Домашнее задание № 1 по эконометрике – 2.

Рассмотрите следующие примеры и выполните задания.

Пример 1. Рассматривается модель линейной регрессии $Y = X\beta + \varepsilon$, где

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix};\ X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix};\ \beta = \begin{bmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix};\ \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Здесь X – некоторая фиксированная $n \times (k+1)$ числовая матрица такая, что $\operatorname{rg} X = k+1$. Вектор ε имеет n-мерное нормальное распределение с нулевым математическим ожиданием и ковариационной матрицей $\sigma^2 \cdot I_{n\times n}$, т.е. $\varepsilon \sim N(0, \sigma^2 \cdot I_{n\times n})$.

Матрицу X генерируем так, что её компоненты $x_{11}, x_{12}, x_{21}, x_{22}, ..., x_{n1}, x_{n2}$ являются реализациями случайной выборки из нормального стандартного распределения.

Положим $n = 20$, $k = 2$, $\alpha = 0$, $\beta_1 = 1$, $\beta_2 = 2$, $\sigma = 2$. При помощи метода Монте-Карло построим ядерную оценку для плотности $\hat{\beta}_1$. Для этого выполним $S = 100000$ симуляций. Опишем алгоритм, согласно которому выполняется симуляции с номером $s = 1, ..., S$.

1. Генерируем $\varepsilon^{(s)} \sim N(0, \sigma^2 \cdot I_{n\times n})$.
2. Рассчитываем $Y^{(s)} = X\beta + \varepsilon^{(s)}$.
3. Находим $\hat{\beta}^{(s)} = (X^T X)^{-1} X^T Y^{(s)}$.
4. Из вектора $\hat{\beta}^{(s)}$ выделяем элемент в позиции 2×1, который соответствует оценке $\hat{\beta}_1^{(s)}$.

После выполнения S симуляций мы получаем набор значений $\{\hat{\beta}_1^{(s)}\}_{s=1}^{S}$, по которому строим искомую ядерную оценку плотности $\hat{\beta}_1$. На рисунке 1 приведены результаты описанных расчетов.

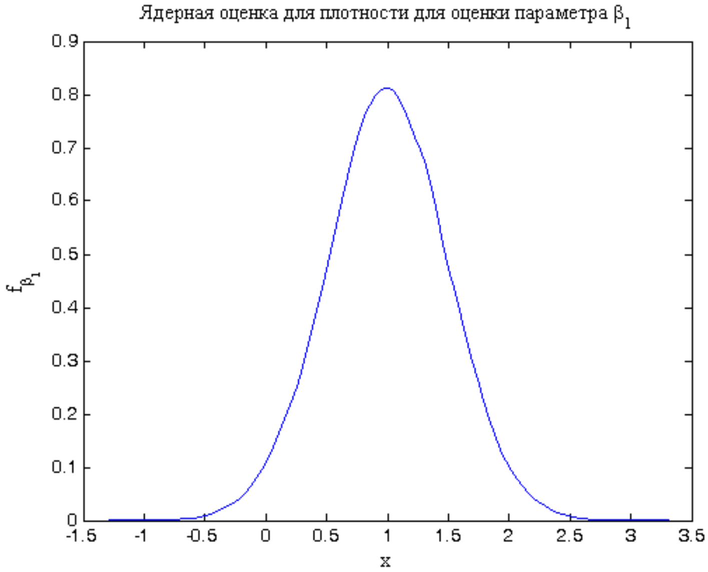

Рис. 1

Код программы в системе MATLAB

```
clc; clear;
n = 20;
k = 2;
X = [ones(n,1), randn(n,k)];
b = [0:k]';
sigma = 2;
S = 100000; % число симуляций
B1 = zeros(S,1);
for s = 1:S
    disp(['starting simulation number ... ', num2str(s)]);
    e = sigma * randn(n,1);
    Y = X * b + e;
    b_hat = (X' * X) \ (X' * Y);
    B1(s,1) = b_hat(2,1)';
end
[f,x] = ksdensity(B1);
plot(x,f);
title('Ядерная оценка для плотности для оценки параметра \beta_1',...
    'FontSize', 11, 'FontName', 'Times New Roman');
xlabel('x', 'FontSize', 11, 'FontName', 'Times New Roman');
ylabel('f_{\beta_1}', 'FontSize', 11, 'FontName', 'Times New Roman');
```

Задание 1.1. Что можно сказать о статистических свойствах оценки $\hat{\beta}_1$, глядя на рисунок 1.

