

Стохастические методы прогнозирования

(краткий конспект без доказательств)

**Александр Михайлович Шурыгин,
МГУ, ВМИК,
a.shurugin@bk.ru**

1. Аналитические основы

Проблема оценивания

Пусть $x \in X \subseteq \mathbb{R}^1$ имеет плотность распределения $f := f(x, \theta)$, зависящую от неизвестного параметра θ . Нам надо найти оценку $\hat{\theta}$ параметра θ по выборке независимых наблюдений x_1, \dots, x_n .

Основы прикладной статистики

- Метод наименьших квадратов (**K. Gauss**, 1794-1795)

- **Оценки максимума правдоподобия (ОМП)**

$$(\mathbf{R. Fisher}, 1912) \theta_0 : \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f(x_i, \theta) \Big|_{\theta=\theta_0} = 0$$

Их недостатки: ►

- **А.Н. Колмогоров (1950)**: «ОМП не универсальны»
- **J.W. Tukey (1960)**: ► ОМП «катастрофически неустойчивы» к малым изменениям п.р.!

► Оценки минимума контраста (ОМК)

(V.P. Godambe 1960, J.Pfanzagle 1969)

- Функция контраста: $\rho(x, \theta), \hat{\theta} : \sum_{i=1}^n \rho(x_i, \hat{\theta}) \rightarrow \min$
- Оценочная функция: $\psi(x, \theta) = c \frac{\partial}{\partial \theta} \rho(x, \theta), \quad c \neq 0,$
- **ОМК**

$$\hat{\theta} : \sum \psi(x_i, \hat{\theta}) = 0.$$

- Условие состоятельности: $E\psi(x, \theta) = 0. \quad (1)$

- Асимптотическая дисперсия:

$$V(f, \psi) = \lim_{n \rightarrow \infty} E n(\hat{\theta} - \theta)^2 = E \psi^2 / (E \dot{\psi})^2. \quad (2)$$

где $\psi := \psi(x, \theta)$ и $\dot{\psi} := \frac{\partial}{\partial \theta} \psi(x, \theta).$ ►

Пример Тьюки

П.р. $f(y) = (1 - \varepsilon)\varphi(y) + \varepsilon\varphi(y/3)/3$, где $y = x - \mu$,

а $\varphi(y) = (2\pi)^{-1/2} e^{-y^2/2}$ - п.р. $N(0,1)$.

Для среднего оценочная функция $\psi(y) = y$, \Rightarrow

$E\psi = -1$, $E\psi^2 = 1$ - это знаменатель в (2).

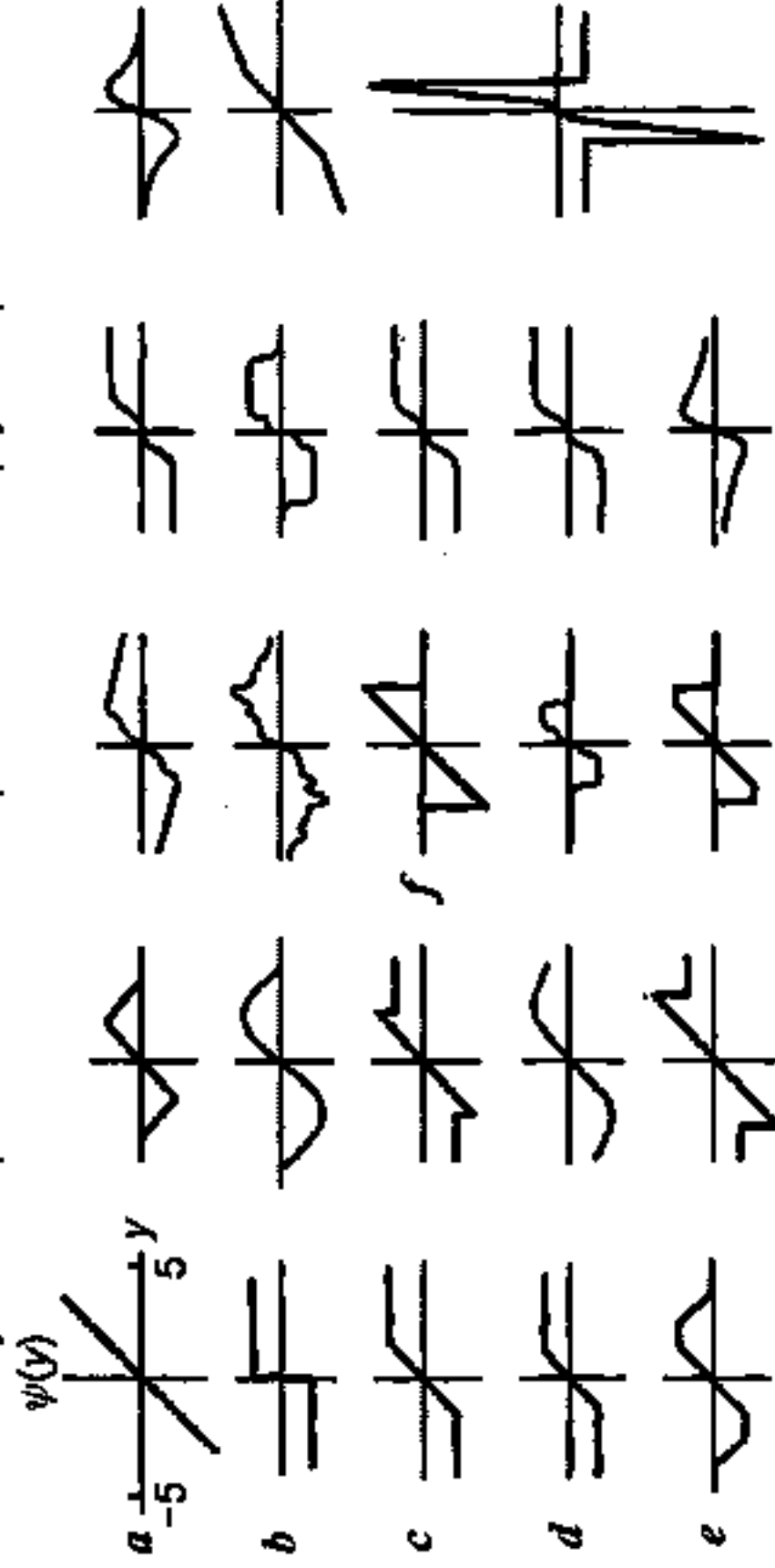
$$\begin{aligned} E\psi^2(y) &= (1 - \varepsilon) \int_{\mathbb{R}^1} y^2 \varphi(y) dy + \varepsilon \int_{\mathbb{R}^1} y^2 \left[\frac{1}{3} \varphi\left(\frac{y}{3}\right) \right] dy = \\ &= (1 - \varepsilon) + 9\varepsilon \int_{\mathbb{R}^1} \left(\frac{y}{3}\right)^2 \varphi\left(\frac{y}{3}\right) d\left(\frac{y}{3}\right) = 1 - \varepsilon + 9\varepsilon = 1 + 8\varepsilon. \end{aligned}$$

Это числитель в (2). Получаем: $V(f, \psi) = 1 + 8\varepsilon$.

При $\varepsilon = 1/8$ дисперсия удваивается!

Теория робастности: поиск устойчивой оценки центра распределения $N(0, 1)$ моделированием точечного загрязнения в Принстонском

университете. Оценочные функции:



- Много устойчивых ОМК было найдено теоретически и эмпирически в Принстонском эксперименте. **Их недостатки:**
- Все оценки зависят от неоцениваемых параметров (дисперсии, степени загрязнения), так что каждая из них является семейством оценок. Хьюбер и Хампель подобрали константы для распределения $N(0,1)$.
 - Не только центр нормального распределения, но и другие параметры различных распределений следует устойчиво оценивать.
 - Авторы ищут устойчивые оценки, не имея определения устойчивости.
 - Надо искать устойчивость к неизбежному отклонению реальной п.р. от модельной. ►

Функциональная оптимизация

Для функции $y(x)$ величина $|dy/dx|$ - хорошая мера неустойчивости y по отношению к вариациям x .

По аналогии используем производную Лагранжа

$$W(f, \psi) = \frac{\partial}{\partial f} V(f, \psi) = \left(\int_x \psi^2 dx \right) / (E \psi)^2 \quad (3)$$

как меру неустойчивости дисперсии оценки при вариациях п.р. f .

Функция ψ определена до умножения на ненулевую константу c , и если

$$c = (E \psi)^{-1} \quad (4),$$

тогда

$$W(f, \psi) = \int_x \psi^2 dx.$$

Теорема. При выполнении условий (1) и (4) для квадратичной по ψ функции $Q(\psi, f)$ функционал

$$G(\psi) = \int_x \left[Q(\psi, f) + \lambda_0 \psi f + \lambda_1^* \psi f \right] dx$$

► достигает минимума на функции

$$\psi(x, \theta) = \frac{\partial}{\partial \psi} Q(\psi, f) + \lambda_0 f + \lambda_1 f.$$

Приложив теорему к функционалу (3), получим, что неустойчивость W достигает минимума W_* на оценочной функции

$$\psi_*(x, \theta) = \frac{\partial}{\partial \theta} f(x, \theta) + \beta f(x, \theta), \quad \beta : E \psi_* = 0.$$

Соответствующую оценку θ_* назовём **оценкой максимальной устойчивости (ОМУ)**.

► Эффективность

$V(f, \psi)$ – положительная мера неэффективности, достигающая минимума V_0 на ОМП.

Эффективность $\text{eff } \hat{\theta} = V_0/V(f, \psi)$

На квадрате $\left[\text{eff } \hat{\theta}, \text{stb } \hat{\theta} \right]$

Условно оптимальные θ_γ с оценочной функцией

$$\psi_\gamma = c \left(\frac{\partial}{\partial \theta} \ln f + \beta \right) / (1 + \gamma/f), \quad \beta: E\psi_\gamma = 0$$

где γ зависит от соотношений между $\text{eff } \theta_\gamma$ и $\text{stb } \theta_\gamma$

Компромиссные θ_c со свойствами:

$$\gamma = V_0/W_* : \left(\text{eff } \hat{\theta} \right)^{-1} + \left(\text{stb } \hat{\theta} \right)^{-1} \rightarrow \min$$

- **Радикальные** θ_r у которых $\text{eff } \theta_r \cong \text{stb } \theta_r$, и

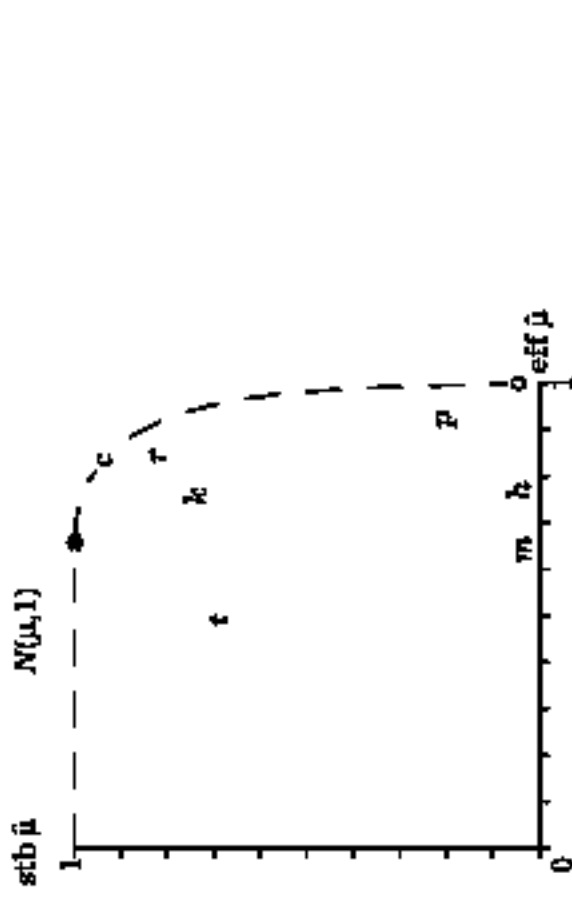
$$\psi_r = c \left(\frac{\partial}{\partial \theta} \ln f + \beta \right) \sqrt{f} \quad \blacktriangleright$$

Устойчивость

$W(f, \psi)$ – положительная мера неустойчивости, достигающая минимума W_* на ОМУ.

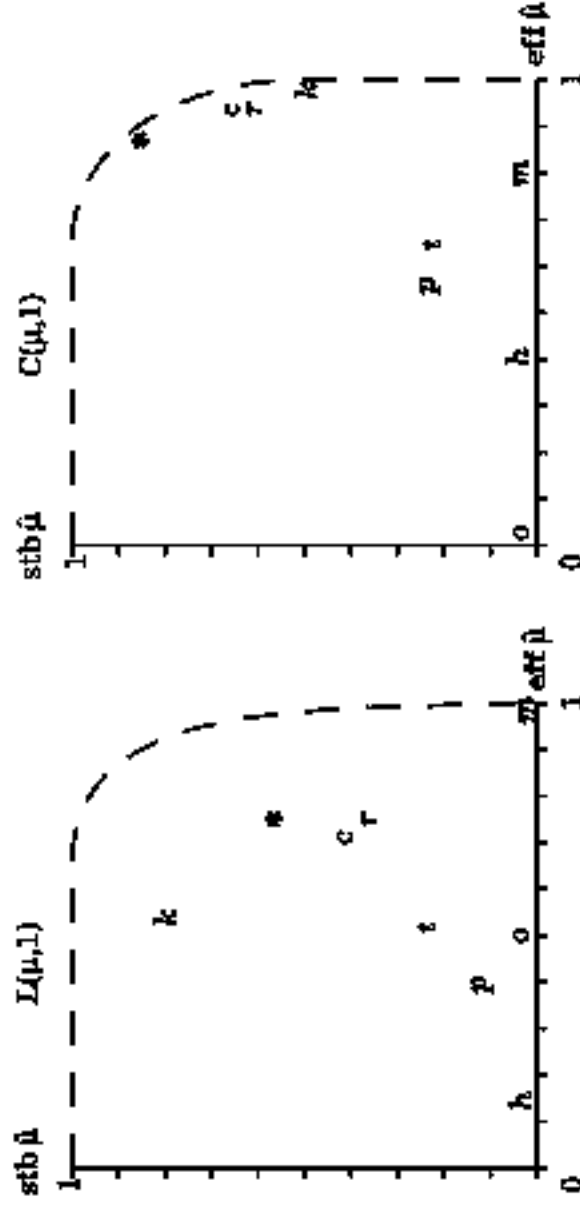
Устойчивость $\text{stb } \hat{\theta} = W_*/W(f, \psi)$

интересны такие оценки.



Обозначение оценок

- o – ОМП (среднее)
 - m – медиана
 - k – ОМП для распределения Коши
 - t – оценка Тьюки
 - h – оценка Хьюбера
 - p – оценка Хампеля
 - $\#$ – ОМУ
 - c – компромиссная
 - r – радикальная
- Прерывистая линия – условно оптимальные оценки \blacktriangleright



► **Радикальность оценок**

Функционал $R(f, \psi) = \left(\int_{x^*}^x \psi^2(x, \theta) \sqrt{f(x, \theta)} dx \right) / (E \dot{\psi}(x, \theta))^2$ достигает минимума R_* на радикальной оценке.

Радикальность оценки $\text{rad} \hat{\theta} = R_*/R(f, \psi)$ - удобная мера полезности в приложениях: ►

$\hat{\mu}$	μ_r	μ_c	μ_k	μ_*	μ_h	μ_p	μ_m	μ_0
$\text{rad} \hat{\mu}$	1	.99	.97	.91	.88	.85	.83	.65

Можно видеть, что ОМГ распределения Коши μ_k мажорирует робастные оценки μ_h и μ_p ; а медиана μ_m лучше, чем ОМГ.

Мультипликативные помехи

встречаются гораздо чаще, чем аддитивные.

Все положительные величины их имеют.

Пусть x_1, \dots, x_{2n} , $y_i = x_{2i} | x_{2i+1} | / c$, $c: E y^2 = 1$.

Оцениваем моделированием $E y (= 0)$ и асимптотическую дисперсию $V(\mu)$ для разных оценок μ :

$$\mu : \mu_0 \quad \mu_m \quad \mu_* \quad \mu_r$$

$$V(\mu) : \mathbf{1,01} \quad \mathbf{0,29} \quad 0,37 \quad 0,45.$$

Видно явное преимущество медианы и отставание ОМГ.

2. Многомерное нормальное распределение $N_p(\mathbf{m}, \mathbf{C})$

Для $\mathbf{x} = (x^{(1)}, \dots, x^{(p)})^T \in N_p(\mathbf{m}, \mathbf{C})$, $\mathbf{C} = \mathbf{C}^T$, $|\mathbf{C}| > 0$,

вектор математических ожиданий $\mathbf{m} = E\mathbf{x}$,

матрица ковариаций $\mathbf{C} = \{c_{ij}\} = E[(\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^T]$,
корреляция $x^{(i)}$ и $x^{(j)}$ $r_{ij} = c_{ij} / \sqrt{c_{ii}c_{jj}}$, $-1 < r < 1$.

При $-1 < r < 0$ увеличение

одного признака ведёт к уменьшению другого,

при $0 < r < 1$ оба признака изменяются вместе.

