

Seminar 8: Линейная регрессия

Carlos Buitrago

Table of contents

1	1. Постановка модели	1
1.1	Идентифицируемость	2
2	2. Метод наименьших квадратов и ортогональная проекция	2
3	3. Несмещённость, ковариация и оценка дисперсии	3
3.1	Теорема Гаусса—Маркова (дополнение)	4
4	4. Разобранный пример из главы: три измерения	4
5	5. Гауссовская модель: проекции и точные распределения	5
5.1	Теорема об ортогональных гауссовских проекциях	5
5.2	Распределения, которые нужно знать	5
6	6. Точные доверительные интервалы и статистические выводы	6
7	7. Взвешенный МНК, коррелированные ошибки и «отбеливание» (дополнение)	6
7.1	Коррелированный шум: накопленные ошибки	7
8	8. Дополнительные инструменты для задач	9
8.1	Геометрия нормальной выборки	9

1 1. Постановка модели

Пусть $X \in \mathbb{R}^n$ — наблюдаемый случайный вектор, $Z \in \mathbb{R}^{n \times k}$ — **известная** матрица плана, $\theta \in \mathbb{R}^k$ — неизвестный вектор параметров, $k < n$, $\text{rank } Z = k$. Рассматриваем

$$X = Z\theta + \varepsilon, \quad \mathbb{E}\varepsilon = 0, \quad \text{Cov}(\varepsilon) = \sigma^2 I_n.$$

В исходной главе параметры модели — θ и σ^2 ; $L = \text{col}(Z)$ — известное k -мерное подпространство, $\ell = Z\theta \in L$ — неизвестное среднее.

! Два разных набора предположений

Для формул несмещённости и ковариации МНК **нормальность не нужна**. Для точных χ^2 , t , F -распределений, независимости проекций и точных доверительных интервалов дополнительно требуется $\varepsilon \sim N(0, \sigma^2 I_n)$.

1.1 Идентифицируемость

Полный столбцовый ранг Z эквивалентен положительной определённости $Z^T Z$ и единственности координат θ . При недостаточном ранге проекция $\hat{\ell}$ всё ещё единственна, но $\hat{\theta}$ может быть неоднозначной.

2. Метод наименьших квадратов и ортогональная проекция

Определение. $\hat{\theta} = \arg \min_{u \in \mathbb{R}^k} \|X - Zu\|^2$. Разложение квадрата даёт

$$\|X - Zu\|^2 = X^T X - 2u^T Z^T X + u^T Z^T Z u.$$

Дифференцируя по u , получаем **нормальные уравнения** $Z^T Z \hat{\theta} = Z^T X$, следовательно,

