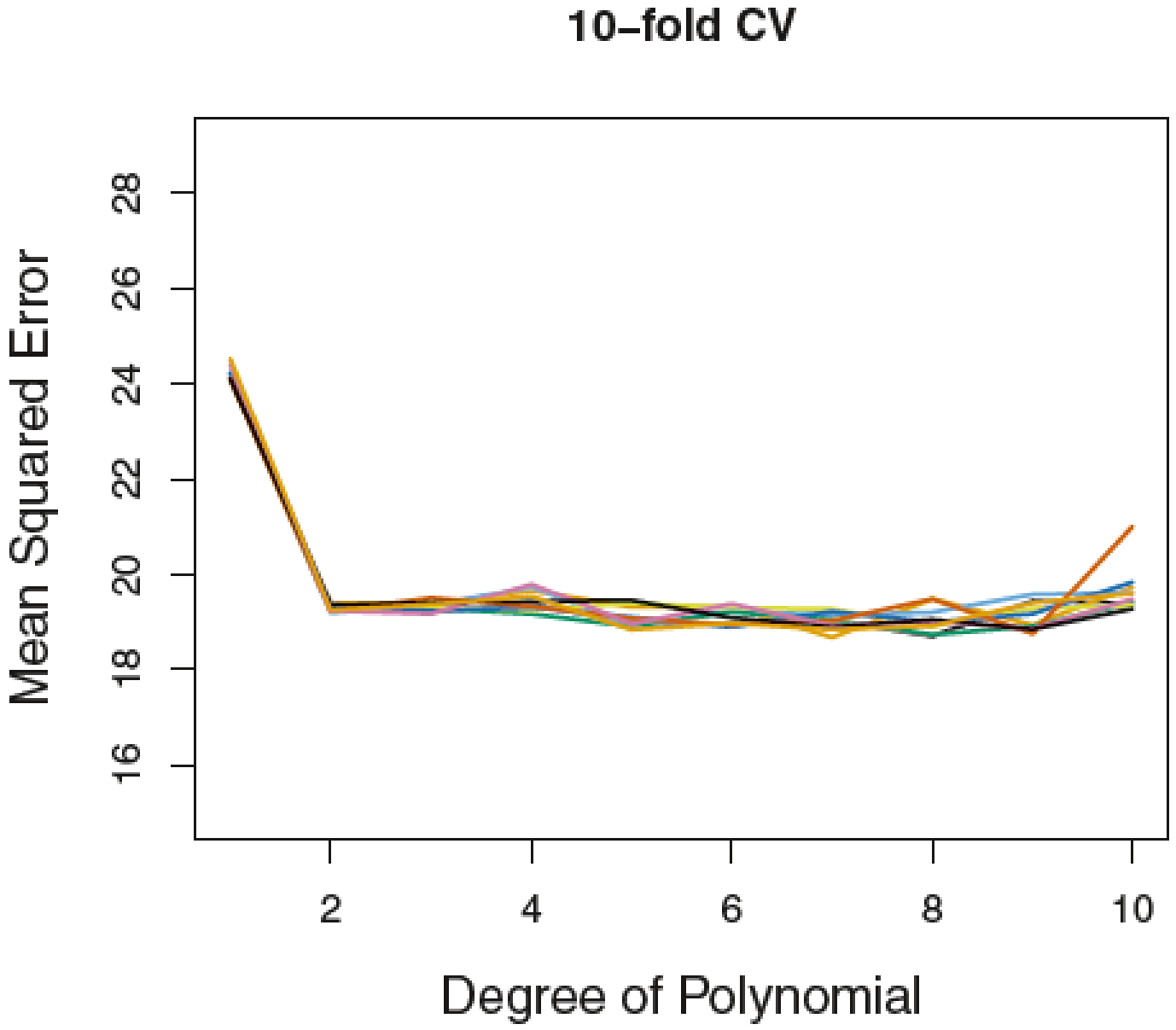
Оцінюємо модель на спостереженнях з k – 1 груп, а тестуємо на іншій групі

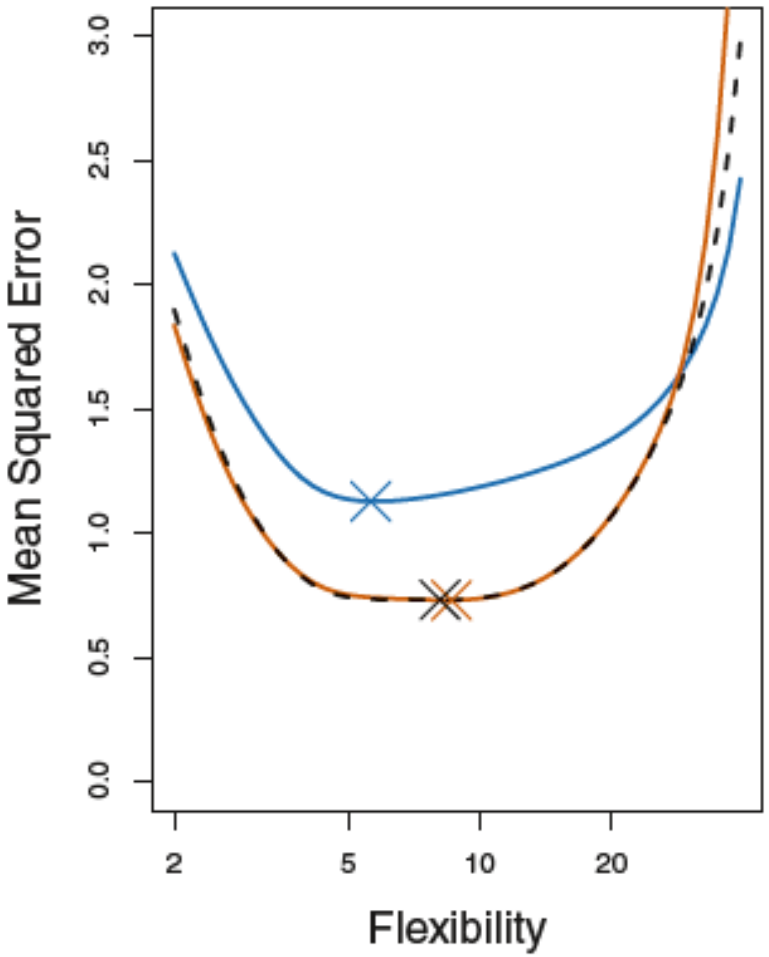
Отримуємо MSE_1,…, MSE_k .

