Санкт-Петербургский политехнический университет
Петра Великого
Институт прикладной математики и механики
Высшая школа прикладной математики и вычислительной
физики

Отчёт по курсовой работе
на тему "Расстояние Фреше"
по дисциплине «Математическая статистика»

Выполнил студент:
Зубкова Дарья
группа:
3630102/70201

Проверил:
к.ф.-м.н., доцент
Баженов Александр
Николаевич

Санкт-Петербург, 2020г.

Содержание

Список иллюстраций

1. Постановка задачи

Рассмотреть расстояние по метрике Фреше между нормальным и равномерным распределениями.

2. Теория

Расстояние Фреше — это мера сходства кривых, принимающая во внимание число и порядок точек вдоль кривых. Расстояние названо по имени французского математика Мориса Фреше.

2.1. Определение

Пусть S — метрическое пространство. Кривая A в пространстве S — это непрерывное отображение единичного отрезка в S, т.е. $A : [0,1] \to S$. Репараметризация α отрезка [0,1] — это непрерывная неубывающая сюръекция $\alpha : [0,1] \to [0,1]$.

Пусть A и B — две кривые в S. Тогда расстояние Фреше между A и B определяется как точная нижняя граница по всем репараметризациям α и β отрезка [0,1] по всем максимумам $\in [0,1]$ расстояний в S между $A(\alpha(t))$ и $B(\beta(t))$. В математических обозначениях расстояние Фреше $F(A,B)$ равно:

$$F(A,B) = \inf_{\alpha,\beta}\ \max_{t\in[0,1]}\ \left\{ d\Big(A(\alpha(t)),\, B(\beta(t))\Big) \right\}$$

где d — функция расстояния пространства S.

Можно сказать, что t - это параметр времени. Тогда $A(\alpha(t))$ является положением собаки, а $B(\beta(t))$ — положением владельца собаки по времени t. Длина поводка между ними в момент времени t будет равно расстоянию между $A(\alpha(t))$ и $B(\beta(t))$. Взятие инфимума по всем возможным репараметризациям отрезка [0,1] соответствует выбору прогулки вдоль кривых, при которой максимальная длина поводка минимизируется. Ограничение, что α и β не убывают, означает, что ни собака, ни её владелец не могут повернуть назад.

Метрика Фреше принимает во внимание течение двух кривых, поскольку пары точек, расстояние между которыми определяет расстояние Фреше, «пробегают» вдоль кривых.

Расстояние Фреше между двумя кривыми — это длина самого короткого поводка, с которым можно пройти кривые. Не самый короткий поводок, при котором можно пройти все пути, а самый короткий, при котором можно пройти этот путь.
Определение симметрично относительно двух кривых.

2.2. Дискретное расстояние Фреше

Определим кривую как непрерывное отображение $f : [a, b] \to V$, где $a, b \in R$ и $a \leq b$ и (V, d) - метрическое пространство.

Даны две кривые $f : [a, b] \to V$ и $g : [a', b'] \to V$, их расстояние Фреше определено в виде:

$$\delta F(f, g) = \inf_{\alpha,\beta} \max_{t\in[0,1]} \left\{ d\Big(f(\alpha(t)),\ g(\beta(t)) \Big) \right\}$$

где α (соответственно β) - произвольная непрерывная неубывающая функция из [0, 1] на [a, b] (соответственно [a',b']).

При вычислении расстояния Фреше между произвольными кривыми обычно аппроксимируют кривые многоугольными кривыми. Многоугольная кривая - это кривая$P : [0, n] \to V$, где n - натуральное число, такое, что для каждого i 0, 1,. , , , n - 1, ограничение P к интервалу [i, i + 1] является аффинным, то есть $P(i + \lambda) = (1 - \lambda)P(i) + \lambda P(i + 1)$.
Пусть $P : [0, n] \to V$ - многоугольная кривая. Обозначим последовательность $(P(0), P(1), ..., P(n))$ конечных точек отрезков линии P на $\sigma(P)$.

Пусть P и Q многоугольные кривые и $\sigma(P) = (u_1, ..., u_p)$ и $\sigma(Q) = (v_1, ..., v_q)$ соответствующих последовательностей. Связь L между P и Q представляет собой последовательность:

$$(u_{a_1}, v_{b_1}), ..., (u_{a_m}, v_{b_m})$$

различных пар из $\sigma(P)x\sigma(Q)$, таких что $a_1 = 1, b_1 = 1, a_m = p, b_m = q$, и для всех $i = 1, ..., q$ мы имеем $a_{i+1} = a_i$ или $a_{i+1} = a_i + 1$, и $b_{i+1} = b_i$ или $b_{i+1} = b_i$.

