ИНФОРМАЦИОННАЯ ЧУВСТВИТЕЛЬНОСТЬ КОМПЬЮТЕРНЫХ АЛГОРИТМОВ И ЕЁ КОЛИЧЕСТВЕННЫЕ МЕРЫ

д.т.н., профессор
М.В. Ульянов
Кафедра
«Управление разработкой программного обеспечения»
НИУ-ВШЭ
«Прикладная математика и моделирование систем»
МГУПечати

ТРЕБОВАНИЯ РАЗРАБОТЧИКОВ АЛГОРИТМИЧЕСКОГО ОБЕСПЕЧЕНИЯ ПРОГРАММ

- Ресурсная эффективность по трудоёмкости и дополнительной памяти
- Прогнозирование временных характеристик на реальном диапазоне длин входов
- Обеспечение стабильности по времени на различных входах фиксированной длины

ОСНОВНЫЕ ОБОЗНАЧЕНИЯ

A – алгоритм;

D_A – множество возможных конкретных проблем для задачи, решаемой с помощью алгоритма A *(множество допустимых входов алгоритма)*;

D – конкретный вход алгоритма;

$f_A(D)$ – функция трудоёмкости для входа D;

$$f_A^{\wedge}(n) = \max_{D \in D_n} \{ f_A(D) \};$$

$$f_A^{\vee}(n) = \min_{D \in D_n} \{ f_A(D) \};$$

$$\overline{f_A}(n) = \sum_{D \in D_n} P(D) \cdot f_A(D).$$

ПРОГНОЗИРОВАНИЕ НА ОСНОВЕ ФУНКЦИИ ТРУДОЁМКОСТИ: НЕДОСТАТКИ СУЩЕСТВУЮЩИХ ПОДХОДОВ

- Прогнозирование по трудоёмкости в худшем случае (гарантированная оценка сверху) даёт почти всегда сильно завышенные результаты
- Прогнозирование по трудоёмкости в среднем не учитывает информацию о размахе варьирования.
- Качество прогноза во многом определяется влиянием различных входов фиксированной длины на трудоёмкость, **информационной чувствительностью** исследуемого алгоритма в рамках выбранной количественной оценки.

