

Оценка погрешностей будущих измерений по имеющимся данным

Поляченко Юрий

19 июля 2020 г.

1 Постановка задачи

Цель – предсказать погрешность ε_0 выдаваемых нашей программой значений x искомого параметра y .

Область изменения $x \in [0; 1]$ разбита на интервалы $[a_j; b_j]$, для каждого из которых есть N_j экспериментов. Считается, что искомая погрешность ε_0 может меняться от интервала к интервалу, но постоянная внутри интервала (т.е. точнее писать ε_{0j}). Фиксируем j и работаем в выбранном интервале, поэтому далее индекс интервала j опущен.

Есть N экспериментов, про которые известно, что в каждом из них истинное значение y_i попало в интервал, т.е. $y_i \in [a; b] \forall i \in \{1, N\}$. На каждый из этих экспериментов у нас есть результат работы нашей программы x_i . Предполагается, что истинное значение y_i распределено по Гауссу со средним x_i и некой дисперсией ε , т.е.

$$\frac{\mathcal{P}(y_i \in [t, t + dt])}{dt} = g(t, x_i, \varepsilon), \quad (1)$$

где $\mathcal{P}(A)$ - вероятность того, что A верно, а

$$g(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right] \quad (2)$$

– Гауссово распределение.

Ищем зависимость

$$\varepsilon_0(a, b, \{x_i\}_{i=1}^N, p_0) \quad (3)$$

такую, что вероятность реализации описанной выше ситуации (т.е. что все истинные значения попали в интервал) $= p_0$, т.е.

$$\mathcal{P}(\forall x \in [a; b] | x - y| < \varepsilon_0) = p_0 \quad (4)$$

Из сторонних соображений считается известным минимально возможная погрешность ε_{min} , т.е. если метод выдает $\varepsilon_0 < \varepsilon_{min}$, то считаем $\varepsilon_0 = \varepsilon_{min}$.

2 Предлагаемое решение

2.1 Идея и приближения

Задав ε , можно посчитать вероятность реализации ситуации, описанной в постановке – попадания всех истинных значений параметра y_i , распределенных по Гауссу каждый около своего x_i , в интервал $[a; b]$. Далее предположение – эта вероятность равна нашей целевой вероятности p_0 . Не очевидно, почему это должно выполняться точно (скорее всего это не выполняется), но для оценки предложено использовать такую модель.

Поясним разумность данного выбора. Будем брать пробные ε и смотреть как от этого зависит ожидаемое поведение истинных значений y_i относительно наших точек x_i . Для примера возьмем весь интервал $[0.2; 0.5]$ и предположим что у нас имеются 5 точек, для которых наша программа выдала ответы 0.22, 0.3, 0.35, 0.4, 0.43. Если предположить, что погрешность наших предсказаний $\varepsilon = 0.02$, то плотность вероятности для каждого из 5 истинных значений будет выглядеть так:

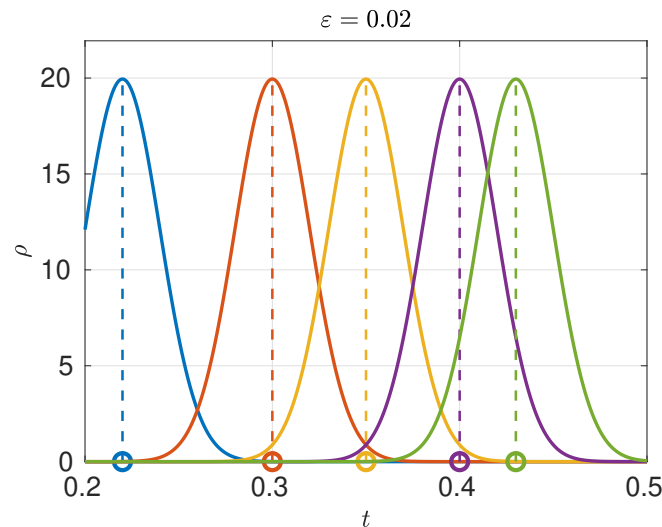


Рис. 1: Разным цветам отвечают разные эксперименты. Для каждого эксперимента: проколотый круг на оси X – наше предсказание ответа, купол – распределение плотности вероятности того, что истинное значение ответа примет значение t , в зависимости от t .