Поверхности равных плотностей – эллипсоиды,
их оси называются главными компонентами

Устойчивые оценки многомерного распределения

- Все одномерные определения переносятся на многомерный случай автоматически.
- Оценки Мешалкина \mathbf{m}_λ и \mathbf{C}_λ единственные устойчивые из предложенных. Они могут быть получены из следующей системы уравнений:

$$\left\{ \begin{array}{l} \Sigma(\mathbf{x}_i - \mathbf{m}_\lambda) e^{-\lambda q_i^2 / 2} = 0, \\ \Sigma [(\mathbf{x}_i - \mathbf{m}_\lambda)(\mathbf{x}_i - \mathbf{m}_\lambda)^T - (1 + \lambda)^{-1} \mathbf{C}_\lambda] e^{-\lambda q_i^2 / 2} = 0, \end{array} \right.$$

где $q_i^2 = (\mathbf{x}_i - \mathbf{m}_\lambda)^T \mathbf{C}_\lambda^{-1} (\mathbf{x}_i - \mathbf{m}_\lambda)$. (5)

Первое уравнение – вектор, второе – матрица. ¹⁵

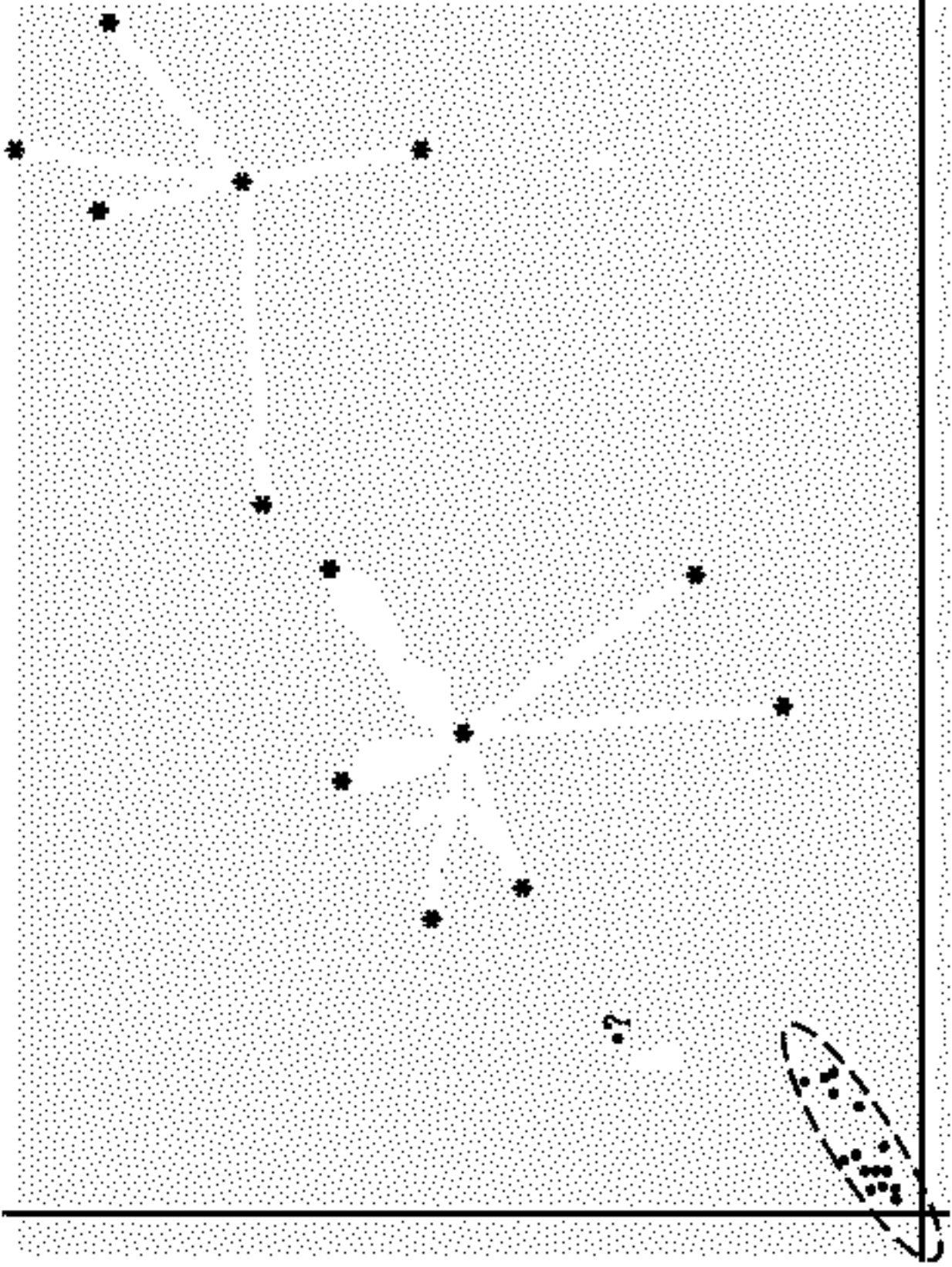
- При $\lambda = 1/2$ получаем **радикальные оценки** $\mathbf{m}_r, \mathbf{C}_r$.
- Если выборка содержит несколько кластеров, то в процессе итераций оценка \mathbf{m}_r сходится к центру большего из них, а матрица \mathbf{C}_r оценивает его ковариацию.

Для проверки однородности выборки предлагается следующий **кластер-тест**:

$$K_\lambda = \frac{1}{np} \Sigma (\mathbf{x}_i - \mathbf{m}_\lambda)^T \mathbf{C}_\lambda^{-1} (\mathbf{x}_i - \mathbf{m}_\lambda). \quad \blacktriangleright$$

- Если элементы выборки имеют нормальное распределение (то есть выборка однородна), то

$$\sqrt{np} \ln K_\lambda \xrightarrow{d} N \left(0, \frac{\left[2 + 4\lambda + (p+2)\lambda^2 \right] (1+\lambda)^{p+2}}{(1+2\lambda)^{p/2+2}} - 2 \right).$$



Кластер-алгоритм

Количество кластеров? Уровень значимости α :

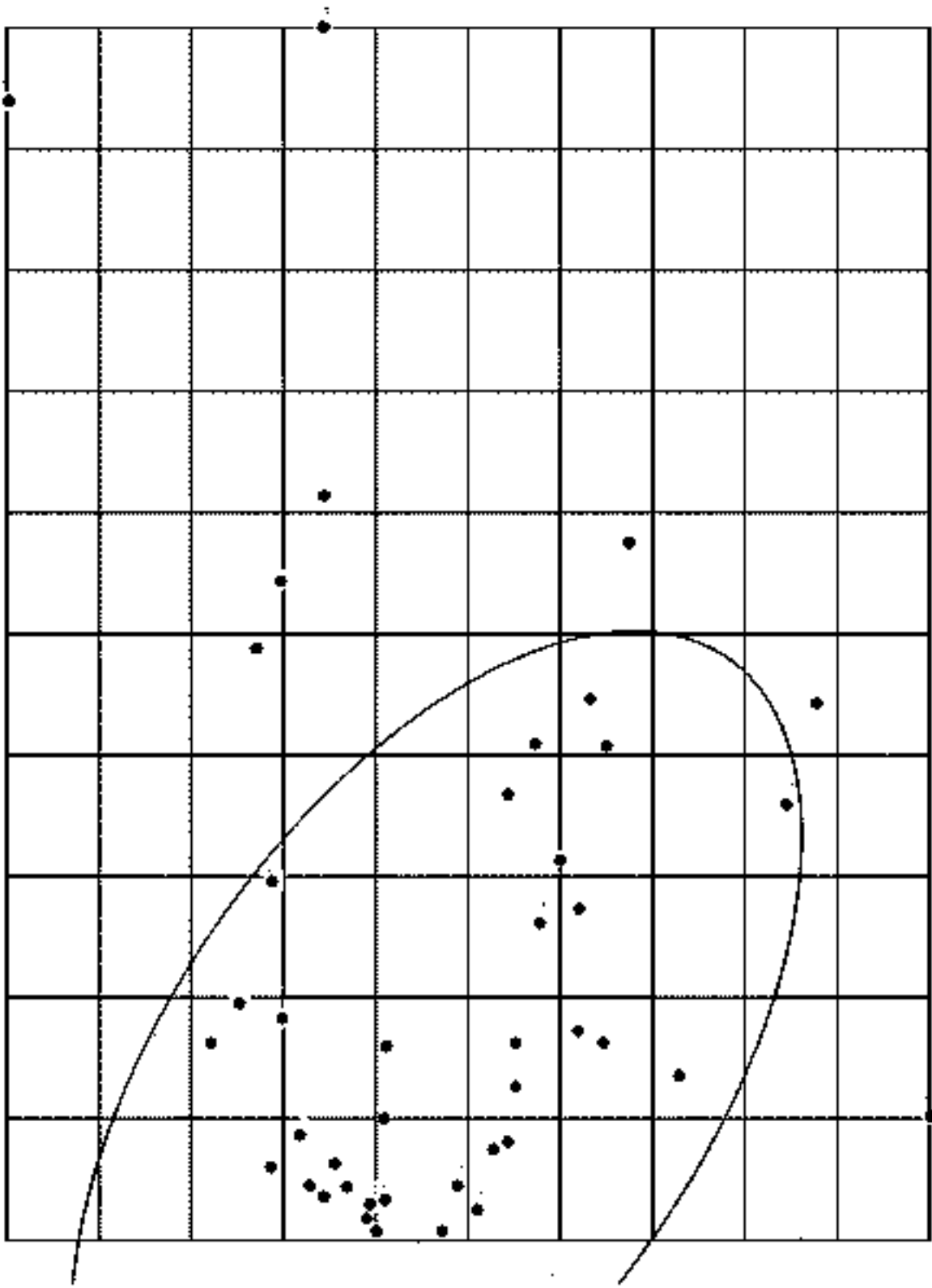
$$P\{\sqrt{(\mathbf{x} - \mu)^T \mathbf{C}^{-1} (\mathbf{x} - \mu)} < a_p\} = 1 - \alpha. \text{ Для } \mathbf{y} \in N_p(0, I)$$

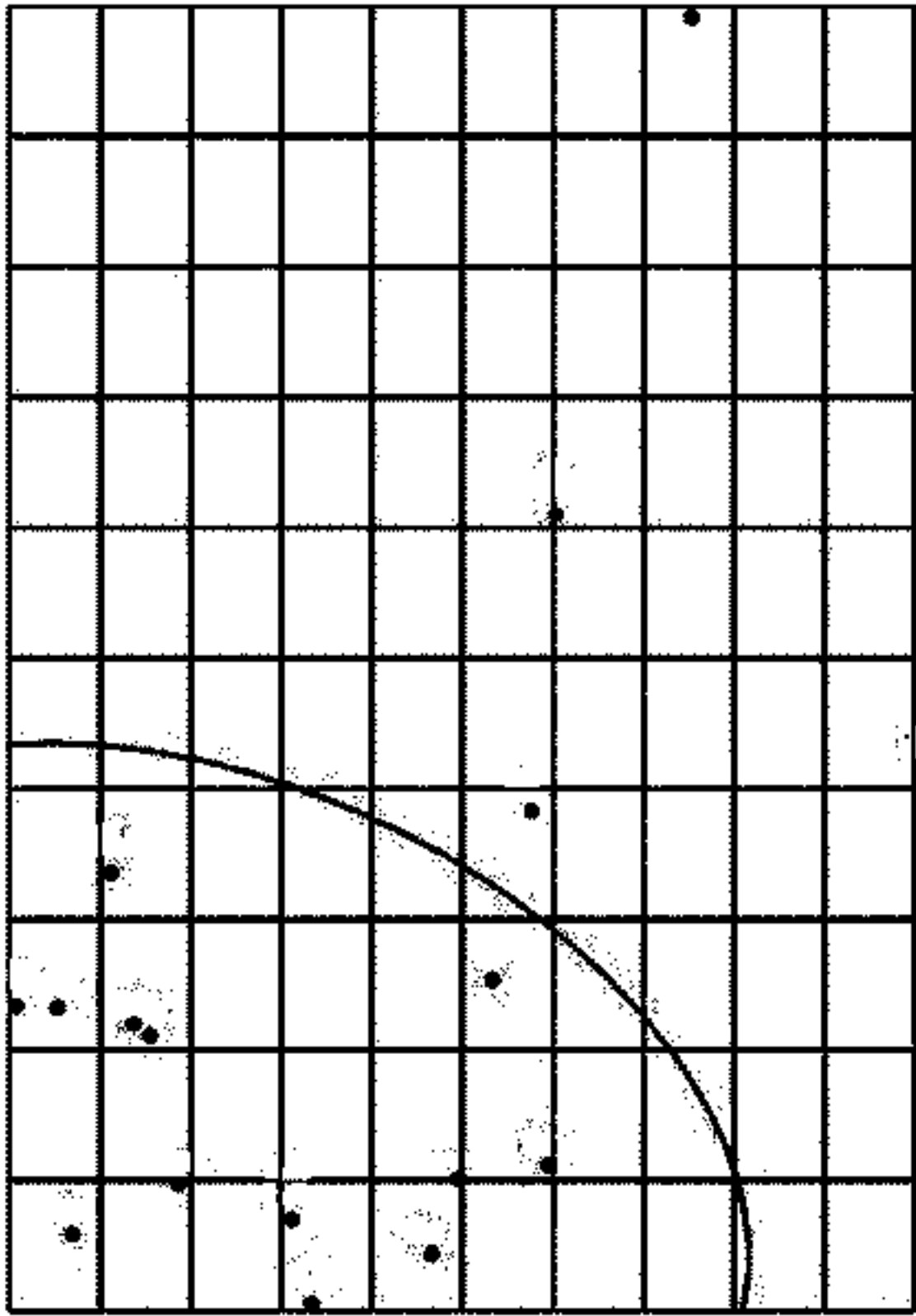
$$1 - \alpha = (2\pi)^{-p/2} \int_0^{b_p} S_p(r) e^{-r^2/2} dr = (2\pi)^{-p/2} S_p(1) \int_0^{b_p} r^{p-1} e^{-r^2/2} dr.$$

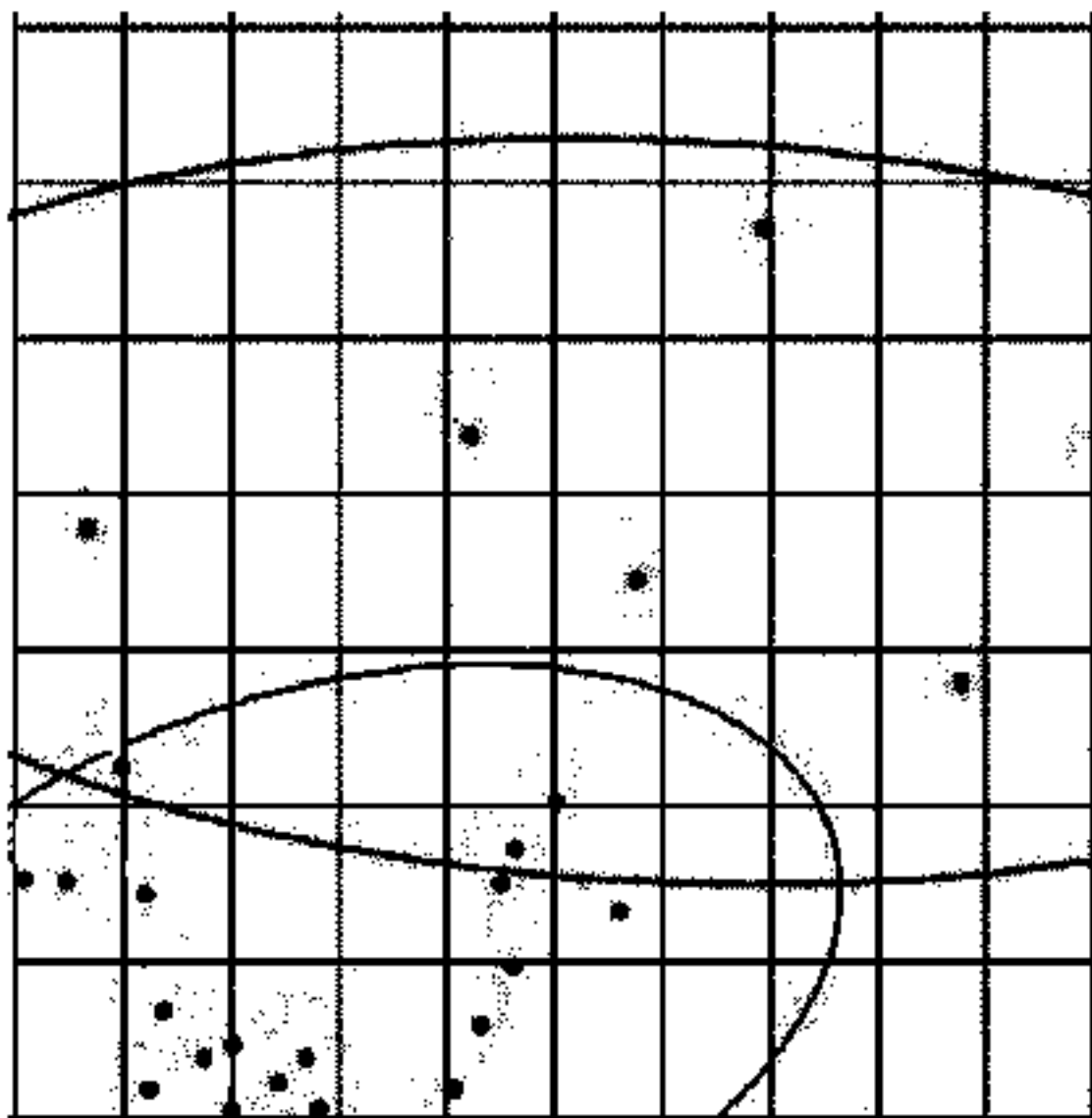
Известно: $S_p(1) = 2\pi^{p/2} / \Gamma(p/2)$