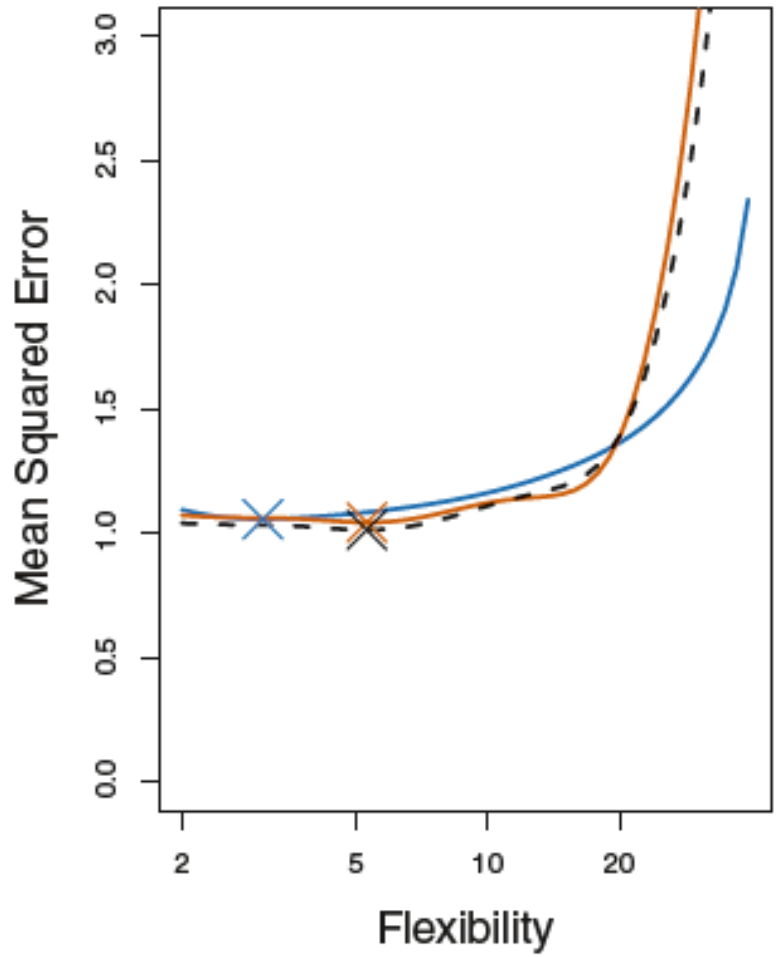
Оцінка для тестової помилки

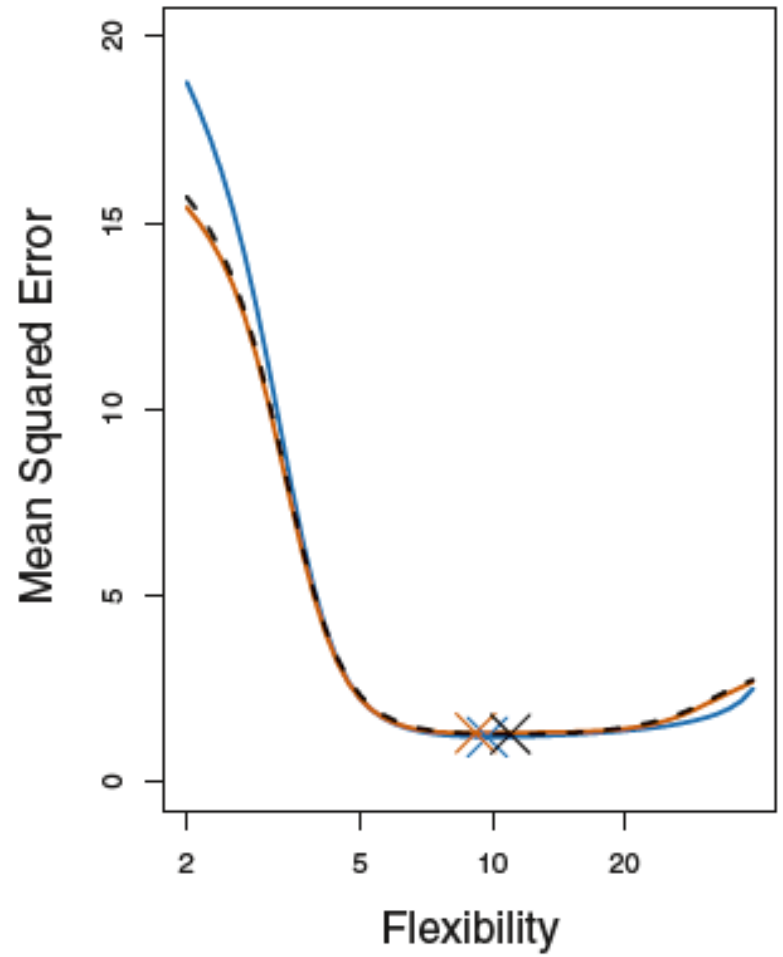
$$\mathrm{CV}_{(k)} = \frac{1}{k}\sum_{i=1}^{k} \mathrm{MSE}_i$$


1 2 3
n
11 76 5
47
11 76 5
47
11 76 5
47
11 76 5
47
11 76 5
47


10−fold CV
Mean Squared Error
16
18
20
22
24
26
28
2
4
6
8
10
Degree of Polynomial


Mean Squared Error
0.0
0.5
1.0
1.5
2.0
2.5
3.0
2
5
10
20
Flexibility


Mean Squared Error
0.0
0.5
1.0
1.5
2.0
2.5
3.0
2
5
10
20
Flexibility


Mean Squared Error
0
5
10
15
20
2
5
10
20
Flexibility

Перехресна перевірка у випадку задач класифікації

Перехресна перевірка у випадку задач класифікації

$$\mathrm{CV}_{(n)} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{Err}_i$$