Таким образом, нужно соблюдать порядок точек в P и Q. Длина $||L||$ - длина самого длинного звена в L, то есть:

$$||L|| = \max_{i=1,...m} d(u_{a_i}, v_{b_i})$$

Для заданных многоугольных кривых P и Q их дискретное расстояние Фреше определяется как:

$$\delta_{dF}(P, Q) = min||L||$$

где $L - PQ$.

3. Реализация

Курсовая работа выполнена с помощью библиотек Numpy, Matplotlib, Similaritymeasures, Math на языке программирования Python в среде разработки PyCharm. Код курсовой работы лежит на GitHub. Ссылка на Git представлена в приложении.

Расстояние Фреше вычисляется посредством динамического пограммирования.

4. Результаты

4.1. Расстояние Фреше

Пусть P будет нормальным распределением, а Q будет равномерным. Тогда зададим эти распределения и разобьем их на малые ломаные.

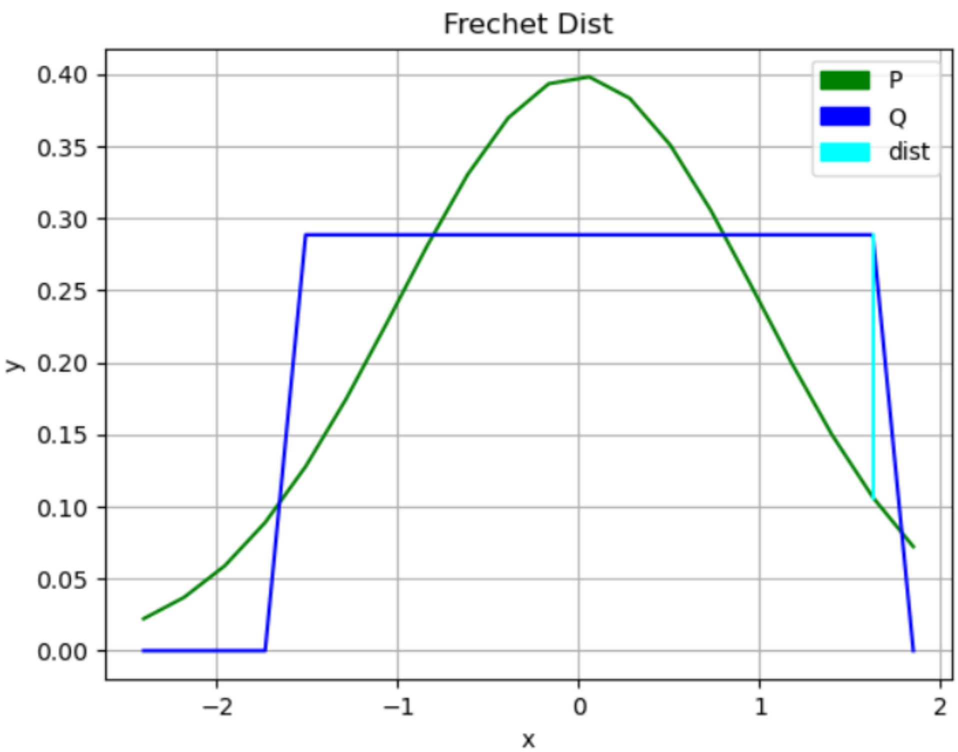

Рис. 1. n = 20, $\sigma = 1$

Расстояние Фреше: $\delta_{dF}(P, Q) = 0.1981767514989412$

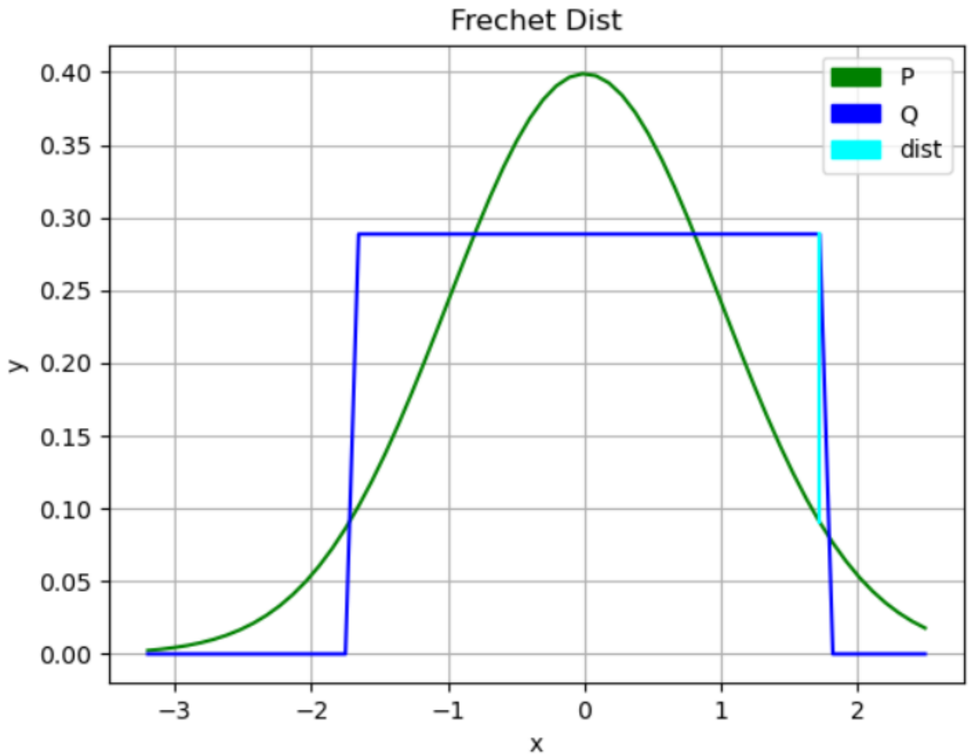

Рис. 2. n = 60, $\sigma = 1$

