
Кинетический подход к анализу нестационарных временных рядов

1. Оптимальный объем выборки и горизонтный ряд
2. Уровень стационарности ряда и разбиение гистограммы
3. Индекс нестационарности временного ряда
4. Индикатор статистической добротности
5. Идентификация текущего распределения
6. Эмпирические уравнения Лиувилля и Фоккера-Планка
7. Генерация пучка траекторий временного ряда

I

Горизонтный ряд и оптимальный объем выборки нестационарного временного ряда

Сходимость выборочной ФР

T1. (Гливенко) Выборочное распределение $F_N(x)$ случайной стационарной величины равномерно по x сходится по вероятности к распределению генеральной совокупности $F(x)$:

$$P \left\{ \lim_{N \rightarrow \infty} \sup_x |F_N(x) - F(x)| = 0 \right\} = 1$$

T2. (Колмогоров) Пусть генеральное распределение $F(x)$ непрерывно. Тогда статистика

$$\sqrt{N} \sup_x |F_N(x) - F(x)|$$

отклонения ВФР от генеральной ФР сходится по вероятности к функции Колмогорова:

$$\lim_{N \rightarrow \infty} P \left\{ 0 < \sqrt{N} \sup_x |F_N(x) - F(x)| < z \right\} = K(z) = \sum_{k=-\infty}^{\infty} (-1)^k \exp(-2k^2 z^2)$$

Трудности анализа нестационарного ряда

- В последовательные моменты времени наблюдаются значения **разных случайных процессов**, поэтому отсутствует понятие **генеральной совокупности**, и связанные с ней предельные теоремы и **статистические критерии**, строго говоря, **не применимы**
- **Эмпирические оценки** вероятностей попадания значений в заданный интервал с увеличением объема выборки **не сходятся** к генеральной совокупности ни в каком смысле
- Если все же требуется оценить эмпирическую вероятность, то по **выборке какого объема** это следует делать, и с **какой точностью** такая оценка будет выполняться на заданном **горизонте прогноза**?

Причины ошибки прогнозирования

$$\langle \Psi[x(t)] \rangle_{\Delta} = \frac{1}{N} \int_{t_0}^{t_0+N} \Psi[x(t)] dt$$

Конечность промежутка N приводит к нерепрезентативности выборки; для уменьшения ошибки следует увеличивать N.

Нестационарность процесса означает изменение статистических свойств ряда на промежутке N; для уменьшения ошибки следует уменьшать N.

Следовательно, существует оптимальный объем выборки, на котором суммарная ошибка прогноза минимальна.

Оценка ошибки прогноза ряда

Ошибка прогнозирования ВПФР в силу нестационарности:

$$\varepsilon(t) = \int_0^1 \left| \hat{f}(x, t) - f(x, t) \right| dx$$

Ошибка прогнозирования среднего значения временного ряда:

$$\left| \hat{\bar{x}} - \bar{x} \right| = \left| \int_0^1 x \hat{f}(x) dx - \int_0^1 x f(x) dx \right| \leq \int_0^1 x \cdot (\hat{f} - f) dx \leq \sup_{[0;1]} x \cdot \int_0^1 |\hat{f} - f| dx = \varepsilon$$

Ошибка прогнозирования собственно временного ряда:

$$\delta^2 = \int_0^1 (x - \hat{x})^2 f(x, t) dx = \int_0^1 (x - \hat{x} + \bar{x} - \bar{x})^2 f(x, t) dx = \sigma^2 + (\bar{x} - \hat{x})^2 \leq \sigma^2 + \varepsilon^2$$

Горизонтный ряд

Расстояние между двумя функциями в L1:

$$\rho(f_1, f_2) = \|f_1(x) - f_2(x)\| = \int |f_1(x) - f_2(x)| dx$$

Функционал близости между двумя ВПФР:

$$V(N, \tau; t) \equiv \rho(f_N(x, t + \tau), f_N(x, t)) = \int |f_N(x, t + \tau) - f_N(x, t)| dx$$

Горизонтным рядом $h(t, \tau; \varepsilon)$ называется такой минимальный объем выборки, что при всех $N \geq h(t, \tau; \varepsilon)$ выполнено условие

$$V(N, \tau; t) \leq \varepsilon$$

Равномерная по времени оценка на функционал нормы:

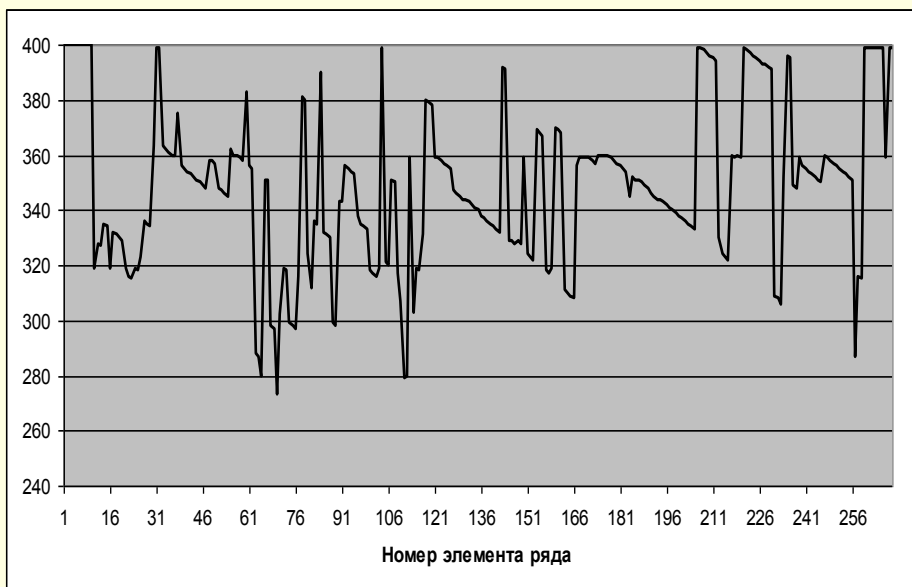
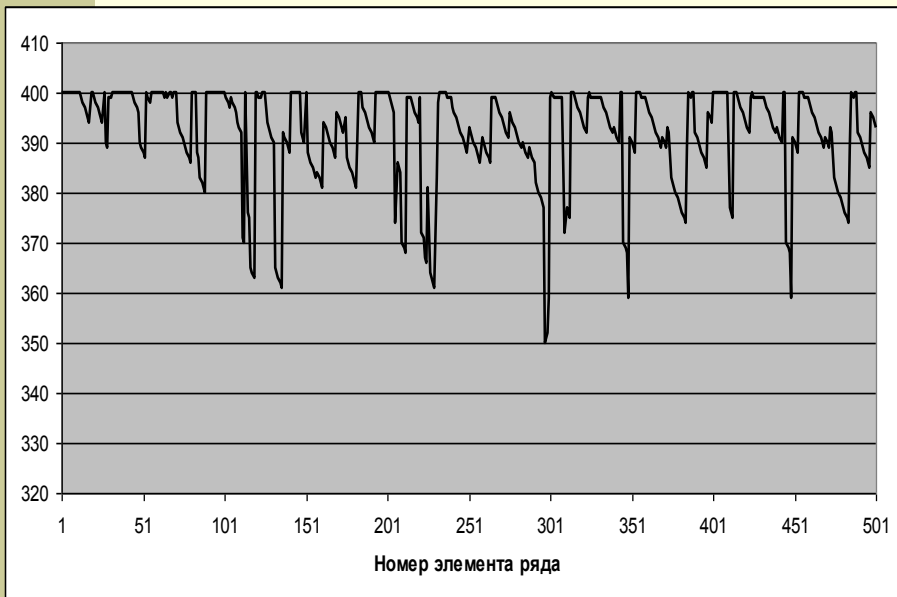
$$0 \leq V(N, \tau; t) \leq \min(2\tau / N; 2)$$

Оптимальный объем выборки

- Если ВПФР ряда $x(t)$ является ε -стационарной, то ВПФР горизонтного ряда $h(t, \tau; \varepsilon)$ также ε -стационарна.
- Относительная дисперсия распределения горизонтного ряда есть величина порядка $O(\varepsilon)$.
- Минимально достаточным объемом выборки для прогноза ВПФР на горизонт τ с точностью ε называется величина
$$h(\tau, \varepsilon) = \max\{ h(t, \tau; \varepsilon) \}.$$
- Оптимальным объемом N для прогнозирования ряда называется величина, доставляющая минимум функционала оценки ошибки прогноза с учетом нестационарности ВПФР:

$$\sigma(h, t)^2 + V^2(h, \tau; t) \rightarrow \min$$

Примеры горизонтных рядов



Горизонтные ряды: для стационарного процесса и для нестационарного (суточного курса евро/доллар) при сдвиге на 10 шагов и точности 0,05

Распределение горизонтного ряда

при больших значениях сдвига τ по времени

Рассмотрим последний промежуток длины $2/\varepsilon$, в котором изменяются значения стационарного горизонтного ряда:

$$\frac{2(\tau - 1)}{\varepsilon} < h(t, \tau; \varepsilon) \leq \frac{2\tau}{\varepsilon}$$

Пусть **A(k)** есть условная вероятность того, что значение горизонтного ряда не превосходит числа **m = [2τ/ε] + τ - k** при условии попадания исходной случайной величины в промежутки

$$\Delta_{i_1}, \Delta_{i_2}, \dots, \Delta_{i_\tau}$$

Вероятность попадания в эти промежутки равна

$$a(i_1, \dots, i_\tau) = \sum_{k=1}^{\tau} p_{i_k} - \sum_{\substack{k, l=1 \\ k \neq l}}^{\tau} p_{i_k} p_{i_l} + \dots + (-1)^{\tau-1} \prod_{k=1}^{\tau} p_{i_k}$$

В стационарном случае условная вероятность $A(k)$ равна

$$A_k = 1 - \frac{Q(1)}{2\pi i} \int_{\Pi} \frac{z^k dz}{Q(z) \cdot (z-1)}$$

где контур Π охватывает все нули характеристического многочлена

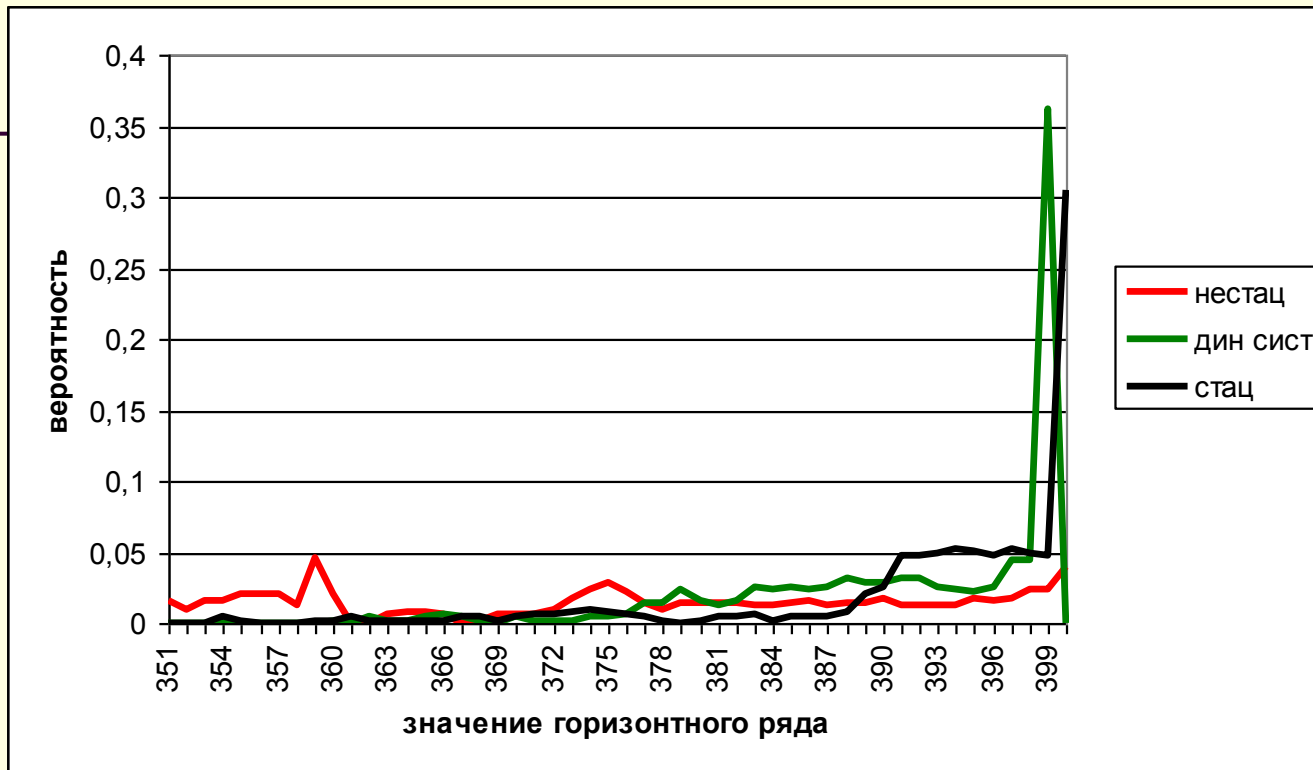
$$Q(\lambda) = \frac{\lambda^{\tau+1} - \lambda^{\tau} + a(1-a)^{\tau}}{\lambda - (1-a)}$$

Следствие 1. Вероятность максимального значения

$$q_{n,\varepsilon}([2n/\varepsilon]) = \left(\sum_i (1-p_i)^n p_i \right)^n$$

Следствие 2. Распределение горизонтного ряда в последней полосе ширины $[2/\varepsilon]$ является равномерным, т.е. не зависит от $m=[2\tau/\varepsilon]+\tau-k$.