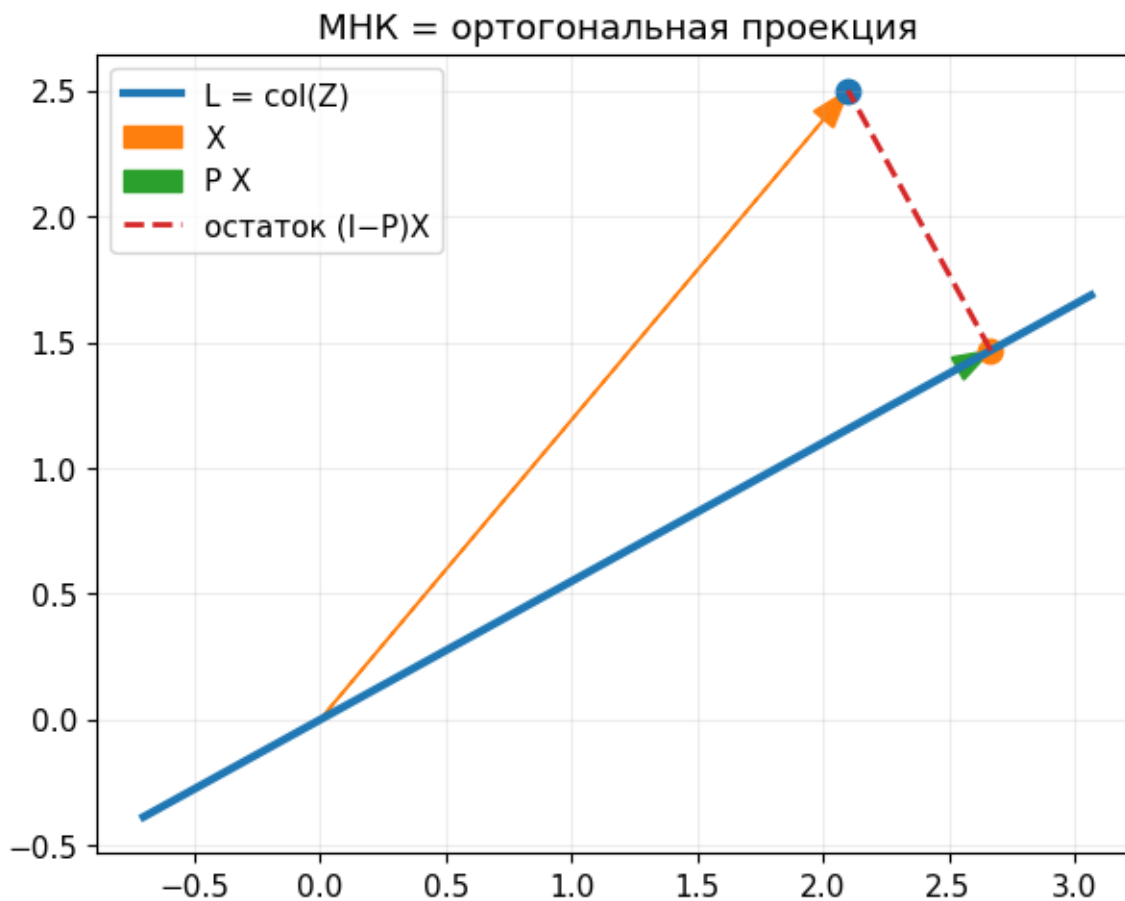
$$\hat{\theta} = (Z^T Z)^{-1} Z^T X, \quad \hat{\ell} = Z \hat{\theta} = P_Z X, \quad P_Z = Z(Z^T Z)^{-1} Z^T.$$

$P_Z^T = P_Z$, $P_Z^2 = P_Z$; значит, P_Z — ортогональный проектор на L . Остатки $e = X - Z \hat{\theta} = (I - P_Z)X$ лежат в $L^\perp$, поэтому $Z^T e = 0$. Теорема Пифагора:

$$\|X\|^2 = \|P_Z X\|^2 + \|(I - P_Z)X\|^2.$$

💡 Как распознать матрицу плана в задаче?

Если $\mathbb{E}X_i = \beta_1 + i\beta_2$, то строка i матрицы Z равна $(1, i)$. Если измеряются суммы $a + b$, строка $(1, 1)$; если только a , строка $(1, 0)$. Матрица плана кодирует **математические ожидания**, а не шум.



Скалярное произведение проекции и остатка: $5.100045734843481\text{e-}16$

3. Несмещённость, ковариация и оценка дисперсии

Подставим $X = Z\theta + \varepsilon$:

$$\hat{\theta} = \theta + (Z^T Z)^{-1} Z^T \varepsilon.$$

Отсюда сразу

$$\mathbb{E}\hat{\theta} = \theta, \quad \text{Cov}(\hat{\theta}) = \sigma^2 (Z^T Z)^{-1}.$$

Обозначим $Q = I - P_Z$. Тогда $QZ = 0$, $Q^2 = Q$, $\text{tr } Q = n - k$, а остаток $e = Q\varepsilon$. Поэтому

$$\mathbb{E}\|e\|^2 = \mathbb{E}(\varepsilon^T Q \varepsilon) = \text{tr}(Q \sigma^2 I_n) = (n - k) \sigma^2.$$

Несмещённая оценка дисперсии:

$$\hat{\sigma}^2 = \frac{\text{RSS}}{n - k}, \quad \text{RSS} = \|X - Z\hat{\theta}\|^2.$$