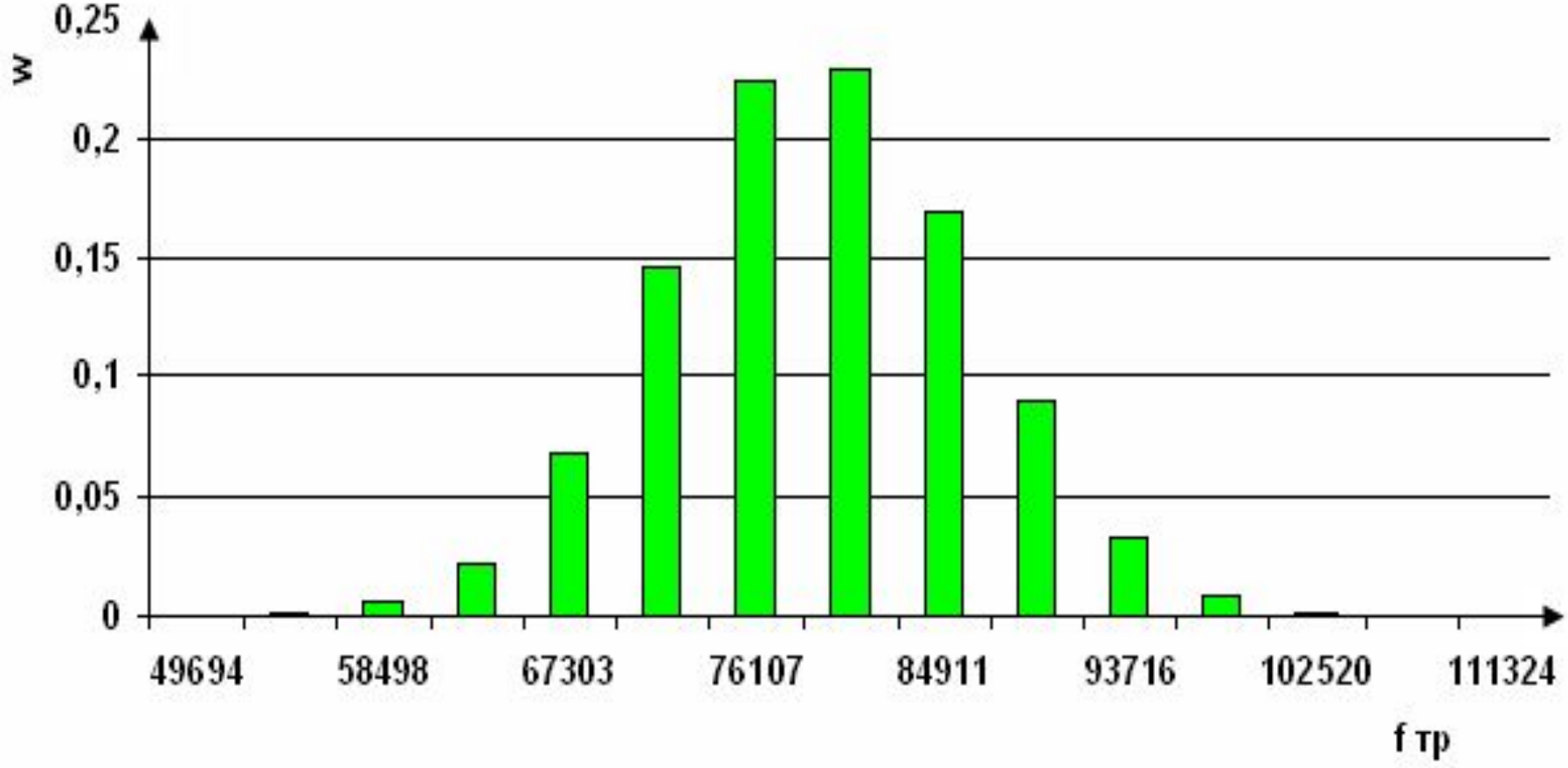
ПОСТАНОВКА ЗАДАЧИ И ПРЕДПОЛОЖЕНИЯ

Задача: Введение и сравнительный анализ различных количественных оценок информационной чувствительности.

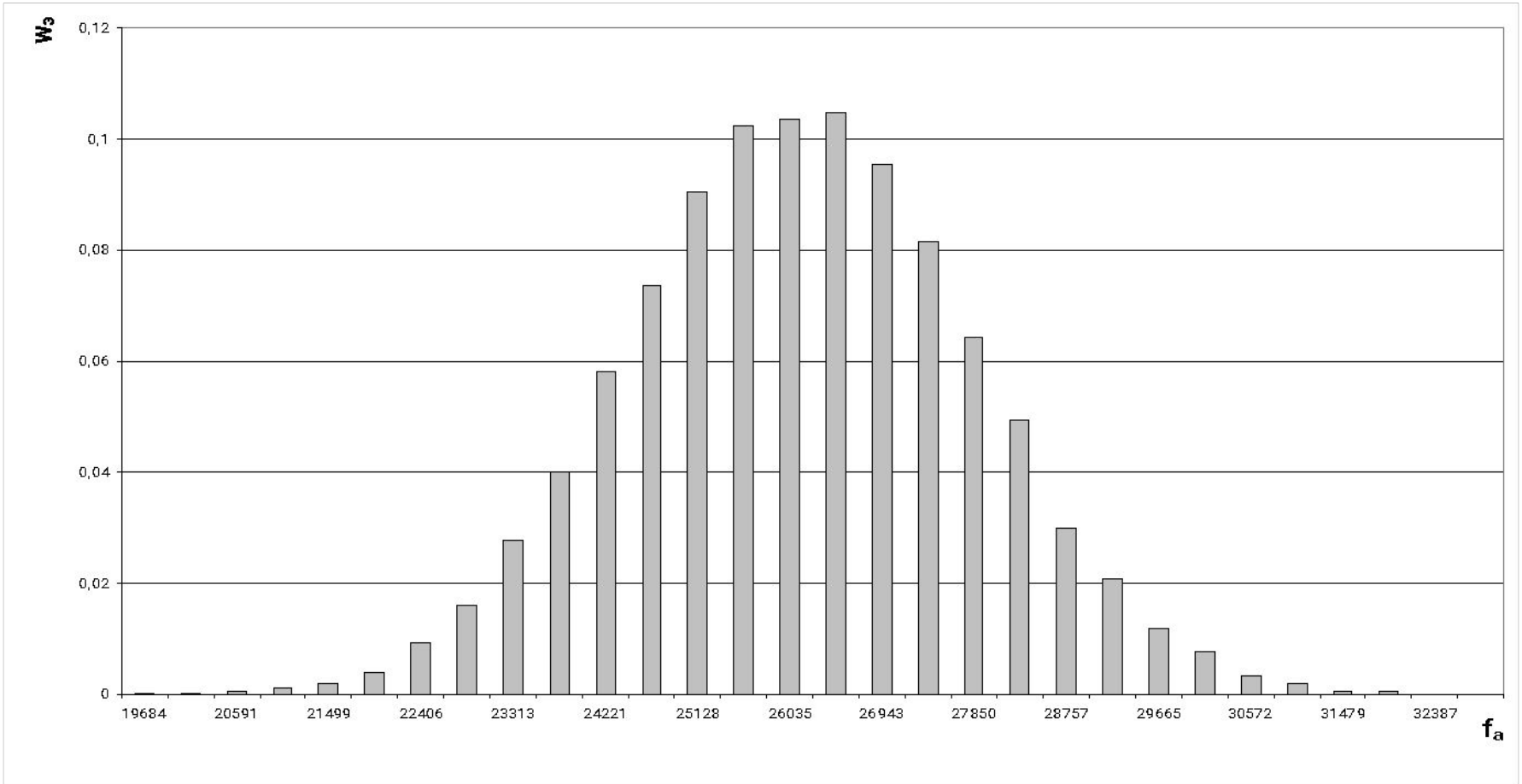
Предположения:

1. Трудоёмкость алгоритма на входах фиксированной длины - дискретная ограниченная случайная величина.
2. Выбор распределения, аппроксимирующего функцию трудоёмкости производится на основе экспериментальных исследований.