Аспекты применения горизонтного ряда



- Распознавание типа случайного процесса
- Индикация хаотизации и консолидации и определение периодов «хаос-порядок»
- Определение характерного объема выборки для прогнозирования распределения с заданной точностью
- Индикация разладки как большой скачок горизонтного ряда

II

Оптимальное равномерное разбиение гистограммы

Оптимальное число n интервалов разбиения

- **Sturges, 1926** $n = 1 + \log_2 N$
- **Mann, Wald, 1942** $n = 4 \left(\frac{3}{4} (N - 1)^2 \right)^{1/5}$
- **Смирнов, 1950** $n = AN^{1/3}$
- **Алексеева, 1975** $n = \frac{1 + \kappa}{6} N^{2/5}$
- **Лернер, 1976** $n = \left(\frac{N |f''(x)|_{\max}}{4 f(x)_{\max}} \right)^{1/5}$
- **Scott, 1979** $n = \left(\frac{N}{6} \int_{-\infty}^{+\infty} f'^2(x) dx \right)^{1/3}$

Точность оценки вероятности по эмпирической частоте

Пусть выбрано равномерное разбиение на n интервалов.
Тогда эмпирическая вероятность по выборке длины N равна

$$f_N(j) = \frac{k_j}{N}, \quad j = 1, 2, \dots, n$$

Дисперсия выборочной вероятности равна

$$s_N^2(j) = f_N(j) \cdot (1 - f_N(j))$$

Если каждая вероятность оценивается с точностью ε , то
выборочное распределение ε -стационарно:

$$\sum_{j=1}^n |f(j) - f^*(j)| \leq \varepsilon \sum_{j=1}^n f^*(j) = \varepsilon$$

Статистика Стьюдента

Если изучаемое распределение стационарно, то статистика

$$t = \sqrt{N-1} \frac{f_N(j) - f^*(j)}{s_N(j)}$$

имеет распределение Стьюдента $\varphi(t, N-1)$ с $N-1$ степенями свободы:

$$\varphi(t; N) = \frac{\Gamma\left(\frac{N+1}{2}\right)}{\sqrt{\pi N} \Gamma(N/2)} \left(1 + \frac{t^2}{N}\right)^{-\frac{N+1}{2}}$$

При больших N распределение Стьюдента сходится к нормальному:

$$\lim_{N \rightarrow +\infty} \varphi(t; N) = \frac{1}{\sqrt{2\pi}} \exp\left(-t^2/2\right)$$

Доверительный интервал оценки вероятности

Интервальная оценка частоты попадания в j -ый классовой промежуток на уровне значимости α имеет вид

$$\left| f^*(j) - f_N(j) \right| \leq t_{1-\alpha/2}(N-1) \frac{s_N(j)}{\sqrt{N}}$$

где $t_\gamma(N)$ есть γ -квантиль распределения $\varphi(t; N)$ (приблизительно квантиль нормального распределения).

Поскольку уровень значимости равен принимаемой ошибке в оценке частоты, то $\alpha = \varepsilon$. Тогда интегральная неопределенность в распределении заведомо не превосходит ε , если

$$t_{1-\varepsilon/2} \frac{s_N(j)}{\sqrt{N}} \leq \varepsilon f^*(j)$$

Согласованная точность оценки

- Если $|f_N(j) - f^*(j)| \leq \varepsilon f^*(j)$, то на согласованном уровне значимости

выполнено неравенство $t_{1-\varepsilon/2} \frac{s_N(j)}{\sqrt{N}} \leq \varepsilon f^*(j)$

Введем средневзвешенную точность:

$$t_{1-\varepsilon/2} = \frac{1}{\Sigma_N(n)} \sum_{j=1}^n s_N(j) t_{1-\varepsilon/2}$$

$$\Sigma_N(n) = \sum_{j=1}^n s_N(j) = \sum_{i=1}^n \sqrt{f_N(j)(1-f_N(j))}$$

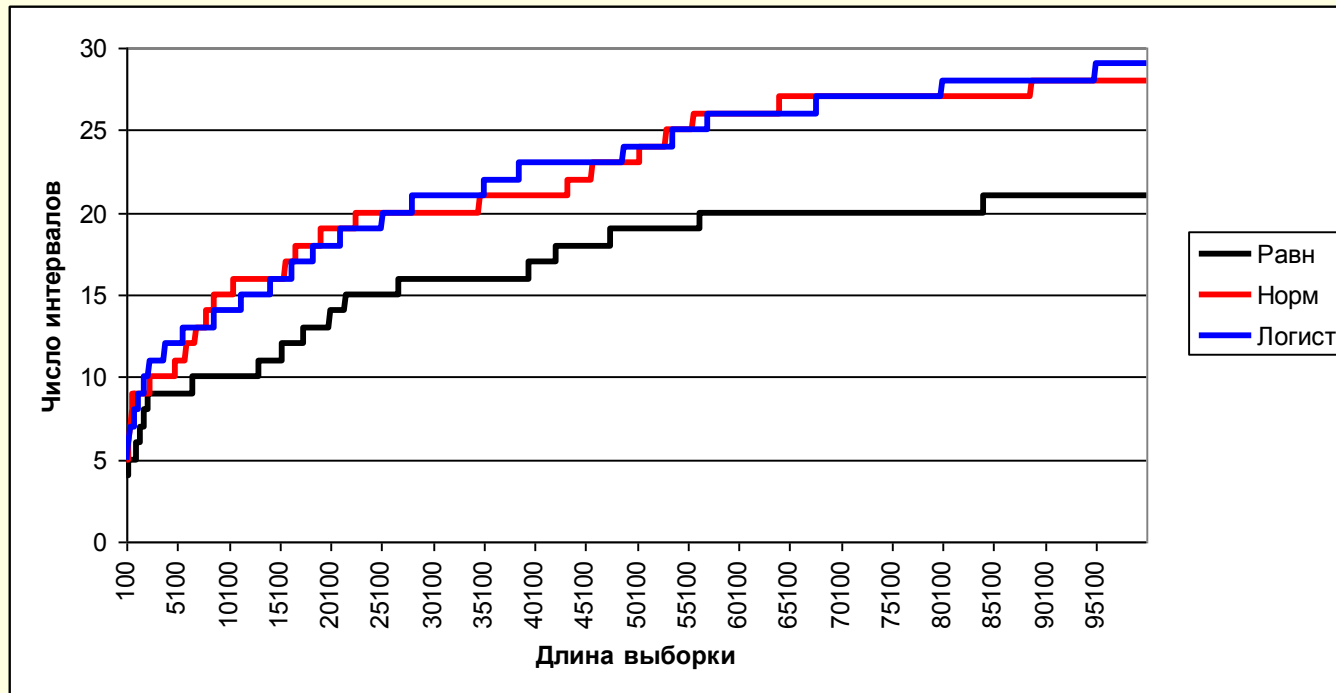
Интегральная точность оценки распределения вероятностей:

$$\frac{t_{1-\varepsilon/2}}{\varepsilon} \leq \frac{\sqrt{N}}{\Sigma_N(n)}$$

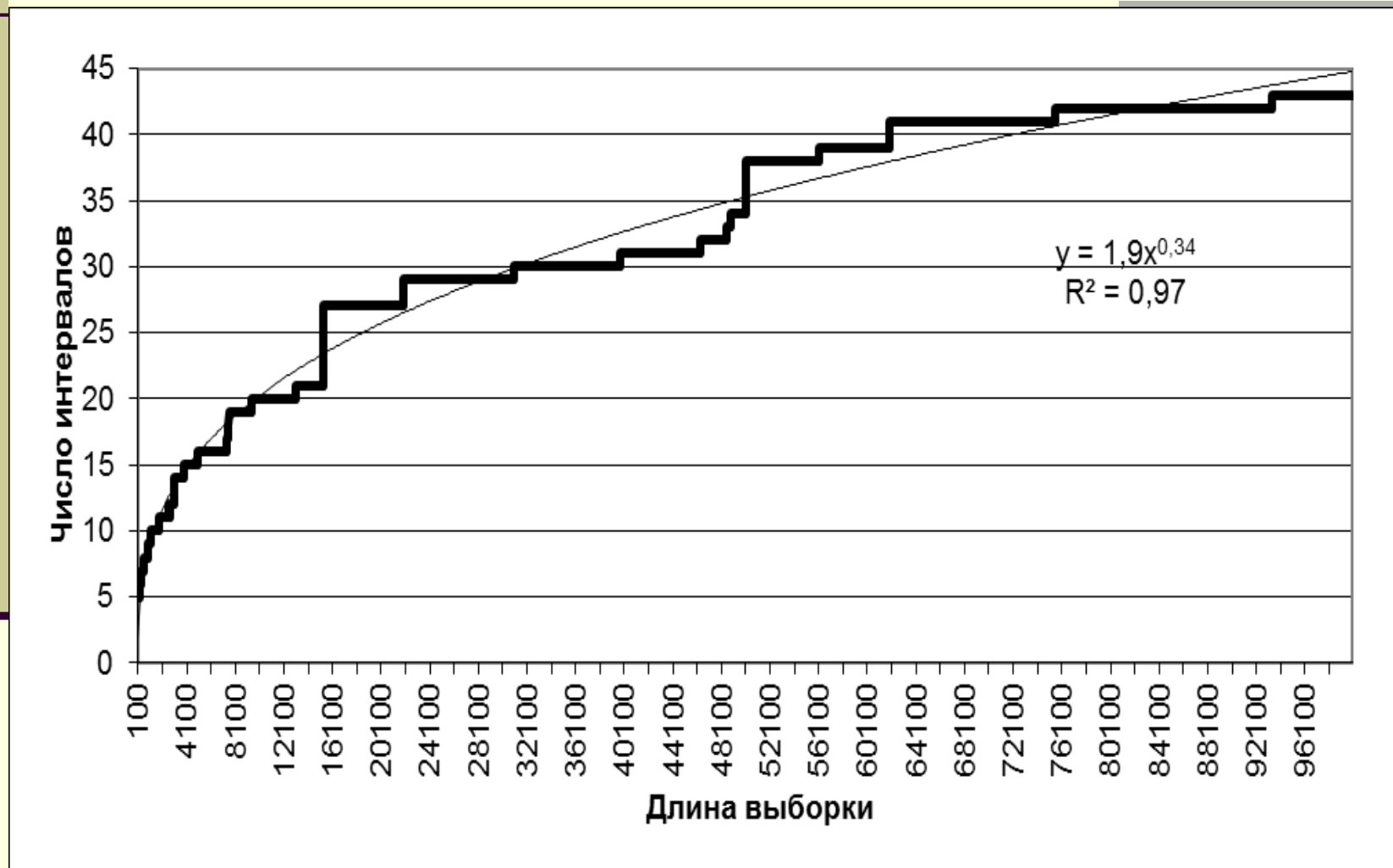
Выбор числа интервалов

- Точность оценки распределения совпадает с точностью измерения случайной величины. Тогда

$$\varphi(\varepsilon) = \frac{t_{1-\varepsilon}}{\varepsilon}, \quad \psi = \varphi^{-1}, \quad \varepsilon = \frac{1}{n} = 2\psi \left(\frac{2\sqrt{N}}{\Sigma_N(n)} \right)$$