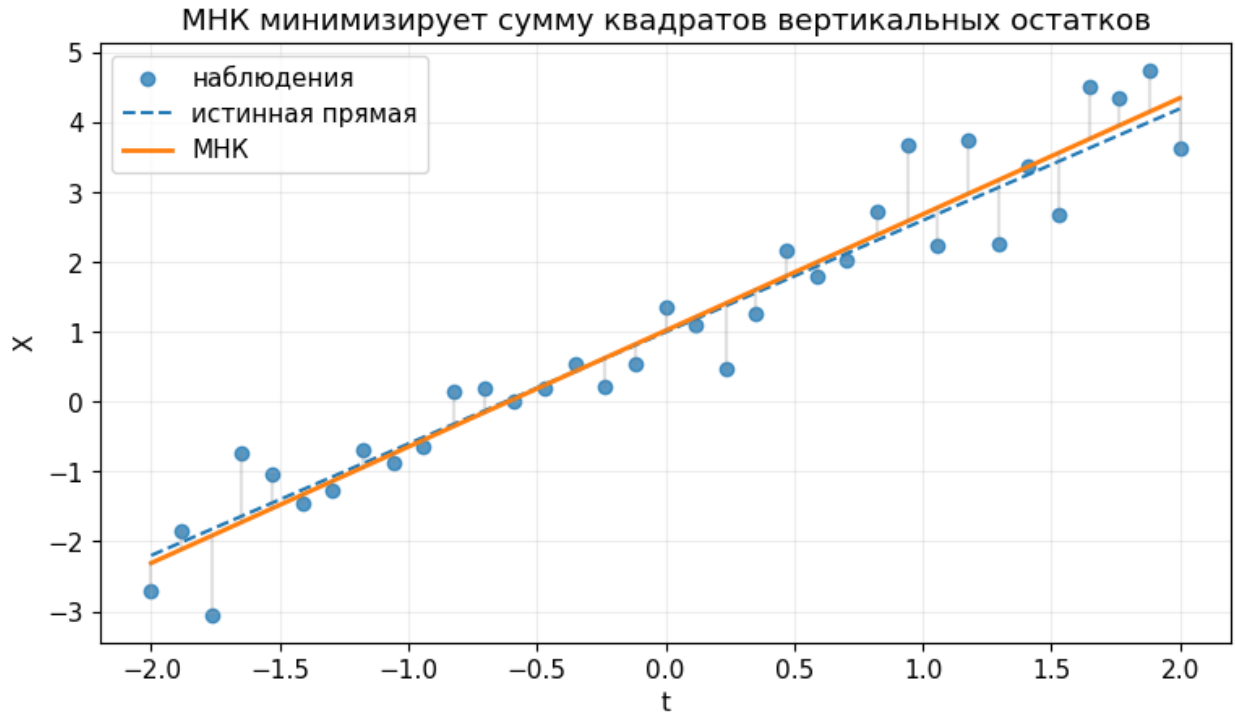
Число $n - k$ — размерность пространства остатков, то есть число остаточных степеней свободы. Не путайте $\hat{\sigma}^2$ с гауссовской MLE RSS/n .

3.1 Теорема Гаусса—Маркова (дополнение)

Среди **линейных несмещённых** оценок θ МНК имеет наименьшую ковариационную матрицу в порядке Лёвнера. Для любой AX с $AZ = I_k$ запишем $A = (Z^T Z)^{-1} Z^T + D$, где $DZ = 0$. Тогда

$$\text{Cov}(AX) - \text{Cov}(\hat{\theta}) = \sigma^2 D D^T \succeq 0.$$

Это утверждение не означает оптимальности среди всех нелинейных оценок без дополнительных условий.



Оценки: [1.02183848 1.66561276] ; остаточная дисперсия: 0.31207222248570765

4 4. Разобранный пример из главы: три измерения

Пусть $\ell = (a, b, a + b)^T$ и $Z = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$. Тогда

$$Z^T Z = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \quad (Z^T Z)^{-1} = \frac{1}{3} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}.$$

Следовательно,

$$\hat{a} = \frac{2X_1 - X_2 + X_3}{3}, \quad \hat{b} = \frac{-X_1 + 2X_2 + X_3}{3},$$

$$\hat{\sigma}^2 = \frac{(X_1 + X_2 - X_3)^2}{3}, \quad (n - k = 1).$$

Последнее равенство следует из того, что $L^\perp$ порождается $(1, 1, -1)^T$.