Задание 1.2. Выполните аналогичные расчеты и постройте графики оценок плотностей для $\hat{\beta}_2$, $\hat{\sigma}$ и $\hat{\sigma}^2$. Прокомментируйте полученные результаты.

Задание 1.3. Выполняя задание 1.2, постройте таблицу,

	α	β_1	β_2
MEAN	$MEAN(\alpha)$	$MEAN(\beta_1)$	$MEAN(\beta_2)$
RMSE	$RMSE(\alpha)$	$RMSE(\beta_1)$	$RMSE(\beta_2)$

где $MEAN(\alpha) := \frac{1}{S}\sum_{s=1}^{S}\hat{\alpha}^{(s)}$, $RMSE(\alpha) := \sqrt{\frac{1}{S-1}\sum_{s=1}^{S}\left(\hat{\alpha}^{(s)} - MEAN(\alpha)\right)^2}$. Остальные элементы таблицы определяются аналогичным образом. Так, например, в наших расчетах получилась следующая таблица.

	α	β_1	β_2
MEAN	0.0009	1.0003	1.9988
RMSE	0.4830	0.3649	0.4025

Пример 2. Как и в примере 1 рассматривается модель линейной регрессии $Y = X\beta + \varepsilon$, где матрица X и вектор ε определяются так же, как и выше.

Положим $k = 2$, $\alpha = 0$, $\beta_1 = 1$, $\beta_2 = 2$, $\sigma = 2$. Последовательно для $n = 10$, 20, 40, 80 и 160 при помощи метода Монте-Карло построим ядерные оценки для плотностей $\hat{\beta}_1(n)$. Для этого генерируем при каждом $n = 10$, 20, 40, 80 и 160 матрицу $X(n)$ так, что её компоненты $x_{11}(n), x_{12}(n), x_{21}(n), x_{22}(n), ..., x_{n1}(n), x_{n2}(n)$ являются реализациями случайной выборки из нормального стандартного распределения.

Выполним $S = 100000$ симуляций. Алгоритм, согласно которому выполняется симуляция соответствующая n и $s = 1,...,S$ приведен ниже.

1. Генерируем $\varepsilon^{(s)}(n) \sim N(0, \sigma^2 \cdot I_{n\times n})$.
2. Рассчитываем $Y^{(s)}(n) = X(n)\beta + \varepsilon^{(s)}(n)$.
3. Находим $\hat{\beta}^{(s)}(n) = (X(n)^T X(n))^{-1} X(n)^T Y^{(s)}(n)$.
4. Из вектора $\hat{\beta}^{(s)}(n)$ выделяем элемент в позиции 2×1, который соответствует оценке $\hat{\beta}_1^{(s)}(n)$.

После выполнения S симуляций для каждого n, мы получаем пять наборов значений $\{\hat{\beta}_1^{(s)}(n)\}_{s=1}^{S}$, по которым строим пять графиков оценок плотностей $\hat{\beta}_1(n)$ в одних координатных осях. На рисунке 2 приведены результаты описанных расчетов.