На Рис. 1 видно, что для всех точек кроме синей почти вся кривая (вероятность = площадь под кривой) находится в исследуемом интервале. Это значит, что при $\varepsilon = 0.02$ для всех точек кроме синей вероятность

того, что истинное значение параметра попадет в интервал, равна почти 100%. Синяя же точка находится на расстоянии $\sim 1\sigma$ (в данном случае 0.02), что значит, что вероятность того, что истинное значение параметра в синем эксперименте попадет в интервал $[0.2; 0.5]$ будет $\approx 16\%$. Попадания истинных значений в интервал – события независимые, поэтому вероятность реализации картины в целом будет произведением вероятностей попадания каждого значения в интервал по отдельности. В нашем случае все вероятности кроме синей ≈ 1 , поэтому общая вероятность \approx синяя вероятность $\approx 84\%$. Это значит, что если бы погрешность нашей программы была 0.02, то вероятность случатся тому что случилось на рассматриваемых 5 экспериментах в совокупности была бы 84 %. Поняв это, можно решить обратную задачу: сказать, что мы верим эксперименту на скажем 95%, и найти такое ε , при котором вероятность его реализации как раз будет 95%. Понятно, что задача решается – если в предыдущем примере мы возьмем $\varepsilon = 0.008$, то распределение вероятностей для истинных значений будет

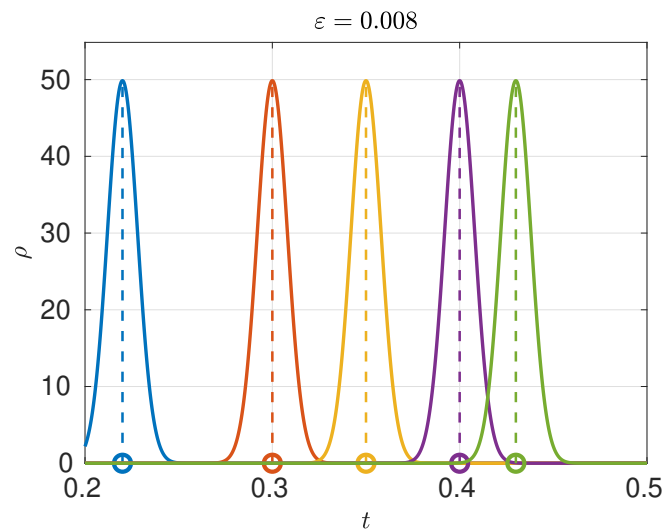


Рис. 2: Обозначения аналогичны Рис. 1. Вероятность реализации эксперимента $>99\%$, что эквивалентно практически полному доверию эксперименту.

Если же все наши экспериментальные точки лежат ближе к центру исследуемой области, то оценка на погрешность выходит грубой. Это понятно из следующего примера. Сдвинем точку 0.22 из предыдущего примера в точку 0.28. График для $\varepsilon = 0.04$ будет выглядеть как

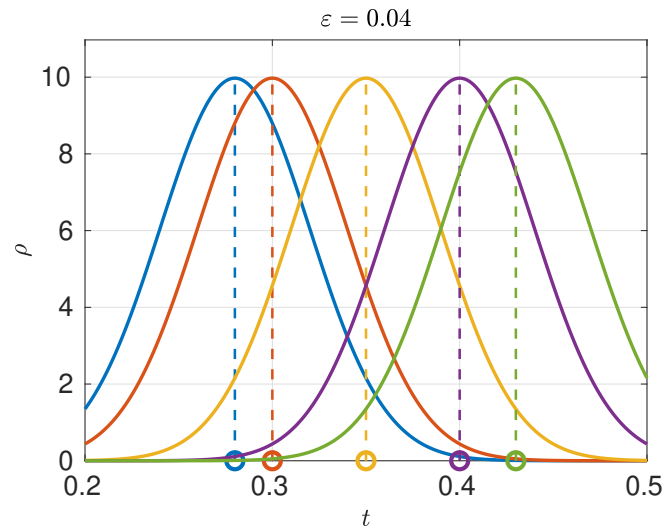


Рис. 3: Обозначения аналогичны Рис. 1. Вероятность реализации эксперимента 92.5%.

Т.е. при наличии точки 0.22, близкой к левой границе исследуемого интервала 0.2, вероятность реализации эксперимента уже при $\varepsilon = 0.02$ была 84%, что говорило о том, что в реальности скорее всего погрешность была меньше. Здесь же даже при $\varepsilon = 0.04$ вероятность все еще $>90\%$. Это на самом деле логично, т.к. если все наши точки у центра интервала, то единственный известный нам факт (на котором и строится вся оценка опгрешности) о том, что все истинные значения попали в интервал, позволяет отбросить только самые большие значения погрешностей.

Заинтересовавшийся читатель может найти весь инходный код проекта и мои контакты для вопросов и предложений [здесь](#). В частности по ссылке лежит программа для оценки ошибок по описываемому здесь методу.