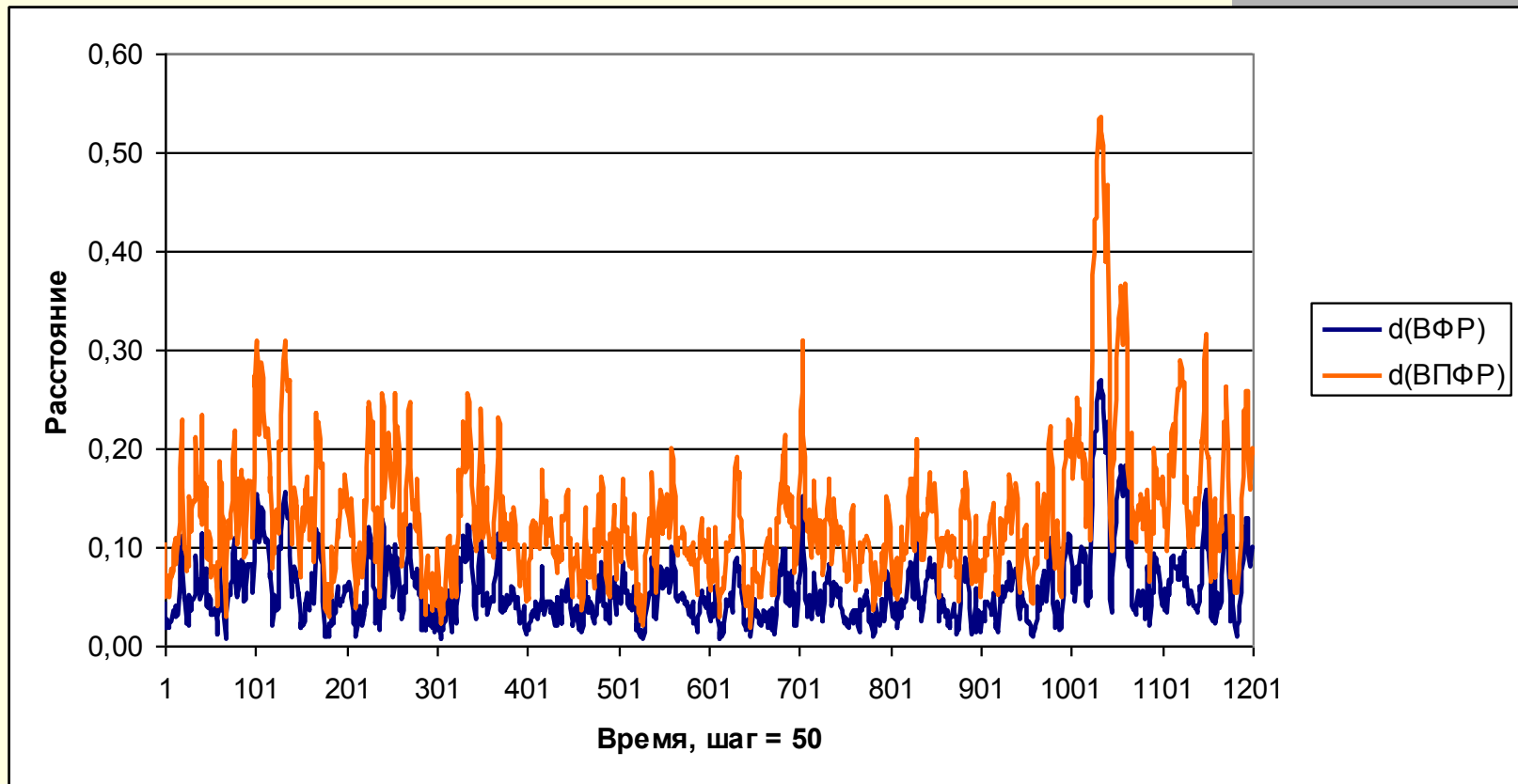
Гистограмма для нестационарного ряда



III

Индикаторы разладки:
согласованный уровень
стационарности,
индекс нестационарности
и индикатор добротности

Ряды расстояний между выборками



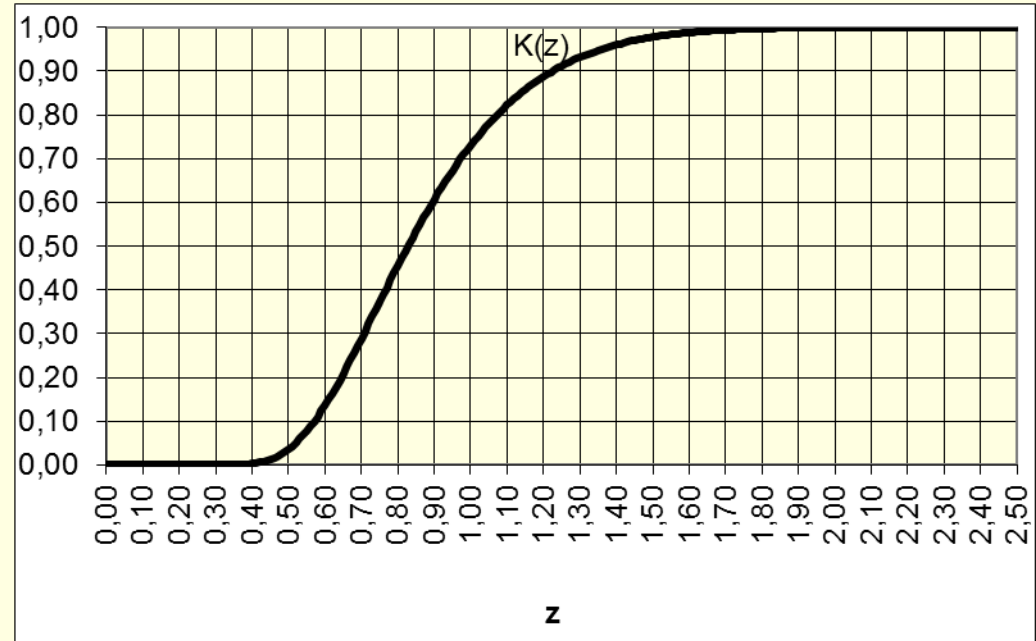
Расстояния в нормах C и $L1$ между выборками длиной 1000 данных из нормального распределения

Непараметрическое сравнение выборок

Пусть случайные величины имеют непрерывное стационарное распределение и являются независимыми. Тогда вероятность того, что две выборки объема N различаются между собой в норме S менее, чем на ε , равна

$$K(\varepsilon\sqrt{N/2})$$

$$S_N = \sup_x |F_{1,N}(x) - F_{2,N}(x)|$$



$$\lim_{N \rightarrow \infty} P \left\{ 0 < \sqrt{\frac{N}{2}} S_N < z \right\} = K(z)$$

Глава 2

Согласованный уровень значимости

На каком расстоянии следует отцепить «хвост» распределения расстояний между распределениями, чтобы оставшийся квантиль был бы равен эмпирически наблюдаемому уровню доверия в задаче распознавания «своих» выборок?

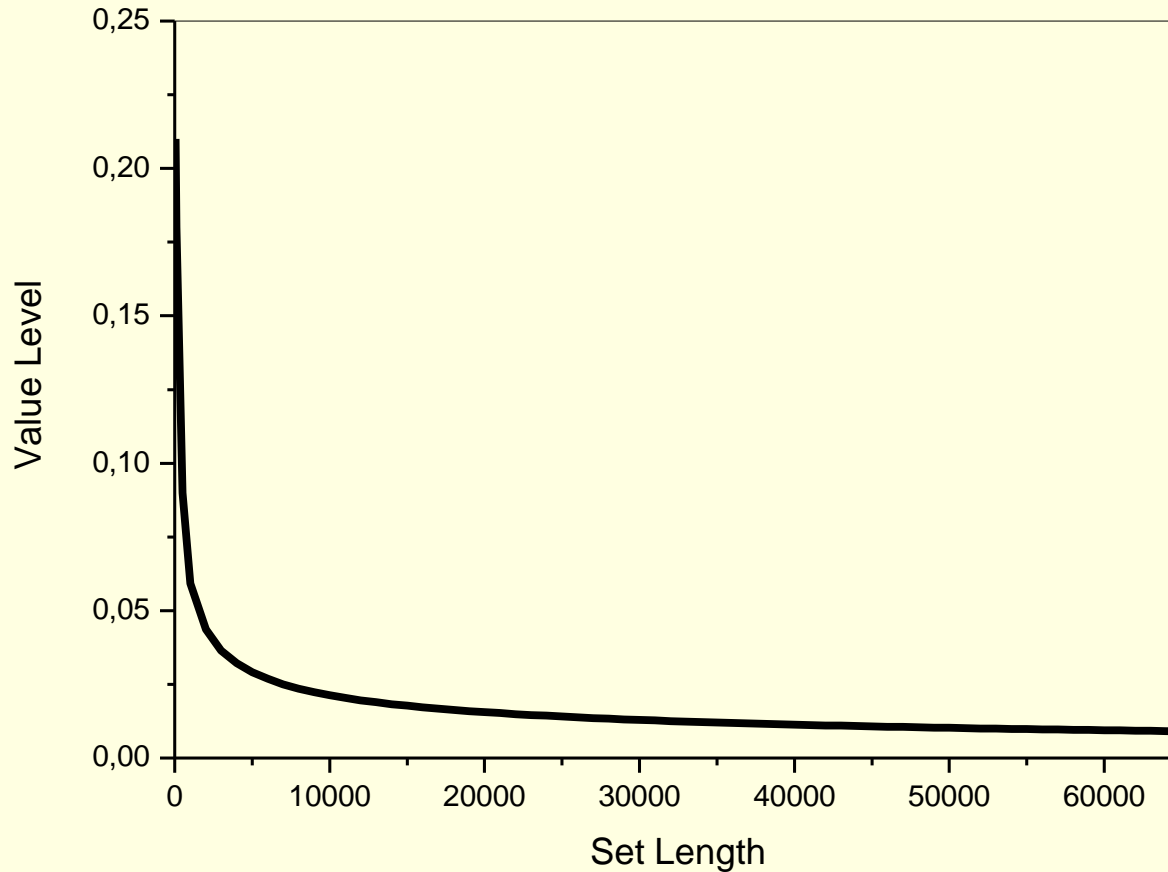
Известно, что если СВ r (это расстояние между выборками) имеет ФР $F(x)$, то СВ $u=F(r)$ (в нашей задаче это уровень значимости) имеет равномерное распределение на $[0;1]$. Следовательно, согласованный с экспериментом уровень значимости как квантиль равномерно распределенной СВ есть функция, линейно зависящая от расстояния между выборками, т.е. $\alpha = \varepsilon$.