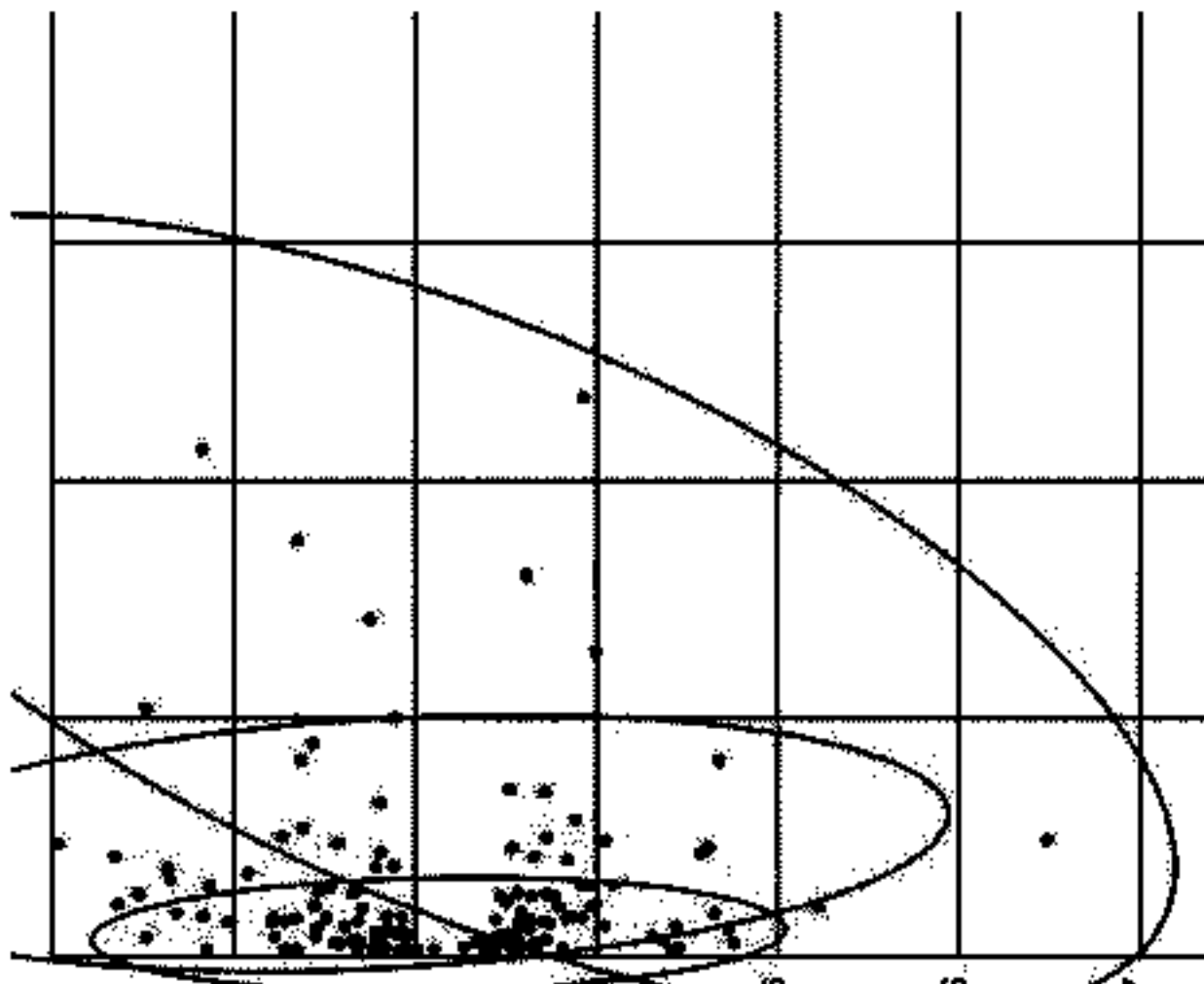
$$1 - \alpha = \frac{1}{2^{p/2-1}} \int_0^{b_p} r^{p-1} e^{-r^2/2} dr / \Gamma(p/2).$$

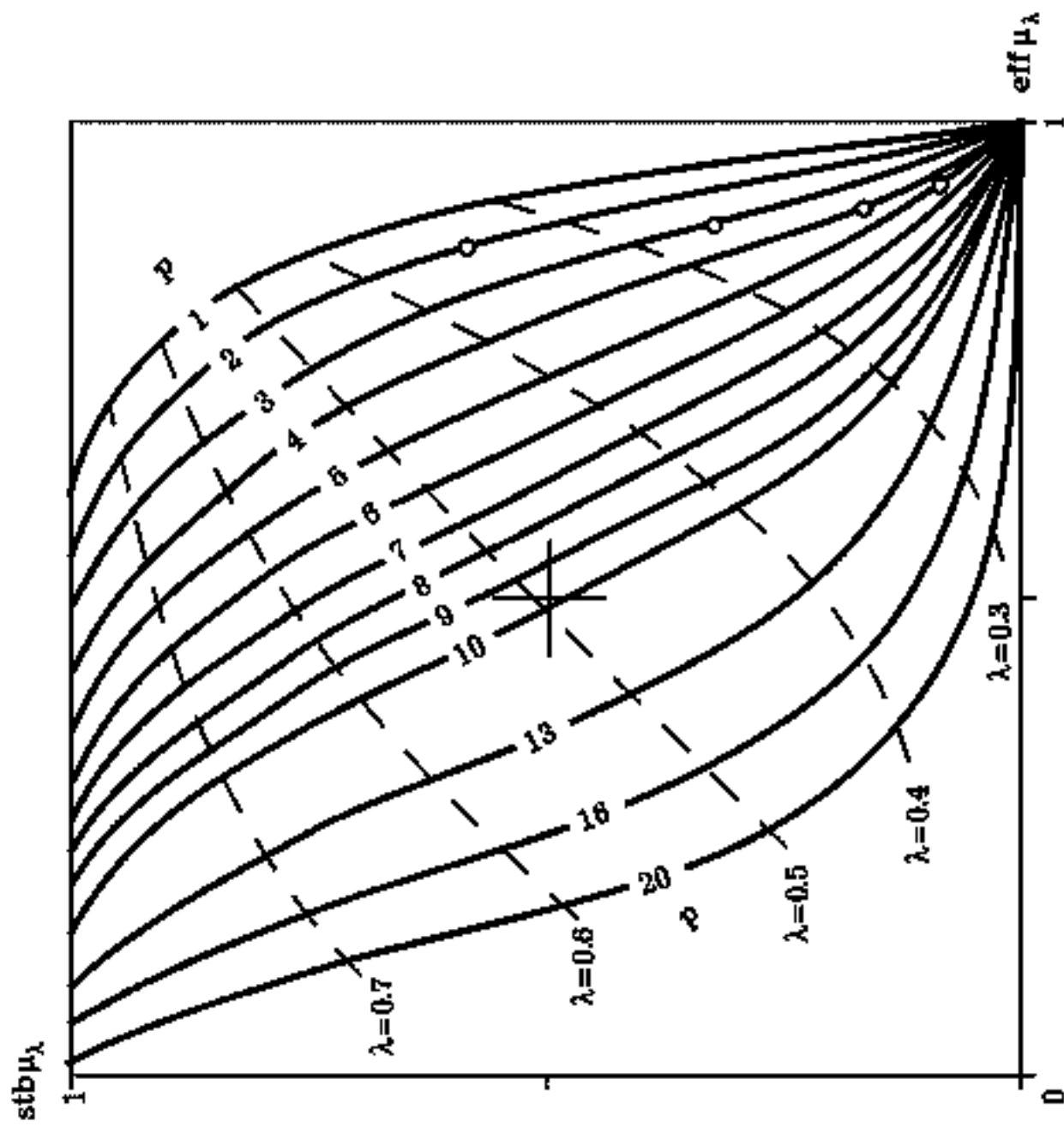
$$P\{\sqrt{\mathbf{y}^T \mathbf{y}} < b_p\} = 1 - \alpha \Rightarrow a_p \approx b_p.$$











Многомерная выборочная медиана

- В одномерном случае выборочная медиана μ_m имеет оценочную функцию

$$\psi(x - \mu_m) = \text{sgn}(x - \mu_m) = (x - \mu_m) / |x - \mu_m|$$

и является решением уравнения

$$\mu_m = (\sum x_i / |x_i - \mu_m|) / (\sum 1 / |x_i - \mu_m|).$$

Решение не изменится, если модули заменить нормированными расстояниями от центра q_i (5).
Получим уравнение для многомерного случая:

$$\mathbf{m}_m = \frac{\sum \mathbf{x}_i / q_i}{\sum 1 / q_i}.$$

Условное многомерное распределение

Центрируем случайный нормальный вектор \mathbf{x} , введя $\mathbf{y} = \mathbf{x} - E\mathbf{x}$, разобьем его на два вектора и соответственно разобьем матрицу ковариаций \mathbf{C} :

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix}.$$

Если компоненты вектора \mathbf{y}_2 фиксированы (известны), то известно и условное распределение компонент вектора \mathbf{y}_1 (**Андерсон**):

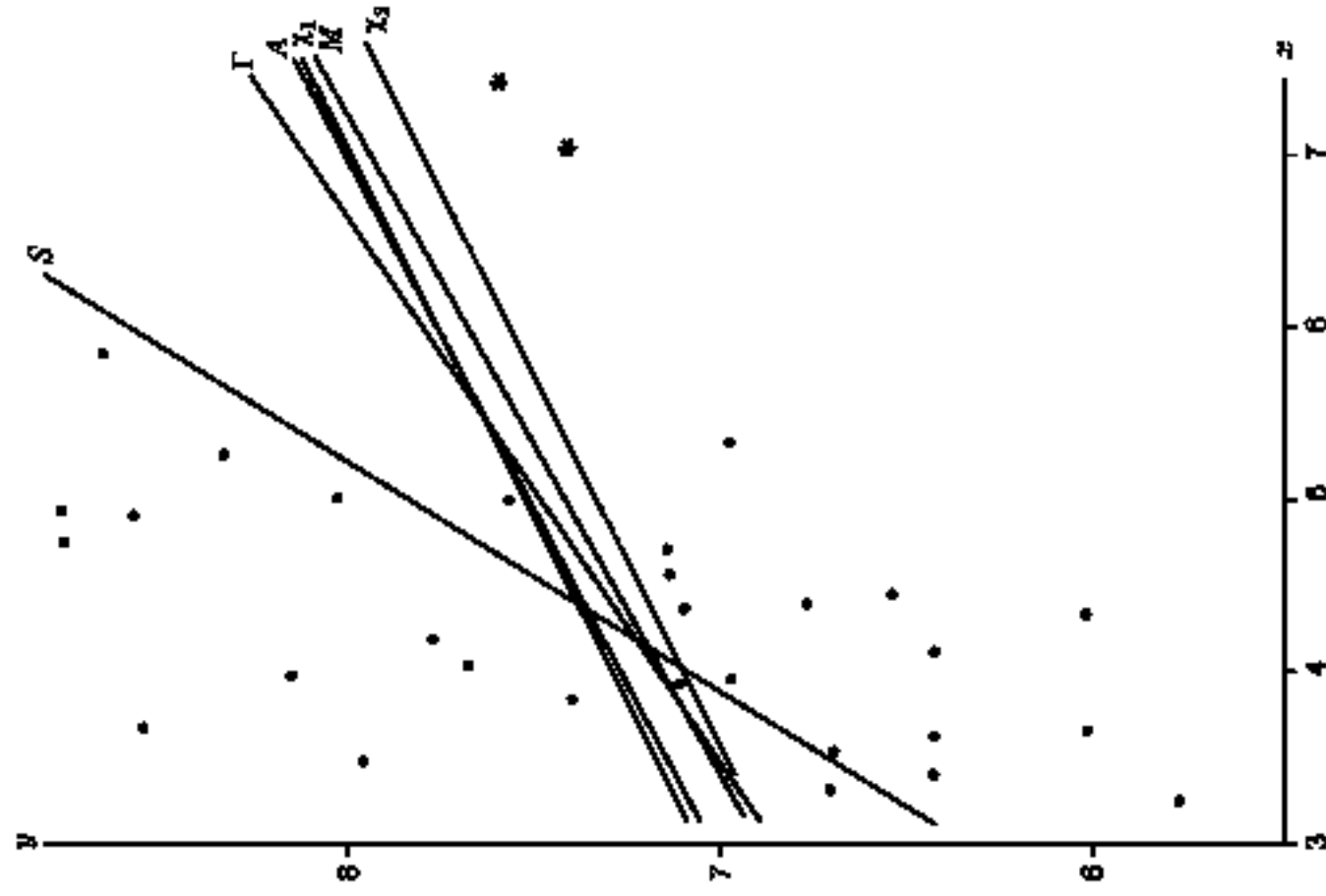
$$\boxed{\mathbf{y}_1 | \mathbf{y}_2 \in N(\mathbf{C}_{12} \mathbf{C}_{22}^{-1} \mathbf{y}_2, \mathbf{C}_{11} - \mathbf{C}_{12} \mathbf{C}_{22}^{-1} \mathbf{C}_{21})}. \quad (6)$$

3. Регрессионный прогноз

- Отклик $y_i = r(\mathbf{x}_i, \Theta) + \varepsilon_i \in \mathbb{R}^1$, $i = 1, \dots, N$, предикторы $\mathbf{x}_i = (x_i^{(1)}, \dots, x_i^{(p-1)})^T \in \mathbb{R}^{p-1}$, остатки $\varepsilon_i \in N(0, \sigma^2)$, функция регрессии $r(\cdot)$, её параметры $\Theta = (\theta_1, \dots, \theta_q)^T$, $\hat{\Theta} = ?$
- Классика: наименьшие квадраты $\sum \varepsilon_i^2 \rightarrow \min$.
- Робастность: оценки центра $\rho(\mathbf{x}_i, \hat{\Theta}) \rightarrow \min$.
- Все эти регрессии неустойчивы.
- **Радикальная по оценочной вектор-функции**

$$\Psi_r(\mathbf{z}, \Theta) = \mathbf{C} \left[\frac{\partial}{\partial \Theta} \ln f(\mathbf{z}, \Theta) \right] \sqrt{f(\mathbf{z}, \Theta)},$$

где $f(\cdot)$ - п.р. вектора $\mathbf{z} = (\mathbf{x}^T, y)$. \blacktriangleright



- – observations
- * – artificial contaminations

Regressions notation:

- Γ – classic
- A – Andrews'
- χ_1, χ_2 – Huber's
- M – Meshalkin's
- S – radical

► Квадратичная ошибка оценки параметров:

$$R = \lim n E \left[r(\mathbf{x}_0, \hat{\Theta}) - r(\mathbf{x}_0, \Theta) \right]^2 = \\ = \text{tr} \left[(E \dot{r} \dot{r}^T) \mathbf{H} (E \Psi \Psi^T) \mathbf{H}^T \right]$$

$$\mathbf{H}^{-1} = E \dot{\Psi},$$

- **Выбор модели** $r(\cdot)$ Квадратичная ошибка предсказания для нового наблюдения \mathbf{x}_0 является суммой «модельной» дисперсии σ^2 и квадратичной ошибки за счёт оценки параметров. Минимизируя их сумму

$$s^2 = \sigma^2 + R/n \rightarrow \min, \quad (7)$$

мы найдём подходящую функцию регрессии. $r(\mathbf{x}, \Theta)$. Как и в многомерной статистике, эффективность и устойчивость оценок снижается при росте размерности. Так что полезно уменьшать размерность.

Простейшие регрессии

Полиномиальная степени p на отрезке.

$$\text{Функция регрессии } r(\mathbf{x}, \Theta) = \sum_{j=0}^p \theta_j x^j, \quad j = 0, \dots, p.$$

Оценочная функция $\psi_j = x^j \varepsilon \exp(-\lambda \varepsilon^2 / 2\sigma^2)$,

$\lambda = 0 \rightarrow$ ОМГ, $\lambda = 1/2 \rightarrow$ радикал, $\lambda = 1 \rightarrow$ макс. устойчив.

Линейная на куб с равном. распредел. $\mathbf{x} = (x^{(1)}, \dots, x^{(p)})^T$

$$r(\mathbf{x}, \Theta) = \theta_0 + \sum_{j=1}^p \theta_j x^{(j)},$$

$$\psi_j = \begin{cases} \varepsilon \exp(-\lambda \varepsilon^2 / 2\sigma^2), \\ x^{(j)} \varepsilon \exp(-\lambda \varepsilon^2 / 2\sigma^2), \end{cases}$$

Линейная на нормальное распределение $N_p(\mathbf{m}, \mathbf{C})$

$$r(\mathbf{x}, \Theta) = \theta_0 + \sum_{j=1}^p \theta_j x^{(j)}, \quad \psi_j = \begin{cases} \varepsilon \exp(-\lambda [\varepsilon^2 / \sigma^2 + q^2] / 2), \\ x^{(j)} \varepsilon \exp(-\lambda [\varepsilon^2 / \sigma^2 + q^2] / 2). \end{cases}$$

При больших размерностях всё ухудшается!

Редуцированная линейная регрессия

Решения ухудшаются, когда растёт размерность P . Поэтому что в многомерной статистике надо оценить $O(p^2)$ параметров, хотя для решения нам нужно только $O(p)$ параметров. Исключение признаков уменьшает информацию. Сведём задачу к последовательности двумерных задач.