Матрица $(Z^T Z)^{-1}$:

```
[ [ 0.66666667 -0.33333333]
  [-0.33333333  0.66666667] ]
```

Проектор остатков:

```
[ [ 0.33333333  0.33333333 -0.33333333]
  [ 0.33333333  0.33333333 -0.33333333]
  [-0.33333333 -0.33333333  0.33333333] ]
```

5 5. Гауссовская модель: проекции и точные распределения

Предположим теперь $X \sim N(Z\theta, \sigma^2 I_n)$.

5.1 Теорема об ортогональных гауссовских проекциях

Если $L_1, \dots, L_r$ попарно ортогональны, то $P_{L_1} X, \dots, P_{L_r} X$ **независимы**, поскольку совместно нормальны и их попарные ковариации равны нулю.

Если $d_j = \dim L_j$, то

$$\frac{\|P_{L_j} X - \mathbb{E}(P_{L_j} X)\|^2}{\sigma^2} \sim \chi_{d_j}^2.$$

Применение к регрессии: L и $L^\perp$ ортогональны, поэтому $\hat{\theta}$ и $\hat{\sigma}^2$ независимы и

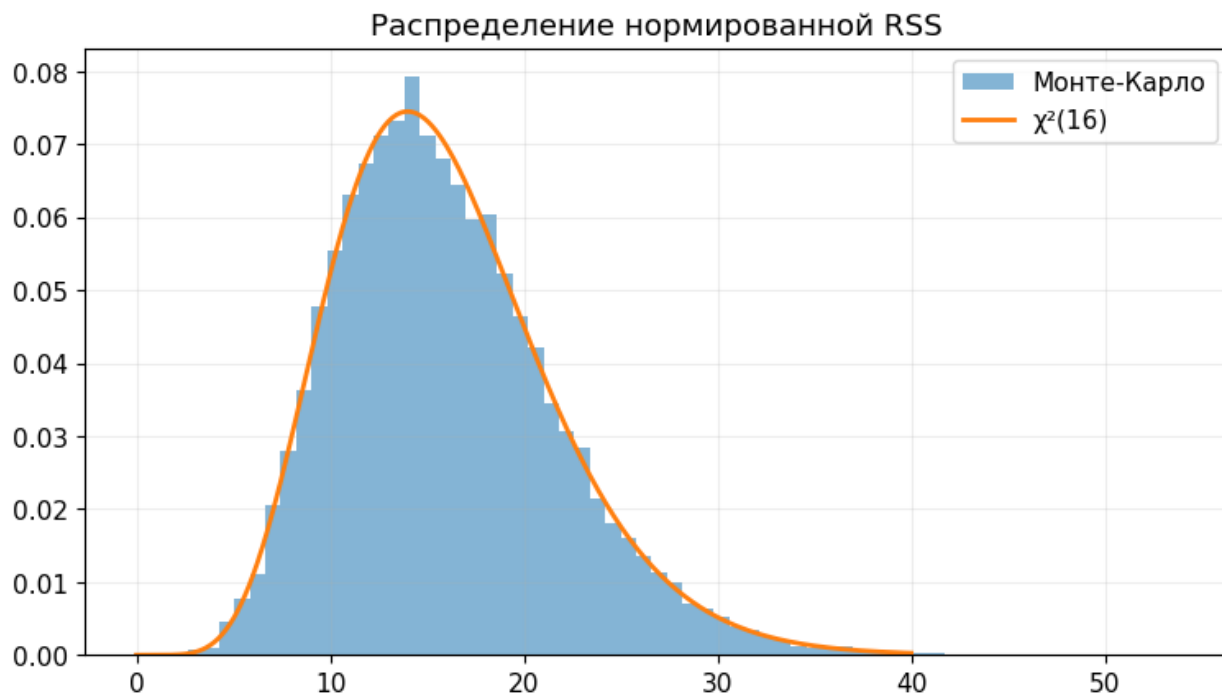
$$\hat{\theta} \sim N_k(\theta, \sigma^2(Z^T Z)^{-1}), \quad \frac{(n-k)\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-k}^2.$$

Почему нужна нормальность?

Некоррелированность проекций следует уже из сферической ковариации, но их **независимость в общем случае нет**. Именно нормальность обеспечивает точные распределения t и F .

5.2 Распределения, которые нужно знать

Если $U \sim \chi_r^2$, $V \sim \chi_s^2$ независимы, то $(U/r)/(V/s) \sim F_{r,s}$. Если $G \sim N(0, 1)$ и $V \sim \chi_\nu^2$ независимы, то $G/\sqrt{V/\nu} \sim t_\nu$.