В норме S две выборки длины N , расстояние между которыми ε , различны на уровне значимости α , если

$$1 - K\left(\sqrt{\frac{N}{2}}\varepsilon\right) < \alpha$$

Уровень стационарного шума в зависимости от длины N выборки



$$1 - \varepsilon = K \left(\sqrt{\frac{N}{2}} \varepsilon \right)$$

Согласованный уровень значимости: доля расстояний, превышающих его, равна уровню значимости критерия

Расстояние между нестационарными ВПФР

$$\rho(N, \tau; t) = \|f_N(x, t) - f_N(x, t + \tau)\| = \sum_{i=1}^n |f_N(i, t) - f_N(i, t + \tau)|$$

$$\rho \in \{0; 2/N; 4/N; \dots; 2\tau/N\}$$

Функция распределения расстояний между распределениями:

$$g_{N, \tau}(k) = P\{\rho = 2k/N\}, \quad G_{N, \tau}(\rho) = \sum_{l=0}^{k(\rho)} g_{N, \tau}(l), \quad k(\rho) = [N\rho/2]$$

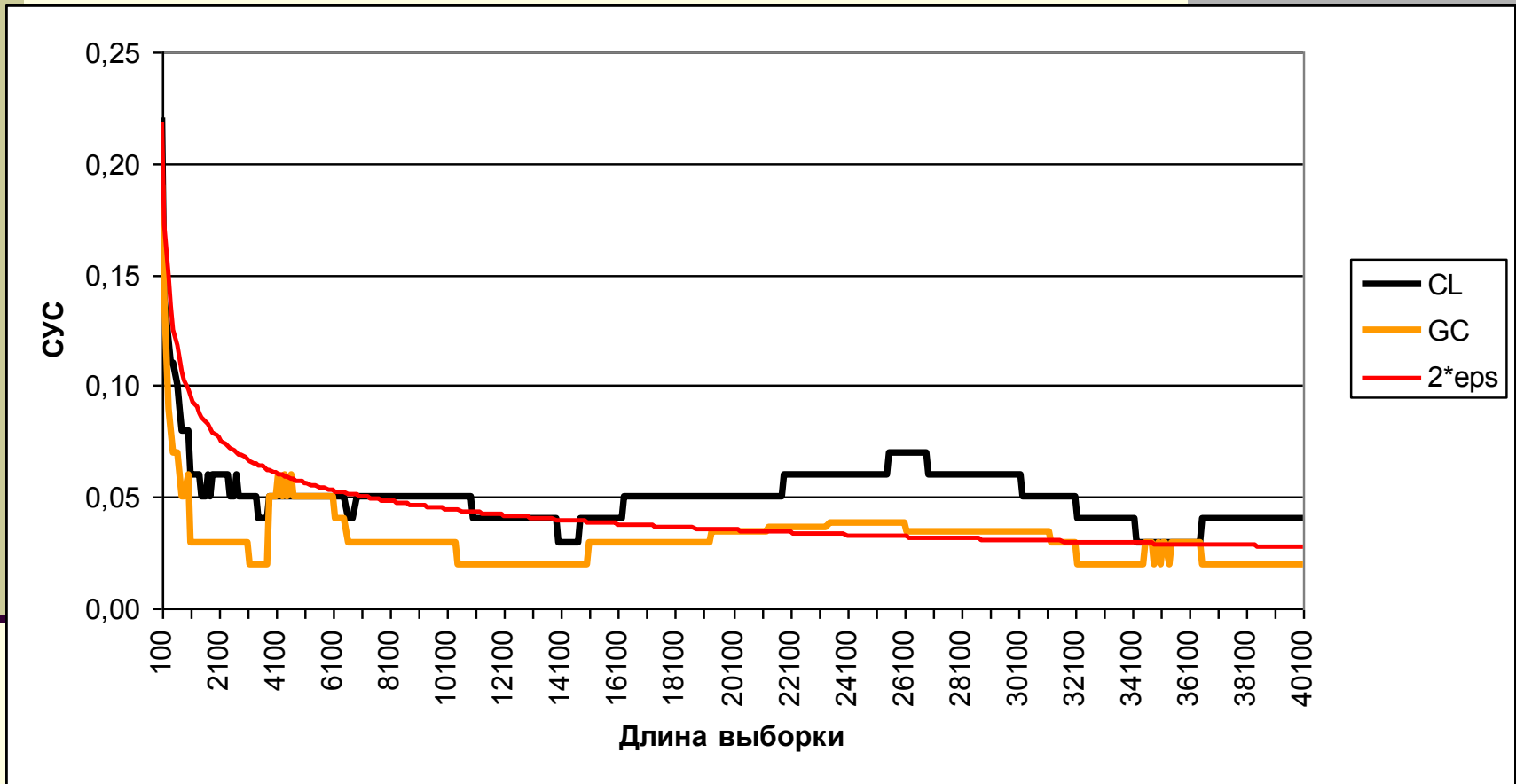
согласованный уровень стационарности (СУС): $\rho^*(N, \tau)$

$$G_{N, \tau}(\rho^*) = 1 - \rho^* \frac{N}{2\tau}$$

стационарность: $\rho^* \leq 2\varepsilon = \frac{2}{n}$

нестационарность: $\rho^* > 2\varepsilon = \frac{2}{n}$

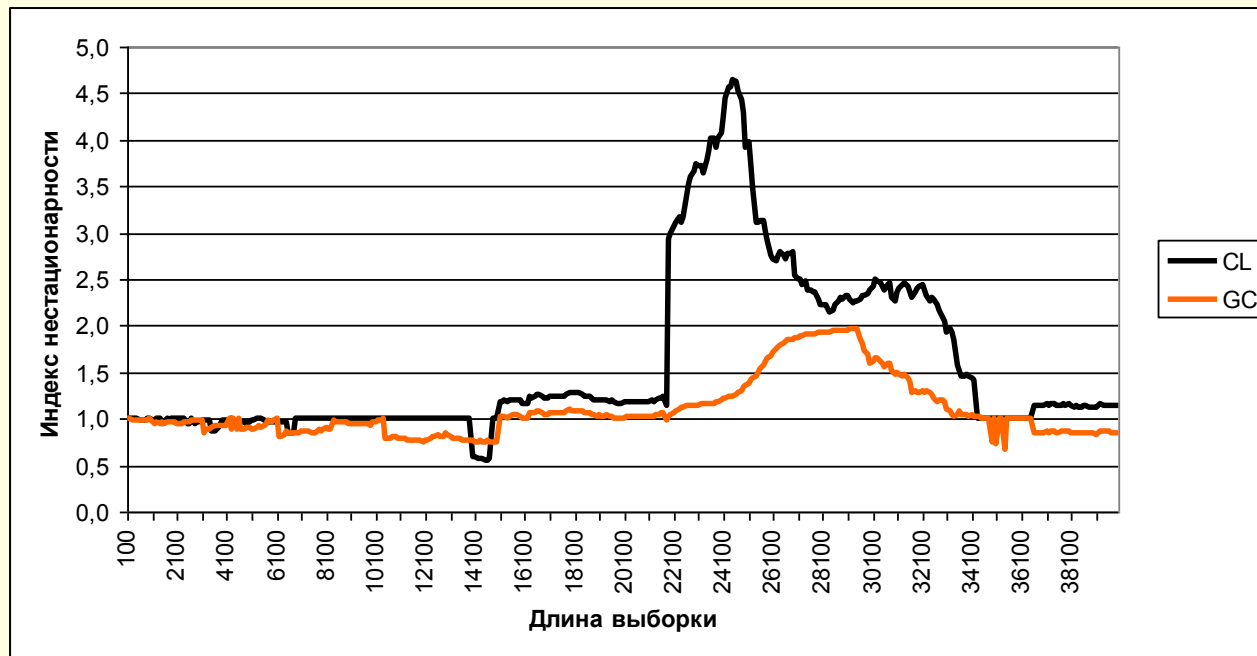
Расстояние между встык-выборками



Индекс нестационарности

Отношение доли расстояний, не превосходящих СУС, к доле расстояний, лежащих в области шума гистограммы:

$$J(N, \tau) = \frac{G_{N, \tau}(\rho^*)}{G_{N, \tau}(2\varepsilon^* \tau / N)}$$



Индикатор добротности

Отношение «сигнал-шум»: доля выборок, расстояние между которыми превосходит СУС, сравнивается с долей выборок, для которых расстояние не превосходит стационарного шума

$$q(N, \tau) = \frac{\rho^* N / (2\tau)}{G_{N, \tau}(2\varepsilon^* \tau / N)} = \frac{1 - G_{N, \tau}(\rho^*)}{G_{N, \tau}(2\varepsilon^* \tau / N)}$$

Оптимальная длина сканирования находится как максимум индикатора добротности; на ней наиболее эффективно определяется разладка как превышение уровня СУС:

$$Q(N, \tau) = q(N, \tau) \left(1 - \rho^* N / (2\tau)\right) = \frac{\left(1 - G_{N, \tau}(\rho^*)\right) \cdot G_{N, \tau}(\rho^*)}{G_{N, \tau}(2\varepsilon^* \tau / N)} \rightarrow \max$$

Оптимальная длина сканирования

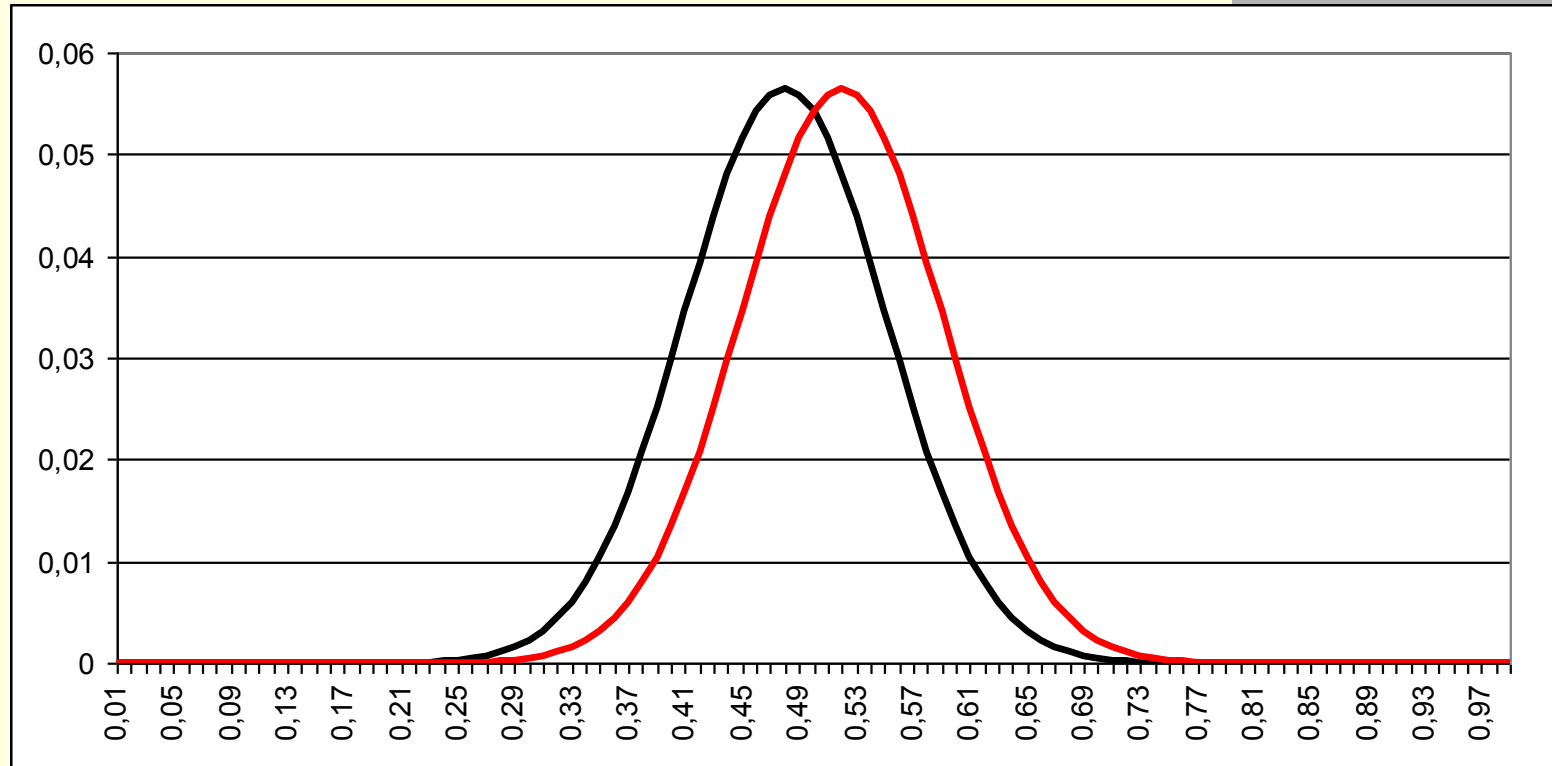


Здесь разладку лучше всего искать на длине выборки 15 тыс. тиков. Разладка трактуется как превышение СУС для пары встык-выборок.