• **Регрессия.** Пусть признаки $\mathbf{x} = (x^{(1)}, \dots, x^{(p-1)})^T$ упорядочены по «информативности». Рассмотрим последовательность двумерных задач, используя s^2

(7) для исключения неинформативных признаков:

$$\begin{aligned} y &= c_0 + c_1 x^{(1)} + \varepsilon^{(1)}, \quad \varepsilon^{(1)} = c_2 x^{(2)} + \varepsilon^{(2)}, \\ &\dots, \varepsilon^{(p-2)} = c_{p-1} x^{(p-1)} + \varepsilon^{(p-1)} + c_p. \end{aligned}$$

Редуцированная дискриминация

В \mathbf{R}^2 с признаками a и b есть два класса точек, разделяемых линейной дискриминантной функцией (ЛДФ), и $n(a, b)$ обозначение множества проекций точек на нормаль к ЛДФ. Упорядочим признаки

$\mathbf{x} = (x^{(1)}, \dots, x^{(p)})^T$ по информативности и составим

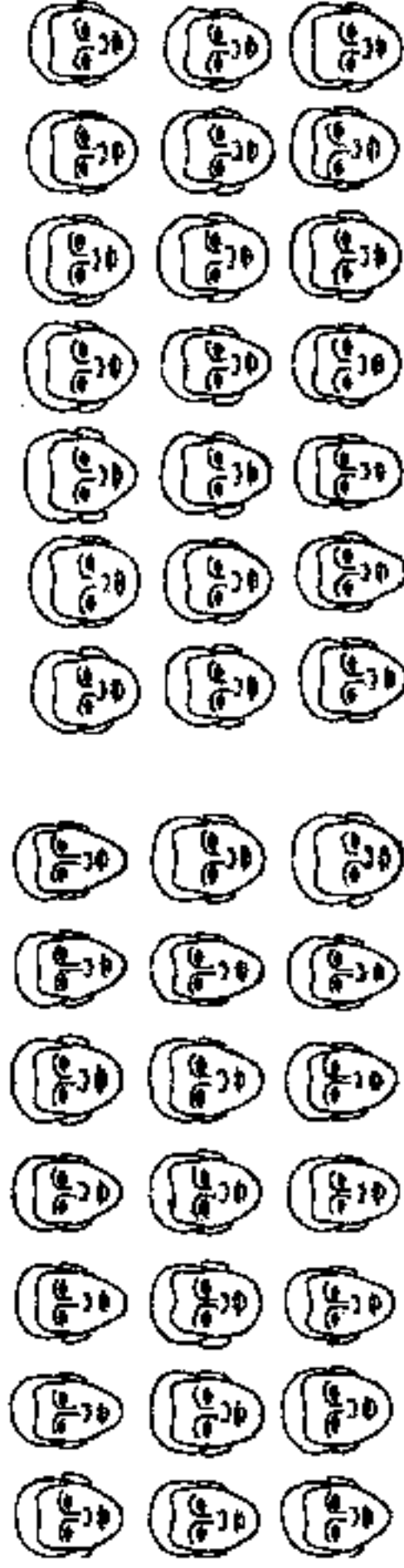
последовательность двумерных задач:

$$Z_2 = n(x^{(1)}, x^{(2)}), Z_3 = n(z_2, x^{(3)}), \dots, Z_p = n(z_{p-1}, x^{(p)}).$$

Признак считается неинформативным, если он не уменьшает ошибку классификации. В результате получаем хорошую ЛДФ. В этой последовательности нормаль к ЛДФ качается по очереди к осям координат. Решение можно улучшить, повторив последовательность в цикле (Колмогоров).

Классификация «по портретам»,

в которых параметры лица соответствуют
признакам объектов.



Сравнение методов дискриминации

В реальной геохимической задаче два класса, размерность $p=12$, количество точек $N_1=N_2=21=15$ (для обучения)+6(для экзамена);

- Ошибок в «типичных» объектах m_t , в случайных - m_r
 - Метод дискриминации
1. Редукция в цикле m_t m_r 2 2
 2. Редукция 2 3
 3. Специалист по таблицам 2 3
 4. Распознавание образов по Бонгарду 1 4
 5. Распознавание «портретов» 2.5 3
 6. ЛДФ с отбором признаков 2 4
 7. То же в «независимых» блоках 2 4
 8. ЛДФ Фишера 3 3
 9. Квадратичная ДФ 4 3
 10. ЛДФ Андерсона-Бахадура 4 4 ▲

4. Прогноз случайных функций

Случайные функции

Будем различать следующие их разновидности.

Случайный процесс x_t , $t \subseteq \mathbf{R}^1$.

Случайное поле x_t , $\mathbf{t} \subseteq \mathbf{R}^p$, $p > 1$.

Аргументы функций (время, пространство) и ниже будут помещаться не в скобках, а в нижних индексах для компактности записей.

Рассмотрим два наиболее часто используемых типа таких процессов.

Стационарный случайный процесс (ССП)

определяется двумя равенствами:

$$\left\{ \begin{array}{l} E x_t = \mu = \text{const}, \\ E(x_s - \mu)(x_t - \mu) = \text{Cov}(x_s, x_t) = c_{|s-t|}. \end{array} \right.$$

Первое показывает, что траектории ССП имеют средний **уровень** μ , около него происходят колебания. Второе равенство указывает на стационарность этих колебаний: **ковариация** значений процесса зависит от расстояния между абсциссами и не зависит от сдвига по оси t .

Ковариационная функция c_τ чётная, достигает максимума, равного дисперсии процесса, при $\tau = 0$. Из чётности следует вещественность преобразования Фурье, дающее **спектральную плотность** (с.п.)

$$f(\omega) = \int_{\mathbf{R}^1} c_\tau e^{-i\omega\tau} d\tau = \int_{\mathbf{R}^1} c_\tau \cos(\omega\tau) d\tau,$$

здесь использовалась формула Эйлера:

$$e^{-iz} = \cos z - i \sin z.$$

С.п. – неотрицательная функция, определённая на \mathbf{R}_+ . Её интеграл ограничен дисперсией процесса:

$$\int_{\mathbf{R}_+} f(\omega) d\omega = c_0.$$

Иногда с.п. называют «энергетической плотностью».

Случайный процесс со стационарными приращениями (СПСП)

Колмогоров (1940):

$$\text{Cov}(x_s, x_t) = (v_{|s|} + v_{|t|} - v_{|s-t|}) / 2.$$

$$\begin{cases} E(x_s - x_t) = 0, \\ E(x_s - x_t)^2 = v_{|s-t|}. \end{cases}$$

За начальную точку может быть взята любая точка траектории. **Структурную функцию** v_τ удобно аппроксимировать степенной: $v_\tau = \omega^2 \tau^\gamma, 0 \leq \gamma \leq 2$ (Яглом). Mandelbrot модель СПСП с такой аппроксимацией назвал **фрактальным броуновским движением**, величину $H = \gamma / 2$ назвал показателем Харста и привёл эффектные примеры.

Локальные свойства СПСП (и ССП)

$$\nu_t = 2(c_0 - c_t)$$

$$c_t = c_0 - \nu_t / 2$$

$$\rho = 2^{\gamma-1} - 1.$$

Корреляция соседних приращений

Непрерывен в среднеквадратичном:

$$\lim_{s \rightarrow t} E(x_t - x_s)^2 = \lim_{s \rightarrow t} \nu_{|s-t|} = 0.$$

Дифференцируем в среднеквадратичном: $\rho_{s \rightarrow t} \rightarrow 1.$

Спрямяем в среднем: $\exists \lim_{m \rightarrow \infty} \sum_1^m E l_i,$

где l_i - длины звеньев аппроксимирующей ломаной.

$\gamma = 0, \rho = -1/2$ «белый шум» с дисперс $\omega^2 / 2$, всюду разрывный,

$\gamma = 1, \rho = 0$ винеровский: непрерыв., не диф., не спрямляема,

$\gamma = 2, \rho = 1$ случайная прямая: непрер., диф., спрямляема

(при $(d/d\tau)\nu(\tau)|_{\tau=0} = 0, \nu(d/d\tau)c(\tau)|_{\tau=0} = 0.$

Прогноз гауссовского ССП

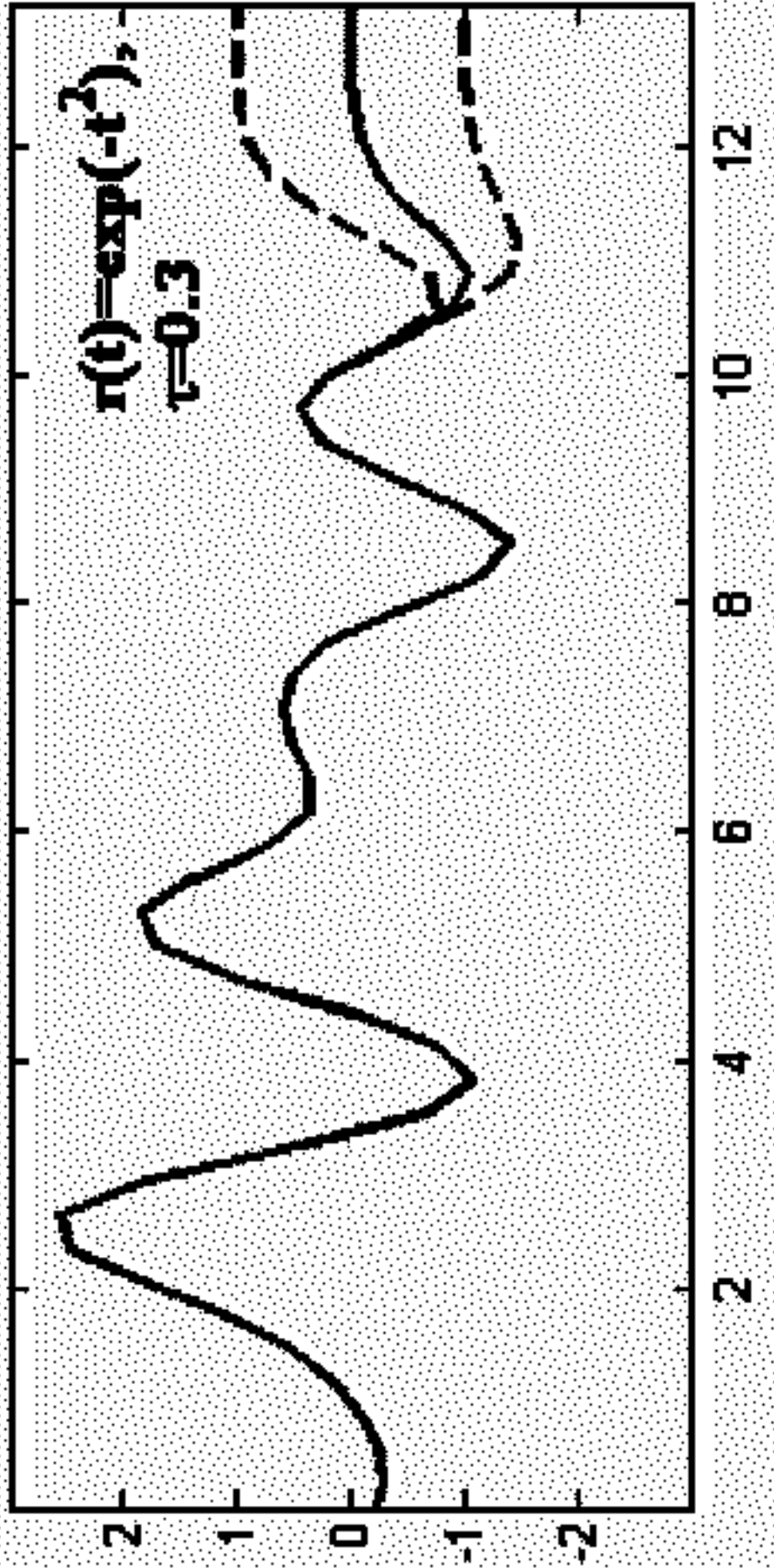
$y_t = x_t - \mu$, шаг h , знаем: $y := (y_{nh}, \dots, y_n, y_0)$,
ковариационную функцию $c := (c_n, \dots, c_{nh}, c_{(n+1)h})$
следующего значения с известными предшествующими
значениями, матрицу ковариаций $C = \{c_{ij}\}_{i,j=0}^n$
известных значений.

Надо найти следующее значение $y_{(n+1)h} = ?$.

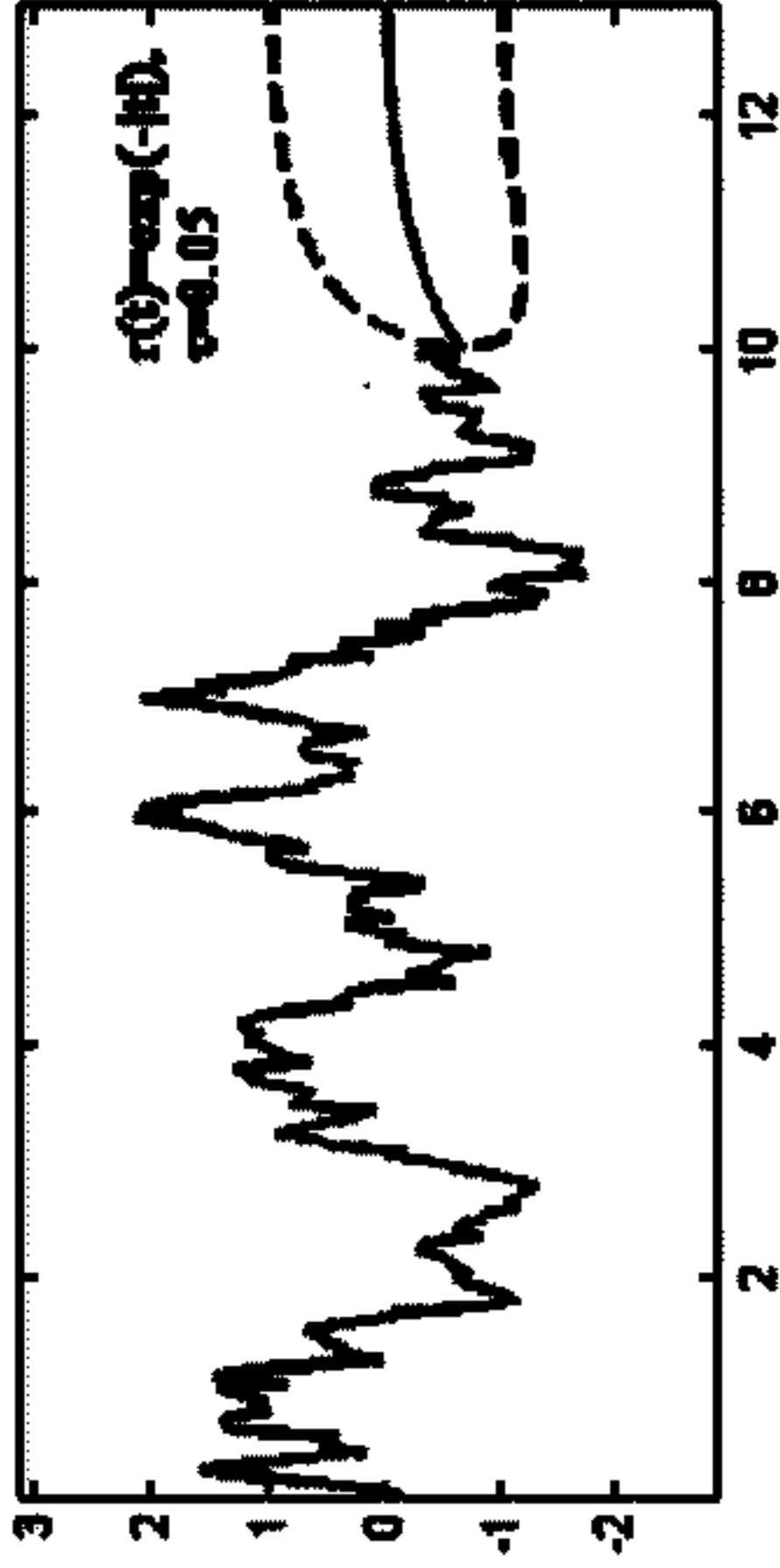
По Андерсону (6):

$$\begin{aligned} E(y_{(n+1)h} | y) &= yC^{-1}c^T, \\ D(y_{(n+1)h} | y) &= c_0 - cC^{-1}c^T. \end{aligned}$$

Прогноз ССП



Прогноз ССП



Прогноз гауссовского СПСП

Приращения $\Delta_k = x_{(k+1)h} - x_{kh}$, в результате $(x_{nh}, x_{(n-1)h}, \dots, x_h, x_0) \rightarrow (\Delta_n, \dots, \Delta_1) =: \Delta$.

Ковариации приращений:

$$c_0 = v_h, \dots, c_k = v_{(k-1)h} / 2 - v_{kh} + v_{(k+1)h} / 2, \dots;$$

для фрактального броуновского движения

$$c_0 = \omega^2 h^\gamma, \dots, c_k = \omega^2 h^\gamma [(k-1)^\gamma / 2 - k^\gamma + (k+1)^\gamma / 2], \dots;$$

матрица $C = \{c_{ij}\}_{i,j=0}^{n-1}$,

ковариации с предыдущими $c := (c_1, \dots, c_n)$.

Прогноз по Андерсону (6)

$$E(\Delta_{n+1} | \Delta) = \Delta C^{-1} c^T,$$

$$D(\Delta_{n+1} | \Delta) = c_0 - c C^{-1} c^T.$$

Авторегрессия (АР)

Последовательность x_1, \dots, x_n , АР k -го порядка:

$$x_{n+1} = \sum_{i=1}^k h_{n-i} x_i + \varepsilon_{n+1}, \quad h_i \text{ коэффициенты, } \varepsilon_i \in N(0, \sigma^2)$$

независимы. Обратная задача: известен прогноз \hat{x}_0 , его дисперсия σ^2 , есть ли соответствующий СП? СПСП не может быть, т.к. его приращение зависит. СПС не может быть по той же причине. Нужна АР с зависимыми приращениями!

Система уравнений:

$$x_{n+1} = \sum_i (a_{n-i} x_i + b_{n-i} y_i) + \varepsilon_{n+1},$$

Нобелевская премия

$$y_{n+1} = \sum_i (c_{n-i} x_i + d_{n-i} y_i) + \varepsilon_{n+1}^*$$

2004 года по экономике.

Изотропное случайное поле

Пусть в области $Q \subset \mathbf{R}^2$ имеется случайная функция x_t , вектор $t \in Q$. В сечениях поля будут СП x_t . Если их вероятностные характеристики совпадают, **поле изотропно**. Для рассмотренных ССП и СПСП будут:

$$\begin{cases} \mathbb{E}x_t = \mu, \\ \text{Cov}(x_s, x_t) = c_{\|s-t\|}; \end{cases} \quad \begin{cases} \mathbb{E}(x_s - x_t) = 0, \\ \mathbb{E}(x_s - x_t)^2 = \nu_{\|s-t\|}. \end{cases}$$

Поле, имеющее сечениями винеровские процессы, называется **полем Леви**. Мандельброт...