Корреляция оценки наклона и RSS: 0.0020482103800153195

6. Точные доверительные интервалы и статистические выводы

Пусть $A = (Z^T Z)^{-1}$ и a_{jj} — его диагональный элемент. Тогда

$$T_j = \frac{\hat{\theta}_j - \theta_j}{\hat{\sigma} \sqrt{a_{jj}}} \sim t_{n-k}.$$

Для доверительной вероятности $\gamma = 1 - \alpha$:

$$\theta_j \in \left[\hat{\theta}_j - t_{1-\alpha/2, n-k} \hat{\sigma} \sqrt{a_{jj}}, \hat{\theta}_j + t_{1-\alpha/2, n-k} \hat{\sigma} \sqrt{a_{jj}} \right].$$

Для дисперсии используем квантили $q_p = \chi_{n-k; p}^2$:

$$\sigma^2 \in \left[\frac{(n-k)\hat{\sigma}^2}{q_{1-\alpha/2}}, \frac{(n-k)\hat{\sigma}^2}{q_{\alpha/2}} \right].$$

Односторонний интервал из пособия: если $z_{1-\gamma} = \chi_{n-k; 1-\gamma}^2$, то $(0, (n-k)\hat{\sigma}^2/z_{1-\gamma})$ имеет доверительную вероятность γ .

7. Взвешенный МНК, коррелированные ошибки и «отбеливание» (дополнение)

В задачах 2, 5 и 7 встречается $\text{Cov}(\varepsilon) = \sigma^2 \Sigma$ с известной положительно определённой Σ . Тогда **обобщённый МНК** минимизирует

$$Q(u) = (X - Zu)^T \Sigma^{-1} (X - Zu),$$

и даёт

$$\hat{\theta}_{\text{GLS}} = (Z^T \Sigma^{-1} Z)^{-1} Z^T \Sigma^{-1} X.$$

Если $C = \Sigma^{-1/2}$, то $Y = CX$, $W = CZ$ и $Y = W\theta + \eta$ с $\text{Cov } \eta = \sigma^2 I$. Таким образом, GLS — обычный МНК после «отбеливания».

$$\text{Cov}(\hat{\theta}_{\text{GLS}}) = \sigma^2 (Z^T \Sigma^{-1} Z)^{-1}, \quad \hat{\sigma}^2 = \frac{(X - Z\hat{\theta})^T \Sigma^{-1} (X - Z\hat{\theta})}{n - k}.$$

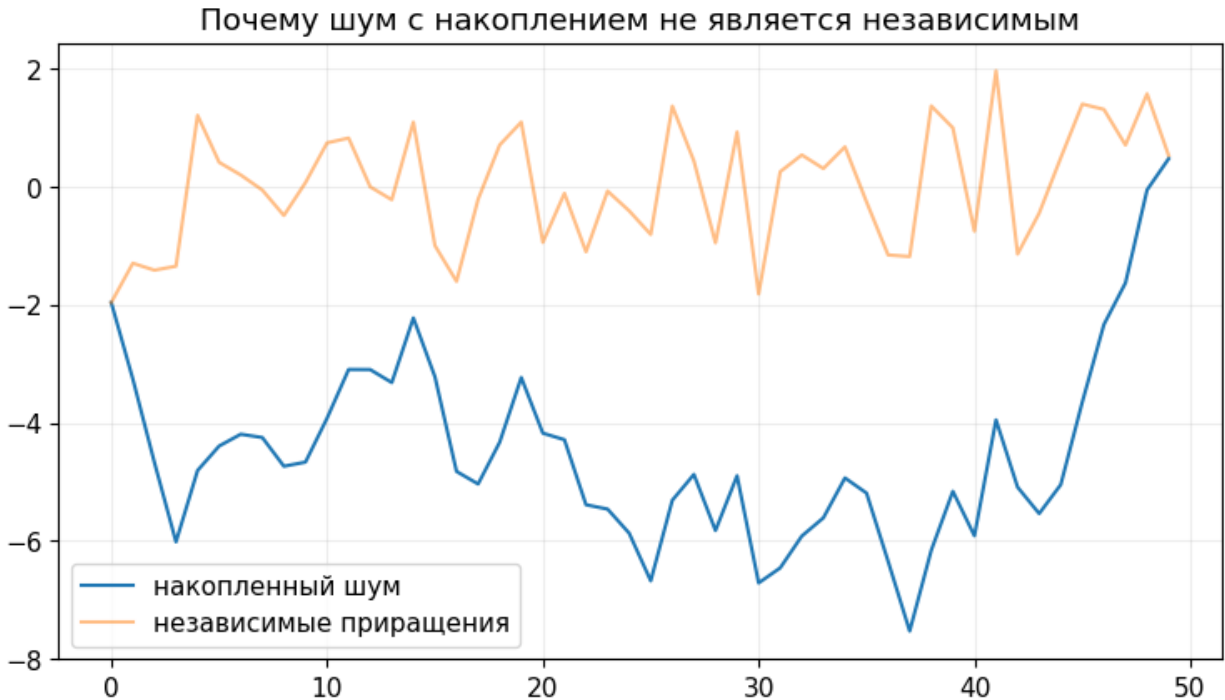
При нормальности нормированный взвешенный RSS имеет χ^2_{n-k} -распределение и независим от GLS-оценки. Если Σ неизвестна, эти точные выводы **нельзя** автоматически переносить на оценённую матрицу.

