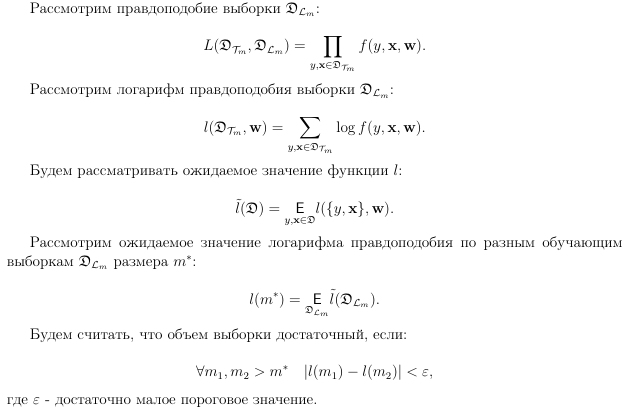
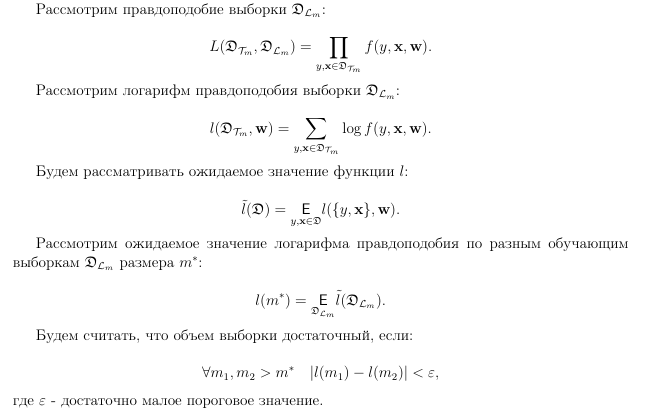
Определение достаточного объема выборки см. в Постановке задачи:



Я использовал выборку Diabetes для задачи линейной регрессии из UCI репозитория, получилась следующий график функции l(m):

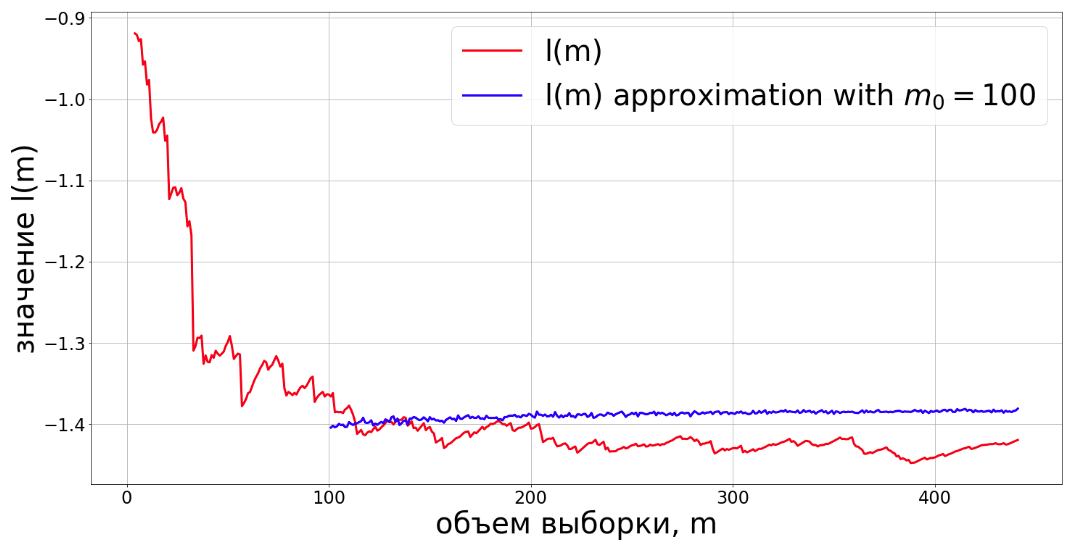


Я попытался построить аппроксимацию функции l(m), используя только первые m0 элементов выборки следующим образом:

1) Используем аппроксимацию распределения вектора параметров для m > m0:

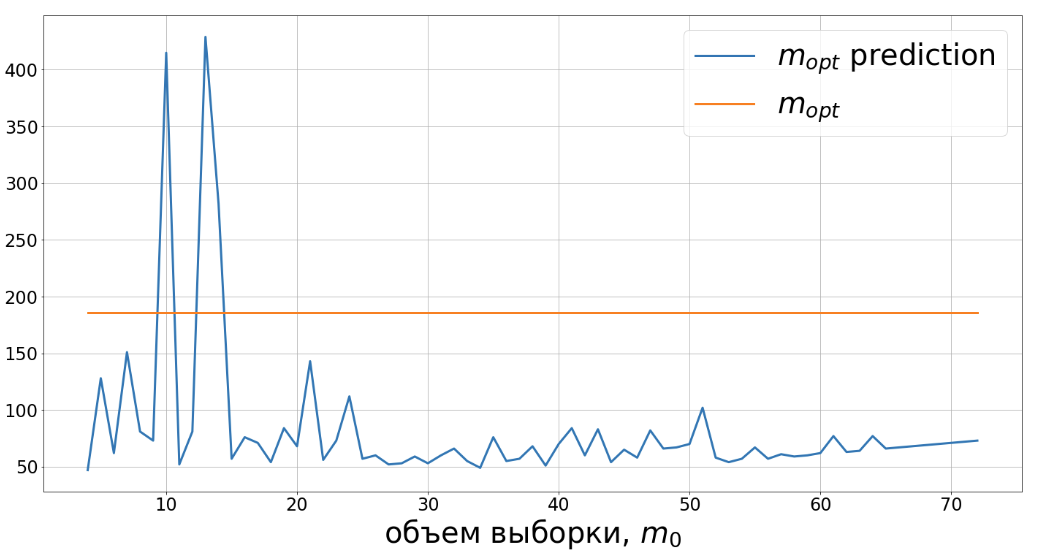
2) Много раз сэмплируем w из заданного распределения и считаем lw(m) по первым m0 элементов, затем усредняем по всем w: l(m) = MEAN(lw(m)).

Например, для m0=100 получается следующая аппроксимация:



Тут видна проблема: функция l(m) очень зависит от выборки, на основе которой она посчитана, и, опираясь только на первые m0 элементов, не получается построить нормальную аппроксимацию l(m).

Если считать m\* по аппроксимации l(m), получается следующий график:



Судя по всему, небходим либо другой способ аппроксимации функции l(m), либо другой критерий достаточности объема выборки.