ГИСТОГРАММА ОТНОСИТЕЛЬНЫХ ЧАСТОТ ДЛЯ АЛГОРИТМА УМНОЖЕНИЯ ДЛИННЫХ ЦЕЛЫХ ЧИСЕЛ В СТОЛБИК (N=100)

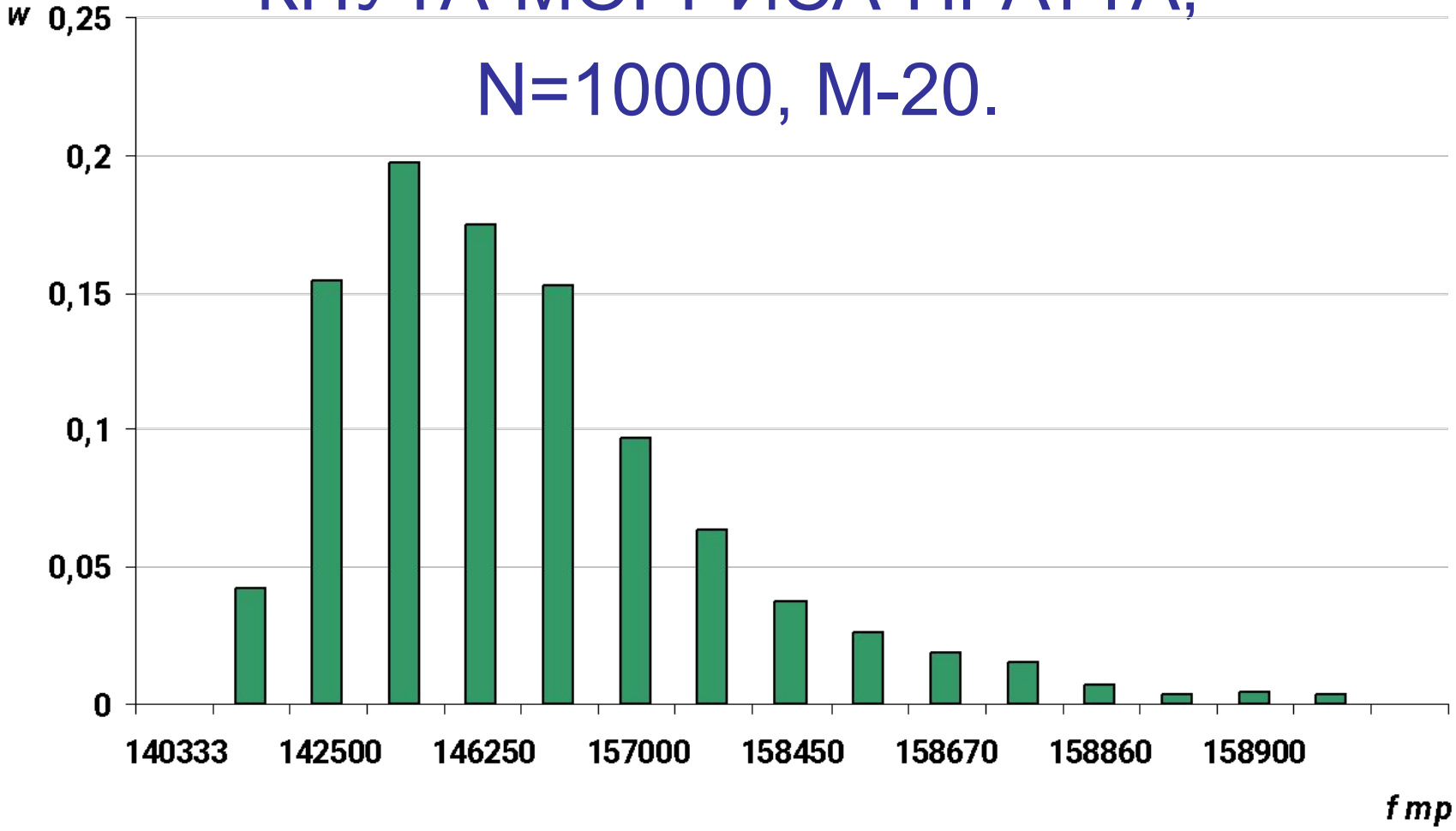


ГИСТОГРАММА ОТНОСИТЕЛЬНЫХ ЧАСТОТ ТРУДОЁМКОСТИ ДЛЯ АЛГОРИТМА СОРТИРОВКИ ВСТАВКАМИ ПРИ N=100, M=20000



ГИСТОГРАММА ОТНОСИТЕЛЬНЫХ ЧАСТОТ ЗНАЧЕНИЙ ФУНКЦИИ ТРУДОЕМКОСТИ ДЛЯ АЛГОРИТМА КНУТА-МОРРИСА-ПРАТТА,

N=10000, M-20.



ПОНЯТИЕ ИНФОРМАЦИОННОЙ ЧУВСТВИТЕЛЬНОСТИ

- Впервые понятие информационной чувствительности алгоритма по трудоёмкости введено М. В. Ульяновым и В. А. Головешкиным. Понятие «**информационная чувствительность**» отражает тот факт, что алгоритм задаёт различное число базовых операций принятой модели вычислений на разных входах , имеющих фиксированную длину . Ключевым для содержательной интерпретации этого термина является выбор количественной оценки (меры), обладающей свойством сопоставимости, т. е. дающей возможность решения задач сравнения алгоритмов и их рационального выбора.