Теперь приведем аналитическое выражение описанной идеи:

2.2 Аналитика

Вероятность попадания i -ой истинной точки в интервал

$$p_i(\varepsilon) = \int_a^b g(x, x_i, \varepsilon) dx = \int_{(a-x_i)/\varepsilon}^{(b-x_i)/\varepsilon} g(x, 0, 1) dx. \quad (5)$$

Введем функцию (известную как функция ошибок):

$$\text{erf}(x) = \int_{-\infty}^x g(t, 0, 1) dt. \quad (6)$$

Попадание каждого истинного значения в интервал - независимое событие, поэтому вероятность реализации нашей совокупности экспериментов

$$p(\varepsilon) = \prod_{i=1}^N p_i(\varepsilon) = \prod_{i=1}^N \left[\text{erf}\left(\frac{b-x_i}{\varepsilon}\right) - \text{erf}\left(\frac{a-x_i}{\varepsilon}\right) \right] \quad (7)$$

Для нахождения желаемого ε_0 решаем уравнение на ε_0 при заданном p_0 .

$$p(\varepsilon_0) = p_0 \quad (8)$$

Уравнения явно не решаются аналитически. Но несложно показать, что функция $p(\varepsilon)$ монотонна, а интервал изменения ε известен и невелик, откуда следует, что уравнение легко решается численно даже самыми простейшими методами вроде деления отрезка пополам. В примерах использован алгоритм, реализованный в функции *fzero* в Matlab и описанный в [1]

2.3 Продолжение примера аналитикой

Продолжим использовать 5 точек из раздела 2.1. В разделе разделе 2.1 был описан алгоритм как мы задавшись определенным ε можем оценить вероятность p , с которой при этом ε реализовались бы имеющиеся у нас экспериментальные данные. Сделав так для многих различных ε , можно для каждого из них получить свое значение $p(\varepsilon)$ (Синяя кривая на Рис. 4).

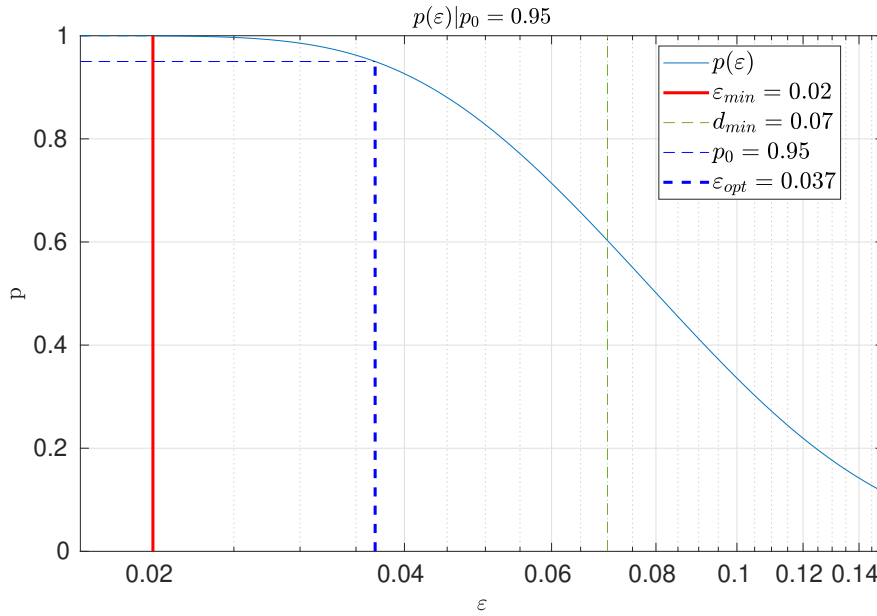


Рис. 4: Зависимость $p(\varepsilon)$ вероятности реализации эксперимента при данной погрешности программы. Красная линия – принятая минимально возможная погрешность $\varepsilon_{min} = 0.02$, Синяя вертикаль – найденная оценка, синяя горизонталь – наш выбор $p_0 = 95\%$, зеленый пунктир - минимальное расстояние точек до границы.

Видно, что наличие множества (> 1) точек позволяет улучшить оценку с очевидного значения минимального расстояния до границы – синяя линия левее зеленой, т.е. оценка по предлагаемому методу лучше чем наивная оценка сверху на глаз.

3 Результат применения

Можно исследовать, как оценка погрешности зависит от количества имеющихся экспериментальных данных в «усредненном» случае, когда ответы нашей программы расположены в интервале на равных промежутках:

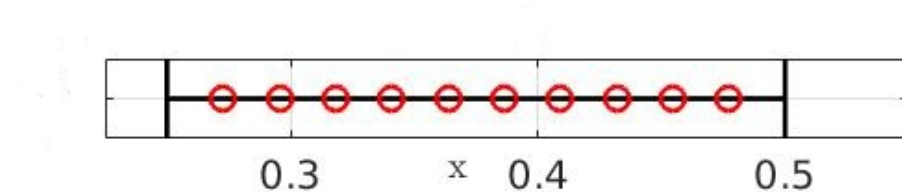


Рис. 5: Равномерное расположение 10 пробных точек в интервале $[0.25; 0.5]$.