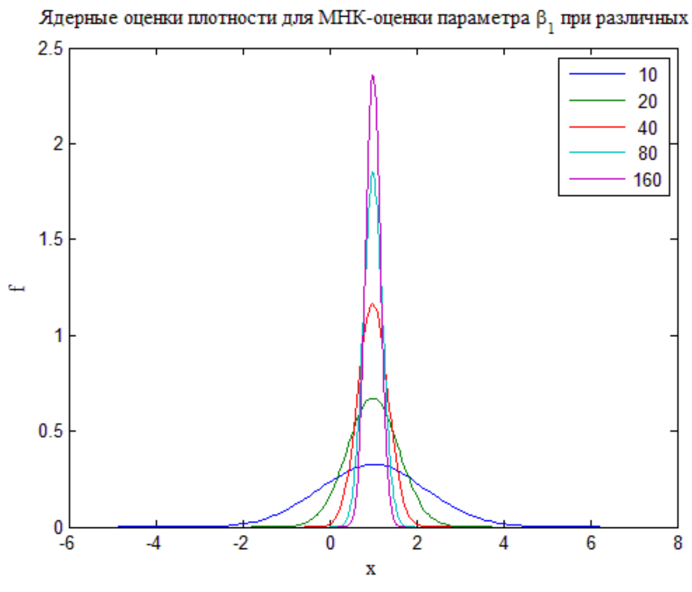

Рис. 2

Код программы в системе MATLAB

```
clc; clear;
Q = 5; % число графиков
x = zeros(100,Q);
f = zeros(100,Q);
for q = [1:Q]
    n = 10 * 2 ^ (q-1);
    k = 2;
    X = [ones(n,1), randn(n,k)];
    b = [0:k]';
    sigma = 2;
    S = 100000; % число симуляций
    B1 = zeros(S,1);
    for s = 1:S
        e = sigma * randn(n,1);
        Y = X * b + e;
        b_hat = (X' * X) \ (X' * Y);
        B1(s,1) = b_hat(2,1)';
    end
    [f(:,q),x(:,q)] = ksdensity(B1);
end
% строим графики...
plot(x,f);
title('Ядерные оценки плотности для МНК-оценки параметра \beta_1 при различных n', ...
    'FontSize', 11, 'FontName', 'Times New Roman');
xlabel('x', 'FontSize', 11, 'FontName', 'Times New Roman');
ylabel('f', 'FontSize', 11, 'FontName', 'Times New Roman');
q = [1:Q]';
n = 10 * 2 .^ (q-1);
n = num2str(n);
legend(n);
```

Задание 2.1. Что можно сказать о статистических свойствах оценки $\hat{\beta}_1(n)$, глядя на рисунок 2.

Задание 2.2. Выполните аналогичные расчеты и постройте аналогичные графики оценок плотностей для $\hat{\beta}_2(n)$, $\hat{\sigma}(n)$ и $\hat{\sigma}^2(n)$. Прокомментируйте полученные результаты.

Задание 2.3. Постройте графики оценок для плотностей

$$\frac{\hat{\beta}_2(n)-\beta_2}{\sqrt{\hat{\mathrm{D}}(\hat{\beta}_2(n))}} \qquad (1)$$

при $n=5$, 10, 15, 20, 25, 30, 35 и 40 в одних координатных осях. Известен ли вам какой-либо факт о предельном поведении величины (1) при $n\to\infty$? Можно ли этот факт наблюдать из полученного вами рисунка?

Пример 3. Приведем пример программы, которая генерирует два коррелированных ряда нормальных случайных чисел. Более точно, данная программа генерирует реализацию двумерной случайной выборки $\{(X_i,Y_i)\}_{i=1}^n$ такой, что каждый случайный вектор (X_i,Y_i) имеет плотность

$$f(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\cdot\exp\left\{-\frac{1}{2(1-\rho^2)}\cdot\left[\frac{x^2}{\sigma_X^2}-\frac{2\rho xy}{\sigma_X\sigma_Y}+\frac{y^2}{\sigma_Y^2}\right]\right\},\ i=1,\ldots,n. \qquad (2)$$

Ниже приведен **код программы в системе MATLAB**.

```
function [X] = gen_2corr_series(n, sigmaX, sigmaY, rho)
V = [sigmaX^2, rho*sigmaX*sigmaY; rho*sigmaY*sigmaX, sigmaY^2];
Z = randn(2,n);
X = (V^0.5) * Z;
X = X';
```