IV

Распознавание типа распределения в различных нормах

Идентификация выборок из двух близких эталонов



Как зависит точность распознавания выборки из близких генеральных совокупностей от длины выборки и функции расстояния между ними?

Функции расстояния

Норма С для ФР или ПФР:

$$\rho_C^f(n) = \max_{x \in [0;1]} |f_1(x;n) - f_2(x;n)|; \quad \rho_C^F(n) = \max_{x \in [0;1]} |F_1(x;n) - F_2(x;n)|$$

Норма L1 для ФР или ПФР:

$$\rho_{L1}^f(n) = \int_0^1 |f_1(x;n) - f_2(x;n)| dx; \quad \rho_{L1}^F(n) = \int_0^1 |F_1(x;n) - F_2(x;n)| dx$$

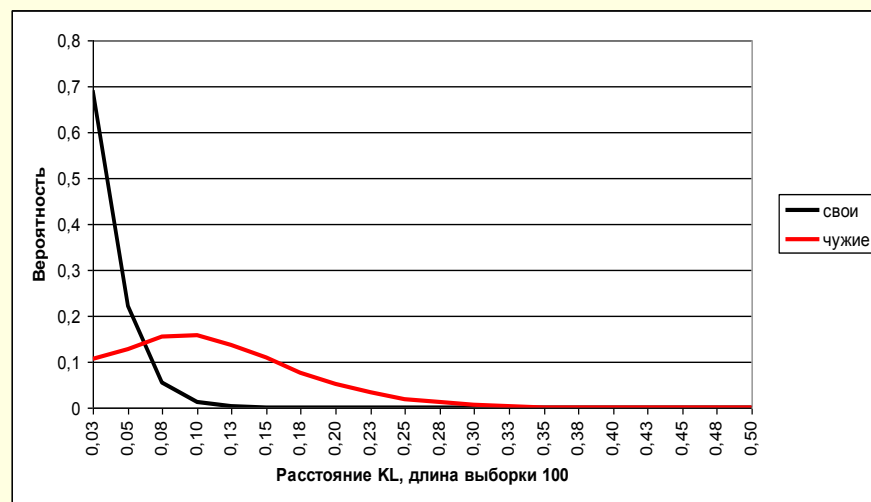
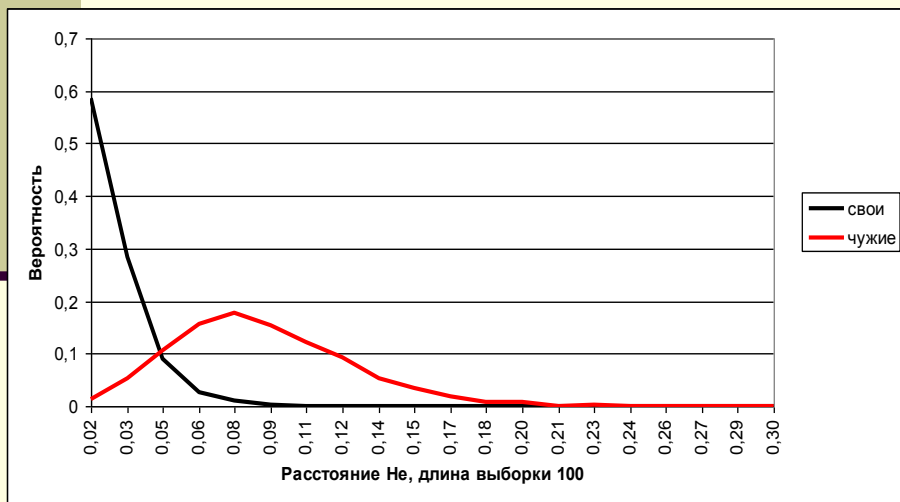
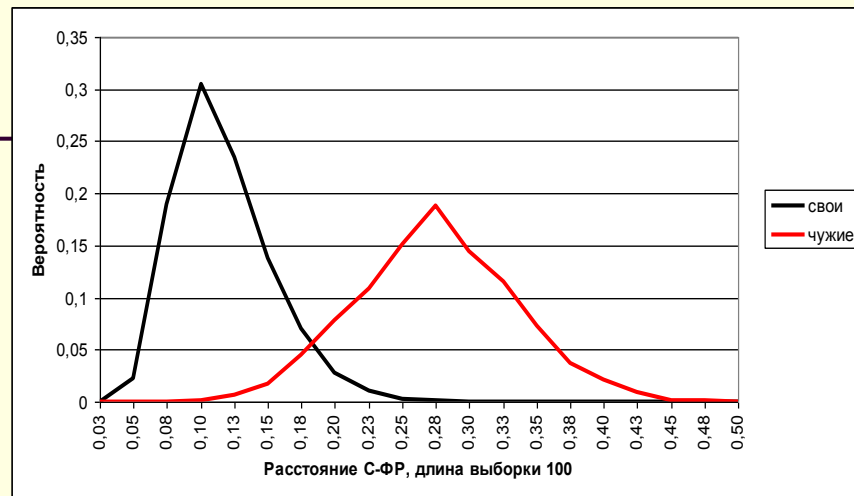
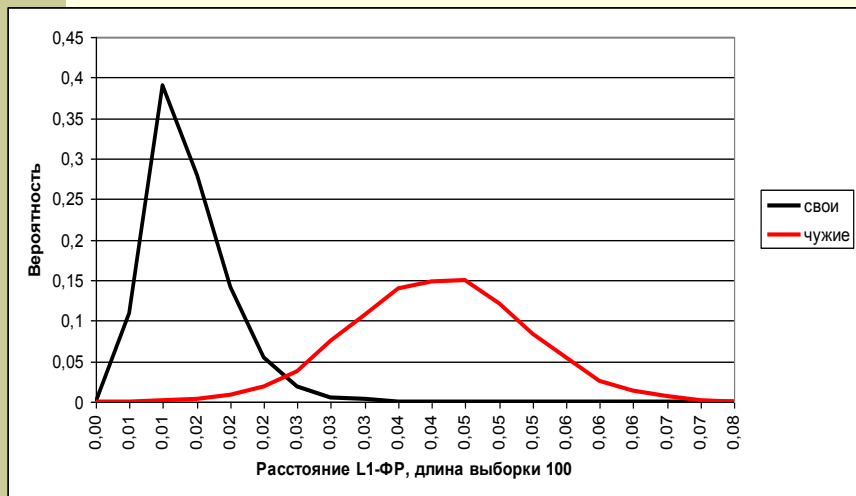
Расстояние Хеллингера для ПФР:

$$\rho_{HE}^f(n) = \int_0^1 \left(\sqrt{f_1(x;n)} - \sqrt{f_2(x;n)} \right)^2 dx$$

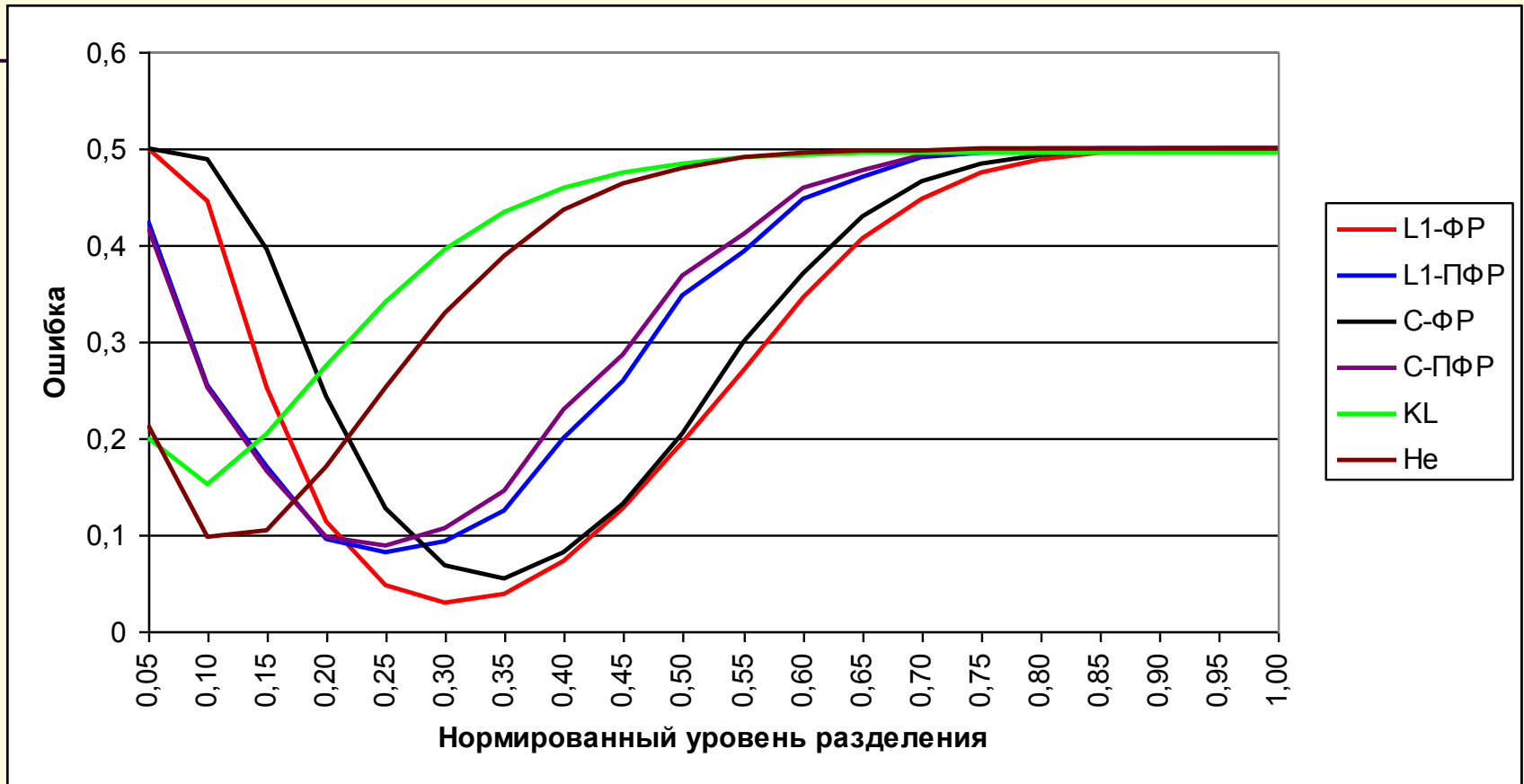
Квазирасстояние Кульбака-Лейблера для ПФР:

$$\rho_{KL}^f(n) = \int_0^1 f_1(x;n) \ln \left(\frac{f_1(x;n)}{f_2(x;n)} \right) dx$$