$$\mathrm{Err}_i = I(y_i \neq \hat{y}_i)$$

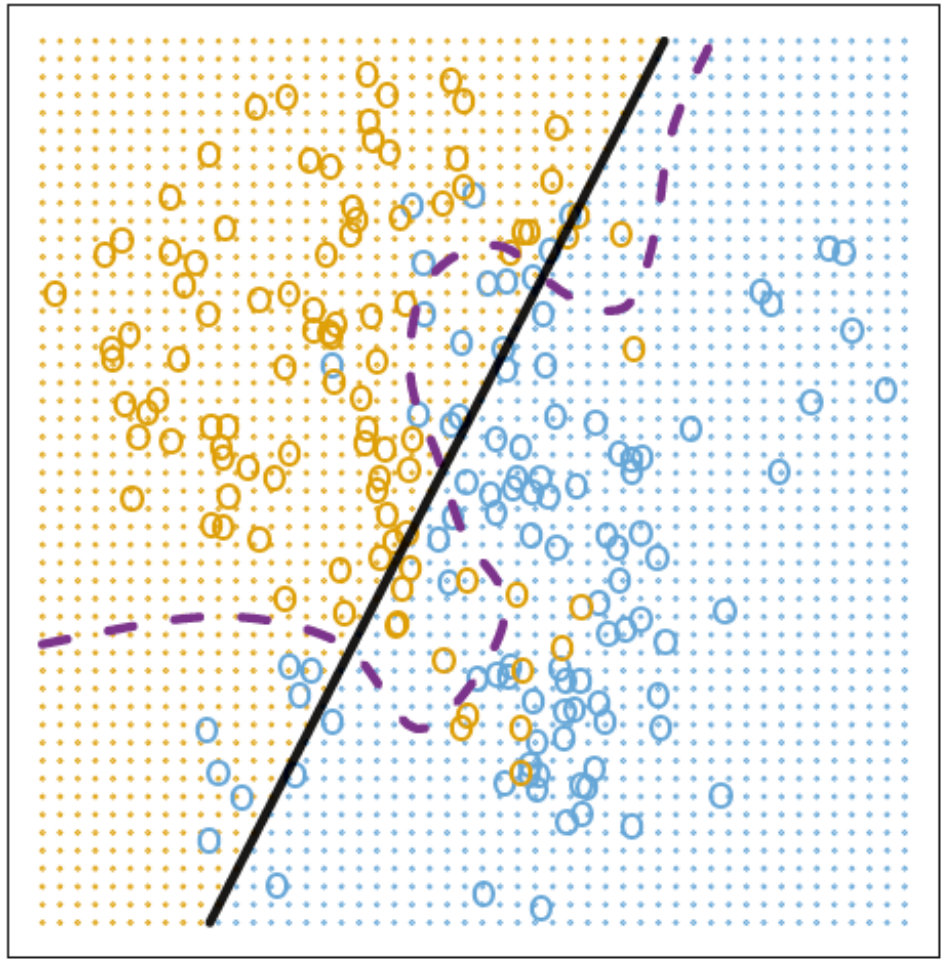
Перехресна перевірка у випадку задач класифікації

$$\mathrm{CV}_{(n)} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{Err}_i$$

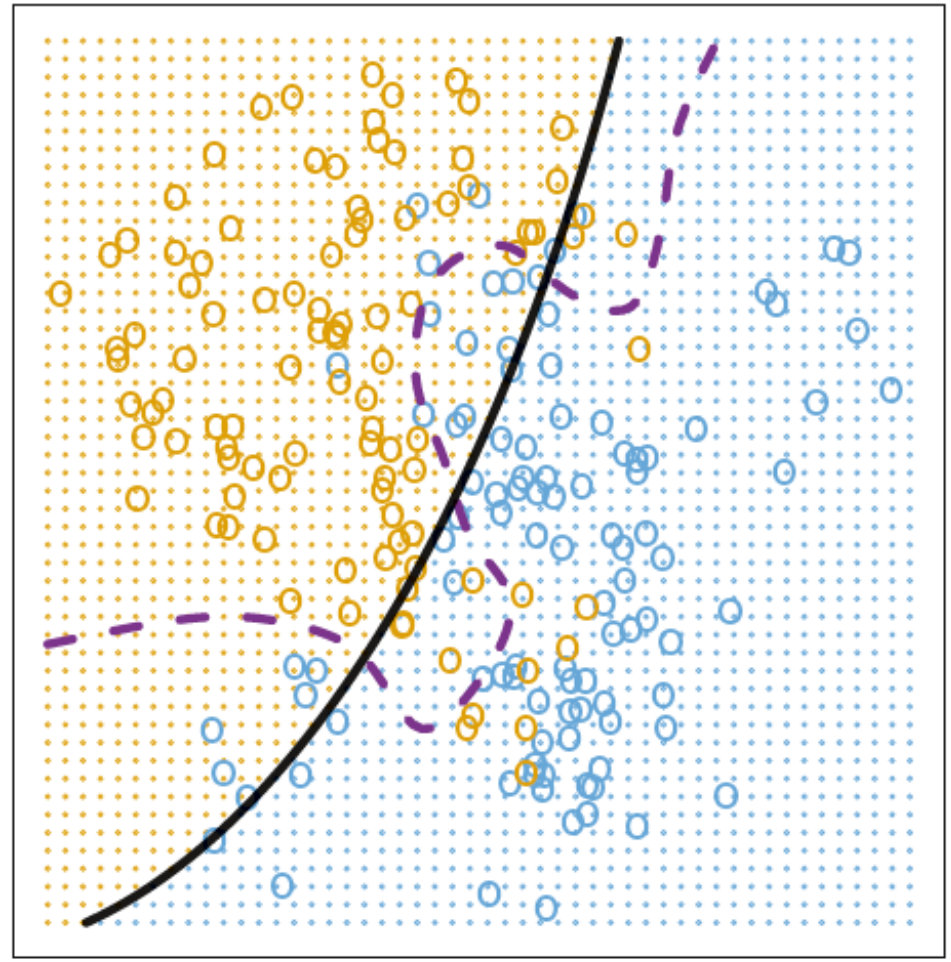
$$\mathrm{Err}_i = I(y_i \neq \hat{y}_i)$$

Для наступного прикладу, маємо: тестові помилки дорівнюють 0.201, 0.197, 0.160 і 0.162. Мінімальна помилка становить 0.133.