Дадим здесь некоторые пояснения. Прежде всего, отметим, что плотность f в формуле (2) соответствует двумерному нормальному случайному вектору с математическим ожиданием $\begin{bmatrix}0 & 0\end{bmatrix}^T$ и ковариационной матрицей

$$V=\begin{bmatrix}\sigma_X^2 & \rho\sigma_X\sigma_Y\\ \rho\sigma_Y\sigma_X & \sigma_Y^2\end{bmatrix}.$$

Известно, что если случайный вектор $Z=\begin{bmatrix}Z_1 & Z_2\end{bmatrix}^T$ имеет нормальное распределение с математическим ожиданием $\begin{bmatrix}0 & 0\end{bmatrix}^T$ и ковариационной матрицей

$$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix},$$

то случайный вектор $X=V^{\frac{1}{2}}\cdot Z$ имеет как раз требуемые параметры: математическое ожидание $\begin{bmatrix}0 & 0\end{bmatrix}^T$ и ковариационную матрицу V. На основе данных соображений написана приведенная выше программа.

Проиллюстрируем работу данной программы на четырёх примерах. Пусть во всех расчетах $n=10000$, $\sigma_X=1$, $\sigma_Y=1$. При этом значениям параметра $\rho=0.7$, 0.9, -0.7, -0.9 соответствуют рисунки с номерами 3, 4, 5 и 6 соответственно.