ОЦЕНКА ИНФОРМАЦИОННОЙ ЧУВСТВИТЕЛЬНОСТИ ПО ЭНТРОПИИ

- Для дискретной случайной величины

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i$$

- Непрерывный аналог - дифференциальная энтропия

$$h(X) = -\int_{-\infty}^{\infty} p(x) \cdot \log_2 p(x) dx$$

- Проблемы: не чувствительность к положению моды, максимум для ограниченной случайной величины – при равномерном распределении, влияние на энтропийную оценку числа сегментов гистограммы

СТАТИСТИЧЕСКАЯ ОЦЕНКА ИНФОРМАЦИОННОЙ УВСТВИТЕЛЬНОСТИ

- Нормированный (относительный) размах варьирования функции трудоемкости

$$R_N(n) = \frac{f_A^{\wedge}(n) - f_A^{\vee}(n)}{f_A^{\wedge}(n) + f_A^{\vee}(n)}$$

- Коэффициент вариации

$$V(n) = s_{f_A}(n) \big/ \overline{f_A}(n), \quad 0 \le V(n) \le 1$$

- $s_{f_A}(n)$ - стандартное отклонение трудоемкости

СТАТИСТИЧЕСКАЯ ОЦЕНКА ИНФОРМАЦИОННОЙ УВСТВИТЕЛЬНОСТИ

- Статистическая количественная оценка информационной чувствительности алгоритма по трудоёмкости

$$\delta_{\mathbf{IS}}(n) = V(n) \cdot R_N(n) \qquad 0 \le \delta_{\mathbf{IS}}(n) \le 1$$

- Её применение возможно в случае отсутствия знаний о законе распределения значений трудоёмкости или какой-либо его аппроксимации
- Недостаток - она не позволяет указать интервал значений трудоёмкости при заданной вероятности,

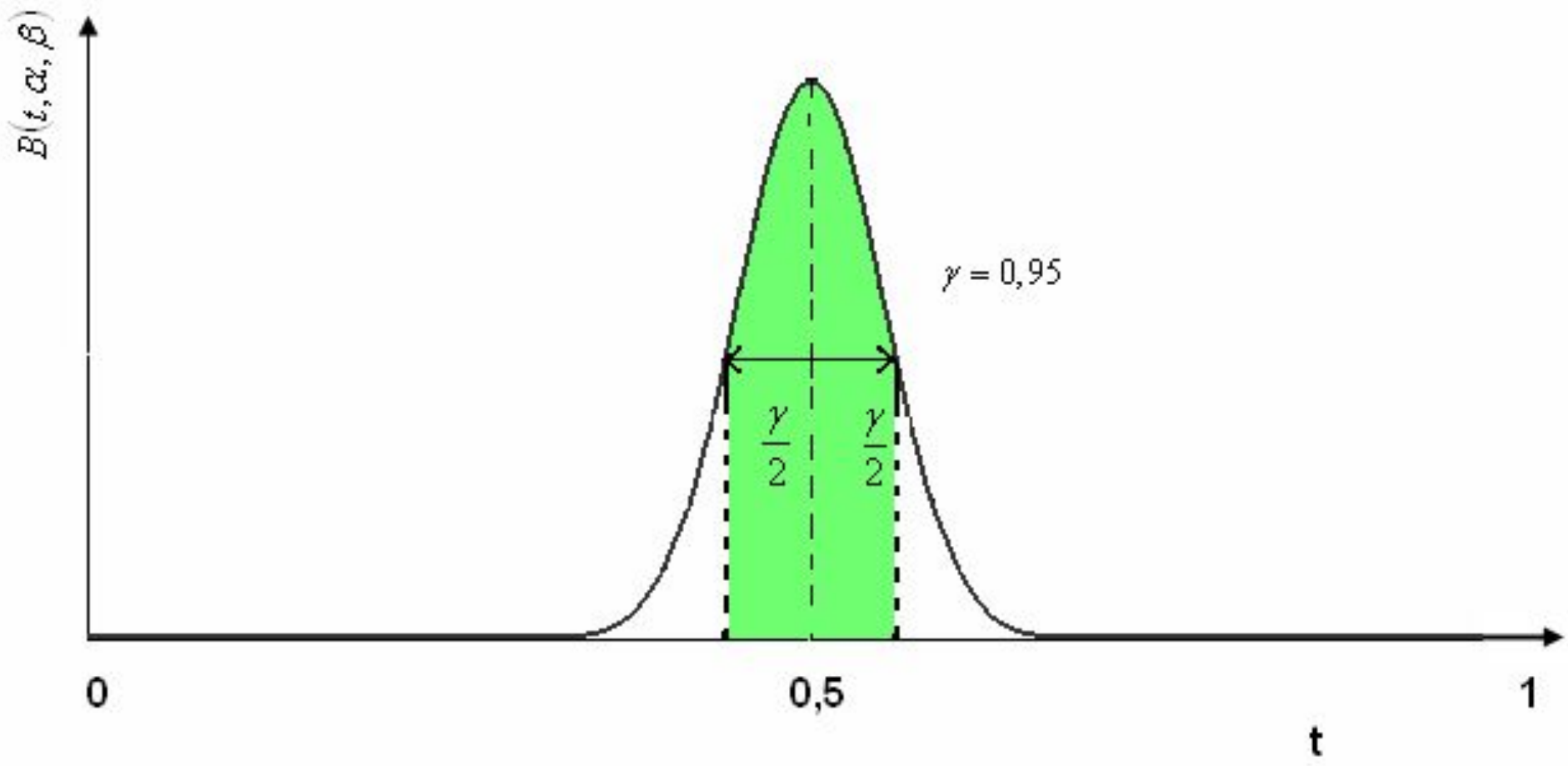
КВАНТИЛЬНАЯ ОЦЕНКА ИНФОРМАЦИОННОЙ УВСТВИТЕЛЬНОСТИ

- Основная идея состоит в определении длины сегмента нормированных значений трудоёмкости, по которому интеграл от функции плотности равен заданной вероятности (надёжности)

