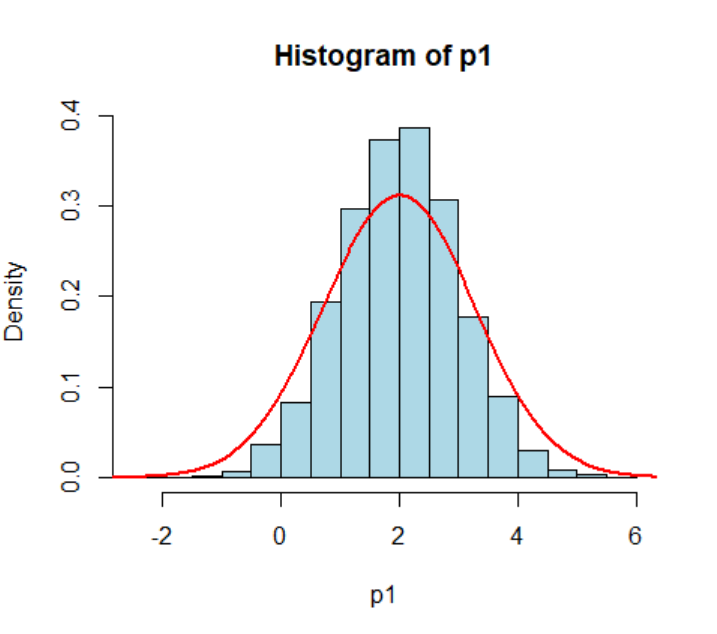
Дз №5.

Часть 1.

Построение ОМП для нормального распределения.

Изображение выглядит как текст

Автоматически созданное описаниеДействительно, похоже на нормальное распределение с параметрами mean = 2, sd = 1.