Распределения расстояний

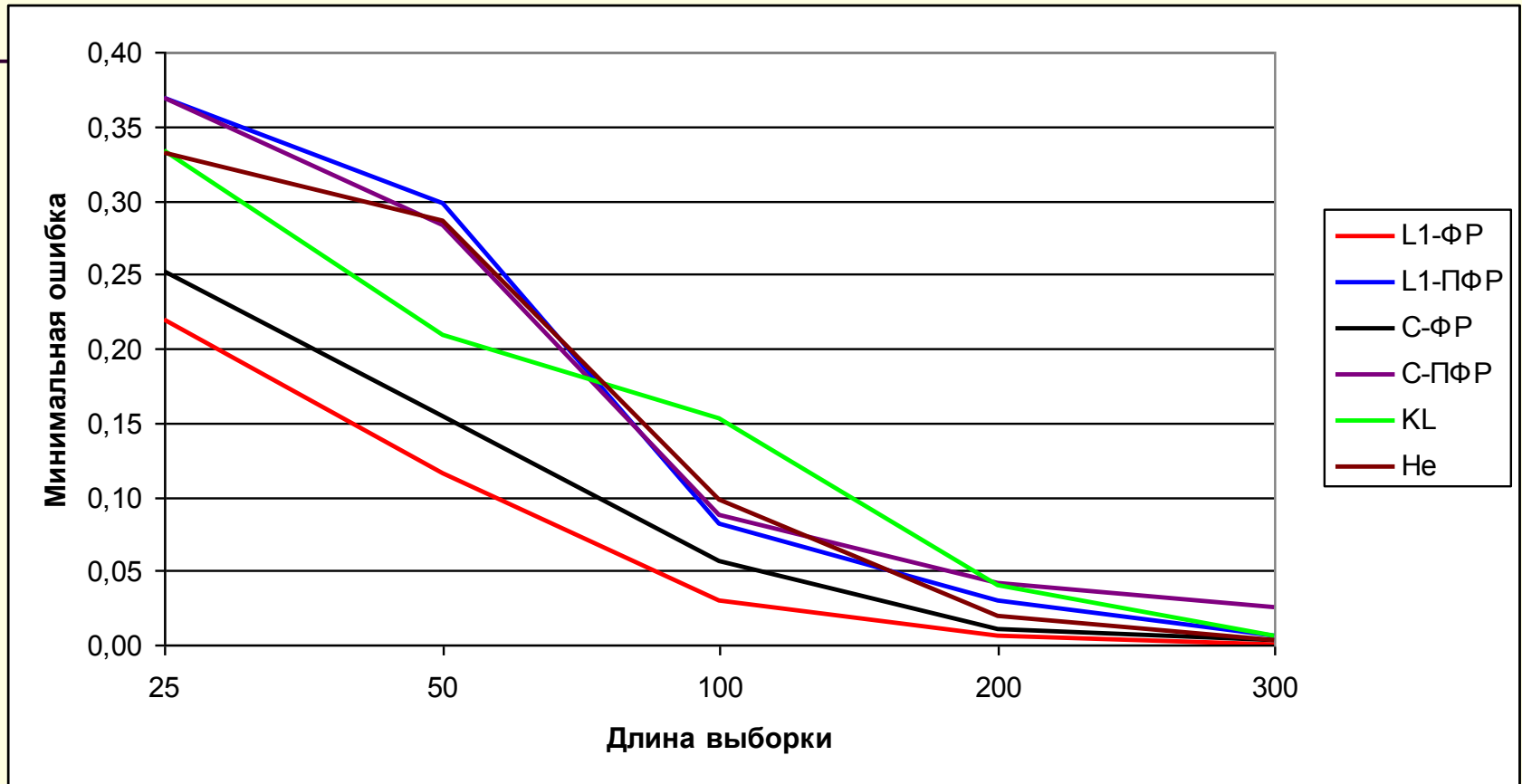


Ошибка идентификации



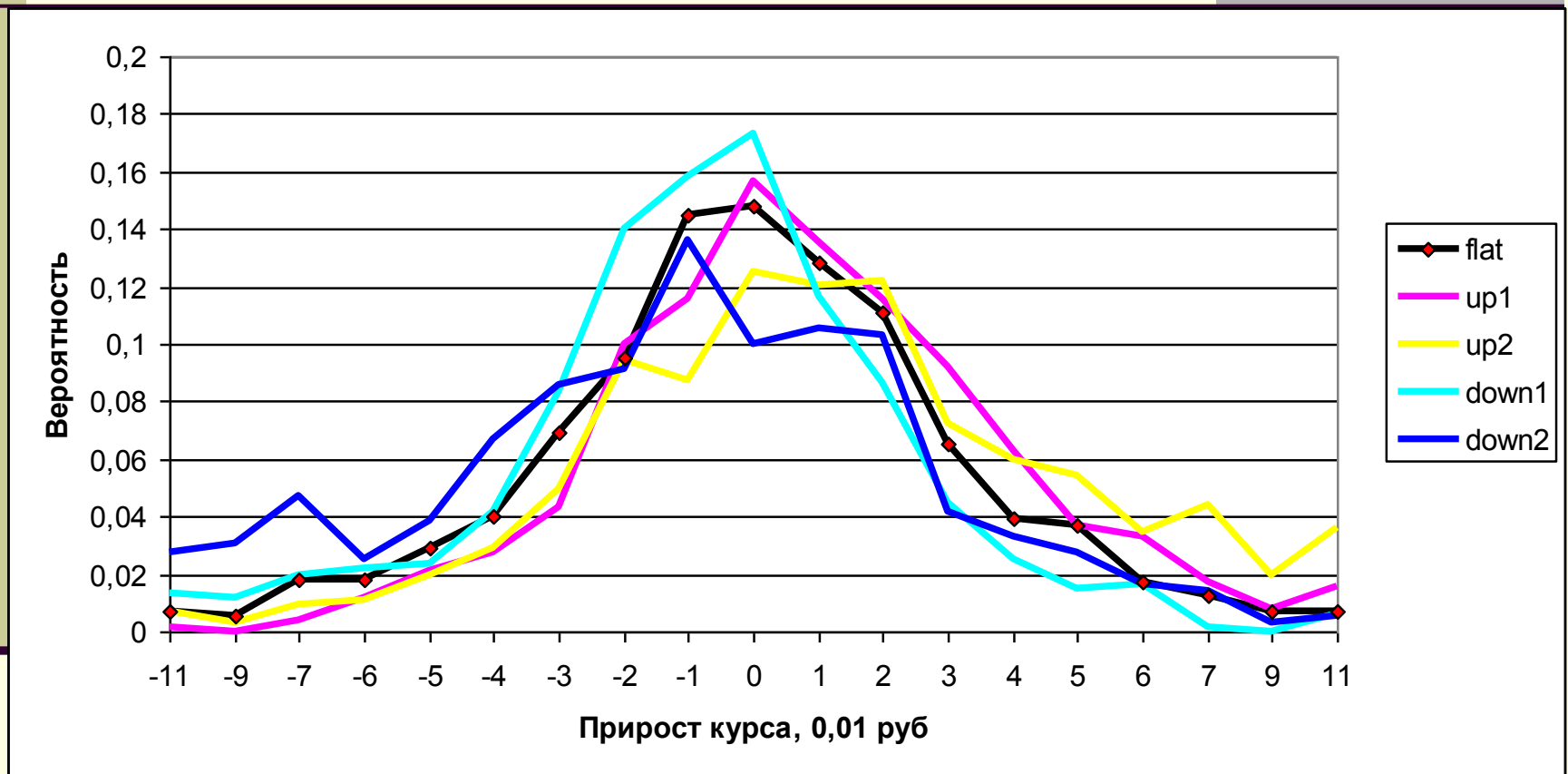
Длина выборки 100. Наилучшей является норма L1 для интегральных функций распределения.

Минимальная ошибка

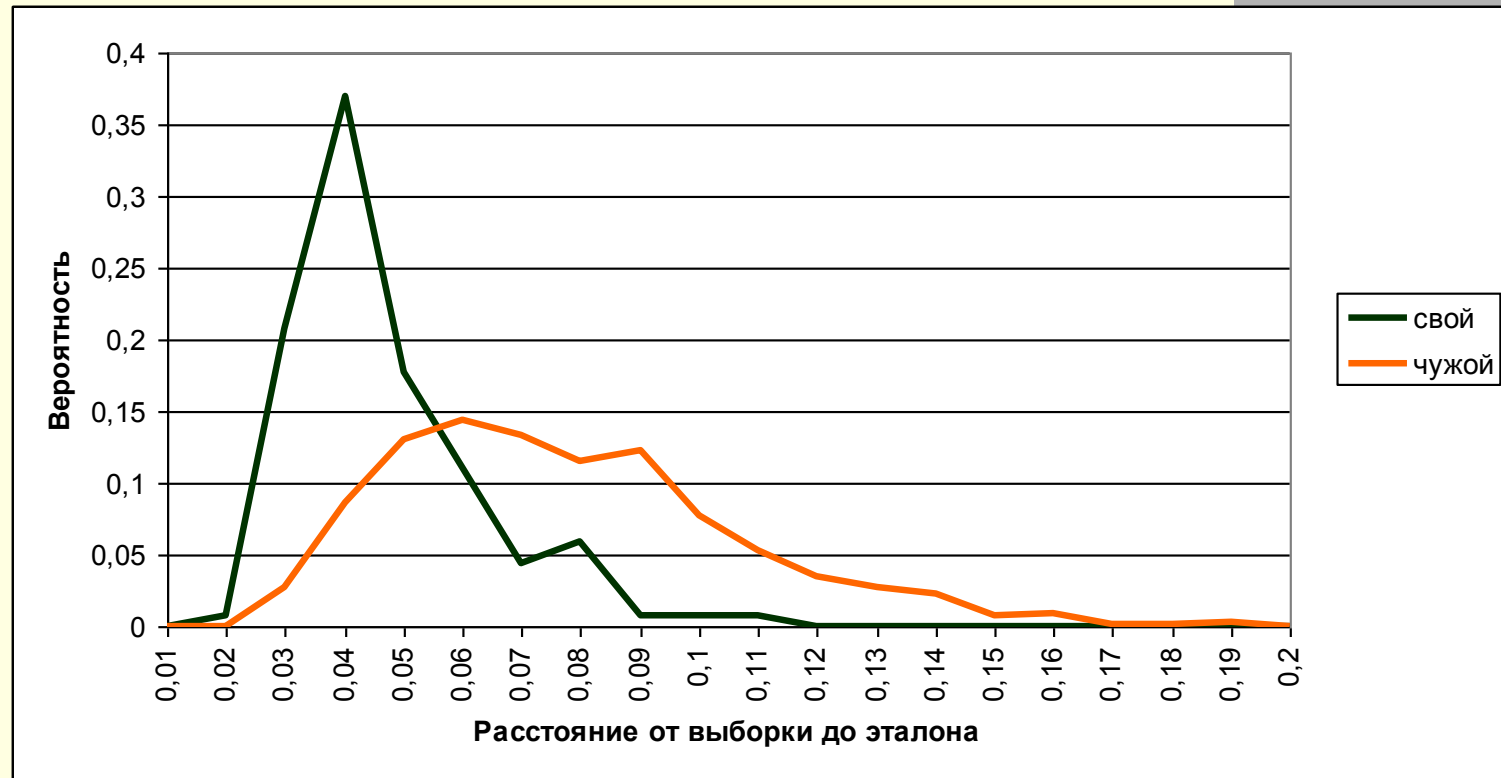


Выборки длины 60 различаются с ошибкой 10%, длины 120 – с ошибкой 2,5%, длины 180 – с ошибкой 1%.

Практические эталоны



Распределения расстояний до эталонов в норме L1-F, длина выборки 30



Выборки длины 30 различаются с ошибкой 20%, длины 60 – с ошибкой 10%, длины 120 – с ошибкой 1%.

V

Эмпирическое уравнение
эволюции ВПФР:
уравнение Лиувилля

Фазовое пространство системы, ассоциированной с временным рядом

- По бесконечному набору значений $x(t)$ можно построить ряды производных $\dot{x}(t) = x(t+1) - x(t)$, $\ddot{x}(t) = \dot{x}(t+1) - \dot{x}(t)$, ... и ввести фазовое пространство с мерой

$$d\Gamma = F_{\infty} \left(x, \dot{x}, \ddot{x}, \dots, x^{(k)}, \dots; t \right) \prod_{k=0}^{\infty} dx^{(k)}$$

- Частичные многомерные ПФР определяются затем по формуле

$$\begin{aligned} f_n \left(x, \dot{x}, \dots, x^{(n)}; t \right) &= \int F_{\infty} \left(x, \dot{x}, \ddot{x}, \dots, x^{(k)}, \dots; t \right) \prod_{k=n+1}^{\infty} dx^{(k)} = \\ &= \int f_{n+1} \left(x, \dot{x}, \dots, x^{(n)}, x^{(n+1)}; t \right) dx^{(n+1)}. \end{aligned}$$

Модель эволюции ВПФР

- Не предполагая, что ВПФР отвечает какой-либо дискретной динамической системе, построим оператор эволюции ВПФР, сохраняющий ее нормировку. Для этого рассмотрим совместное выборочное распределение $F_N(\xi, t)$ величин

$$x(t), \dot{x}(t) = x(t+1) - x(t), \ddot{x}(t) = \dot{x}(t+1) - \dot{x}(t), \dots$$

так что $f_N(x, t) = \int F_N(x, \dot{x}, t) d\dot{x}$

- Формальное уравнение эволюции ВПФР, полученное на основе теоремы Лиувилля, имеет вид

$$\frac{\partial F_N(\xi, t)}{\partial t} + \operatorname{div}_\xi (\dot{\xi} F_N(\xi, t)) = 0, \quad \xi = (x, \dot{x}, \dots), \quad \dot{\xi} = (\dot{x}, \ddot{x}, \dots)$$

- Поскольку объем выборки N конечен, то число компонент ξ должно быть ограниченным. Обрыв цепочки позволяет получить модельное уравнение эволюции ВПФР с локально меняющейся скоростью.