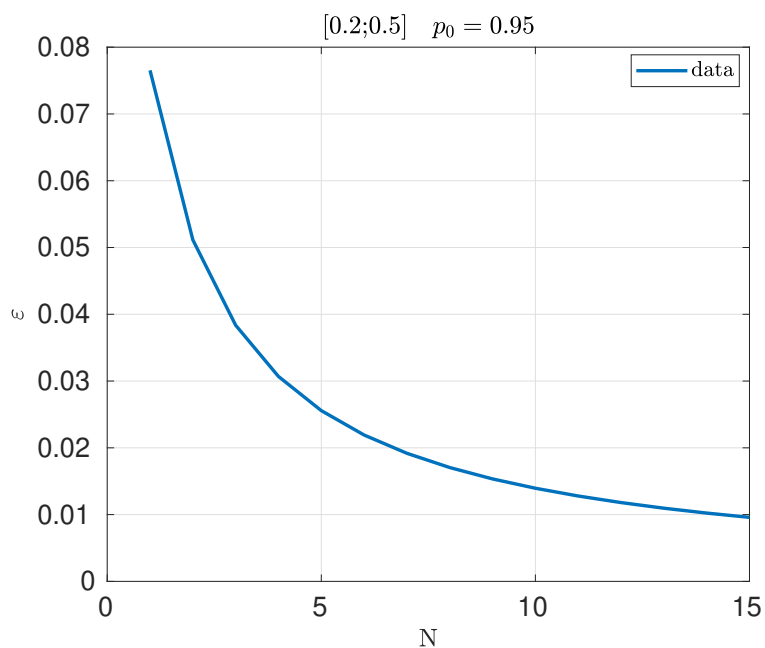


Рис. 6: Зависимость $\varepsilon(N)$ оценки погрешности программы от количества экспериментальных точек при их равномерном распределении в интервале.

На глаз зависимость на Рис. 6 близка к $1/N$, что ожидаемо, т.к. погрешность в основном определяется минимальным расстоянием до границы, которое при выбранной расстановке точек убывает как $1/N$.

Можно проверить отклонения от закона $1/N$ – рис.(7).

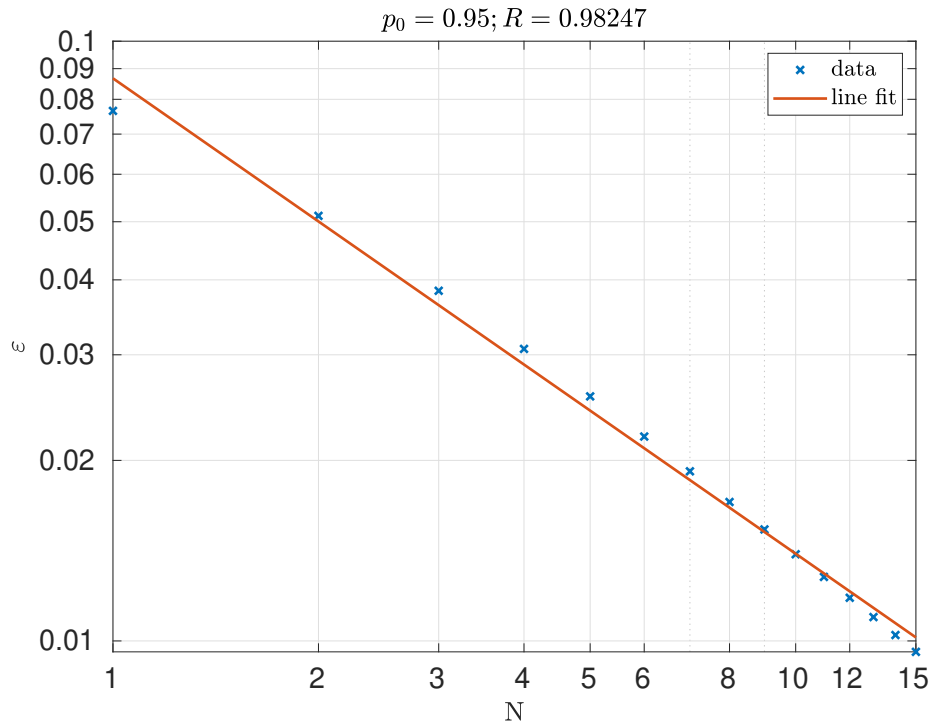


Рис. 7: $\varepsilon(N)$, логарифмический масштаб, попытка линеаризации

Видно, что наклон с правда близок к -1 , но небольшие отклонения от линейности есть.

4 Описание и рекомендации к программе

Для удобства применения изложенного метода создана программа, позволяющее получить оценку погрешности для любого заданного набора экспериментов.