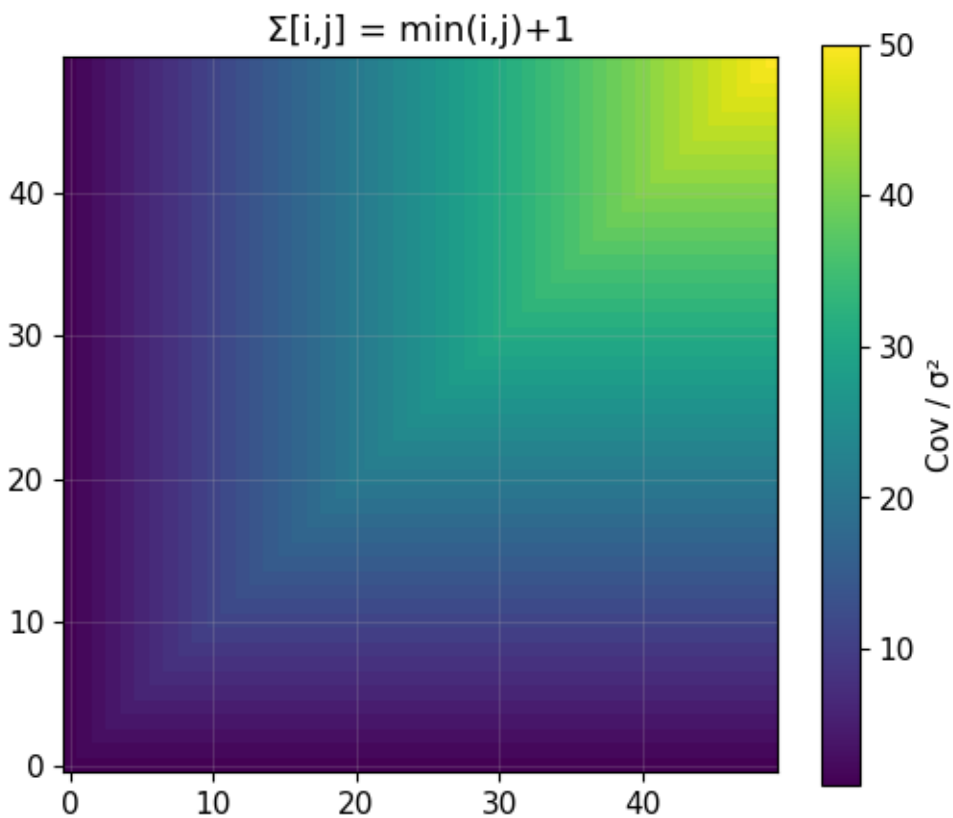
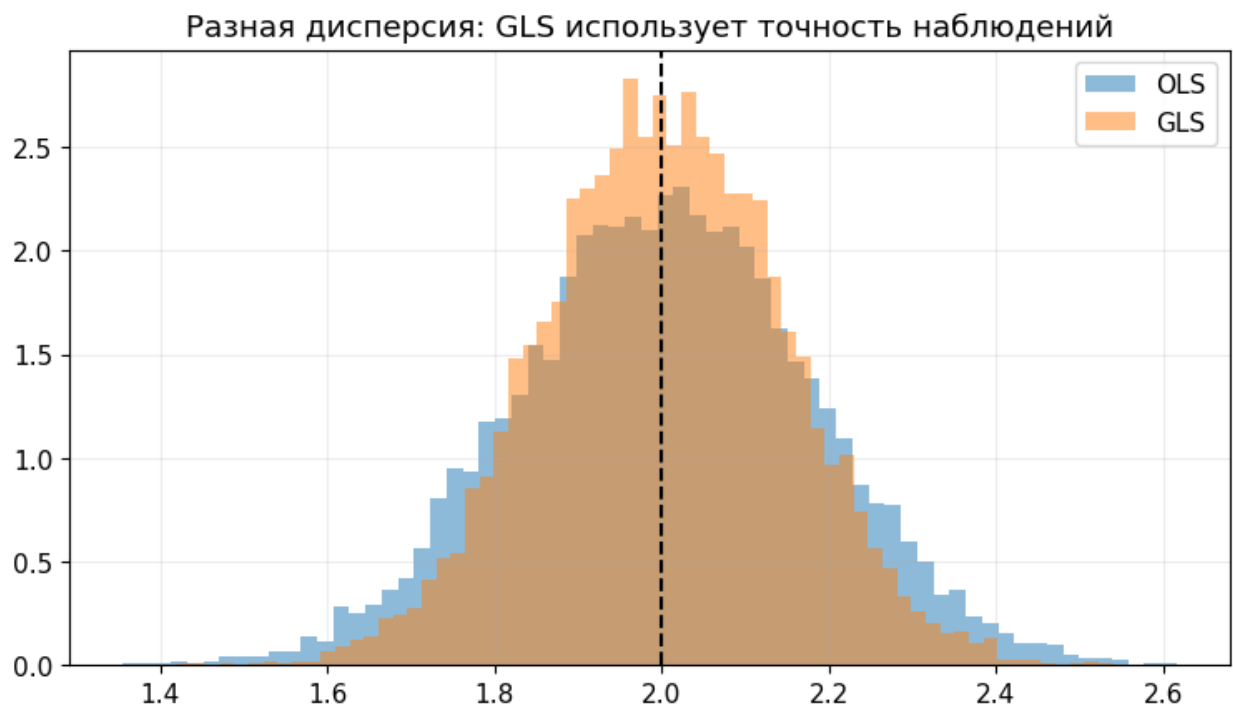
7.1 Коррелированный шум: накопленные ошибки