Уравнение Лиувилля

- Рассматривается совместное выборочное распределение случайной величины и ее приращений. Вводится средняя локальная скорость

$$u_N(x,t)f_N(x,t) = \int \dot{x}F_N(x, \dot{x}, t)d\dot{x}$$

- Записывается уравнение эволюции ВПФР (**эмпирическое уравнение Лиувилля**)

$$\frac{\partial f_N(x,t)}{\partial t} + \frac{\partial(u_N(x,t)f_N(x,t))}{\partial x} = 0$$

- Аналогично вводится среднее локальное ускорение (и т.д.)

$$w_N(x, \dot{x}, t)F_N(x, \dot{x}, t) = \int \ddot{x}\Phi_N(x, \dot{x}, \ddot{x}, t)d\ddot{x}$$

Статистический смысл скорости изменения ВПФР

$$u_N(x, t) f_N(x, t) = \int \dot{x} F_N(x, \dot{x}, t) d\dot{x}$$

■ Эмпирическое уравнение Лиувилля

$$\frac{\partial f_N(x, t)}{\partial t} + \frac{\partial (u_N(x, t) f_N(x, t))}{\partial x} = 0$$

$$\frac{dm(t)}{dt} = m(t+1) - m(t) = \frac{1}{N} \sum_{k=1}^N x(t - N + k + 1) -$$

$$- \frac{1}{N} \sum_{k=1}^N x(t - N + k) = \frac{1}{N} \sum_{k=1}^N \dot{x}(t - N + k) = \int v F(x, v, t) dx dv =$$

$$= \int x \frac{\partial f(x, t)}{\partial t} dx = \int u(x, t) f(x, t) dx \equiv U(t)$$

Эволюция эмпирической скорости

$$\begin{aligned}\frac{\partial}{\partial t} u(x, t) f(x, t) &= \int \dot{x} \frac{\partial}{\partial t} F(x, \dot{x}, t) d\dot{x} = \\ &= - \int \dot{x} \left(\frac{\partial}{\partial x} \dot{x} F(x, \dot{x}, t) + \frac{\partial}{\partial \dot{x}} w(x, \dot{x}, t) F(x, \dot{x}, t) \right) d\dot{x} = \\ &= - \frac{\partial}{\partial x} e(x, t) f(x, t) + W(x, t) f(x, t) \\ e(x, t) f(x, t) &= \int \dot{x}^2 F(x, \dot{x}, t) d\dot{x} \\ W(x, t) f(x, t) &= \int w(x, \dot{x}, t) F(x, \dot{x}, t) d\dot{x}\end{aligned}$$

Замыкание кинетической модели

Дополним уравнение Лиувилля уравнением эволюции для скорости:

$$\left(\frac{\partial u}{\partial t} - u \frac{\partial u}{\partial x} \right) f - u^2 \frac{\partial f}{\partial x} = - \frac{\partial (ef)}{\partial x} + Wf$$

$$e(x,t)f(x,t) = \int \dot{x}^2 F(x, \dot{x}, t) d\dot{x}, \quad W(x,t)f(x,t) = \int w(x, \dot{x}, t) F(x, \dot{x}, t) d\dot{x}$$

Уравнение для $u(x,t)$ требуется вследствие того, что, по построению, скорость известна в предыдущий момент по сравнению с ВПФР:

$$u(i+1,t) = \frac{u(i,t)f(i,t) - f(i,t) + f(i,t+1)}{f(i+1,t)}$$

Если известны $e(x,t)$ и $W(x,t)$, то система замкнется. В противном случае добавляются уравнения для этих величин, которые зависят от моментов высших порядков и распределений более высокой размерности. Обрыв цепочки на каком-нибудь порядке приводит к замкнутым моделям эволюции.

«Гидродинамическая» модель

$$\frac{\partial f(x,t)}{\partial t} + \frac{\partial}{\partial x} f(x,t)u(x,t) = 0; \quad u(x,t)f(x,t) = \int \dot{x}F(x, \dot{x}, t)d\dot{x}$$

$$\left(\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} \right) f = \frac{\partial}{\partial x} (u^2 - e)f; \quad e(x,t)f(x,t) = \int \dot{x}^2 F(x, \dot{x}, t)d\dot{x}$$

$$\left(\frac{\partial e}{\partial t} + u \frac{\partial e}{\partial x} \right) f = \frac{\partial}{\partial x} (ue - e_3)f, \quad e_3 f = \int \dot{x}^3 F(x, \dot{x}, t)d\dot{x}.$$

Идеология модели: надо оборвать цепочку на первом уравнении (нулевые средние ускорения), но учитывать эволюцию высших моментов. Простейшим вариантом замыкания является наивный прогноз относительно момента третьего порядка. Тогда система рекуррентно разрешима.

Многомерные ВПФР

Уравнение Лиувилля для r -мерных ВПФР имеет вид

$$\frac{\partial F_r(x, \dot{x}, \dots, x^{(r-1)}; t)}{\partial t} + \frac{\partial}{\partial x} \dot{x} F_r + \frac{\partial}{\partial \dot{x}} \ddot{x} F_r + \dots + \frac{\partial}{\partial x^{(r-1)}} w(x, \dot{x}, \dots, x^{(r-1)}; t) F_r = 0,$$

$$w(x, \dot{x}, \dots, x^{(r-1)}; t) F_r(x, \dot{x}, \dots, x^{(r-1)}; t) = \int x^{(r)} F_{r+1}(x, \dot{x}, \dots, x^{(r)}; t) dx^{(r)}.$$

Замыкание на ВПФР порядка r состоит в том, что среднее ускорение следующего порядка полагается равным нулю:

$$W(x, \dot{x}, \dots, x^{(r-1)}, x^{(r)}; t) = 0$$

В этом случае эволюция $w(x, \dot{x}, \dots, x^{(r-1)}; t)$ будет зависеть только от моментов F_{r+1} , но не от высших ускорений

r-мерная ЭВОЛЮЦИОННАЯ МОДЕЛЬ

$$\frac{\partial F_r(x, \dot{x}, \dots, x^{(r-1)}; t)}{\partial t} + \frac{\partial}{\partial x} \dot{x} F_r + \frac{\partial}{\partial \dot{x}} \ddot{x} F_r + \dots + \frac{\partial}{\partial x^{(r-1)}} w(x, \dot{x}, \dots, x^{(r-1)}; t) F_r = 0,$$

$$w(x, \dot{x}, \dots, x^{(r-1)}; t) F_r(x, \dot{x}, \dots, x^{(r-1)}; t) = \int x^{(r)} F_{r+1}(x, \dot{x}, \dots, x^{(r)}; t) dx^{(r)}.$$

$$\begin{aligned} & \left(\frac{\partial w}{\partial t} + \frac{\partial \langle \dot{x} x^{(r)} \rangle}{\partial x} + \frac{\partial \langle \ddot{x} x^{(r)} \rangle}{\partial \dot{x}} + \dots + w \frac{\partial w}{\partial x^{(r-1)}} \right) F_r = \\ & = \left(w \dot{x} - \langle \dot{x} x^{(r)} \rangle \right) \frac{\partial F_r}{\partial x} + \left(w \ddot{x} - \langle \ddot{x} x^{(r)} \rangle \right) \frac{\partial F_r}{\partial \dot{x}} + \dots + \\ & + \left(w x^{(r-1)} - \langle x^{(r-1)} x^{(r)} \rangle \right) \frac{\partial F_r}{\partial x^{(r-2)}} + \frac{\partial}{\partial x^{(r-1)}} \left(w^2 - \langle (x^{(r)})^2 \rangle \right) F_r. \end{aligned}$$

VI

Эмпирическое уравнение Фоккера-Планка для описания эволюции выборочных распределений

Уравнение Колмогорова

Условная вероятность $P(x', t' | x, t)$

Вероятность перехода $w(x, x'; t) = \lim_{\tau \rightarrow 0} \frac{1}{\tau} (P(x', t + \tau | x, t) - \delta(x' - x))$

Совместная вероятность $F(x', t'; x, t) = P(x', t' | x, t) \cdot f(x, t)$

ОКУ $\frac{\partial f(x, t)}{\partial t} = \int w(x', x; t) f(x', t) dx'$

$\forall t'', t \leq t'' \leq t' : P(x', t' | x, t) = \int P(x', t' | x'', t'') P(x'', t'' | x, t) dx''$

Уравнение Фоккера-Планка

$$\begin{aligned} \frac{\partial P(x, t | x_0, t_0)}{\partial t} &= \lim_{\tau \rightarrow +0} \frac{P(x, t + \tau | x_0, t_0) - P(x, t | x_0, t_0)}{\tau} = \\ &= \lim_{\tau \rightarrow +0} \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x^n} \left(\frac{m_n(x, t; \tau)}{\tau} P(x, t | x_0, t_0) \right). \end{aligned}$$

Если отличны от нуля только первые два момента, то

$$\frac{\partial f(x, t)}{\partial t} = -\frac{\partial}{\partial x} (u(x, t) f(x, t)) + \frac{1}{2} \frac{\partial^2}{\partial x^2} (\lambda(x, t) f(x, t))$$

$$\lambda(x, t) = \lim_{\tau \rightarrow +0} \frac{m_2(x, t; \tau)}{\tau} = \lim_{\tau \rightarrow +0} \frac{1}{\tau} \int_{-\infty}^{\infty} (x' - x)^2 P(x', t + \tau | x, t) dx' \geq 0$$

Эволюция выборочной дисперсии

$$\sigma^2(t) = \int (x - m(t))^2 f(x, t) dx$$

$$\frac{d\sigma^2}{dt} = -\int (x - m(t))^2 \frac{\partial(uf)}{\partial x} dx + \frac{\lambda}{2} \int (x - m(t))^2 \frac{\partial^2 f}{\partial x^2} dx = \lambda + 2 \int uxf dx - 2mU$$

$$\begin{aligned} \text{cov}_{x,v}(t) &= \int xvF(x, v, t) dx dv - \int xF(x, v, t) dx dv \cdot \int vF(x, v, t) dx dv = \\ &= \int xu(x, t) f(x, t) dx - m(t)U(t) \equiv \text{cov}_{x,u}(t) \end{aligned}$$

$$\frac{d\sigma^2}{dt} = \lambda + 2 \text{cov}_{x,u}$$

Согласованность уравнений эволюции для моментов

$$\frac{\partial f(x,t)}{\partial t} + \frac{\partial}{\partial x} (u(x,t) f(x,t)) - \frac{\lambda(t)}{2} \frac{\partial^2 f(x,t)}{\partial x^2} = 0;$$

$$u(x,t) = \frac{1}{f(x,t)} \int v F(x,v,t) dv, \quad v = \frac{dx}{dt};$$

$$f(x,t) = \int F(x,v,t) dv;$$

$$\begin{aligned} \lambda(t) &= \sigma^2(t+1) - \sigma^2(t) - 2 \operatorname{cov}_{x,u}(t) = \\ &= \frac{1}{T} \sum_{k=t-T+1}^t (x(k) - x(k+1))^2 - \frac{1}{T^2} (x(t+1) - x(t-T+1))^2 \geq 0 \end{aligned}$$

Выборочные среднее и дисперсия временного ряда изменяются так же, как и моменты ВФР в силу уравнения Фоккера-Планка, если снос и диффузия определены как написано выше.