Расстояние Фреше: $\delta_{dF}(P, Q) = 0.1980977419778272$

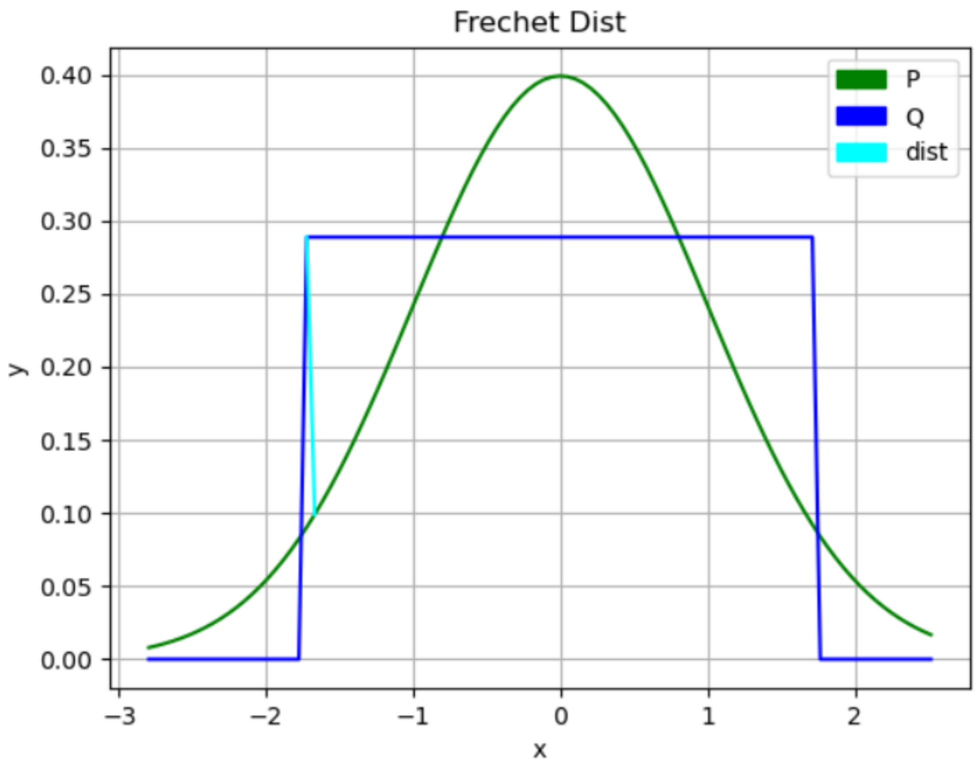

Рис. 3. n = 100, $\sigma = 1$

Расстояние Фреше: $\delta_{dF}(P, Q) = 0.19709660691043254$

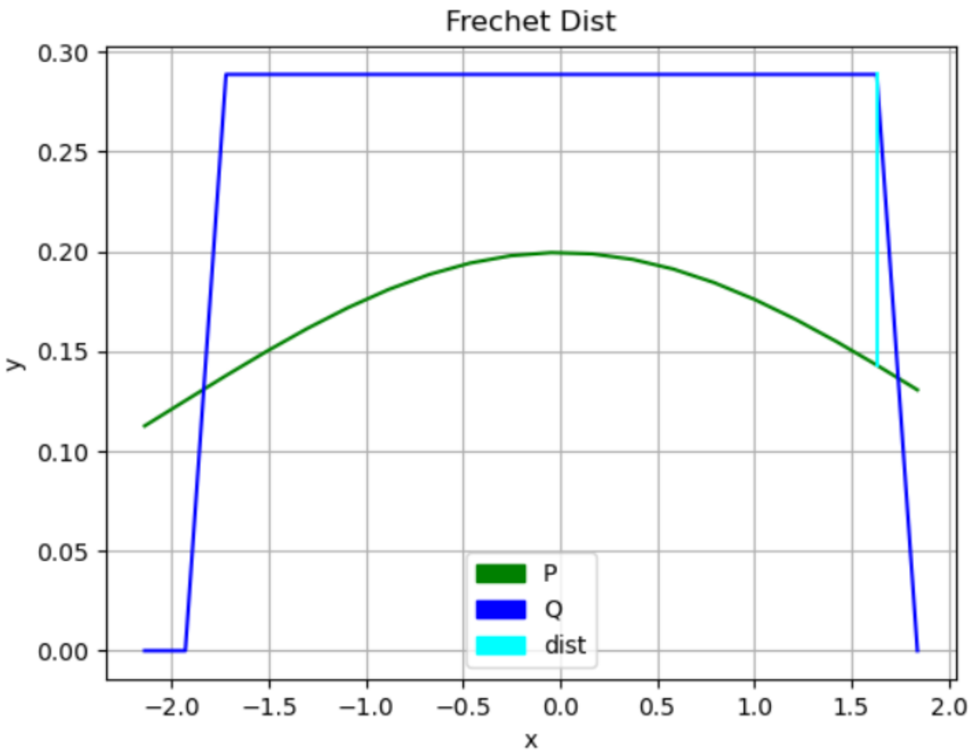

Рис. 4. n = 20, $\sigma = 2$

Расстояние Фреше: $\delta_{dF}(P, Q) = 0.1508654068467085$

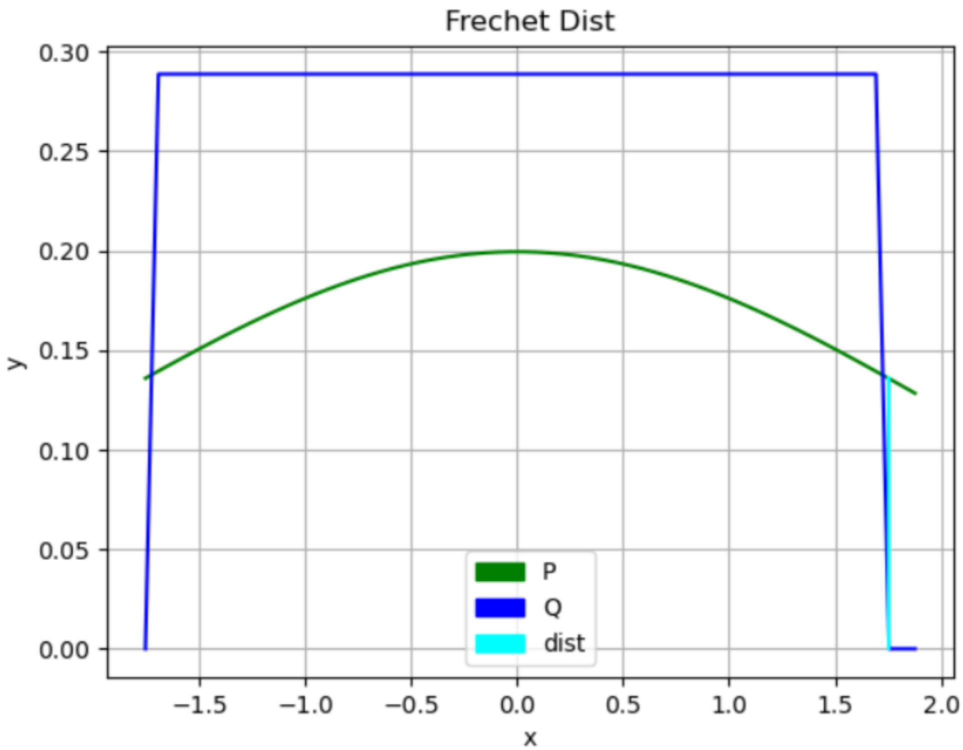

Рис. 5. n = 60, $\sigma = 2$

Расстояние Фреше: $\delta_{dF}(P, Q) = 0.1492918552142273$

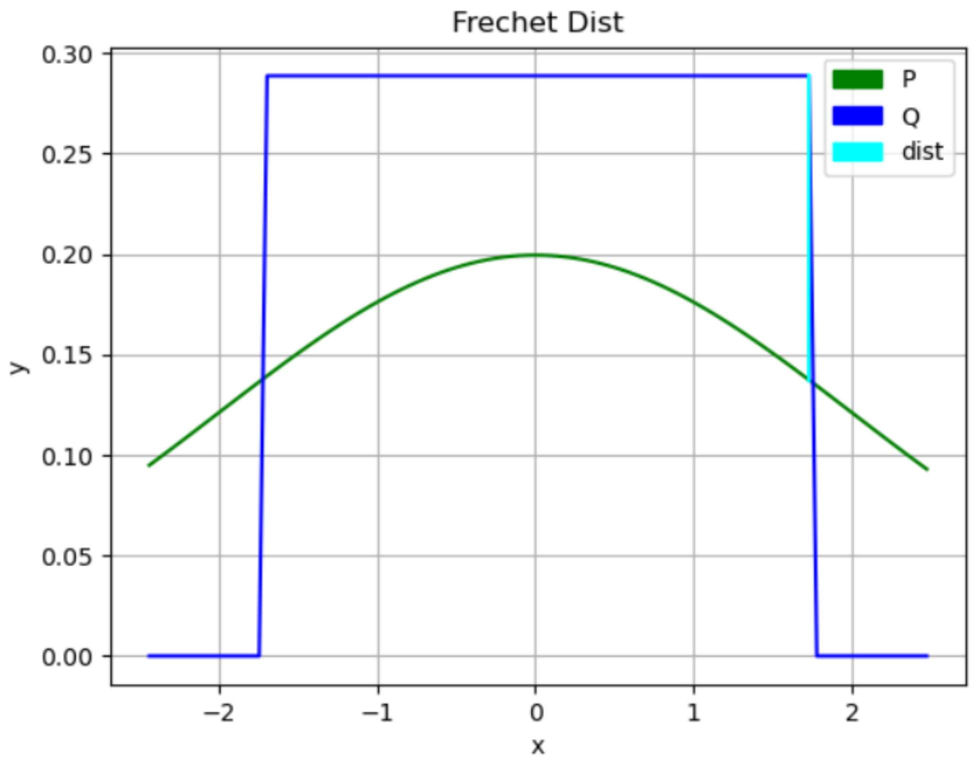

Рис. 6. n = 100, $\sigma = 2$

Расстояние Фреше: $\delta_{dF}(P, Q) = 0.1512990631656247$

5. Обсуждение

Алгоритм вычисления расстояния Фреше – рекурсивный, поэтому с большими n работать не может.

Для $[a, b] = [-\sqrt{3}, \sqrt{3}]$ в равномерном распределении расстояние Фреше до нормального распределения с $\sigma = 1$ больше, чем $\sigma = 2$.

Для $[a, b] = [-\sqrt{2}, \sqrt{2}]$ в равномерном распределении расстояние Фреше до нормального распределения с $\sigma = 1$ будет почти таким же, как с $\sigma = 2$:

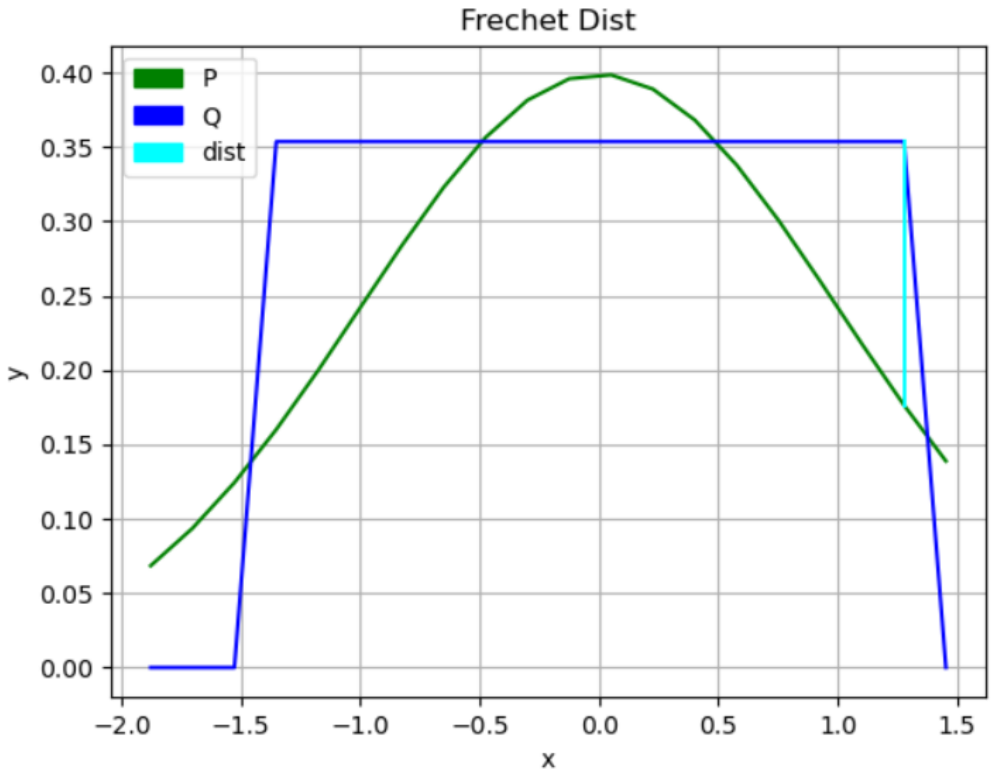

Рис. 7. n = 20, $\sigma = 1$, $[-\sqrt{2}, \sqrt{2}]$

Расстояние Фреше: $\delta_{dF}(P, Q) = 0.19354122524926806$

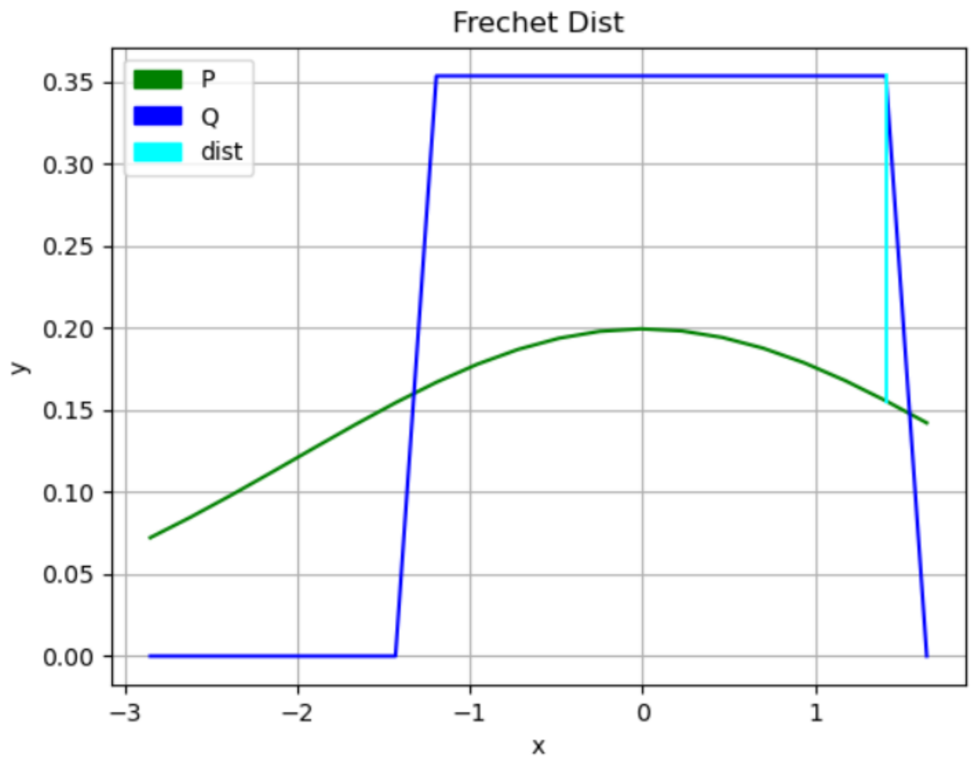

Рис. 8. n = 20, $\sigma = 2$, $[-\sqrt{2}, \sqrt{2}]$

Расстояние Фреше: $\delta_{dF}(P, Q) = 0.19796336393599687$

Также расстояния Фреше необязательно должно быть единственным. При выполнении одного из следующих трёх условий решениее пара точек, между которыми найдено расстояние Фреше, не является единственной:

$$\delta_{dF}(P[i], Q[j]) = c[i-1, j] = \delta_{dF}(P[i-1], Q[j])$$

$$\delta_{dF}(P[i], Q[j]) = c[i-1, j-1] = \delta_{dF}(P[i-1], Q[j-1])$$

$$\delta_{dF}(P[i], Q[j]) = c[i, j-1] = \delta_{dF}(P[i], Q[j-1])$$

При $n = 100, sigma = 1, [-\sqrt{3}, \sqrt{3}]$, при расстоянии Фреше $\delta_{dF}(P, Q) = 0.1948993947662015$ нашлось 3 пары точек:

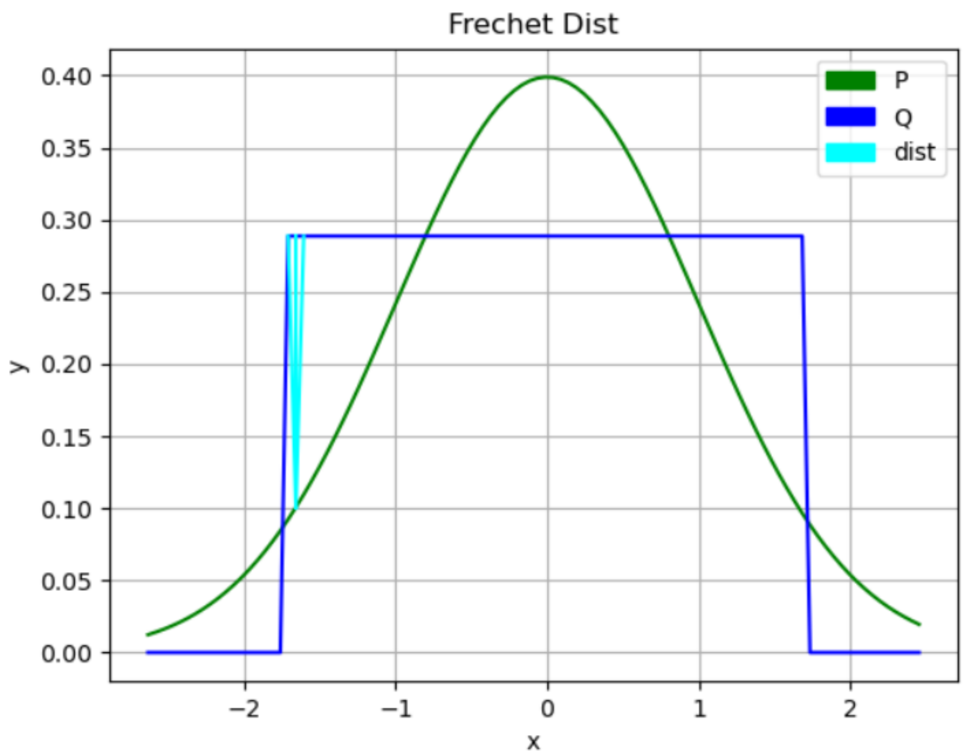

Рис. 9. n = 100, $\sigma = 1$, $[-\sqrt{3}, \sqrt{3}]$

Расстояние Фреше между двумя кривыми — это длина самого короткого поводка, с которым можно пройти именно этот путь.