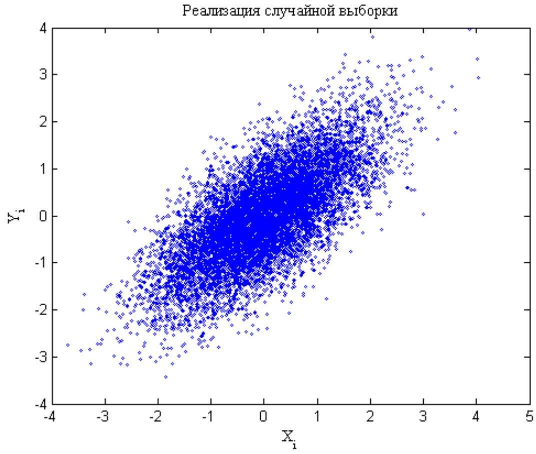

Рис. 3

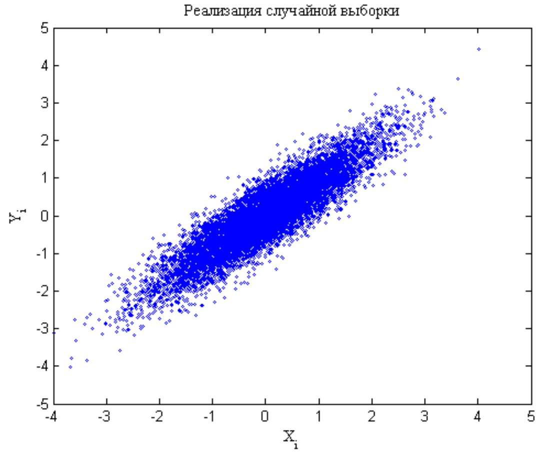

Рис. 4

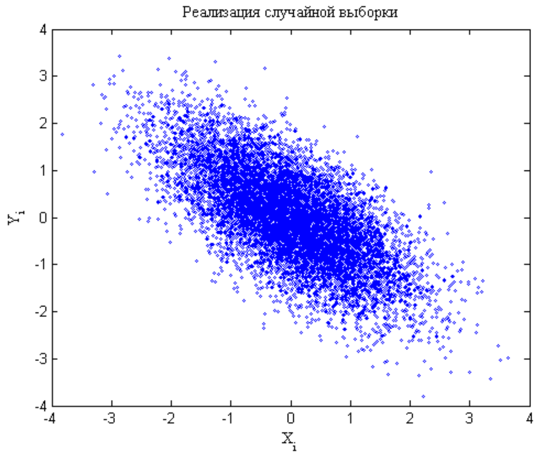

Рис. 5

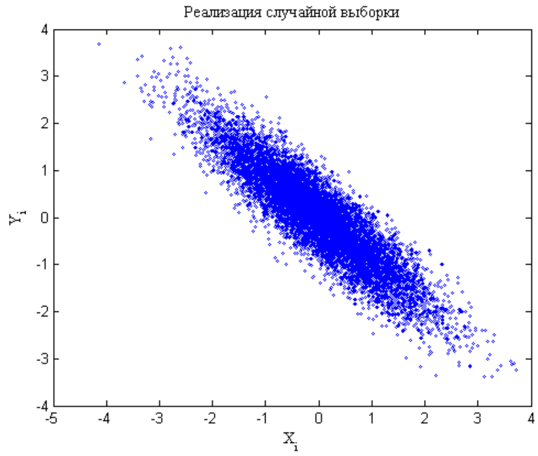

Рис. 6

Задание 3.1. Допишите несколько строк в программе `gen_2corr_series`, которые позволят воспроизвести приведенные выше рисунки 3 – 6.

Задание 3.2*. На рисунках из задания 3.1 постройте 99%-ые эллипсы рассеивания (см. [1] стр. 80-83).

Задание 3.3. Напишите программу, которая генерирует реализацию d-мерной случайной выборки $\left\{\left(X_i^{(1)},\ldots,X_i^{(d)}\right)\right\}_{i=1}^{n}$, каждый случайный вектор которой имеет d-мерное нормальное распределение с математическим ожиданием μ и ковариационной матрицей Σ, $\det\Sigma\neq 0$.

Пример 4. Данный пример посвящен изучению последствий, к которым приводят ошибки при спецификации модели регрессии. Здесь будут рассмотрены три случая:

1. в оцениваемую модель *не включена существенная* переменная;
2. оцениваемая модель *специфицирована правильно*;
3. в оцениваемую модель *включили несущественную переменную*.

Задание 4.1. Разберите следующий **код программы, написанный в системе MATLAB**, и опишите алгоритм работы программы так, как это сделано в примерах 1 и 2.

```
clc; clear;
[X] = gen_2corr_series(20, 1, 1, 0.9);
X = [ones(20,1), X(:,1), randn(20,1), X(:,2)];
[n,k] = size(X);
X_U = X(:,1:k-2);
X_T = X(:,1:k-1);
X_O = X(:,1:k);
sigma = 2;
S = 100000;
BETA1_U = zeros(S,1);
BETA1_T = zeros(S,1);
BETA1_O = zeros(S,1);
for s = 1:S
    disp(['starting simulation number  ', num2str(s), ' ... ']);
    b_T = [0 10 20]';
    e = sigma * randn(n,1);
    Y = X_T * b_T + e;

    b_hat_U = (X_U' * X_U) \ (X_U' * Y);
    BETA1_U(s,1) = b_hat_U(2);

    b_hat_T = (X_T' * X_T) \ (X_T' * Y);
    BETA1_T(s,1) = b_hat_T(2);

    b_hat_O = (X_O' * X_O) \ (X_O' * Y);
    BETA1_O(s,1) = b_hat_O(2);
end
[f_U,x_U] = ksdensity(BETA1_U);
[f_T,x_T] = ksdensity(BETA1_T);
[f_O,x_O] = ksdensity(BETA1_O);
plot(x_U,f_U,x_T,f_T,x_O,f_O);
title('Оценки плотностей для \beta_1 при различных спецификациях модели', 'FontSize', 11, ...
    'FontName', 'Times New Roman');
xlabel('x', 'FontSize', 11, 'FontName', 'Times New Roman');
ylabel('f', 'FontSize', 11, 'FontName', 'Times New Roman');
legend('under', 'true', 'over');
```

На рисунке 7 приведены результаты расчетов, которые выполняет данная программа.