Окно программы выглядит так:

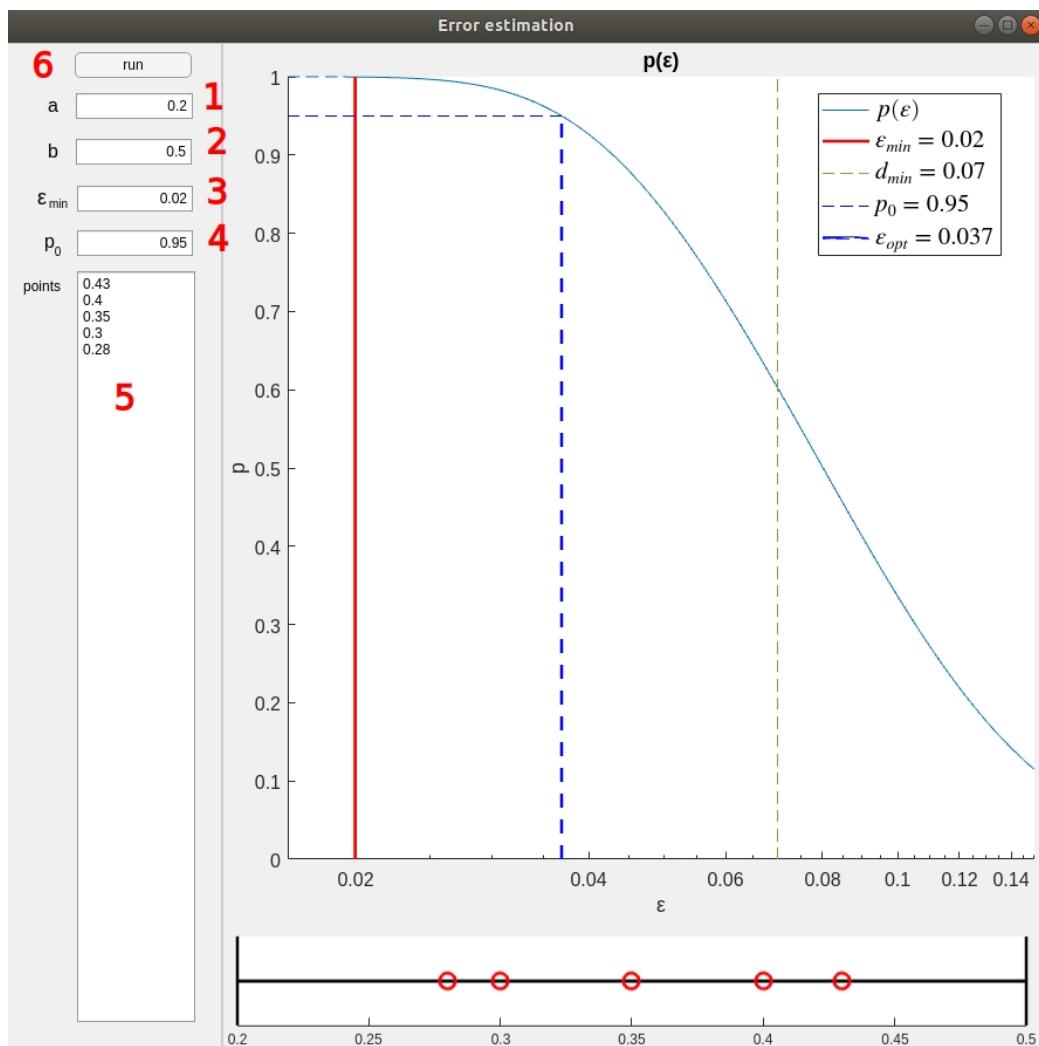


Рис. 8: Окно программы. 1,2 – границы анализируемого интервала. 3 – априорная минимально допустимая погрешность. 4 – степень доверия эксперименту. 5 – имеющиеся результаты работы программы. 6 – сделать расчет с введенными данными.

Поясним, что изображено на рисунках:

4.1 Исследуемый интервал, ввод имеющихся данных

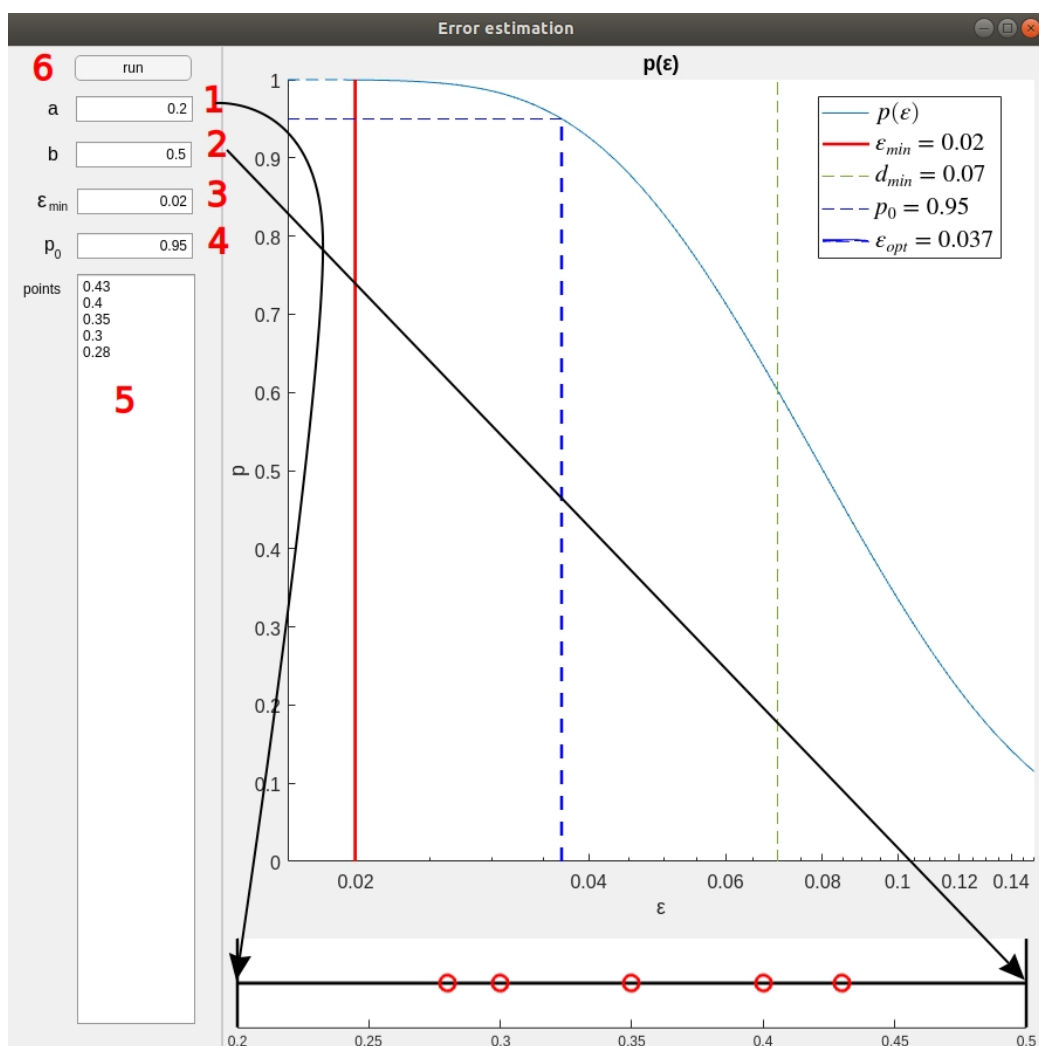


Рис. 9: Значения полей 1 и 2 – значения a и b , т.е. границ исследуемого интервала.