Условный случайный процесс (УСП) и подсчёт запасов (ПЗ) рудных месторождений

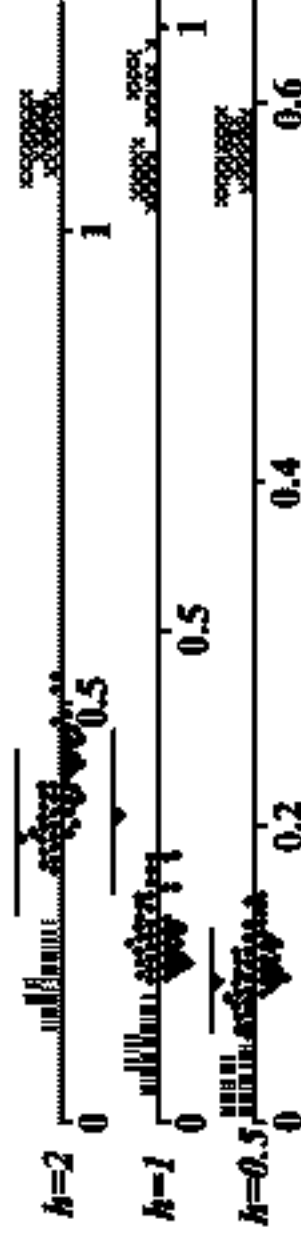
- На отрезке T концентрация x_t , линейные запасы $w = \int_0^T x_t dt$ надо оценить по пробам с ошибками u_i , $Du_i \stackrel{\text{def}}{=} \sigma_i^2$ известны.

Традиционный способ: линейная интерполяция, Погрешность по статистике: $D\bar{x} = Dx/n$.

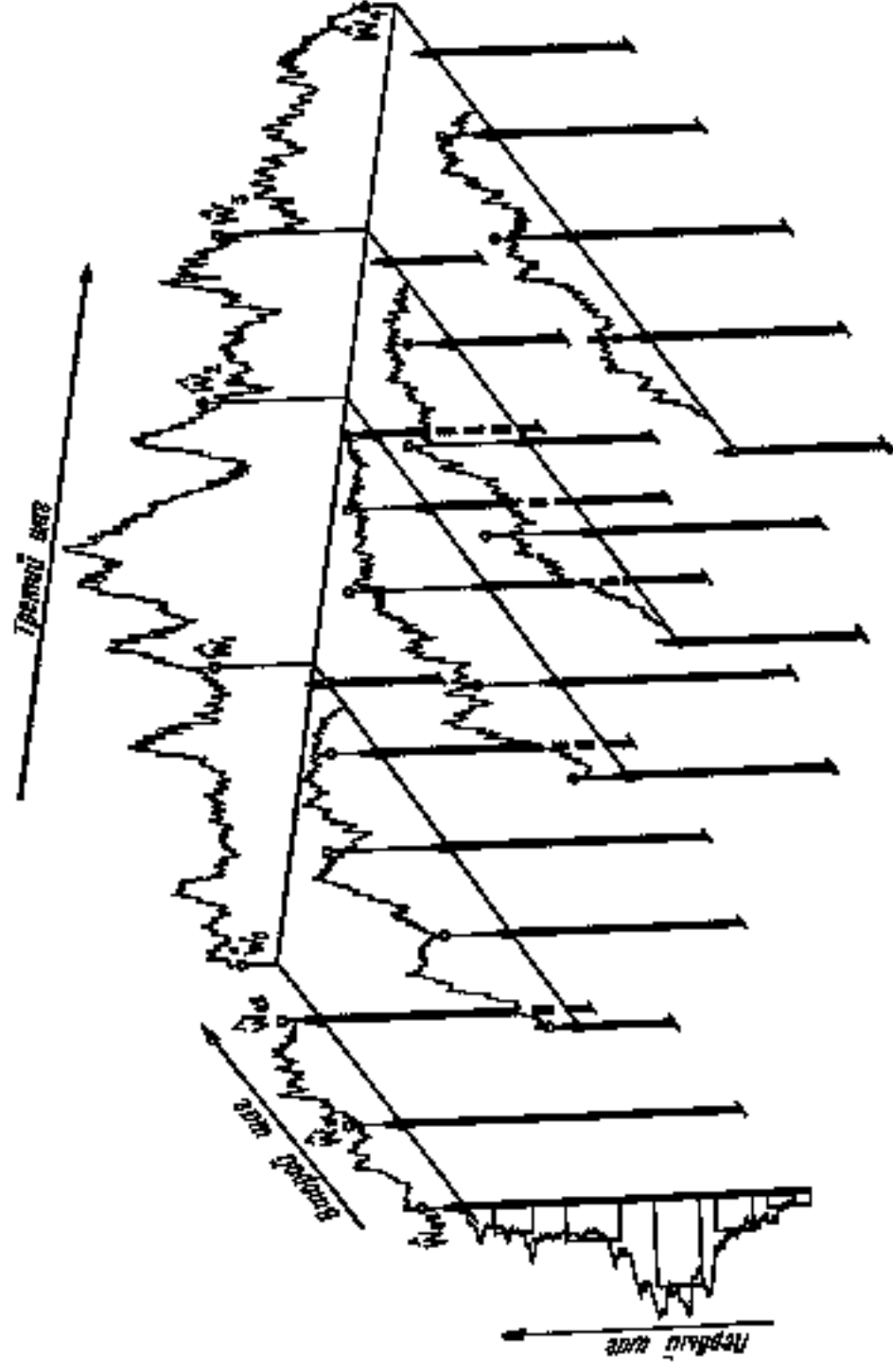
Численное интегрирование гладких функций...

УСП: в точках опробования СП имеет известные (фиксированные) значения (см. «Условное многомерное нормальное распределение»).

Оценки погрешности подсчёта площади рудной жилы по имитации разведки: | численного интегрирования, х-традиционные, точки внизу – условный винеровский процесс, точки вверху – условный СПСП. Над осями доверительные интервалы.



Подсчёт запасов в объёме месторождения, разведанного профилями скважин



Проблема ураганных проб. Концентрации x_t многих руд распределены резко асимметрично, и ПЗ по среднему даёт завышенные значения, хотя среднее является несмещённой оценкой мат. ожидания.

Пусть $x_t = \exp y_t$, а y_t - ССП и $E y_t = \mu$. Тогда траектории УСП между фиксированными точками будут выгнуты вниз, если они выше e^μ , и вверх, если они ниже. В месторождении концентрации аномальны по сравнению с фоновыми, поэтому интеграл по УСП будет меньше, чем даст линейная интерполяция.



Условное случайное поле и ПЗ нефте-газовых месторождений

Скважины располагаются не по профилям, надо рассмотреть плоское поле $Q \in t$, обычно x_t - поле Леви, параметр которого определяется по межскважинным разностям. Разбив Q на треугольники, по возможности изометричные, получаем мозаику условных полей, каждое из них определяется скважинами в вершинах треугольника, одну из которых можно принять за начальную, и решать задачу при известных значениях в двух других численно по формулам условного многомерного распределения.

Выбор шага наблюдений («дисперсия» Аллана)

Если в наблюдаемом процессе Z_t на низкочастотный сигнал с амплитудой ν наложен «белый шум» с дисперсией $\omega^2 / 2$, то при шаге наблюдений τ

$$E(\Delta z_\tau)^2 = \nu^2 \tau^2 / 2 + \omega^2.$$

Разделив эту функцию на τ , получим функцию

$$E(\Delta z_\tau)^2 / \tau = \nu^2 \tau / 2 + \omega^2 / \tau =: a_\tau.$$

Эта функция достигает минимума по τ при таком значении τ_* , при котором дисперсии приращений сигнала и «белого шума» равны.

Выбор количества слагаемых при разложении функции

Гладкая функция $f(x)$ наблюдается в точках $X = \{x_0, \dots, x_n\}$ с помехами: вместо $f_i := f(x_i)$ фиксируются $y_i = f_i + \varepsilon_i$, помехи ε_i одинаково распределены, нормальны и независимы. Ломаная $y(X)$ аппроксимируется суммой ортогональных полиномов $g_j(X)$: $y(X) = \sum_{j=0}^{\infty} a_j g_j(X)$. Остановиться, если

- 1) оценки дисперсии «остатков» d_i $s_1^2 = \sum (d_i - \bar{d})^2 / n$ и $s_2^2 = \sum_{i=0}^{n-1} (d_i - d_{i+1})^2 / (2n)$ примерно равны,
- 2) корреляция соседних остатков равна $-1/2$.

5. Прогноз точечных полей

Поля зависимых точек

В приложениях часто ищут п.р. зависимых точек.

Пусть x и x^* имеют п.р. $f(x) \equiv f(x^*)$ и между ними

действует сила $\xi(x - x^*)$. В условии равновесия

$$g(x) = \int_{\mathbf{R}^1} \xi(x - x^*) f(x^*) dx^* = f * f \equiv 0 \quad - \text{ свёртка.}$$

Преобразование Фурье F превращает свёртку в произведение: $F(g(x)) = F(\xi(x))F(f(x)) \equiv 0$. Противоречие: либо $\xi(x) \equiv 0$, либо $f(x) \equiv 0$. n зависимых точек могут иметь п.р. в выборочном пространстве \mathbf{R}^n .

Распределение межточечных расстояний (МТР)

Выборка $x_1, \dots, x_n \in Q \subset \mathbf{R}^p$, $r_{\alpha\beta} = \|x_\alpha - x_\beta\| \in (0, \rho]$, их $m = C_n^2 = n(n-1)/2$. Разделим $[0, \rho]$ точками $\rho_0 = 0 < \rho_1 < \dots < \rho_k = \rho$ на k интервалов. Пусть m_i - количество МТР, для которых $\rho_{i-1} < r < \rho_i$, а $\delta_i = P\{\rho_{i-1} < r < \rho_i\}$ при равномерном распределении x в Q . Последовательность

$$g_i := m_i / (\delta_i m), \quad i = 1, \dots, k$$

будет гистограммой МТР, нормированной равномерным распределением. Очевидно, $E m_i = \delta_i m$. $D m_i = ?$

Теорема. Пусть $\mathbf{x}_1, \dots, \mathbf{x}_n \in Q \subset \mathbf{R}^p$, $p \geq 1$, имеют в Q распределение,

абсолютно непрерывное относительно меры Лебега. Тогда МТР между ними асимптотически ($n \rightarrow \infty$) попарно независимы (и при зависимых точках).

$$n = 10 \Rightarrow$$

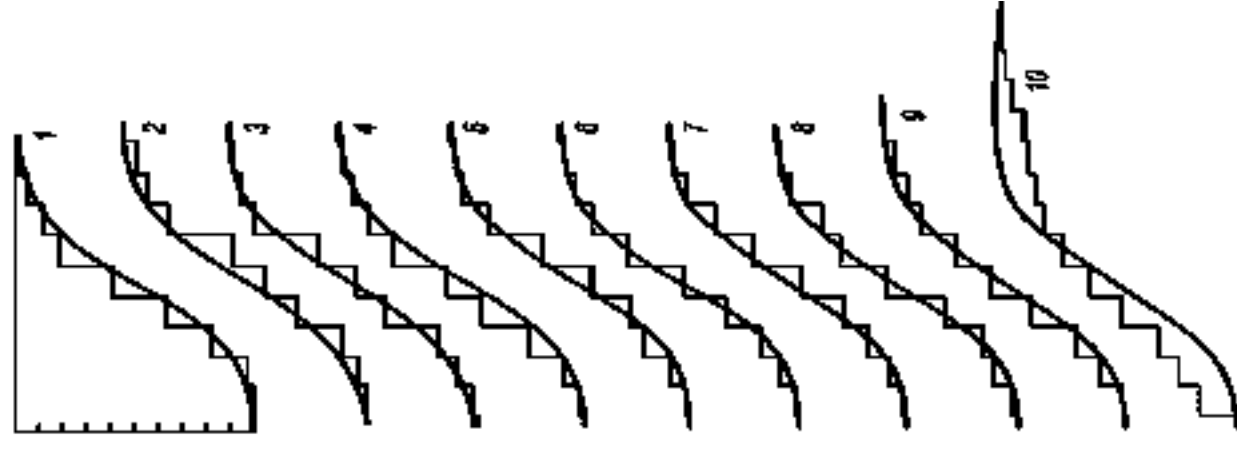
Следствие. Если для всех i

$\delta_i = cn^{-1-\zeta} \rightarrow 0$, $0 < \zeta < 1$, то при равномерном распределении \mathbf{x} в Q вектор из

величин $b_i := (m_i - \delta_i m) / \sqrt{\delta_i (1 - \delta_i) m}$

имеет предельным распределением

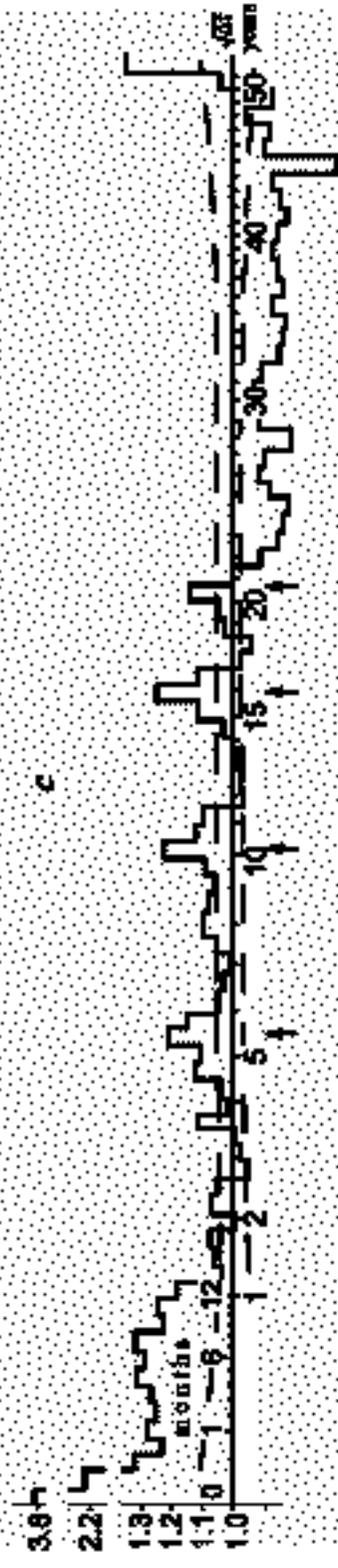
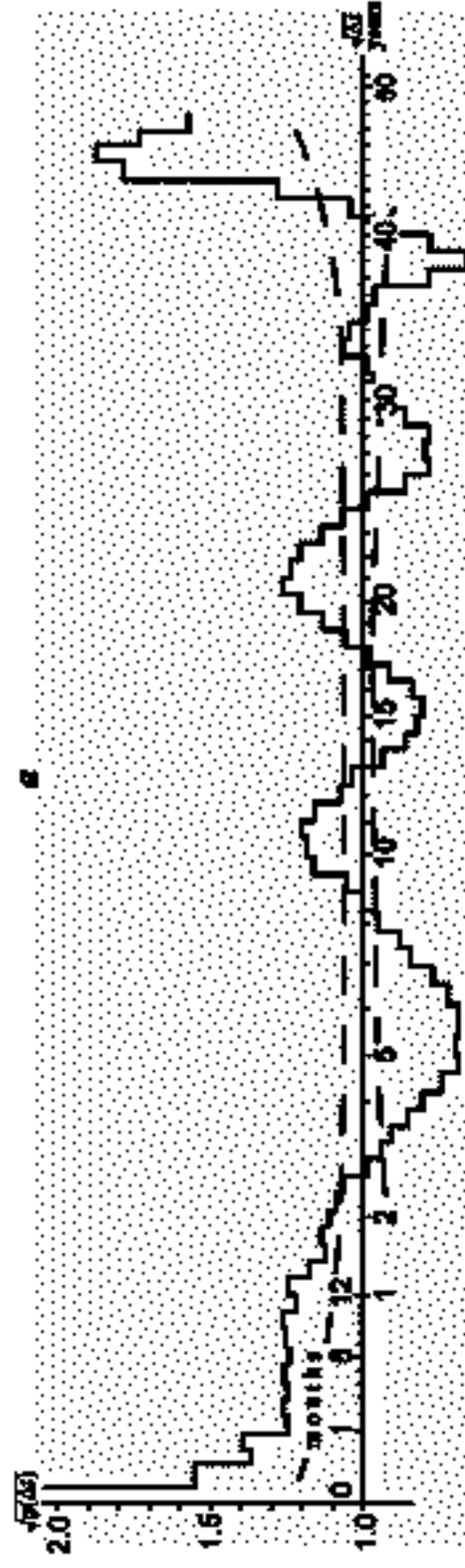
$$N_k(0, I).$$



Функции g (и доверительный интервал)

для солнечных вспышек,

глубокофокусных и нормальных землетрясений



Распределение со стационарными расстояниями (РСРС)

Выборка $\{\mathbf{x}_1, \dots, \mathbf{x}_n\} =: \mathbf{X}_n$ при равномерном распределении \mathbf{x} в Q имеет п.р.м(\mathbf{X}_n), а МТР п.р. $\delta(r)$.