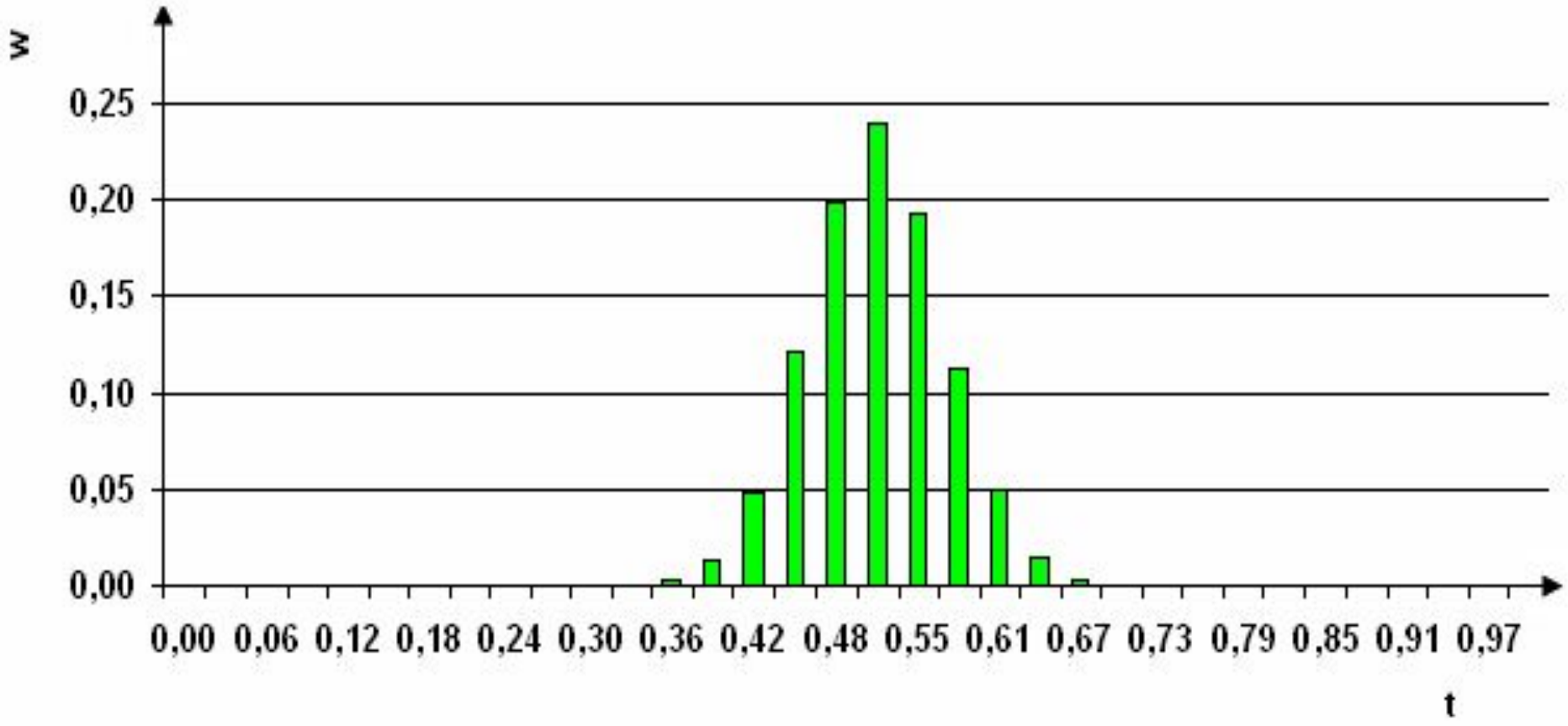
$$\delta_{\mathrm{IQ}}(\gamma, n) = F^{-1}\left(n, \frac{1}{2} + \frac{\gamma}{2}\right) - F^{-1}\left(n, \frac{1}{2} - \frac{\gamma}{2}\right)$$

- Отметим, так же, что эта оценка не содержит информацию о положении гамма-квантиля закона распределения на нормированном сегменте.

КВАНТИЛЬНАЯ МЕРА ИНФОРМАЦИОННОЙ ЧУВСТВИТЕЛЬНОСТИ – ИНТЕРВАЛ, СИММЕТРИЧНЫЙ ОТНОСИТЕЛЬНО МЕДИАНЫ РАСПРЕДЕЛЕНИЯ



НОРМИРОВАННАЯ ВЕЛИЧИНА Т ДЛЯ АЛГОРИТМА УМНОЖЕНИЯ (N=100)



ВЫБОР ФУНКЦИИ ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

$$b(t,\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)}t^{\alpha-1}(1-t)^{\beta-1},\quad t\in[0,1]$$

$$\Gamma(x)=\int_0^{+\infty}t^{x-1}e^{-t}dt$$ — гамма функция Эйлера

$\boldsymbol{\alpha;\beta}$ — параметры функции плотности бета-распределения

$$M(T)=\frac{\alpha}{\alpha+\beta},\quad D(T)=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

$$\alpha=\frac{\bar{t}}{s^2}\cdot\left(\bar{t}-(\bar{t})^2-s^2\right),\quad \beta=\frac{1-\bar{t}}{s^2}\cdot\left(\bar{t}-(\bar{t})^2-s^2\right)$$