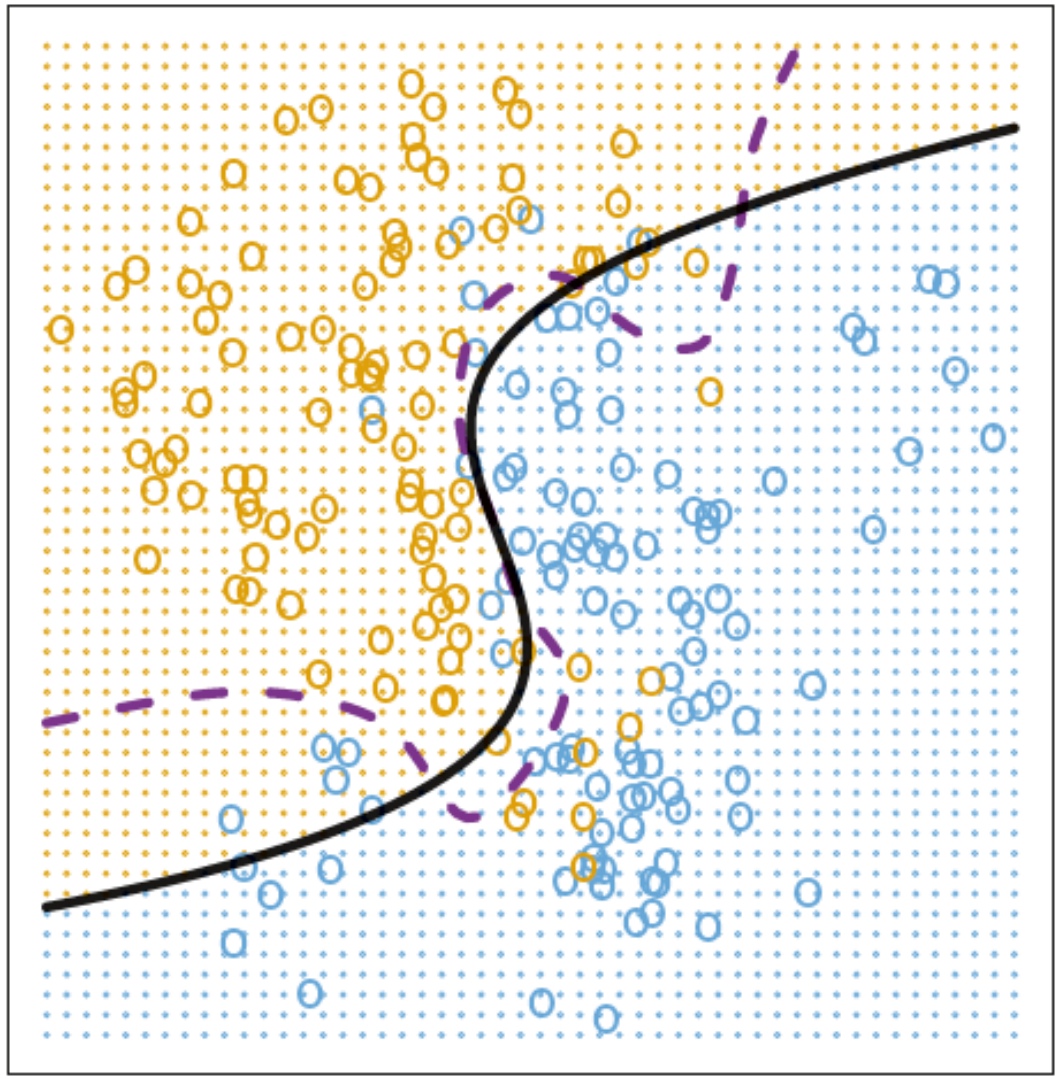
Degree=1



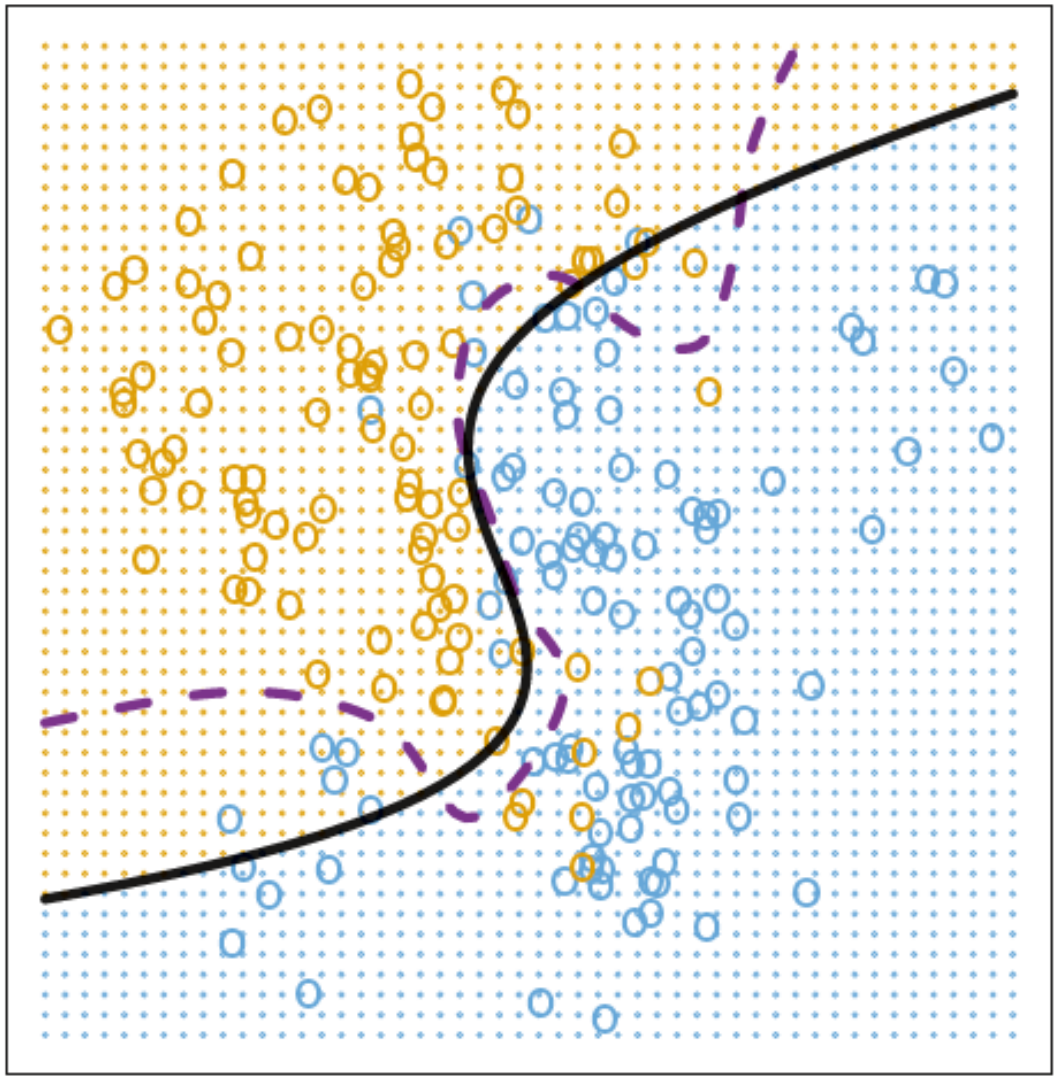
Degree=2

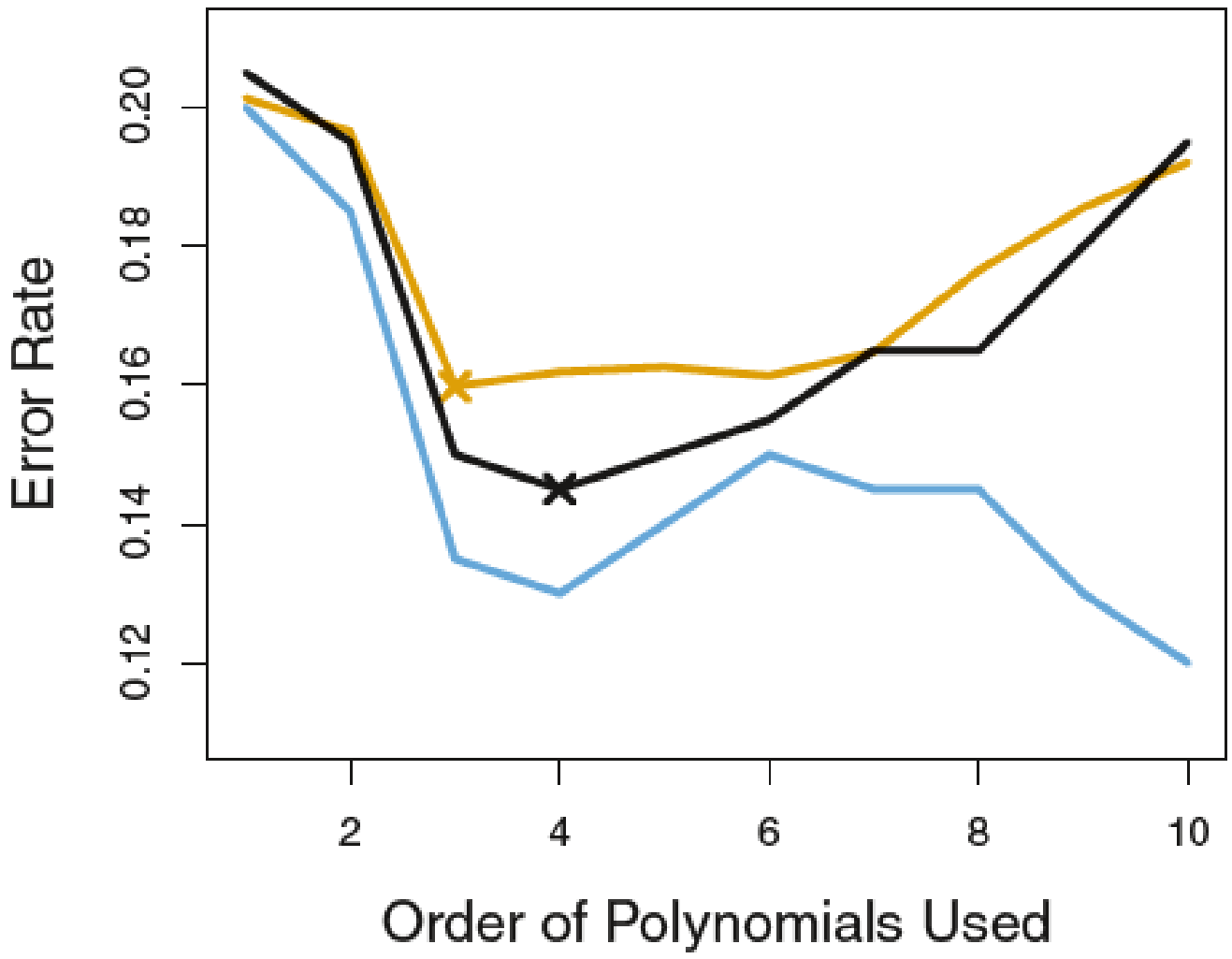


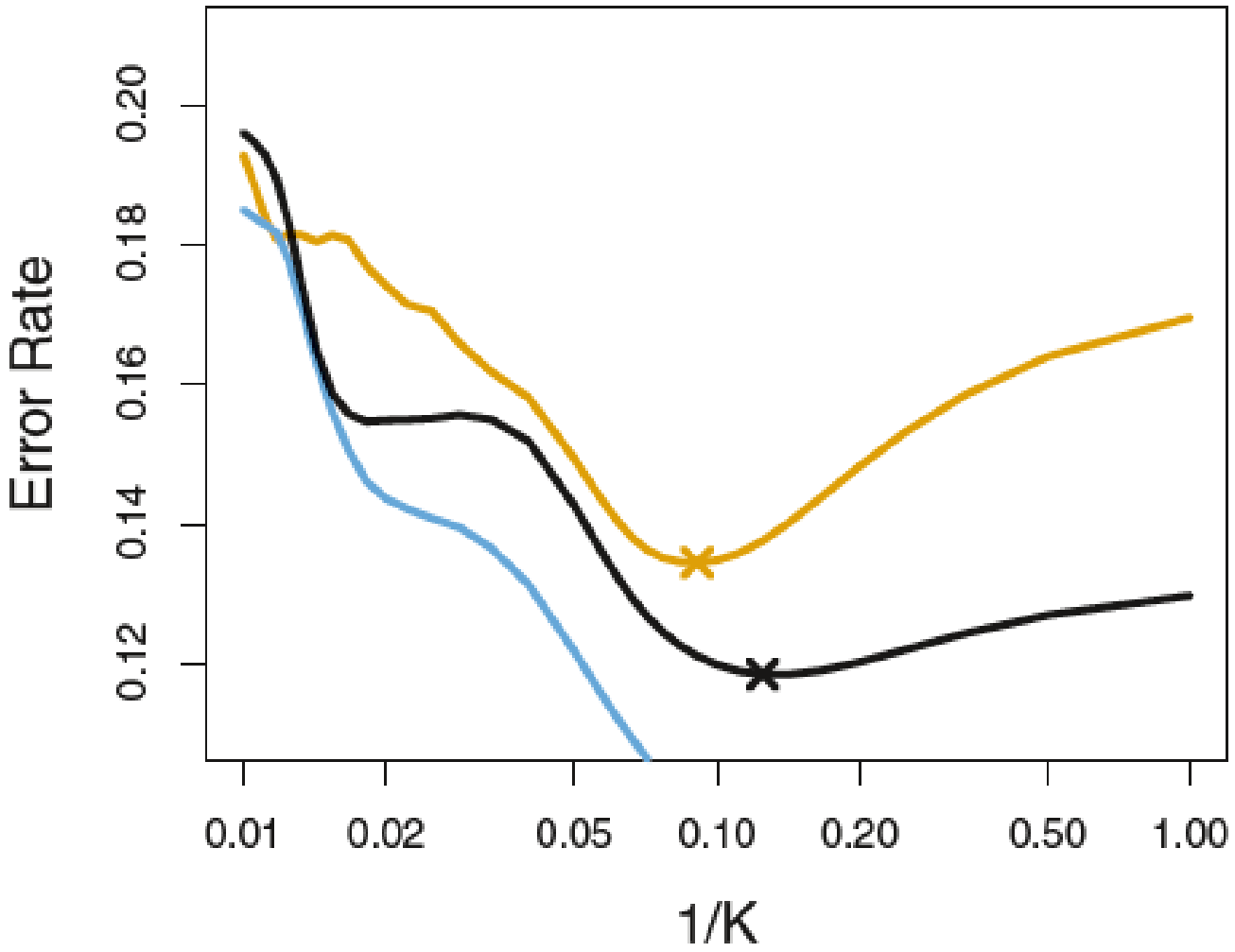
Degree=3



Degree=4




Error Rate
0.12
0.14
0.16
0.18
0.20
2
4
6
8
10
Order of Polynomials Used


Error Rate
0.12
0.14
0.16
0.18
0.20
0.01
0.02
0.05
0.10
0.20
0.50
1.00
1/K

Бутстрап

Бутстрап