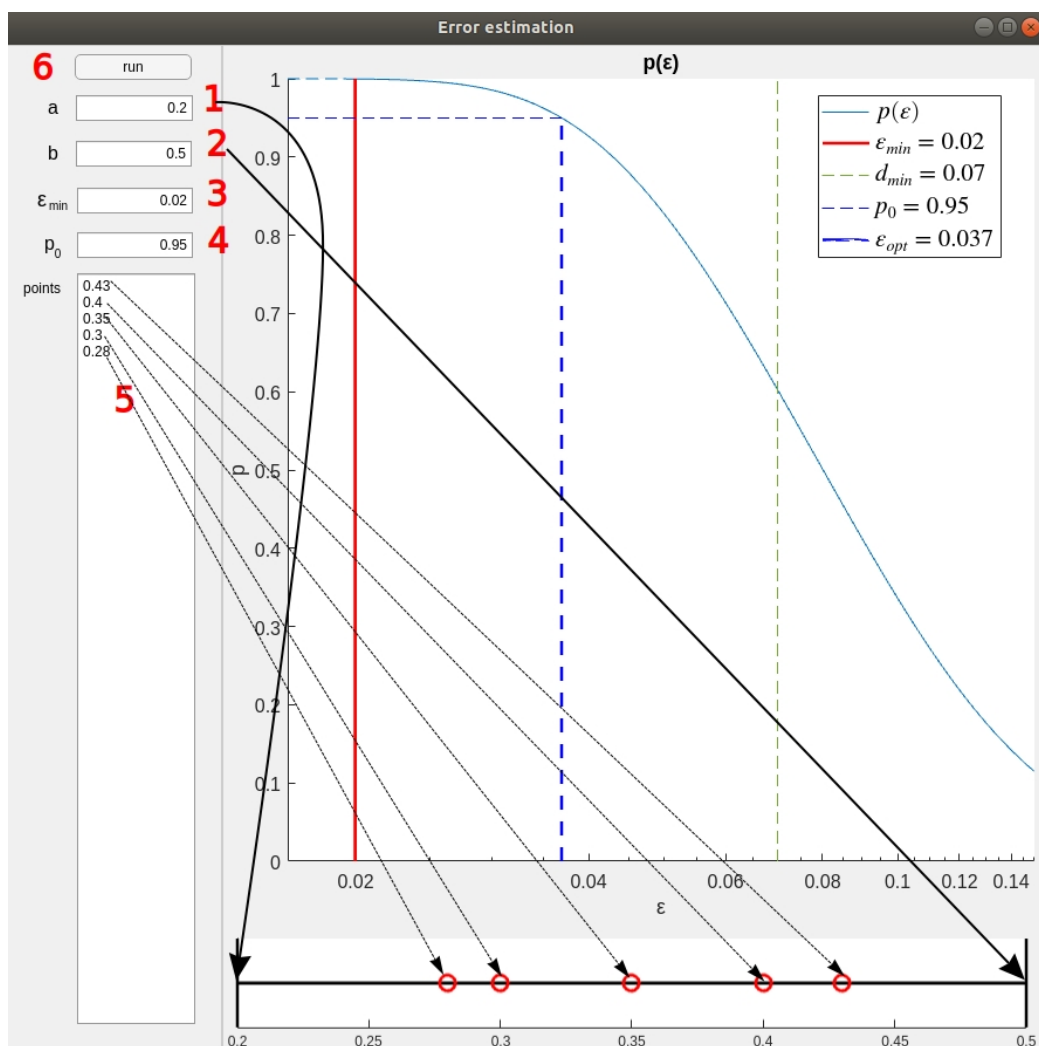


Рис. 10: Таблица 5 – имеющиеся ответы программы.

4.2 Основной рисунок

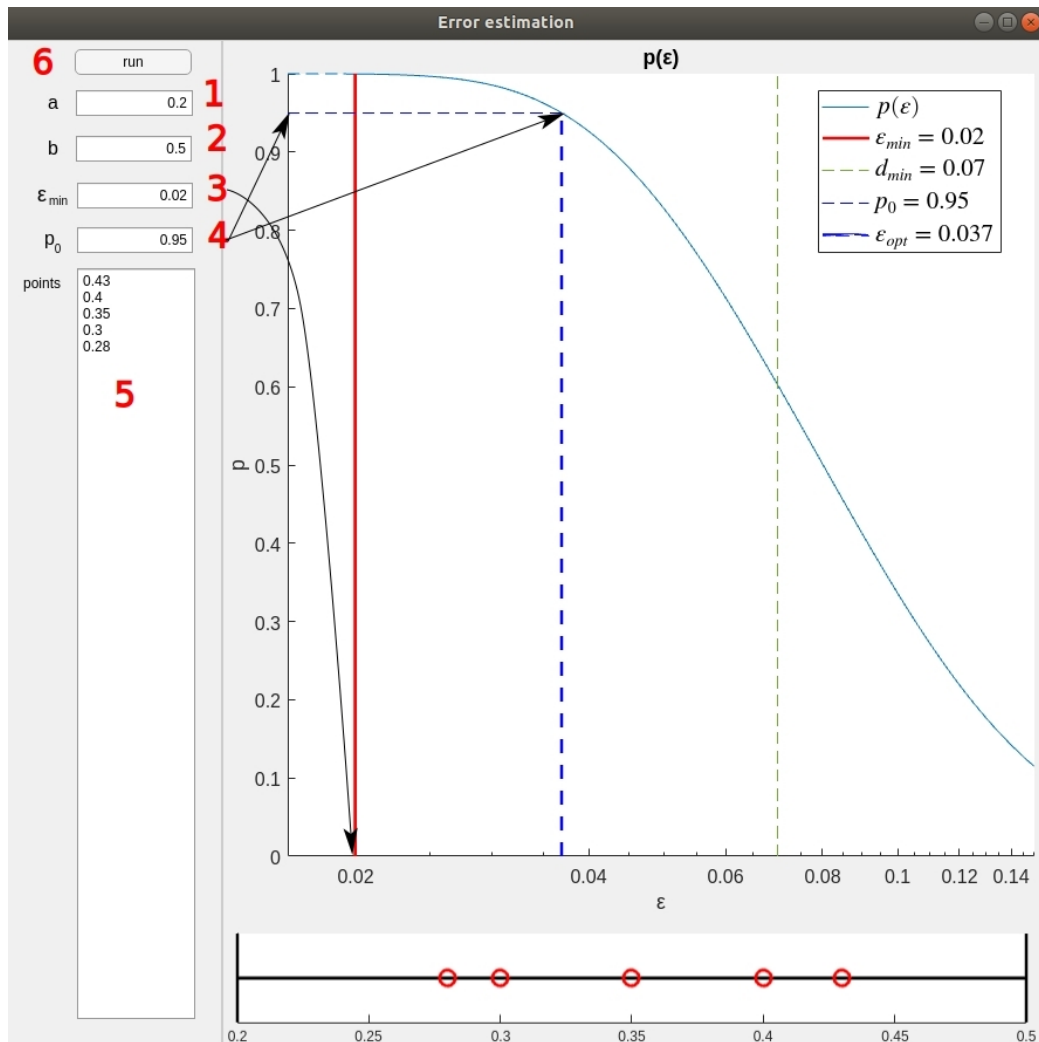


Рис. 11: Поле 3 – априорное значение минимально допустимой погрешности. Оно задается по соображениям пользователя исходя из сторонней информации. Поле 4 – степень доверия эксперименту. Если мы уверены в эксперименте на 95% – пишем 0.95. Основная задача программы состоит в использовании этой величины – по точкам 5 и интервалу 1,2 программа строит голубую кривую и находит тот X (т.е. тот ϵ), при котором кривая принимает значение в поле 4 (т.е. вероятность реализации эксперимента совпадает с нашими ожиданиями).