VII

Генерация траекторий
нестационарных временных
рядов с определенным
эмпирическим законом
эволюции ВПФР

Генерация нестационарного ряда

1. По имеющимся данным строится ряд первых разностей ряда $a(t)$

$$x(t) = a(t + 1) - a(t)$$

2. Выбирается длина N , на которой индекс нестационарности максимальный, после чего строятся распределения встык-выборок

$$f_N(x, t) \quad f_N(x, t - N) \quad f_N(x, t - 2N) \dots$$

3. Определяются параметры эмпирического уравнения Фоккера-Планка для ВПФР (коэффициенты сноса и диффузии) и строится его численное решение. Тем самым находится прогнозная ВПФР

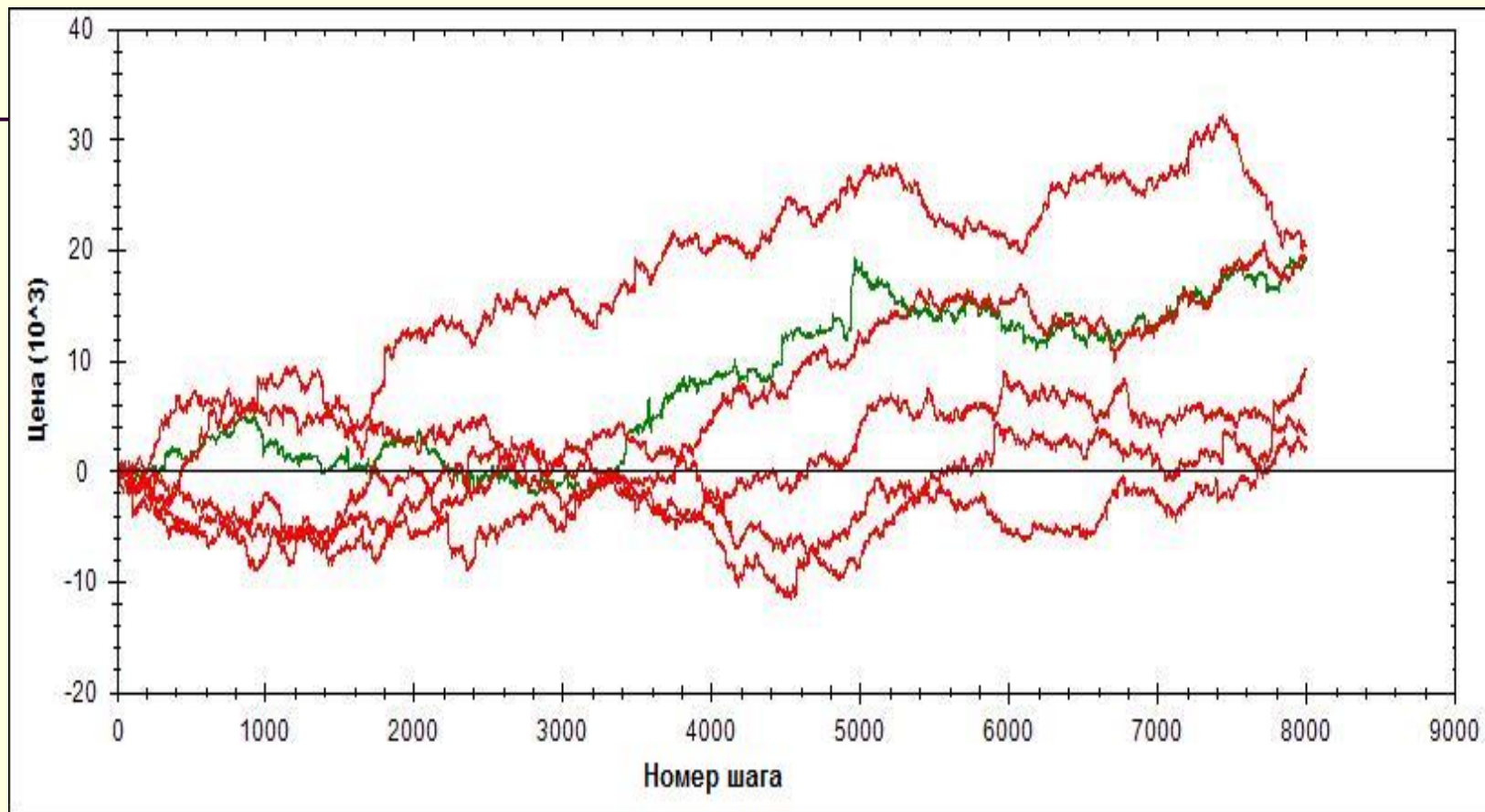
$$f_N(x, t + k), \quad k = 1, 2, \dots, N$$

4. Генерируется равномерно распределенный ряд $\{y_k\} \in [0; 1]$

5. Строится прогнозная траектория приростов цен по формуле

$$y_k = F_N(x_k, t + k)$$

Пример пучка траекторий



Разные траектории отличаются одна от другой и от фактической как различные реализации одного и того же процесса; расстояния между выборками, сдвинутыми на окно N , имеют СУС данного ряда.

Проверка корректности генерации ансамбля

Пусть сгенерированы s равномерно распределенных рядов длиной N : $\{y_k\}_j, j = 1, \dots, s$. Каждая j -ая траектория порождает на отрезке $[t_0 + 1; t_0 + N]$ ВПФР $\tilde{f}_N(\{y\}_j; x, t_0 + N)$, отличную от факта $f_N(x, t_0 + N)$.
Рассмотрим расстояния:

$$r = \left\| \tilde{F}_N(\{y\}; x, t_0 + N) - F_N(x, t_0) \right\|$$

$$\rho = \left\| \tilde{F}_N(\{y\}; x, t_0 + N) - F_N(x, t_0 + N) \right\|$$

$$\tilde{\rho} = \left\| \tilde{F}_N(\{y\}; x, t_0 + N) - \tilde{F}_N(\{y'\}; x, t_0 + N) \right\|$$

СУС r^* должен быть равен СУС встык-выборок исходного ряда.

СУС ρ^* должен быть равен СУС $\tilde{\rho}^*$ и оба они меньше СУС r^* .

Тестирование управляющего функционала

Пусть на выборке длины \mathbf{N} задан функционал $\Psi\{x(t-T+1), \dots, x(t)\}$.

Вместо одной исторической траектории, уходящей в не актуальное прошлое, для тестирования функционала можно задать ансамбль траекторий. Пусть Ψ_j есть значение функционала на j -ой выборке.

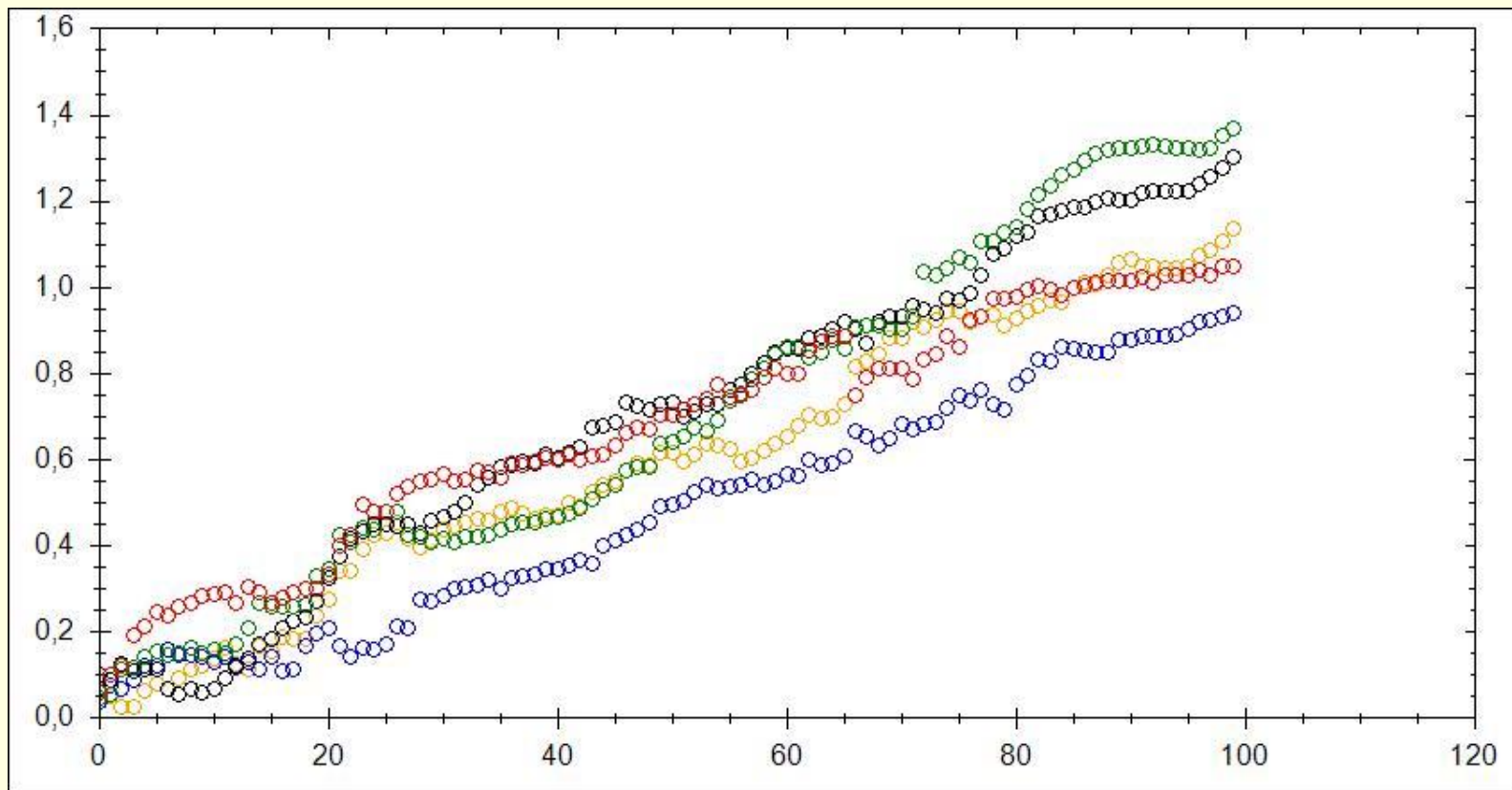
Тогда

$$\bar{\Psi} = \frac{1}{N} \sum_{j=1}^N \Psi_j, \quad \sigma_{\Psi}^2 = \frac{1}{N} \sum_{j=1}^N (\Psi_j - \bar{\Psi})^2, \quad S_{\Psi} = \frac{\bar{\Psi}}{\sigma_{\Psi}}$$

Пусть некий параметр λ уравнения Ф-П меняется, так что меняется и среднее значение функционала. Тогда можно протестировать чувствительность функционала по параметру:

$$\Lambda_{\Psi} = \frac{\partial \ln \bar{\Psi}}{\partial \ln \lambda}$$

На каждой траектории запускается торговая система с одним и тем же набором параметров, после чего вычисляется среднее по ансамблю. Меняя параметры, можно получить их оптимальный набор – более правильный, чем при тестировании в прошлом.



ВЫВОДЫ

ОСНОВНЫЕ ЭТАПЫ АНАЛИЗА

1. Составление горизонтной статистики, выявление типа временного ряда, нахождение оптимального объема выборки для прогнозирования с заданной точностью на заданный горизонт
2. Конструирование индикаторов разладки
3. Составление и решение кинетического уравнения для моделирования эволюции выборочной плотности
4. Генерация пучка нестационарных траекторий с заданными свойствами эволюции выборочных распределений
5. Тестирование и оптимизация функционала доходности, анализ его устойчивости к вариации параметров кинетической модели

СПАСИБО ЗА ВНИМАНИЕ