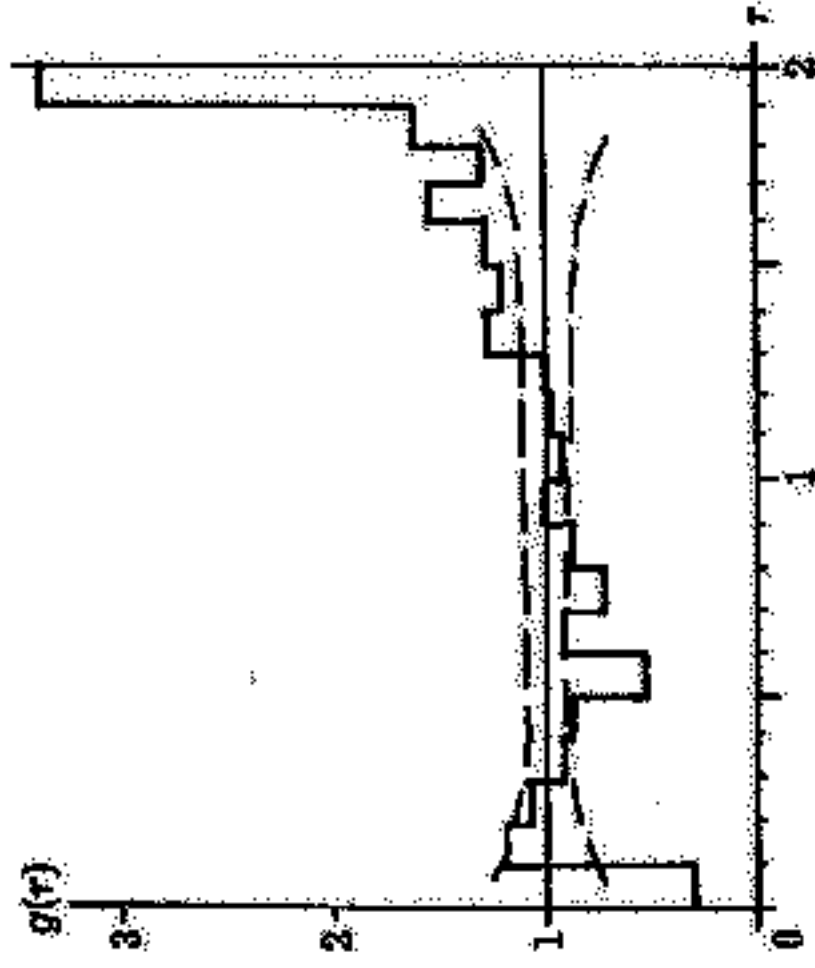
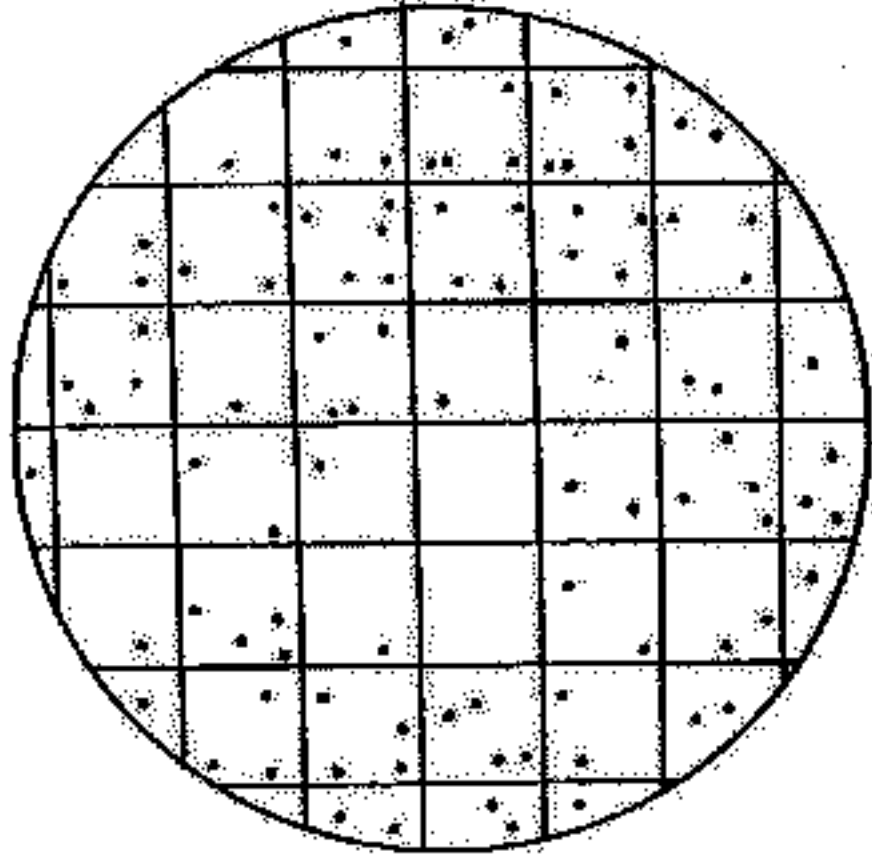
В РСРС п.р. $\sum_{\mathbf{X}_n \in Q^n}$ равна

$$w(\mathbf{X}_n) = \mu(\mathbf{X}_n) \prod_{\alpha < \beta} \gamma(r_{\alpha\beta}), \quad \gamma(r) \geq 0, \quad \int_0^{\rho} \gamma(r) \delta(r) dr = 1.$$

Теорема. Если параметрическая функция $\gamma(r)$ кусочно-постоянная: $\gamma(r) = \gamma_i$ при $\rho_{i-1} < r < \rho_i$, и $\sum \gamma_i \delta_i = 1$, то нормированная гистограмма g_1, \dots, g_k является ОМП для величин $\gamma_1, \dots, \gamma_k$.

Изотропное поле точек со стационарными расстояниями

- Бактерии в чашечке Петри по Феллеру



Прогноз изотропных полей со стационарными расстояниями

- Выборка $\{\mathbf{X}_1, \dots, \mathbf{X}_n\} = \Sigma_n$ с п.р. $w(\Sigma_n)$.

П.р. новой точки \mathbf{X}_{n+1} при известных точках Σ_n :

$$h(\mathbf{X}_{n+1} | \Sigma_n) = \frac{w(\Sigma_{n+1})}{w(\Sigma_n)} = c_n \prod_{\alpha=1}^n \gamma(r_{\alpha, n+1}).$$

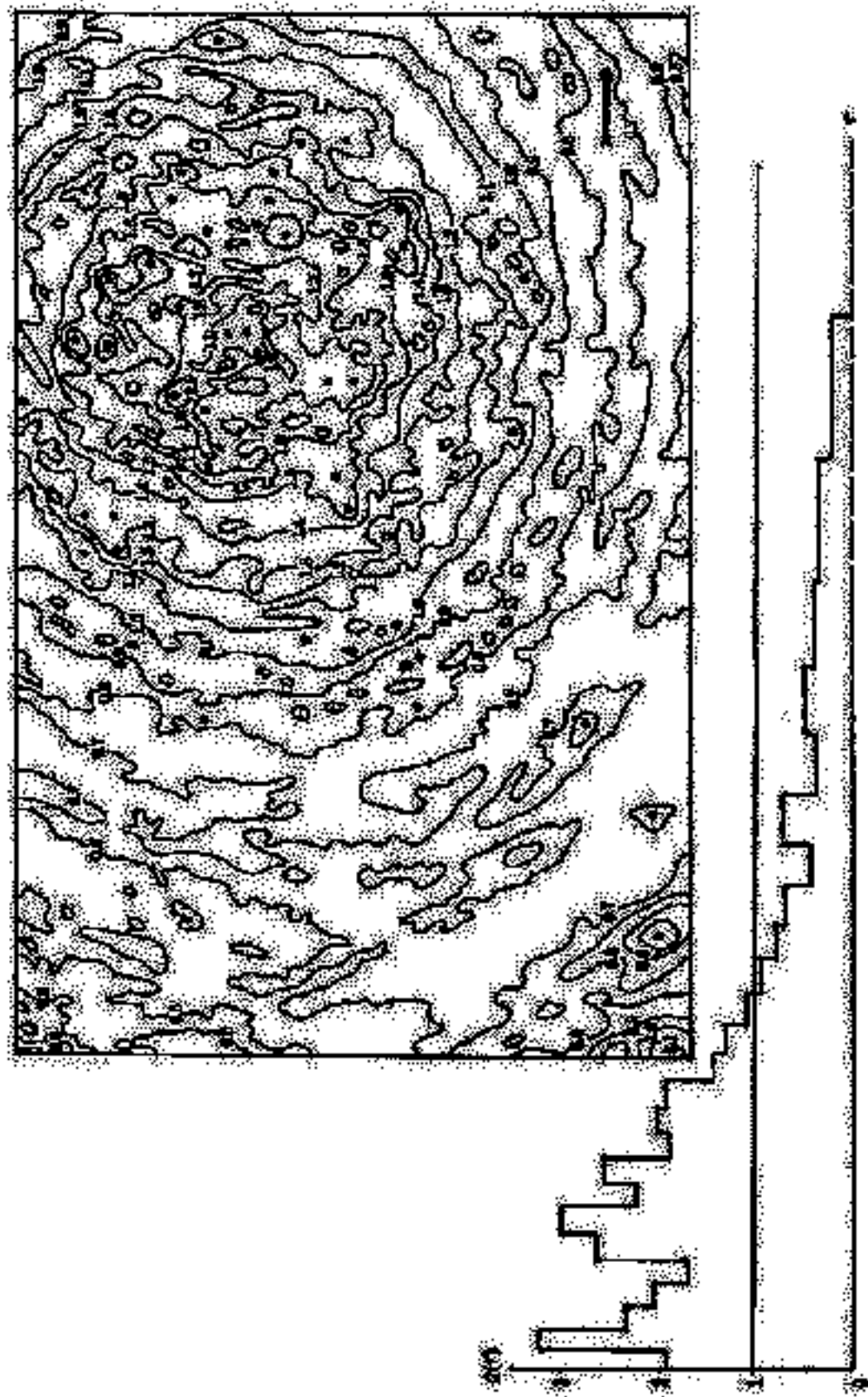
Она достигает максимума там же, где её логарифм

$$\log h(\mathbf{X}_{n+1} | \Sigma_n) = \log c_n + \sum \log \gamma(r_{\alpha, n+1}).$$

Константа $\log c_n$ не зависит от \mathbf{X}_{n+1} , её опустим. Чтобы не зависеть от n , осредним логарифмы и получим прогнозную функцию

$$f(\mathbf{X}_{n+1} | \Sigma_n) = \prod_{\alpha=1}^n [\gamma(r_{\alpha, n+1})]^{1/n}.$$

Прогноз кимберлитовых трубок в алмазоносном поле



Анизотропное поле точек со стационарными разностями

- При $p \geq 2$ разности $\mathbf{h}_{\alpha\beta} = \mathbf{X}_\alpha - \mathbf{X}_\beta \in \mathcal{X}$.

Модель анизотропного поля точек со стационарными разностями строится аналогично. П.р, точек равна $w(\Sigma_n) = u(\Sigma_n) \prod_{\alpha < \beta} \gamma(\mathbf{h}_{\alpha\beta})$, $\gamma(\mathbf{h}) \geq 0$, $\int_{\mathcal{X}} \gamma(\mathbf{h}) \delta(\mathbf{h}) d\mathbf{h} = 1$, где $\delta(\mathbf{h})$ есть п.р. вектора \mathbf{h} в \mathcal{X} .

Теорема об ОМП справедлива, как и условная п.р.

$$h(\mathbf{x}_{n+1} | \Sigma_n) = w(\Sigma_{n+1}) / w(\Sigma_n) = c_n \prod_{\alpha=1}^n \gamma(\mathbf{h}_{\alpha, n+1}).$$

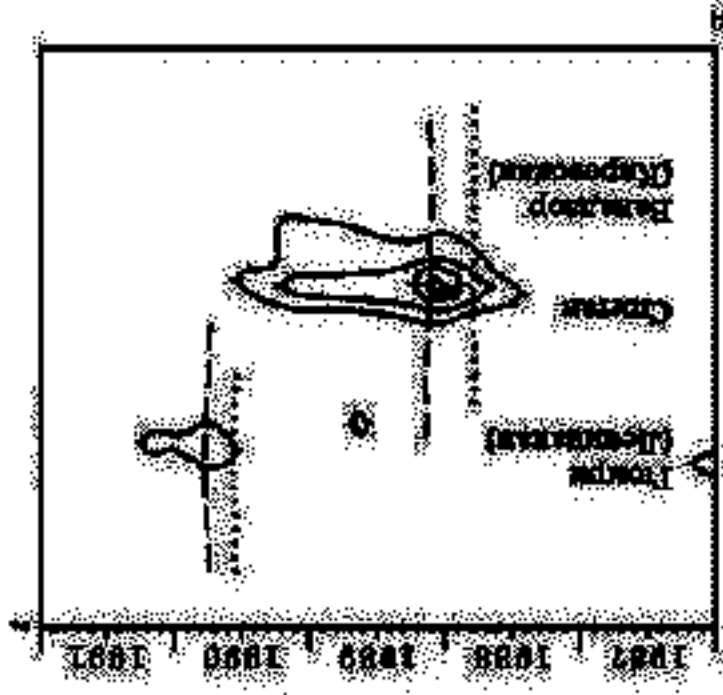
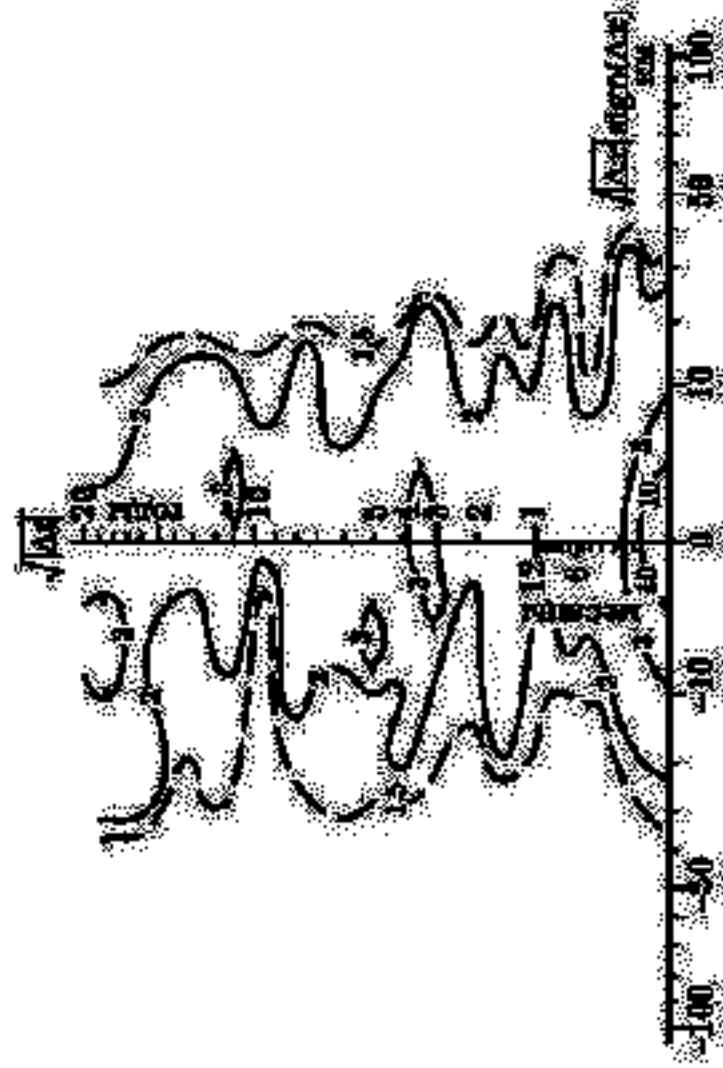
Из неё получаем прогнозную функцию

$$f(\mathbf{x}_{n+1} | \Sigma_n) = \left[\prod_{\alpha=1}^n \gamma(\mathbf{h}_{\alpha, n+1}) \right]^{1/n}.$$

Спитак, 1988

g-функция

ретропрогноз



Личный рекорд

**Директор Ин-та вулканологии акад. С.А.Федотов
Ректору МГУ академику В.А.Садовничему.**

«... 10 декабря 1998 г. А.М.Шурыгин со студентами продемонстрировали мне на дисплее компьютера зависящие от времени прогнозные функции для разных зон Камчатки на 1999 г. и обратили внимание на опасность для Петропавловска-Камчатского 8 марта 1999 г.. Именно в этот день там произошли два землетрясения с магнитудами 7,0 и 5,7. Они были сильнейшими землетрясениями на Камчатке за прошедший год. Такая точность сделанного прогноза --- указана дата за три месяца --- является высокой для мировой практики прогноза землетрясений...»

Примеч.: это землетрясение сильнейшее до сих пор!

Некоторые выводы

Сильные землетрясения предопределены последовательностью предшествовавших умеренных.

Никакие подкоровые движения не могут работать

с такой точностью, но может небесная механика, что было ясно после работ Кельвина в XIX веке.

Классическая механика не объясняет механизм

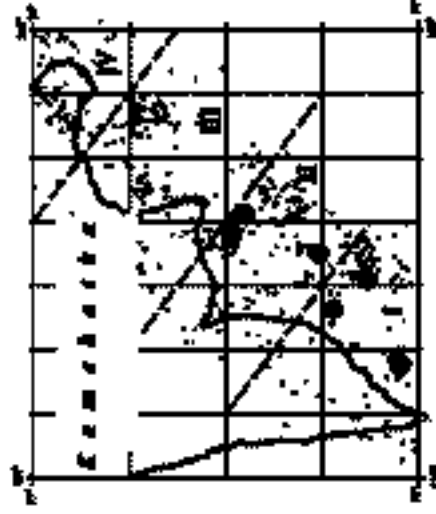
образования землетрясений, а Воронков может объяснить, что такое землетрясение.

Приливная теория Ньютона требует коррекции.

Можно искать лучшую линейную функцию землетрясений – предикторов.

Internet: www.ccas.ru/cito/eq.

СКОПЬЗЯЩИЙ НЕЙРОН И ПРОГНОЗ СИЛЬНЫХ ЗЕМЛЕТРЯСЕНИЙ

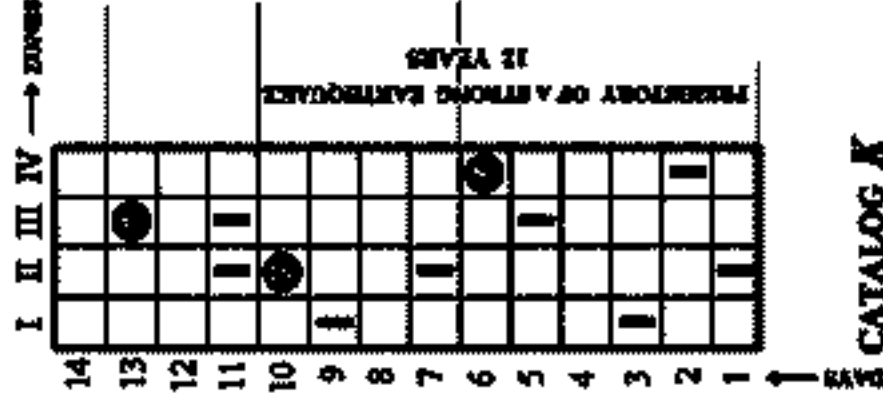


Задача:

Линейный регион с сильными и умеренными землетрясениями (ЗТ). Предсказать сильные ЗТя по предшествовавшим умеренным.