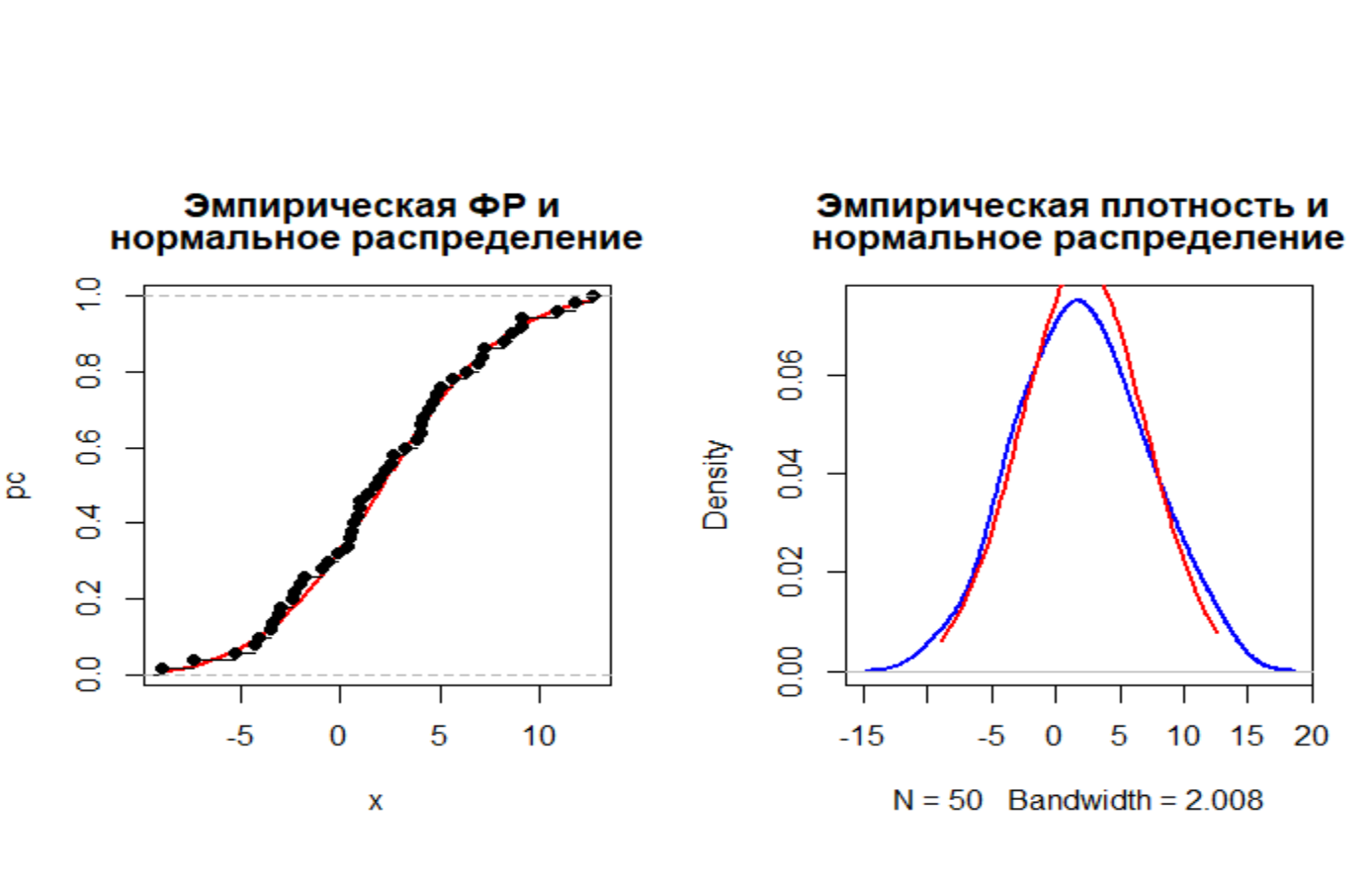
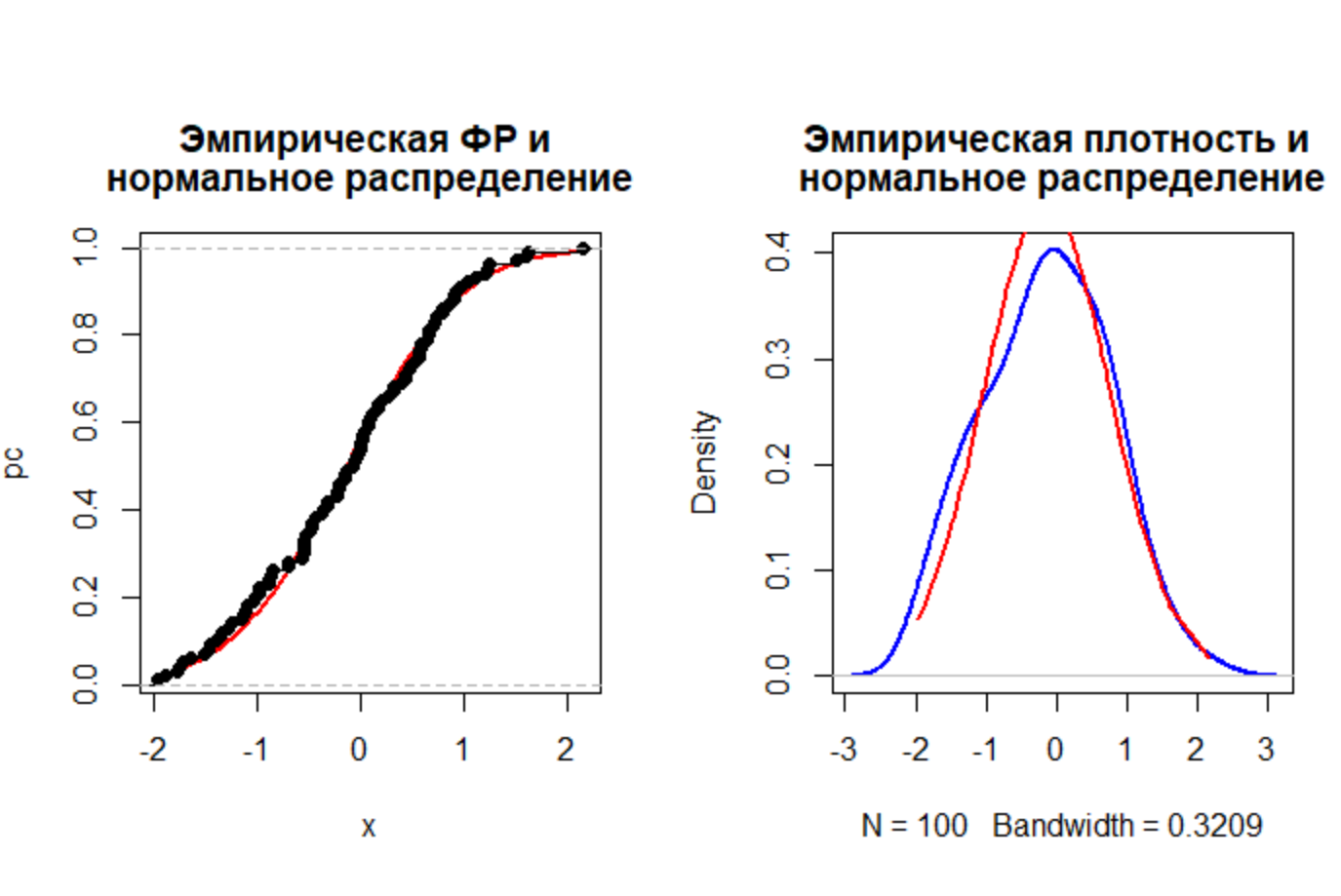
Посмотрим, что скажет наш итерационный метод(итерация начинается с начального приближения mean = 0, sd = 5)