КОЛИЧЕСТВЕННАЯ МЕРА ИНФОРМАЦИОННОЙ ЧУВСТВИТЕЛЬНОСТИ

$$\delta(\gamma, n) = B^{-1}\left(\frac{1}{2} + \frac{\gamma}{2}, \alpha(n), \beta(n)\right) - B^{-1}\left(\frac{1}{2} - \frac{\gamma}{2}, \alpha(n), \beta(n)\right)$$

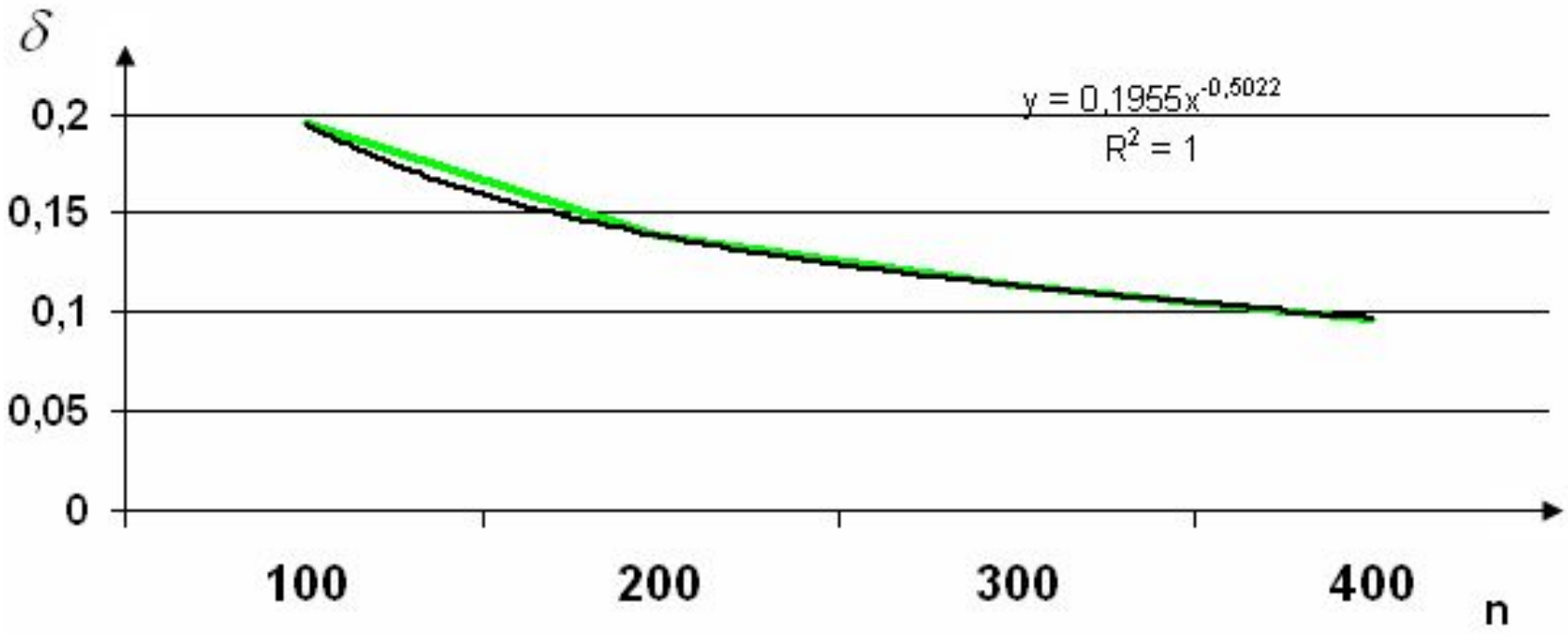
количественная мера информационной чувствительности является функцией от длины входа и вероятности.

γ - доверительная вероятность

$B^{-1}(\gamma, \alpha, \beta)$ - функция, обратная интегралу плотности распределения

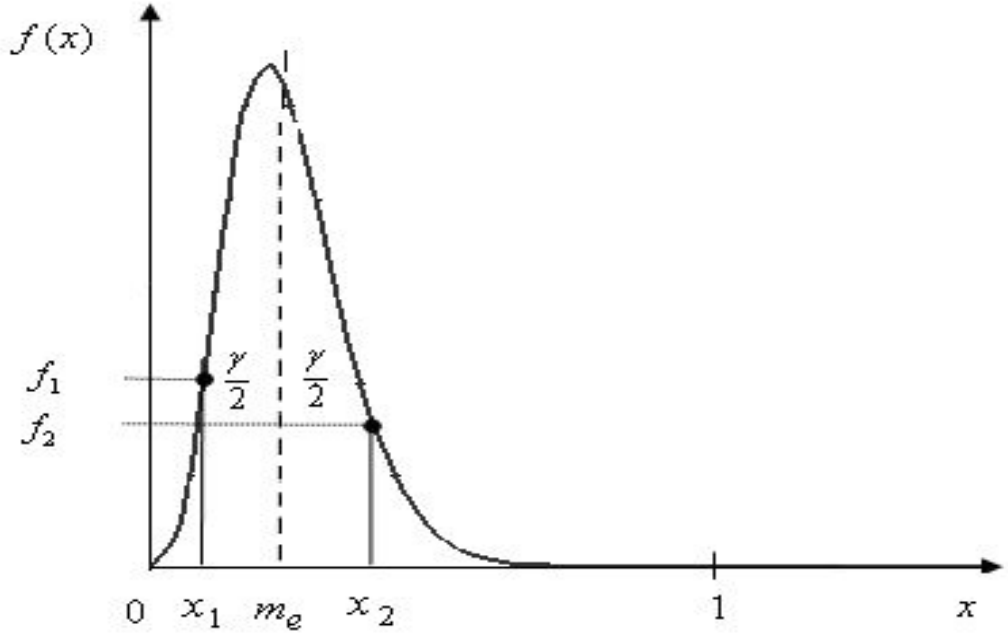
$\alpha(n), \beta(n)$ - параметры бета-распределения