Гипотеза:

Сильные ЗТя имеют похожую предысторию

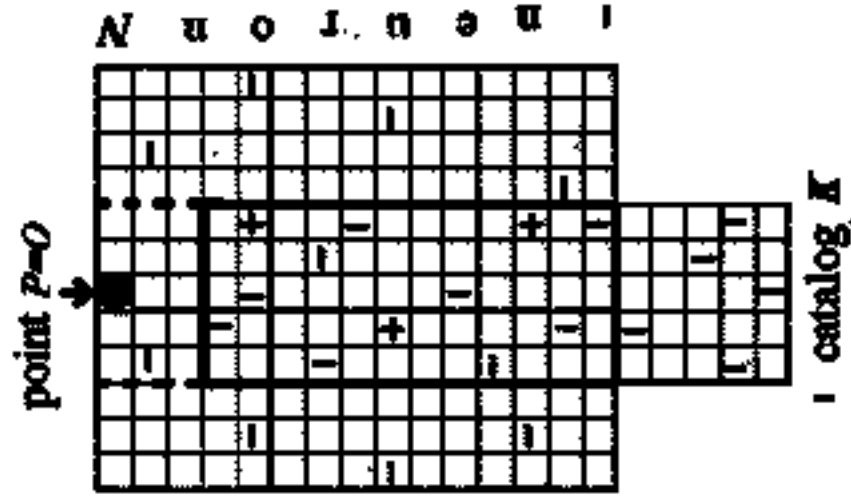


Простейшее

решение.

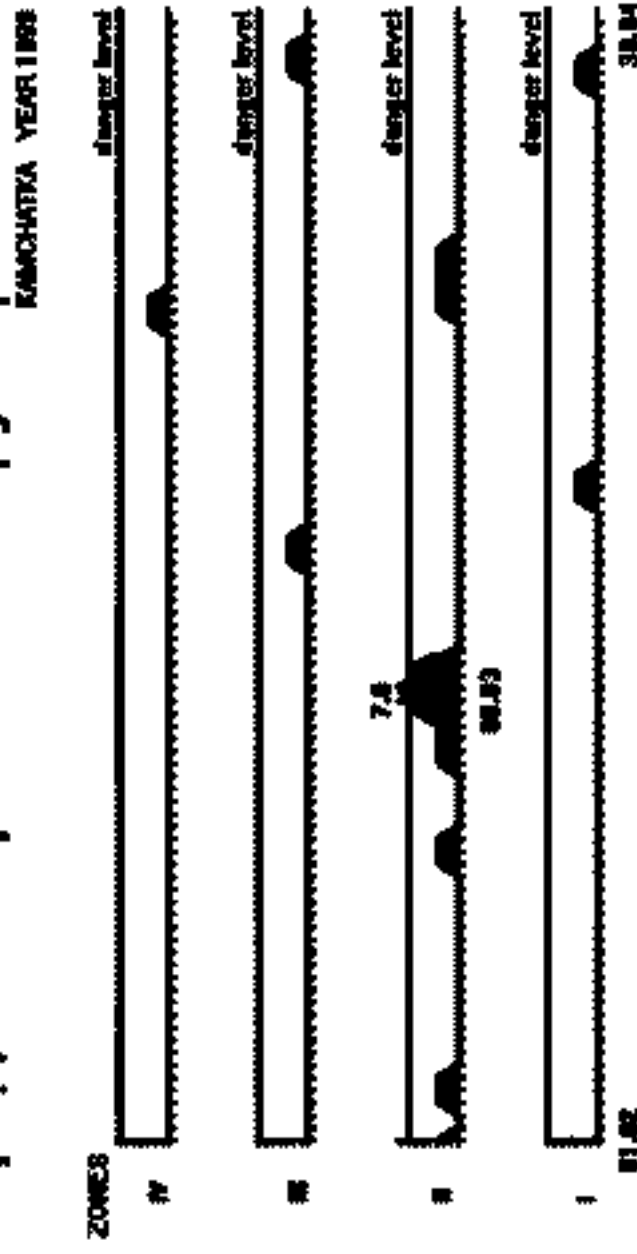
Множество ЗТий (зона, день) образует булеву матрицу-каталог →

Прогноз в точке скользящим нейроном



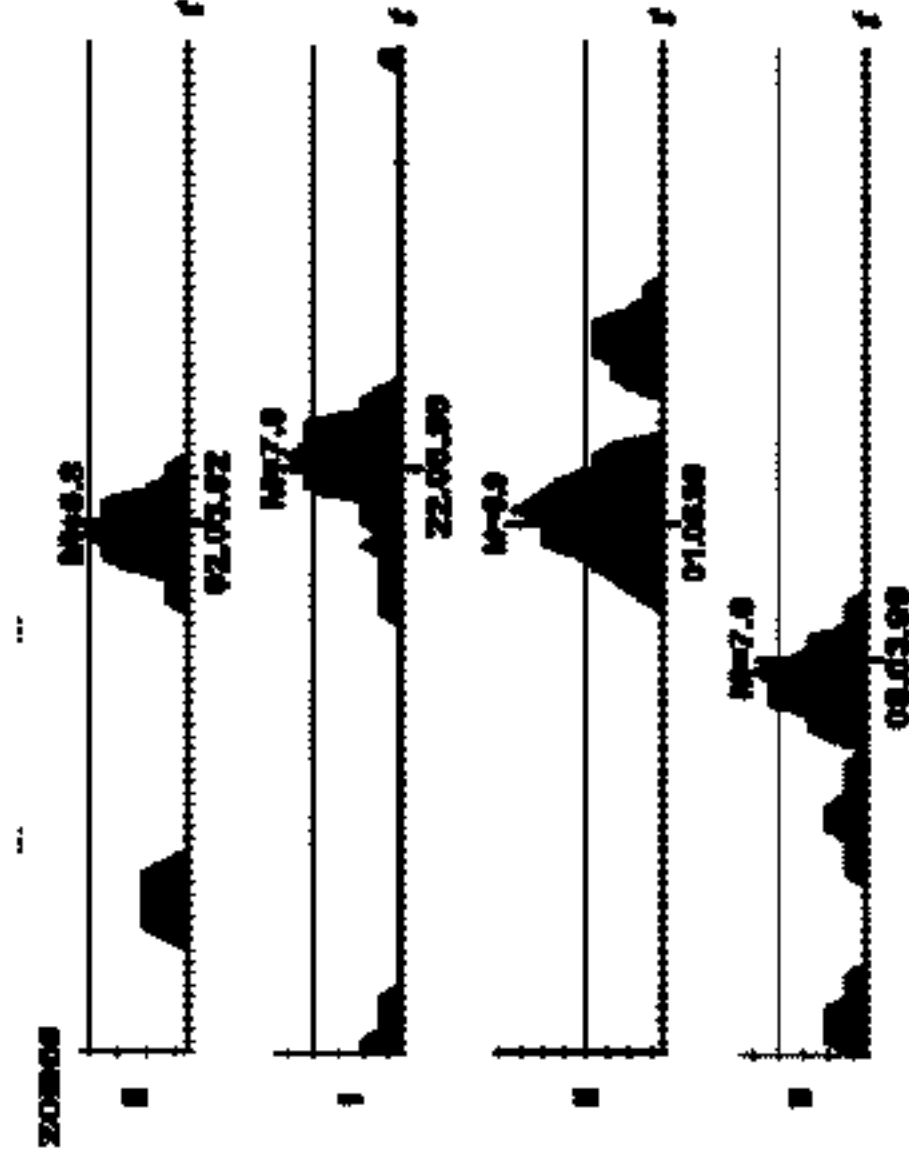
Прогнозная функция $d(P) = c \cdot K \otimes N$,
 \otimes знак скалярного умножения

со скользящим окном W ,
 с определяет уровень опасности.
 Нейрон скользят точкой O в будущем,
 определяя прогнозные функции зон:



Камчатка:

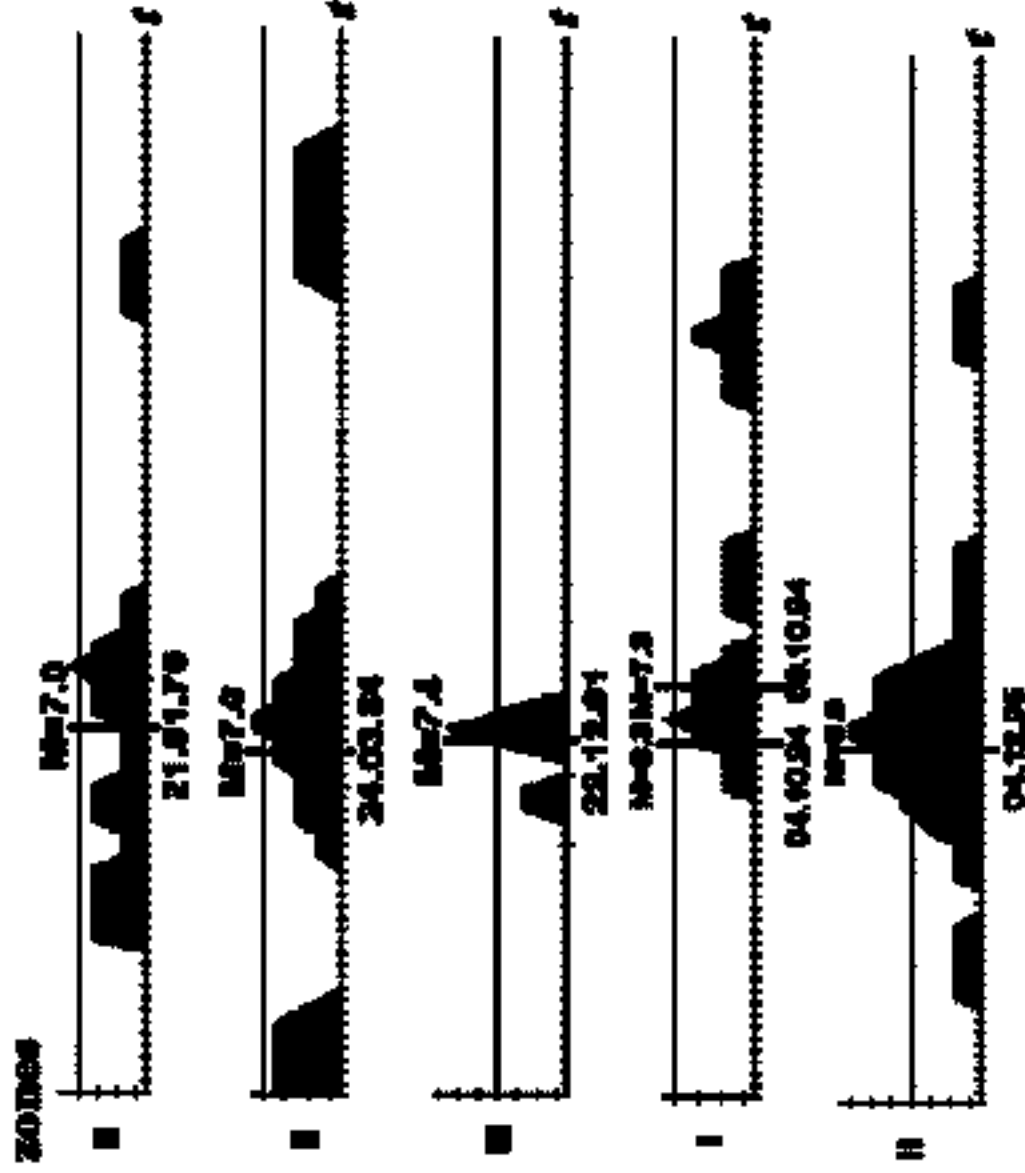
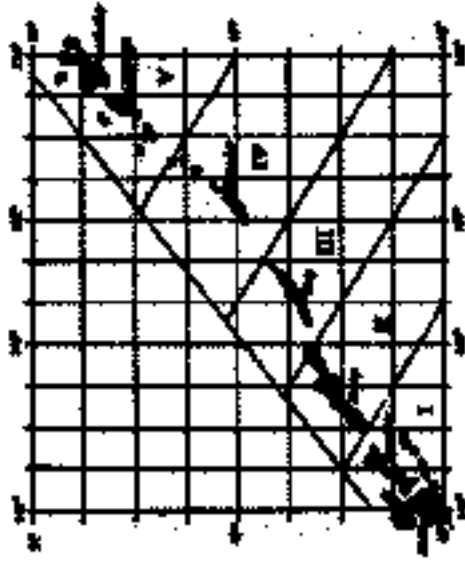
Прогнозные функции для предсказуемых землетрясений. Горизонтальные Оси t градуированы в сутках.

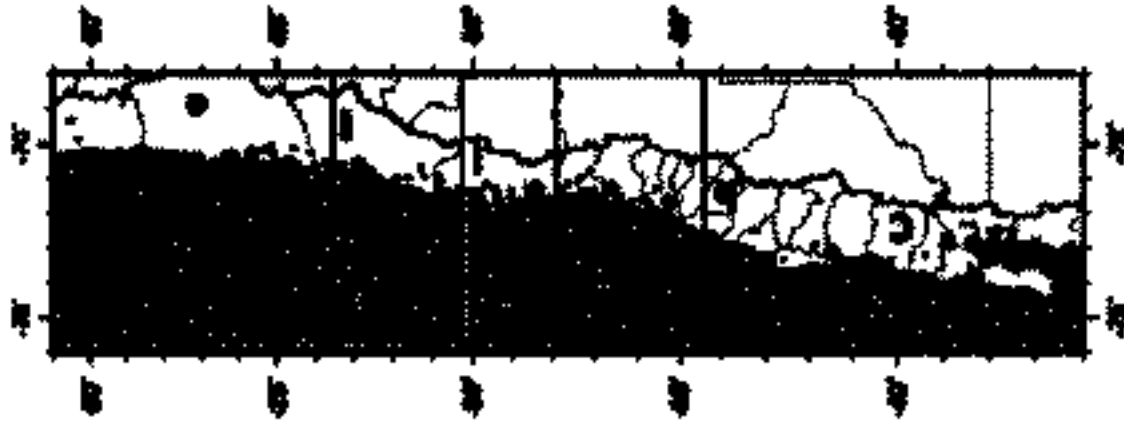


Курильские острова

острова

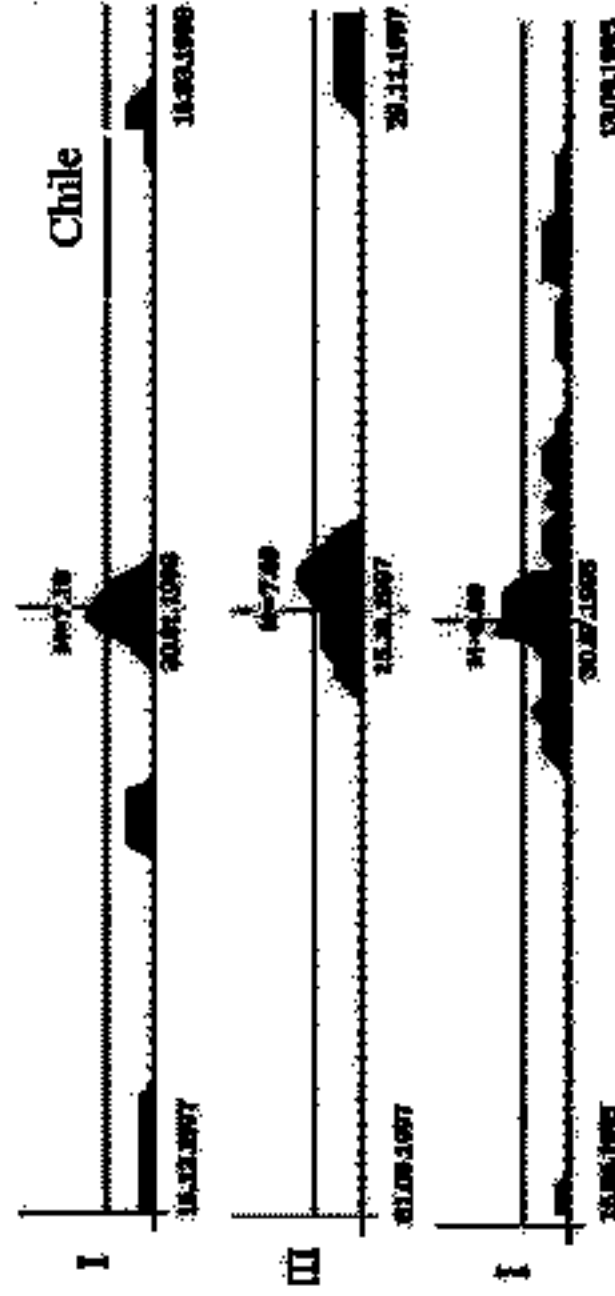
Прогнозные функции для предсказуемых землетрясений. Горизонтальные рованы в сутках.