Приклад: Припустимо, що ми хочемо вкласти фіксовану суму грошей у два фінансових активи, що мають дохідність X та Y відповідно, де X та Y випадкові величини. Ми вкладемо частку α наших грошей у X, а решту 1 - α в Y. Ми хочемо вибрати α, щоб мінімізувати загальний ризик наших інвестицій. Іншими словами, ми хочемо мінімізувати

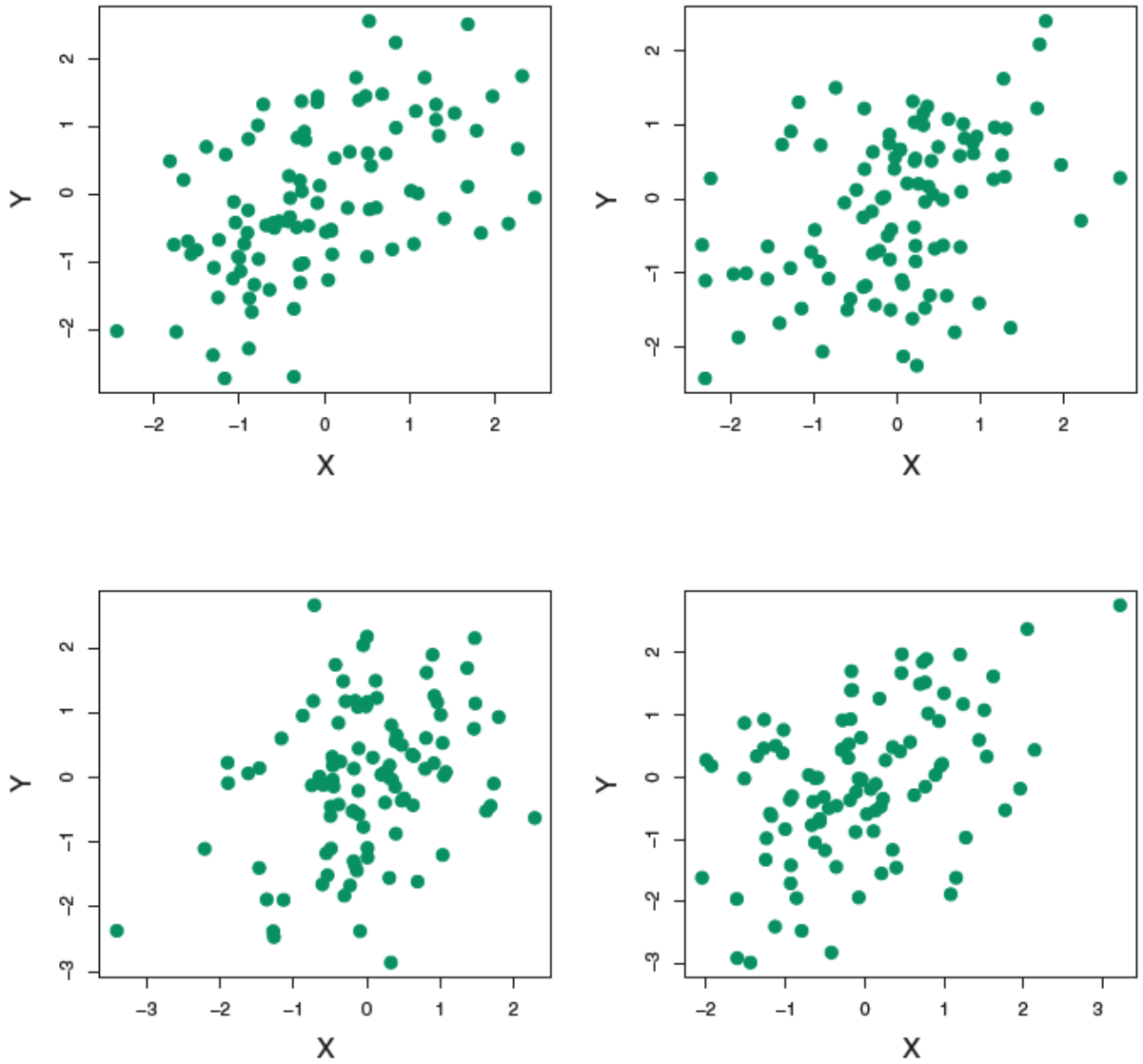
$$\text{Var}\,(\alpha X + (1 - \alpha)\, Y).$$

Бутстрап

Приклад: Припустимо, що ми хочемо вкласти фіксовану суму грошей у два фінансових активи, що мають дохідність X та Y відповідно, де X та Y випадкові величини. Ми вкладемо частку α наших грошей у X, а решту 1 - α в Y. Ми хочемо вибрати α, щоб мінімізувати загальний ризик наших інвестицій. Іншими словами, ми хочемо мінімізувати