ЗАВИСИМОСТЬ ИНФОРМАЦИОННОЙ ЧУВСТВИТЕЛЬНОСТИ АЛГОРИТМА УМНОЖЕНИЯ ОТ ДЛИНЫ ВХОДА ПРИ γ=0.95

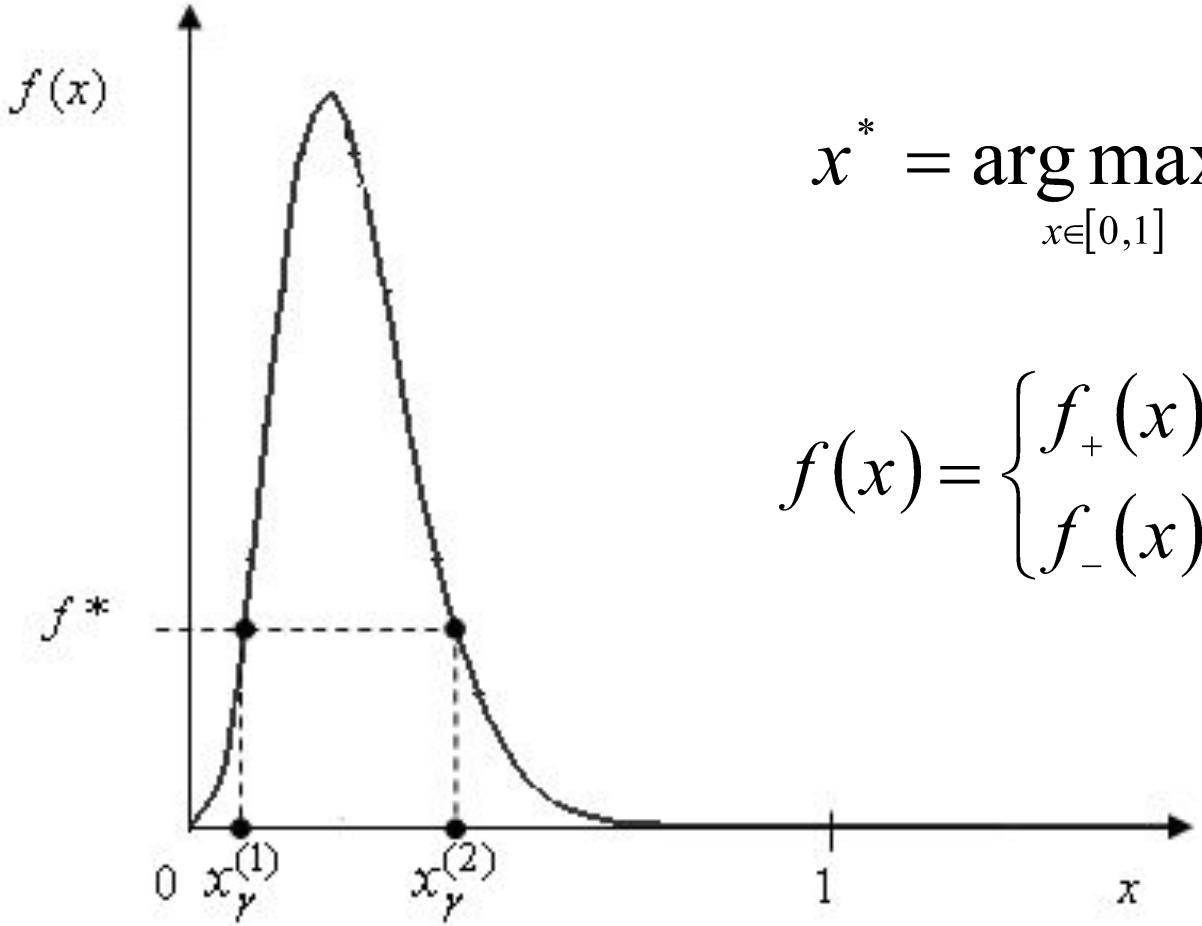


СИММЕТРИЧНАЯ ПО ПЛОТНОСТИ ВЕРОЯТНОСТЕЙ ОЦЕНКА ИНФОРМАЦИОННОЙ ЧУВСТВИТЕЛЬНОСТИ

- Квантильная мера для ассиметричных функций плотности не является симметричной — происходит выбрасывание из сегмента, более вероятных величин в пользу менее вероятных



ИДЕЯ СИММЕТРИЧНОЙ ОЦЕНКИ



$$x^* = \arg\max_{x \in [0,1]} f(x)$$

$$f(x) = \begin{cases} f_+(x), & 0 \le x \le x^* \\ f_-(x), & x^* \le x \le 1 \end{cases}$$

СИММЕТРИЧНАЯ ПО ПЛОТНОСТИ ВЕРОЯТНОСТЕЙ ОЦЕНКА ИНФОРМАЦИОННОЙ ЧУВСТВИТЕЛЬНОСТИ

$$I(s) = F\left(f_-^{-1}(s)\right) - F\left(f_+^{-1}(s)\right), \quad 0 \le s \le f\left(x^*\right)$$