Если $X_i = \beta_1 + i\beta_2 + \varepsilon_0 + \dots + \varepsilon_i$ и независимые $\varepsilon_j \sim N(0, \sigma^2)$, то

$$\text{Cov}(X_i, X_j) = \sigma^2 (\min(i, j) + 1), \quad i, j = 0, \dots, n.$$

Матрица $\Sigma_{ij} = \min(i, j) + 1$ — **не диагональная**. Альтернатива: взять первые разности $X_i - X_{i-1} = \beta_2 + \varepsilon_i$ для $i \geq 1$; при этом $X_0 = \beta_1 + \varepsilon_0$. Это полезное преобразование, но для оптимальной совместной оценки β_1, β_2 важно учитывать весь преобразованный вектор.





Дисперсия оценки наклона OLS: 0.03122071327540109 GLS: 0.020795964561147996

8. Дополнительные инструменты для задач

8.1 Геометрия нормальной выборки

Для $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} N(a, \sigma^2)$ возьмите $L = \text{span}\{\mathbf{1}\}$. Тогда $P_L X = \bar{X}\mathbf{1}$, а $\|(I - P_L)X\|^2 = \sum_i (X_i - \bar{X})^2$. Из теоремы о проекциях следует независимость $\bar{X}$ и $S^2 = n^{-1} \sum (X_i - \bar{X})^2$, а также

$$\frac{nS^2}{\sigma^2} \sim \chi_{n-1}^2, \quad \frac{\sqrt{n}(\bar{X} - a)}{\sqrt{nS^2/(n-1)}} \sim t_{n-1}.$$