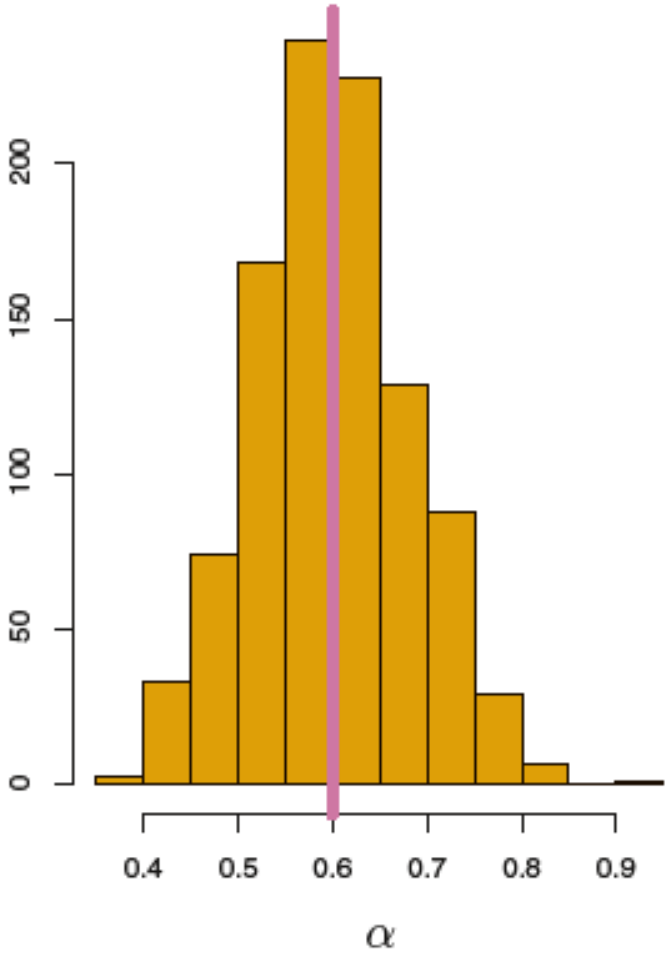
$$\text{Var}\,(\alpha X + (1 - \alpha)\, Y).$$

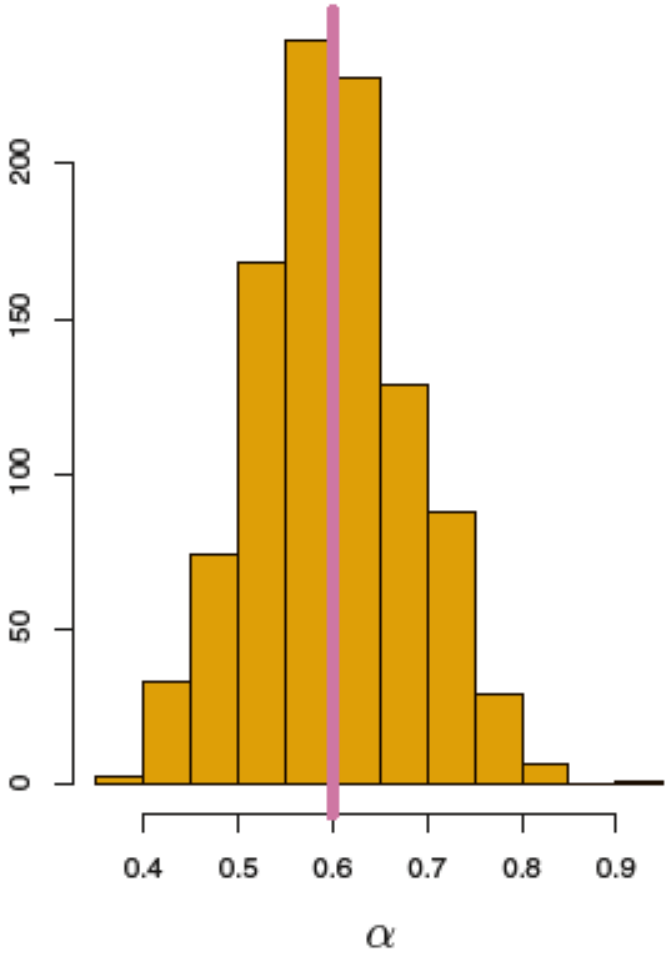
$$\alpha = \frac{\sigma_Y^2 - \sigma_{XY}}{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}}$$



Отримані значення для α становлять 0.576, 0.532, 0.657 і 0.651.

Точні значення $\sigma_X^2 = 1$, $\sigma_Y^2 = 1.25$, $\sigma_{XY} = 0.5$, звідки отримуємо $\alpha = 0.6$.