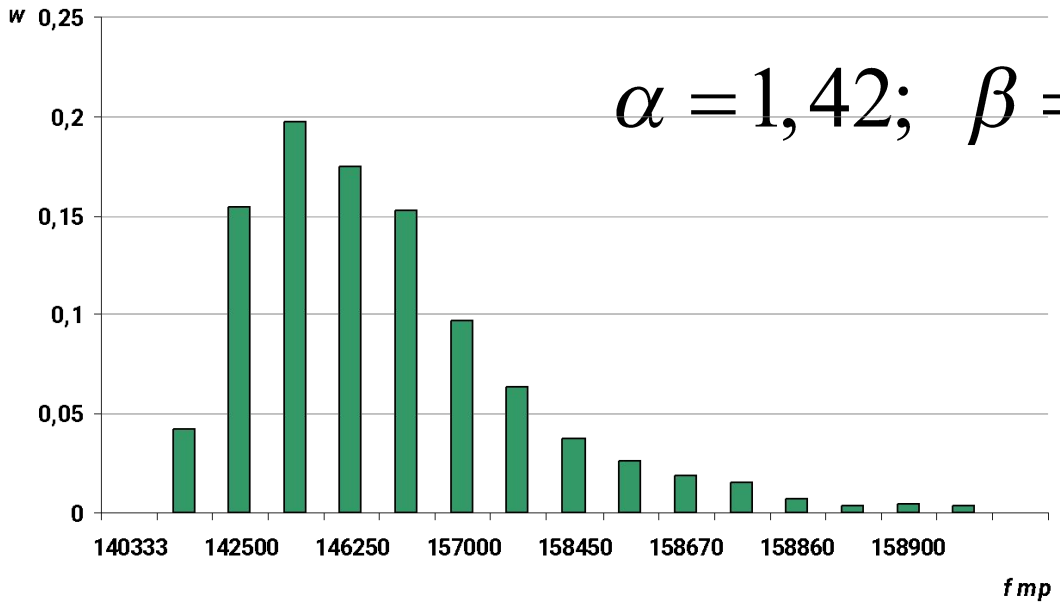
$$x_\gamma^{(1)} = f_+^{-1}\left(f_\gamma\right), \quad x_\gamma^{(2)} = f_-^{-1}\left(f_\gamma\right), \quad f_\gamma : I\left(f_\gamma\right) = \gamma$$

$$\delta_{\mathbf{IP}}(\gamma) = f_-^{-1}\left(I^{-1}(\gamma)\right) - f_+^{-1}\left(I^{-1}(\gamma)\right)$$

$$\delta_{\mathbf{IP}}^{\ *}(\gamma, n) = \left(x_\gamma^{(1)}(n),\ x_\gamma^{(2)}(n),\ x_\gamma^{(2)}(n) - x_\gamma^{(1)}(n)\right)$$

МОДЕЛЬНЫЙ ПРИМЕР

- Результаты экспериментального исследования алгоритма Кнута-Морриса-Пратта для поиска подстроки в строке
- Методом моментов были определены параметры аппроксимирующего бета-распределения



$$\alpha = 1,42; \quad \beta = 18,47$$

МОДЕЛЬНЫЙ ПРИМЕР

Результаты расчётов:

$$\delta_{\mathbf{IQ}}(\gamma) = 0{,}2104 \qquad x_1 = 0{,}0049 \quad x_2 = 0{,}2153$$

$$b(x_1, \alpha, \beta) \approx 0{,}008; \;\; b(x_2, \alpha, \beta) \approx 0{,}0007.$$

$$\delta_{\mathbf{IP}}(\gamma) = 0{,}1785 \qquad x_1 = 0{,}00065 \quad x_2 = 0{,}1792$$

$$b(x_1, \alpha, \beta) = b(x_2, \alpha, \beta) = 0{,}0015$$

$$\delta_{\mathbf{IP}}^{*}(0{,}95, 10000, 20) =$$

$$= \left(x_\gamma^{(1)} = 0{,}00064,\; x_\gamma^{(2)} = 0{,}1792,\; \delta_{\mathbf{IP}}(0{,}95) = 0{,}1785 \right)$$

ЗАКЛЮЧЕНИЕ

- Введённая симметричная по плотности вероятностей количественная оценка информационной чувствительности может быть использована как более достоверная в случае асимметрии аппроксимирующей функции плотности при прогнозировании временной эффективности компьютерных алгоритмов, а также при решении задачи рационального выбора алгоритмов по критерию стабильности по времени, равно как и при решении других задач прикладной теории алгоритмов

ПУБЛИКАЦИИ ПО ТЕМЕ

- 1. **Головешкин В.А., Ульянов М.В.**
  Теория рекурсии для программистов.
  М.: Издательство «Наука ФИЗМАТЛИТ», 2006 г.

- 2. **Ульянов М.В.**
  Ресурсно-эффективные компьютерные алгоритмы.
- Разработка и анализ.
  М.: Издательство «Наука ФИЗМАТЛИТ», 2008 г.

- 3. **Петрушин В.Н., Ульянов М.В.**
  Планирование экспериментального исследования
  трудоёмкости алгоритмов на основе бета-распределения.
  Информационные технологии и вычислительные системы.
  2008. №2. С. 81–91.

- 4. **Петрушин В.Н., Ульянов М.В.**
- Информационная чувствительность компьютерных алгоритмов
- М.: Издательство «Наука ФИЗМАТЛИТ», 2010 г.