

Практика 4. Сравнение МО двух нормальных случайных величин в случае известных дисперсий.

Достоверность показаний датчиков с известными случайными погрешностями

Часть 1. Подготовка генераторов данных

1. Рассматривается измерение одного физического параметра двумя отдельными измерительными приборами (датчиками). У каждого прибора своя систематическая и случайная погрешность.
2. Реализовать генератор показаний датчика в виде **отдельной** функции.

```
function x = generate_sensor_data(n, true_value,  
                                systematic1, dispersion1, systematic2, dispersion2)
```

Исходными данными является объем измерительной выборки n , систематические погрешности, дисперсии случайных составляющих погрешности, а также истинное значение измеряемого параметра, которое принять постоянным, но не равным нулю. Результатом функции является сгенерированная выборка.

Код данного генератора должен в коде фигурировать ровно один раз в этой написанной функции. Глобальные переменные в данной функции запрещены. В дальнейшем генерация измерений для обоих датчиков обязательно только вызывает эту функцию. Код с нарушением этого требования не принимается.

3. Записать вероятностную модель генератора «до измерений» с учетом истинного значения параметра, аналитически вывести его МО, дисперсию.
4. Методом Монте-Карло ($N = 1000$ или больше) построить распределение значений из генератора. На него же наложить теоретический график плотности распределения, взяв теоретические МО, дисперсию из предыдущего пункта.

Часть 2. Функции проверки гипотез

1. Реализовать функцию расчета статистики e и z по двум выборкам x_1, x_2 :

```
function e = e_statistic(x1, x2)  
function z = z_statistic(x1, x2, D1, D2)
```

Результатом обеих функций являются скалярные величины e, z , учитывающие усреднение по выборке. Формулы расчета обеих статистик должны в коде фигурировать ровно один раз в этих функциях. Глобальные переменные в данных функциях запрещены. В дальнейшем при расчете статистик e, z обязательно вызывать эти функции. Код с нарушением этого требования не принимается.

2. Осознать, почему на входе функции расчета e -статистики не нужны дисперсии случайных составляющих погрешности, а z -статистики нужны. Осознать, почему не передается объем выборки n .
3. Реализовать функцию проверки гипотезы о равенстве МО двух СВ на основе статистик e и z .

```
function [h, border_low, border_high] = test_equal_means_e(e, D1, D2, n, alpha)  
function [h, border_low, border_high] = test_equal_means_z(z, alpha)
```

h – результат проверки гипотезы

$border_low, border_high$ – границы области принятия основной гипотезы по e или z статистикам.

Далее везде в коде для проверки гипотез использовать только эти функции. Код с нарушением этого требования не принимается.

4. Осознать, почему в функции расчета e -статистики не нужны дисперсии случайных составляющих погрешности, а для z -статистики нужны.

Часть 3. Распределения статистик в условиях H_0 и H_1

1. Задаться объемом измерительной выборки $5 < n < 20$.
2. Для статистики e на **одном графике** построить:
 - a. теоретическую плотность распределения статистики Критерия в условиях основной гипотезы (показания датчиков «в среднем» равны, отличия обусловлены случайной составляющей погрешности и носят чисто случайный характер);
 - b. теоретическая плотность распределения статистики Критерия в условиях альтернативной гипотезы;
 - c. в тех же условиях методом статистического моделирования Монте-Карло ($N \geq 1000$) построить полигон относительных частот статистики Критерия (`hist_density`);
 - d. методом статистического моделирования ($N \geq 1000$) построить полигон относительных частот статистики Критерия при условии альтернативной гипотезы;
 - e. осознать, чем отличается n от N .
3. Построить то же самое для статистики z .
4. На построенные в этой части задания графики (их должно быть два) нанести нижнее и верхнее пороговые значения для принятия или отклонения основной гипотезы. Пороги брать из функций `test_equal_means_e`, `test_equal_means_z`.

Часть 4. Исследование статистических свойств критерия

1. Задаться объемом измерительной выборки $5 < n < 20$.
2. Методом Монте-Карло оценить ошибку **первого** рода при принятии гипотезы о равенстве MO нормальных СВ на основе e -статистики.
Для этого сгенерировать $N \geq 1000$ реализаций e -статистики в условиях основной гипотезы и посчитать долю случаев, в которых функция `test_equal_means_e` принимает альтернативную гипотезу.
3. Аналогичным образом оценить ошибку **второго** рода при принятии гипотезы о равенстве MO нормальных СВ на основе e -статистики
4. Посчитать ошибки первого и второго рода для z -статистики.
5. Увеличить объем измерительной выборки n в два раза. Проверить ошибки первого и второго рода. Обосновать их изменение или, наоборот, неизменность.

Часть 5. Спроектировать систему проверки достоверности показаний датчиков

Имеется интерфейс связи, совместимый с Матлабом, по которому приходят измерения от продублированных датчиков. Случайные составляющие погрешностей датчиков известны.

Как можно определить само понятие достоверность показаний в данном случае?

Как можно написать на Матлабе систему проверки достоверности показаний датчиков, используя наработки кода из данной работы?

Наводящее соображение. Пусть есть два показания двух датчиков на один момент времени. Проверьте гипотезу об отсутствии систематических погрешностей у этих датчиков.

Вопросы к защите

- Вероятностная модель измерения двумя датчиками одного физического параметра. Как соотносить термины случайная и систематическая погрешности с параметрами распределения случайной величины?
- Каковы распределения статистики e в условиях основной и альтернативной гипотезы?
- Каковы распределения статистики z в условиях основной и альтернативной гипотезы?

- Как надо модифицировать расчет статистики e , чтобы учесть изменение физического параметра? Почему нельзя использовать расчет по учебникам?
- Как связана дисперсия средневывборочного с дисперсией исходной СВ?
- Что такое случайная и систематическая погрешности?
- Что такое статистика критерия? Какие статистики использовались в данном ДЗ?
- Что такое критическая область и область принятия гипотезы?
- Почему в функции расчета e -статистики не нужны дисперсии случайных составляющих погрешности, а для z -статистики нужны?
- Что такое ошибка первого и второго рода? Как посчитать ошибку первого и второго рода для e и z -статистик?
- Что такое простая и сложная гипотеза? Какая из двух гипотез рассматриваемых критерием является простой, а какая сложной?
- Как связаны плотность распределения вероятности и функция распределения вероятности. Что такое инверсная функция распределения вероятностей? Как с ее помощью посчитать пороги для e ?
- Как влияет объем выборки на ошибку первого рода?
- Как влияет объем выборки на ошибку второго рода?