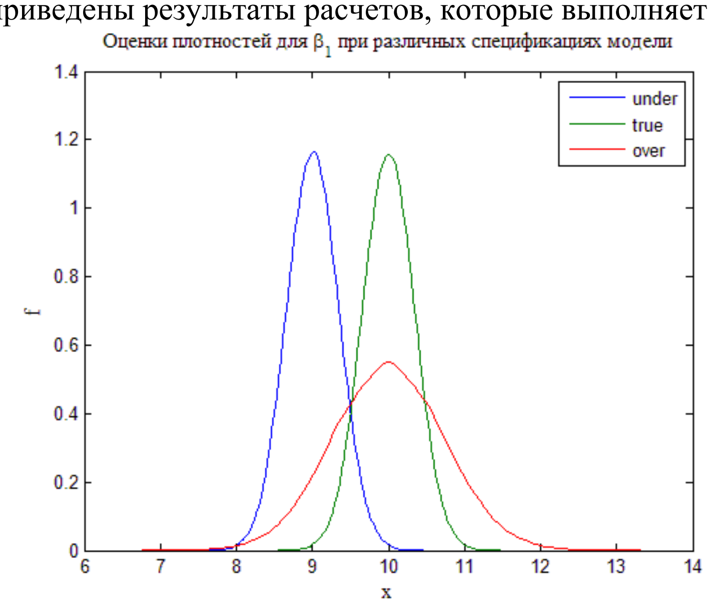

Рис. 7

На рисунке выше кривая, помеченная ‘under’, соответствует ситуации, в которой в оцениваемую модель *не включена существенная* переменная. Кривая, помеченная ‘true’, соответствует случаю, когда оцениваемая модель *специфицирована правильно*, а график плотности, помеченный ‘over’ – ситуации, когда в оцениваемую модель *включили несущественную переменную*.

Если во второй строке приведенной выше программы исправить параметр $\rho = 0.9$ на $\rho = 0$, т.е. во второй строке вместо `[X] = gen_2corr_series(20, 1, 1, 0.9);` написать `[X] = gen_2corr_series(20, 1, 1, 0);`, то получим следующий рисунок.

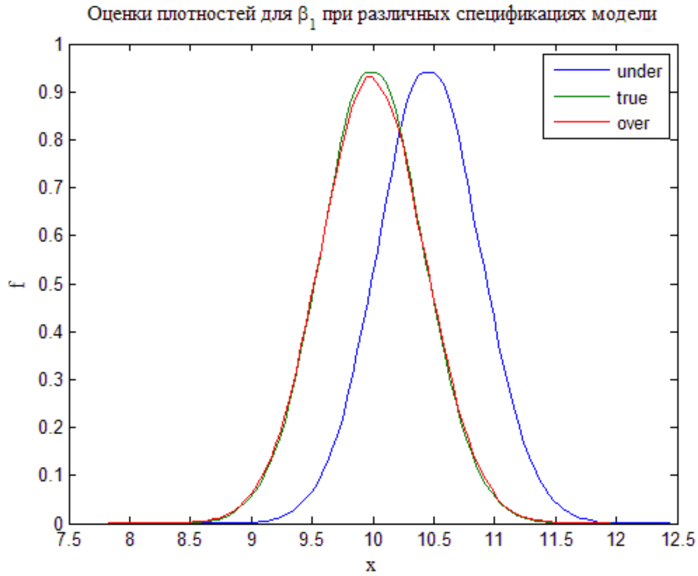

Рис. 8

Задание 4.2. Попытайтесь объяснить причину, по которой возникло такое существенное различие в рисунках 7 и 8.

Задание 5*. Рассмотрите модель линейной регрессии $Y_i = \alpha + \beta \cdot x_i + \varepsilon_i$, $i = 1,\ldots,n$, для случая, когда $\{x_i\}$ *не являются детерминированными*. Предположим $\mathrm{corr}(x_i, \varepsilon_i) = \rho \neq 0$ при $i = 1,\ldots,n$. Аналогично тому, как это сделано в примере 2, исследуйте на несмещенность и состоятельность оценку $\hat{\beta}(n)$.

Список литературы

[1] Шведов А.С., Теория вероятностей и математическая статистка – 2 (промежуточный уровень). Гос. У-нт – Высшая Школа Экономики. – М.: ТЕИС, 2007.