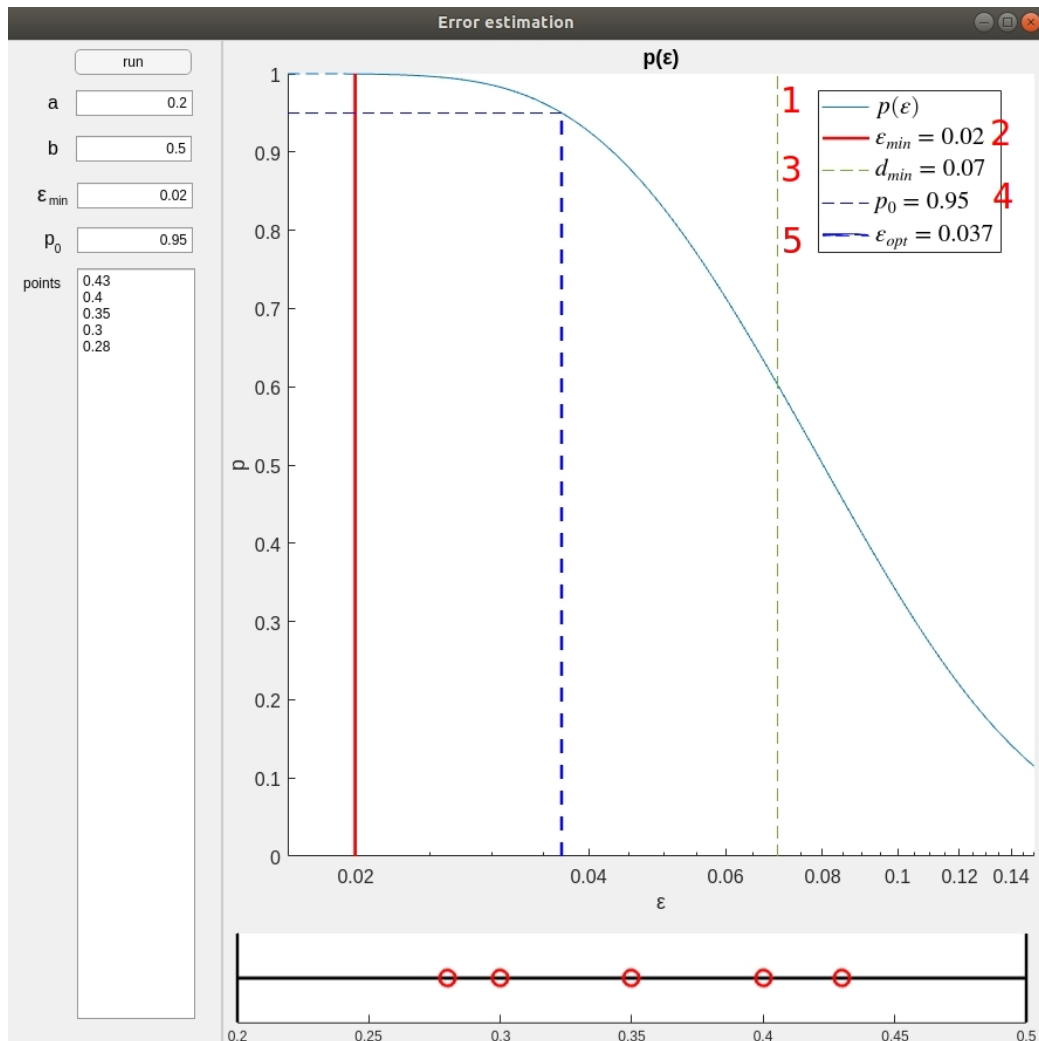


Рис. 12: Описание легенды. 1 – зависимость вероятности реализации имеющейся совокупности экспериментов от предполагаемой погрешности предсказаний. 2 – априорно заданная минимальная погрешность. 3 – минимальное расстояние от заданного множества точек до ближайшей границы интервала. 4 – степень доверия эксперименту. 5 – ответ программы, получанная оценка погрешности предсказаний.

Список литературы

- [1] F Grund. “Forsythe, GE/Malcolm, MA/Moler, CB, Computer Methods for Mathematical Computations. Englewood Cliffs, New Jersey 07632. Prentice Hall, Inc., 1977. XI, 259 S”. В: *ZaMM* 59.2 (1979), с. 141—142.