Рассмотрим расстояние Фреше для двух распределений с n = 10:

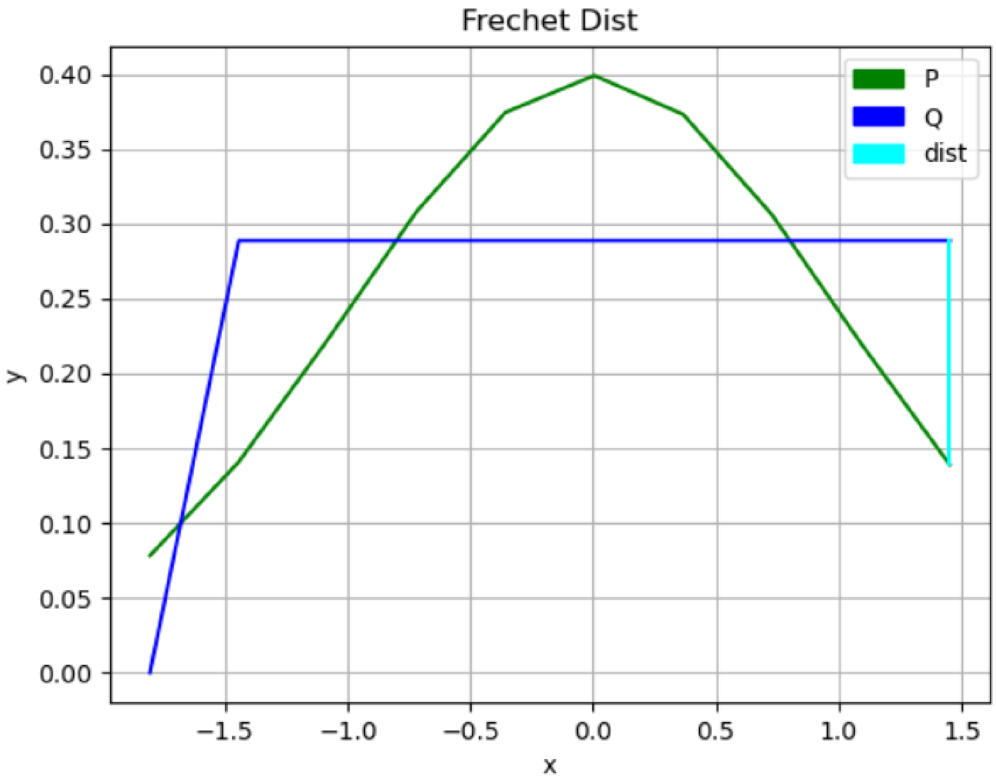

Рис. 10. n = 10

Расстояние Фреше: $\delta_{dF}(P, Q) = 0.1497743904590718$

Рассмотрим расстояние Фреше для двух кривых: ядерной оценки плотности для равномерного распределения и самого равномерного распределения при n = 100:

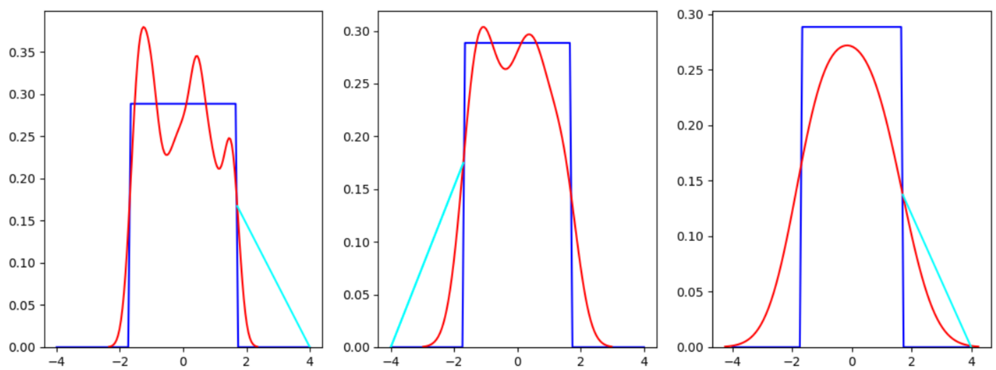

Рис. 11. n = 100, Kernel uniform and uniform distribution

Здесь расстояние Фреше получается таким:
для $h_n/2$, $\delta_{dF}(P,Q) = 2.2965233902629834$
для h_n, $\delta_{dF}(P,Q) = 2.2948883897868826$
для $2 * h_n$, $\delta_{dF}(P,Q) = 2.294546249392843$

6. Приложение

Ссылка на код: `https://github.com/DariaZubkova/MathStat/tree/master/Freshe`

Список литературы

[1] Расстония Фреше.
URL:`https://en.wikipedia.org/wiki/Frechet_distance`

[2] Дискретное расстояние Фреше, вычисление.
URL:`http://www.kr.tuwien.ac.at/staff/eiter/et-archive/cdtr9464.pdf`