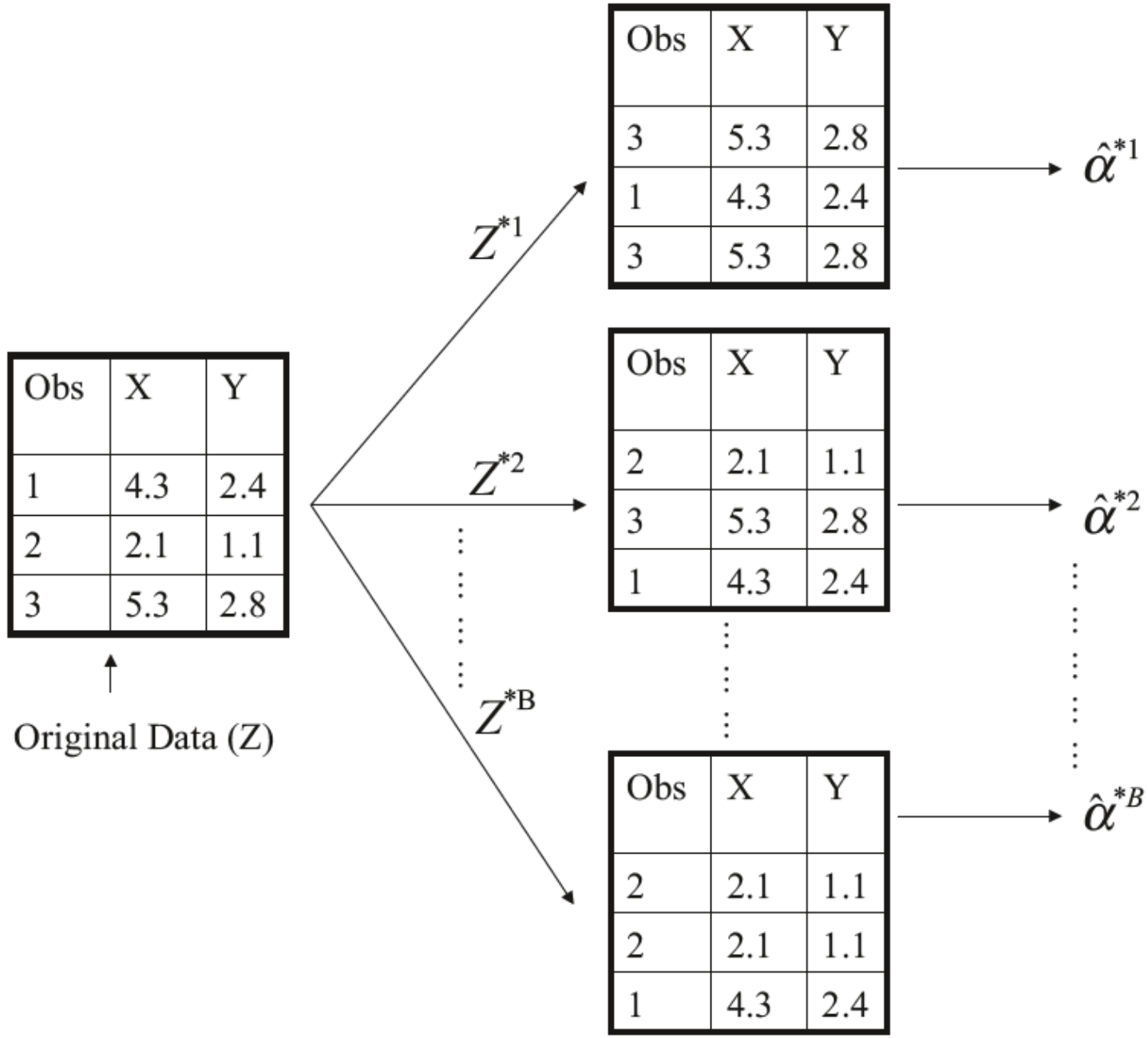




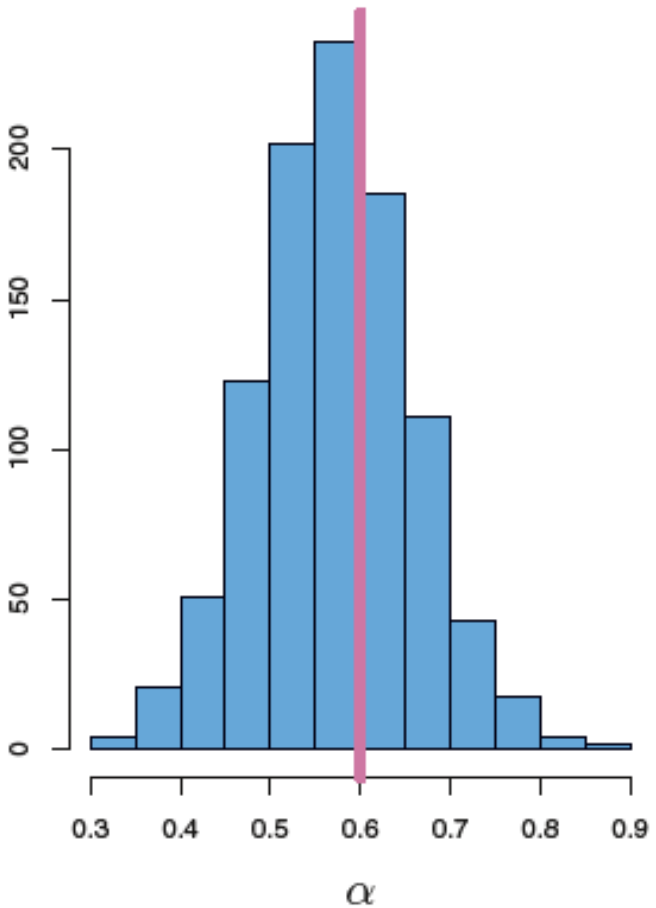
Точні значення $\sigma_X^2 = 1$, $\sigma_Y^2 = 1.25$, $\sigma_{XY} = 0.5$, звідки отримуємо $\alpha = 0.6$.

$$\bar{\alpha} = \frac{1}{1,000} \sum_{r=1}^{1,000} \hat{\alpha}_r = 0.5996$$

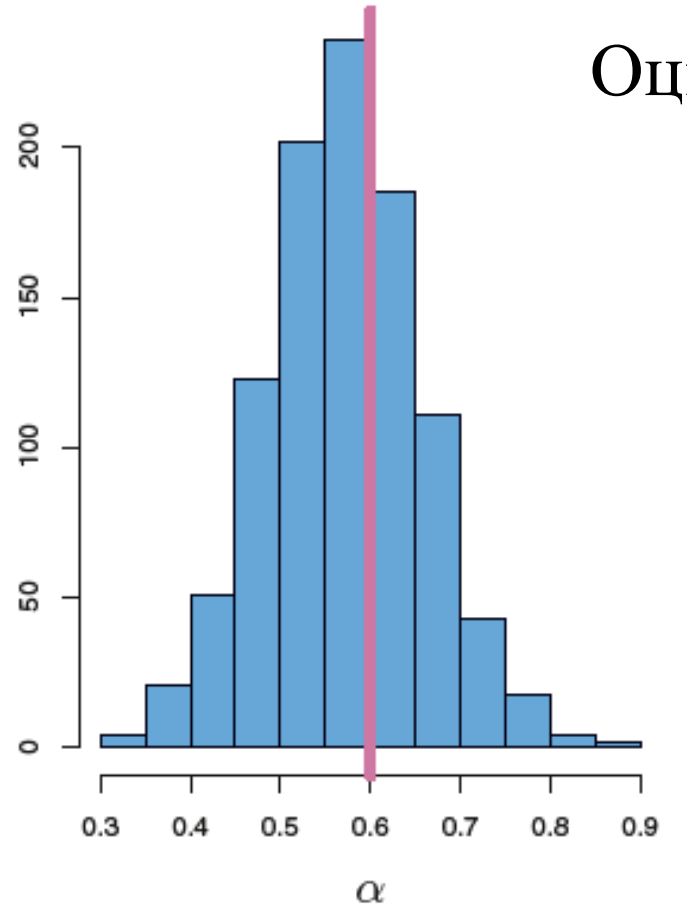
$$\sqrt{\frac{1}{1,000-1} \sum_{r=1}^{1,000} (\hat{\alpha}_r - \bar{\alpha})^2} = 0.083$$


Obs X Y
1 4.3 2.4
2 2.1 1.1
3 5.3 2.8
Original Data (Z)
Z^{*1}
Z^{*2}
Z^{*B}
Obs X Y
3 5.3 2.8
1 4.3 2.4
3 5.3 2.8
Obs X Y
2 2.1 1.1
3 5.3 2.8
1 4.3 2.4
Obs X Y
2 2.1 1.1
2 2.1 1.1
1 4.3 2.4
$\hat{\alpha}^{*1}$
$\hat{\alpha}^{*2}$
$\hat{\alpha}^{*B}$

$$\mathrm{SE}_B(\hat{\alpha}) = \sqrt{\frac{1}{B-1}\sum_{r=1}^{B}\left(\hat{\alpha}^{*r} - \frac{1}{B}\sum_{r'=1}^{B}\hat{\alpha}^{*r'}\right)^2}.$$



$$\mathrm{SE}_B(\hat{\alpha}) = \sqrt{\frac{1}{B-1}\sum_{r=1}^{B}\left(\hat{\alpha}^{*r} - \frac{1}{B}\sum_{r'=1}^{B}\hat{\alpha}^{*r'}\right)^2}.$$



Оцінка дисперсії становить 0.087