Изображение выглядит как текст

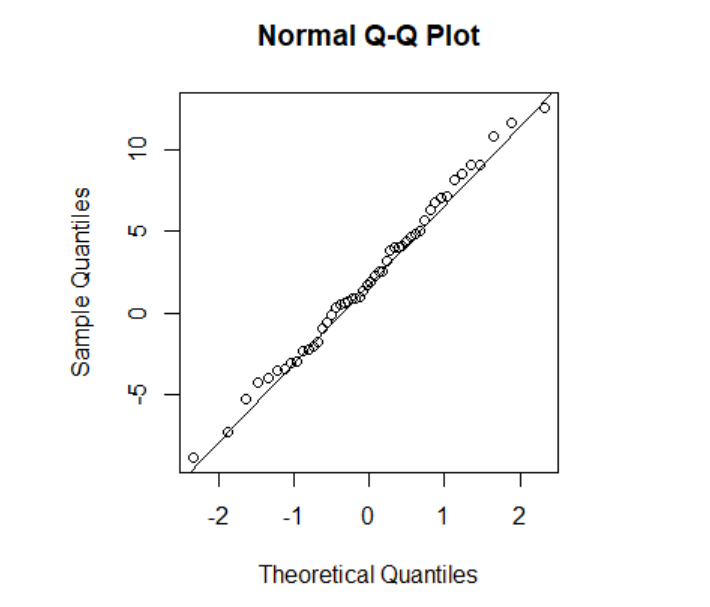
Автоматически созданное описаниеКак мы видим построенная оценка очень близка к исходным параметрам(погрешность меньше 1%). Можно сделать вывод, что с помощью функции mle можно оценивать параметры с помощью ОМП. Формируя для каждой задачи свою функцию nll мы будем получать хорошие оценки параметров.

Изображение выглядит как текст

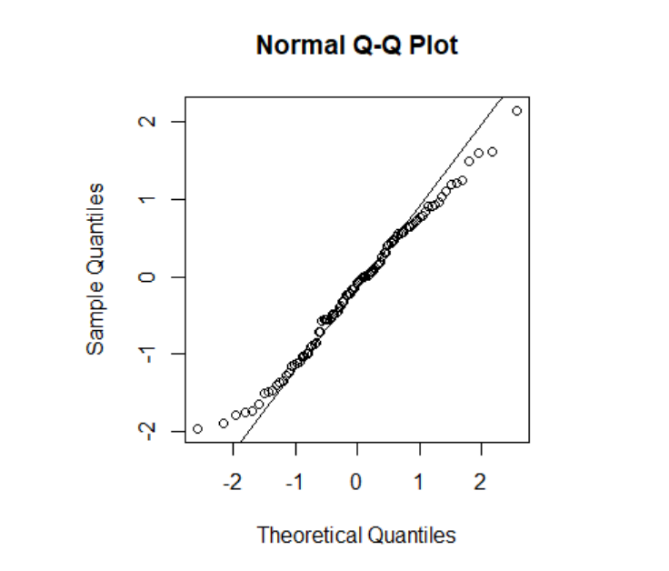
Автоматически созданное описаниеЧасть 2. Анализ выборки на нормальность распределения.



Что для случая N = 50, что для N = 100 графики плотности и функций распределения(красные графики) нормального распределения с оцененными параметрами действительно похожи на эмпирическую плотность(синяя density) и эмпирическую функцию распределения(черные точки) сгенерированной выборки соответственно. Эти графики позволяют нам сказать, что сгенерированная выборка – похожа на выборку из нормального распределения.

По квантилям:  
Изображение выглядит как текст

Автоматически созданное описание

 N = 100 N = 50

На мой взгляд QQ график дает хуже представление о распределении, чем непосредственно сами графики. Причем примечательно, что для N = 100 точки в среднем лежат дальше от прямой, чем для N = 50, скорее всего это связано с тем, что выборка для N = 100 не стандартизована( соответственно в концах расходится).

Метод огибающих:

Изображение выглядит как текст, монитор, экран, снимок экрана

Автоматически созданное описаниеДля N = 100 Для N = 50

А вот с помощью аналога метода огибающих можно более уверенно сказать, что данные распределения – нормальные распределения. Ни одна точка не заходит за доверительный интервал.

Тесты. Критерии их применимости.

Колмогоров – Смирнов: n > 25

Шапиро тест: 3 < n < 5000

Андерсон-Дарлинг: n > 7

Крамера фон Мизеса: n > 7

Изображение выглядит как текст

Автоматически созданное описаниеКолмогорова-Смирнова в модификации Лиллиефорса: n > 4

Шапиро-Франсия: 5 < n < 5000

Изображение выглядит как текст

Автоматически созданное описаниеИзображение выглядит как текст

Автоматически созданное описание

Для выборки N = 100. Все тесты показывают вероятность нормального распределения > 50%. Самой большую вероятность выдал тест Колмогорова – Смирнова.

Изображение выглядит как текст

Автоматически созданное описаниеИзображение выглядит как текст

Автоматически созданное описание Однако для N = 50 все тесты показывают почти 100% вероятность. Скорее всего это связано с тем, что выборка из 50 элементов бралась из стандартного нормального распределения, а выборка из 50 – нет.

Вывод: если есть подозрения что нормальное распределение стандартизировано можно использовать любой из этих 6 тестов(с поправкой на критерии). В целом для небольшой выборки лучше использовать тест Колмогорова – Смирнова.