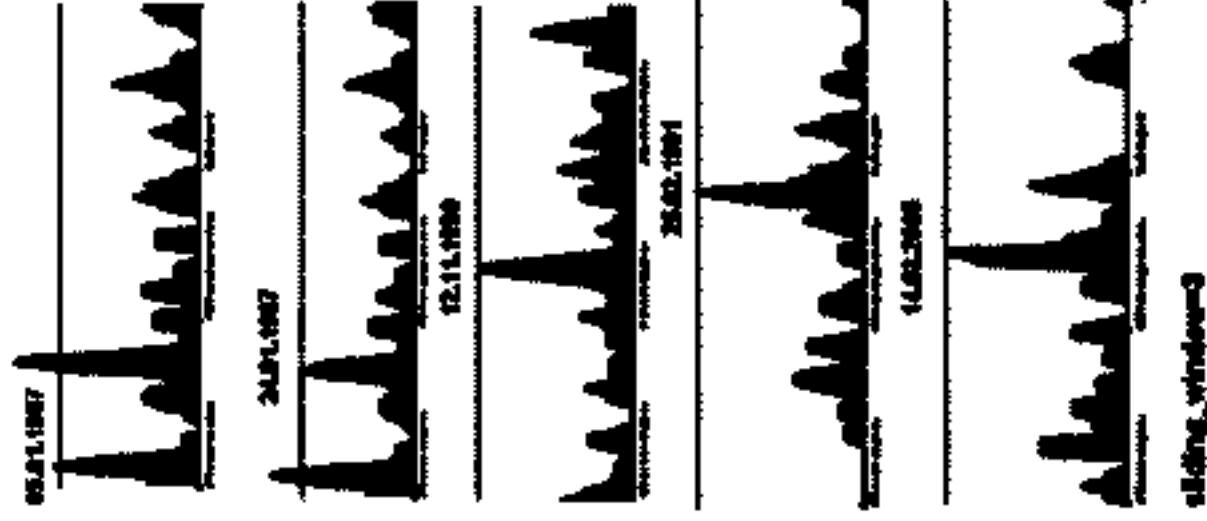
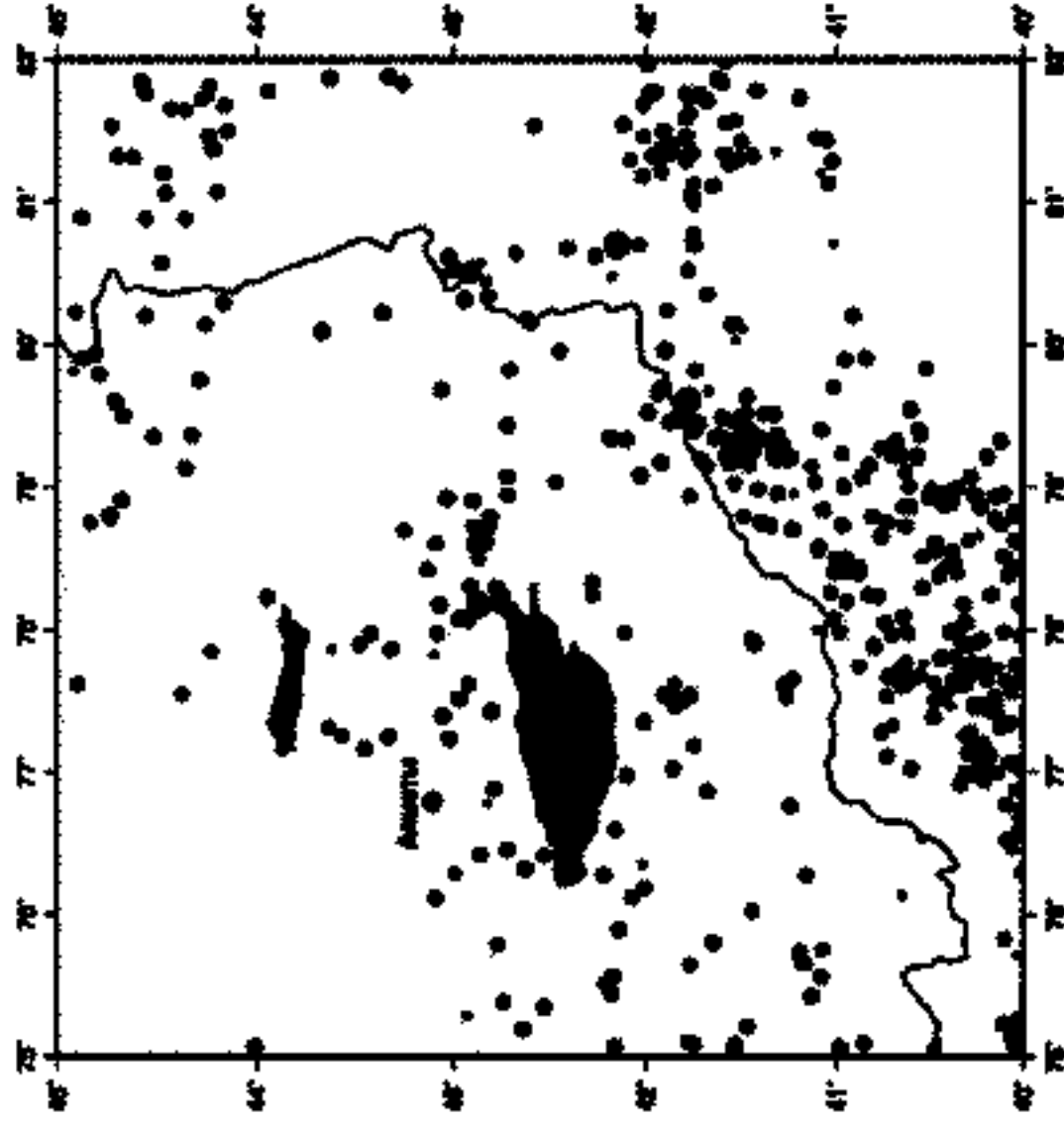


ЧИЛИ

Прогнозные функции для предсказуемых землетрясений. Горизонтальные оси t градуированы в сутках.



Южный Казахстан



Результаты скользящего экзамена

регион	количество сильных землетрясений	количество прогнозируемых сильных землетрясений	ответственный за регион
Камчатка	6	4	М.Стригунова
Курилы	9	6	
Алеуты	8	6	
Чили	5	4	Т.Крупнова
Ю. Казахстан	6	5	Р.Сатаев

70% сильных ЗТй прогнозируются на скользящем экзамене

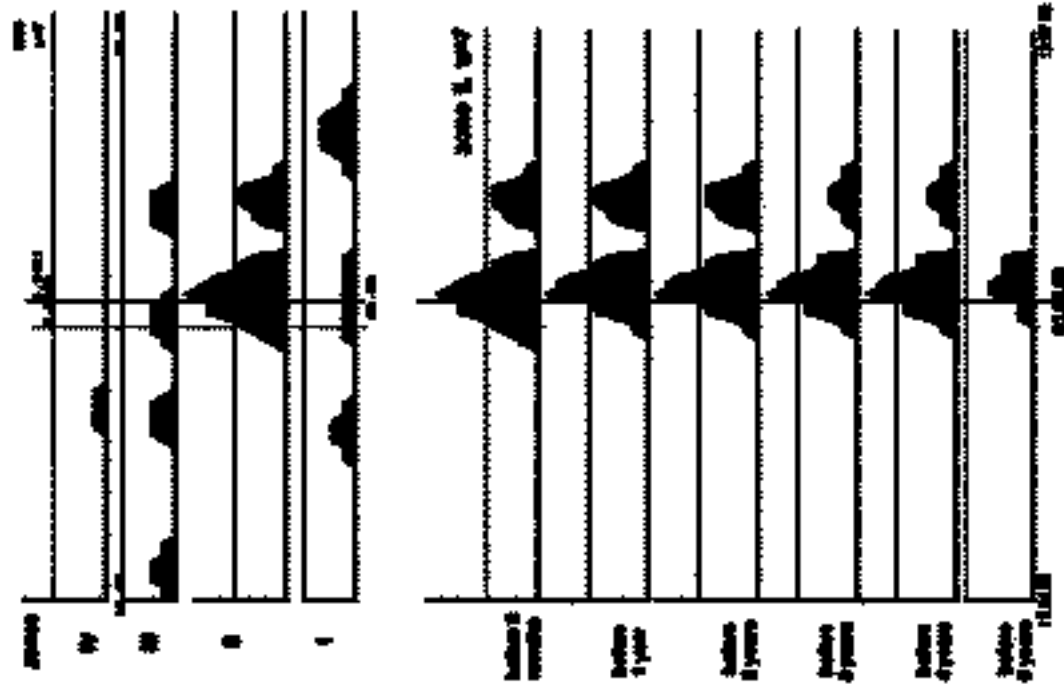
Прогноз с точностью до номера зоны.

Максимальная ошибка по времени ± 3 дня.

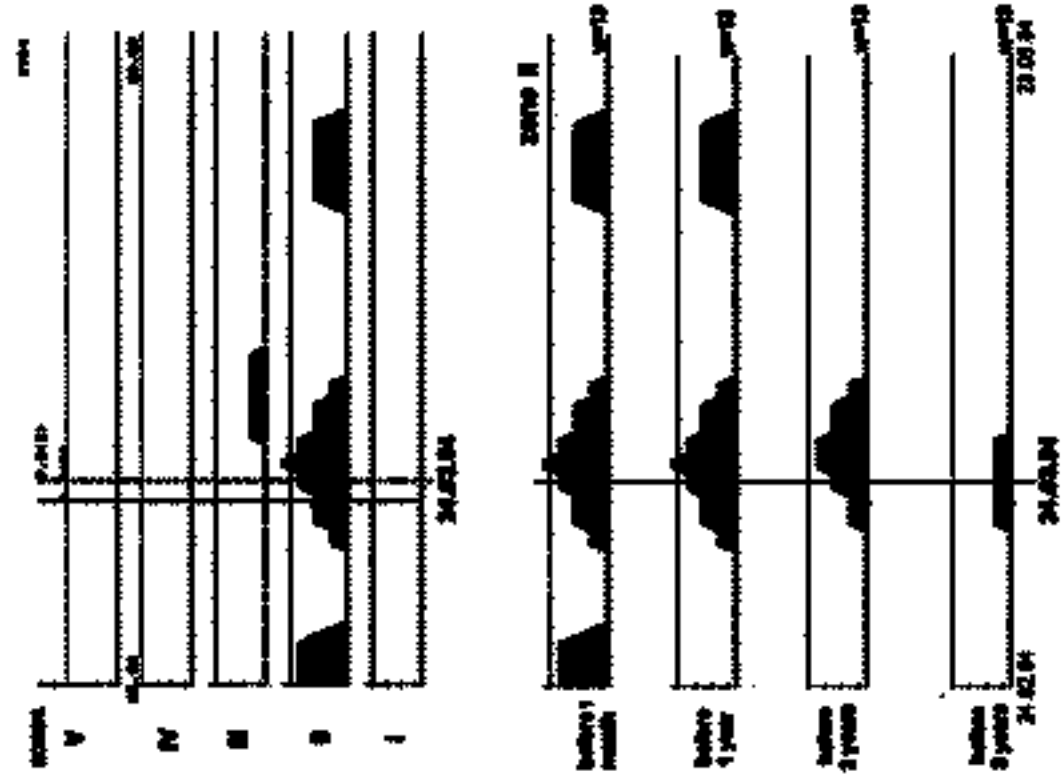
Нет ложных тревог!

Долгосрочный прогноз

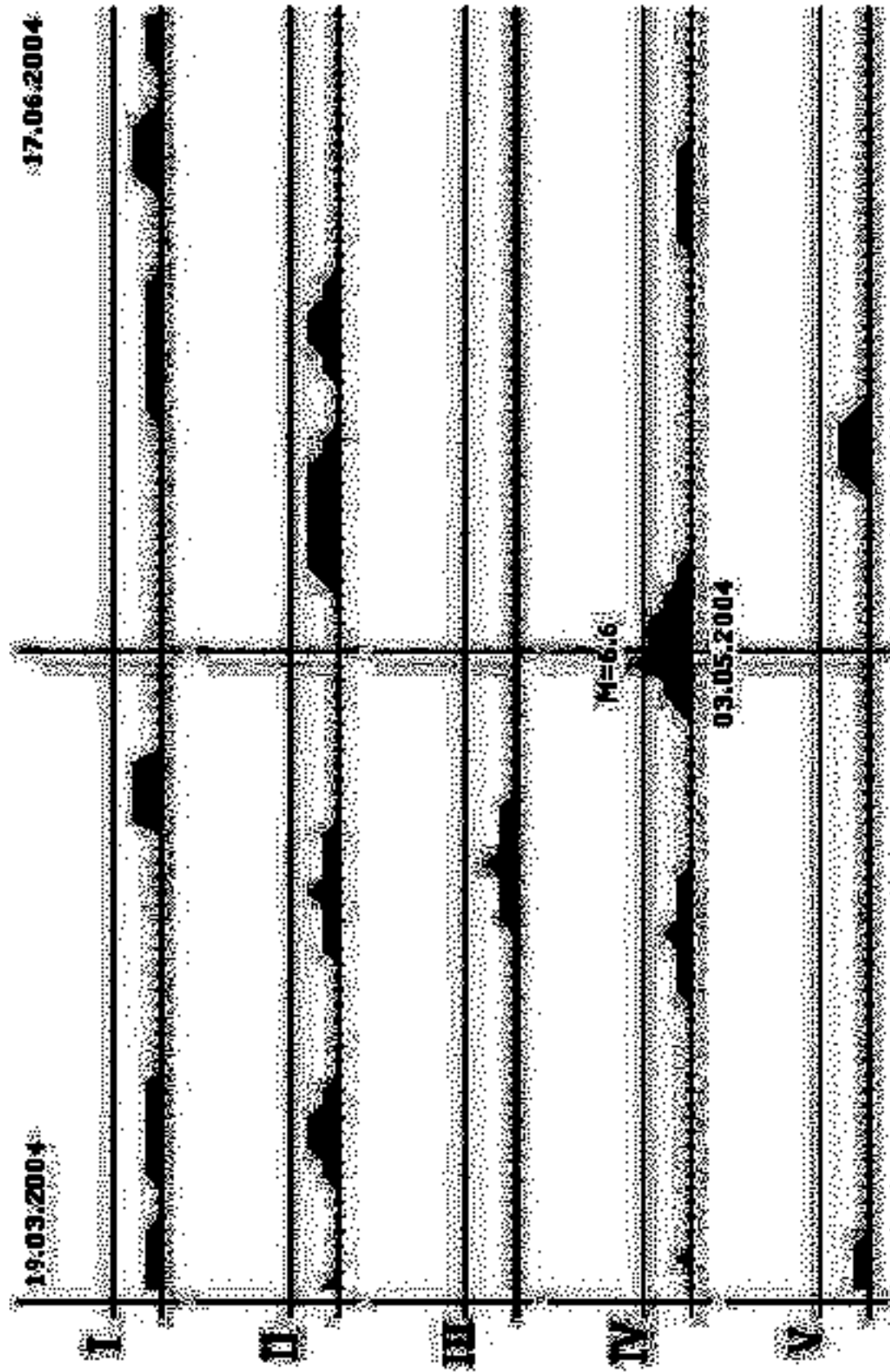
Камчатка



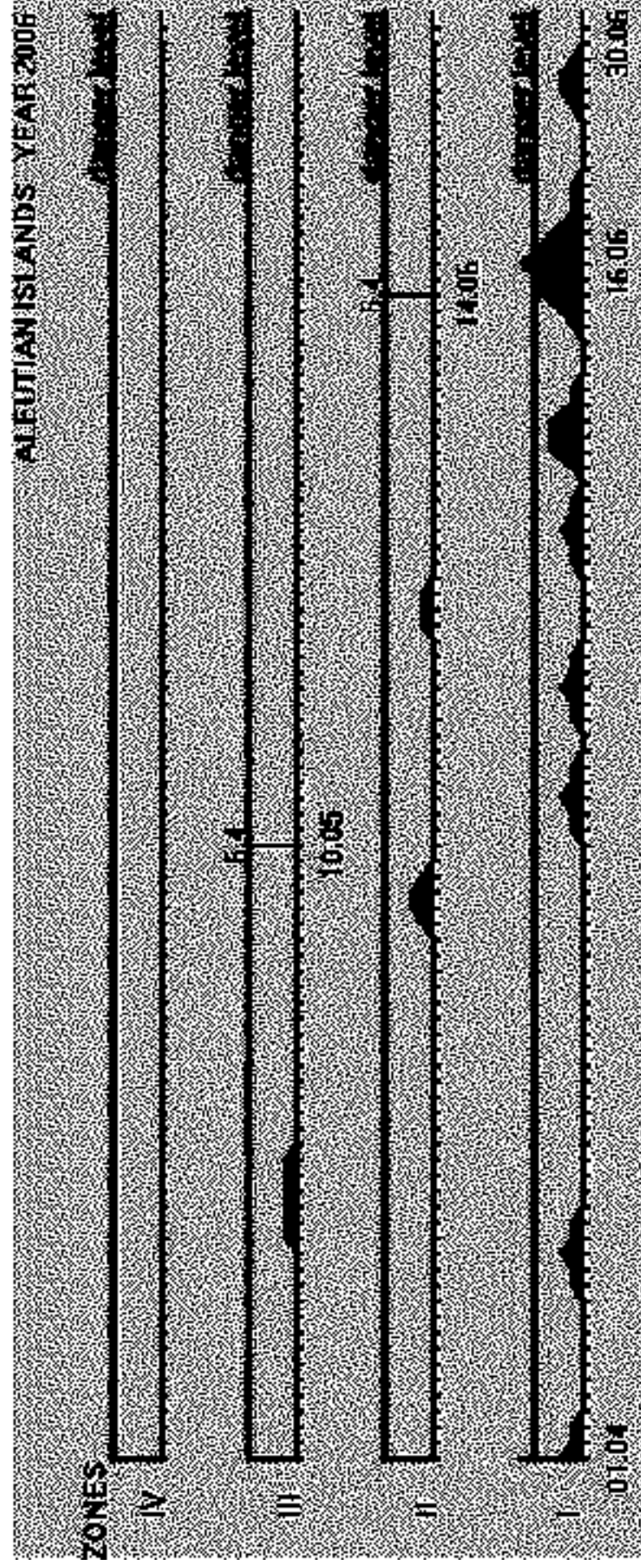
Курильские острова



Землетрясение 03.05.2004 в Чили (по Т. Крупновой)



Землетрясение 14.06.2006 на Алеутских островах (по М.Стригнуновой)



Оценка плотности распределения

Симметричная п.р. Составляет $f(x) = f(-x)$,

её МТР $\chi_{\alpha} = \chi_{\beta}$ имеют п.р.

$$\begin{aligned}v(h) &= \int_{\mathbb{R}^1} f(x)f(x-h)dx = \\ &= \int_{\mathbb{R}^1} f(x)f(h-x)dx = f * f.\end{aligned}$$

$$F(v) = F(f * f) = [F(f)]^2.$$

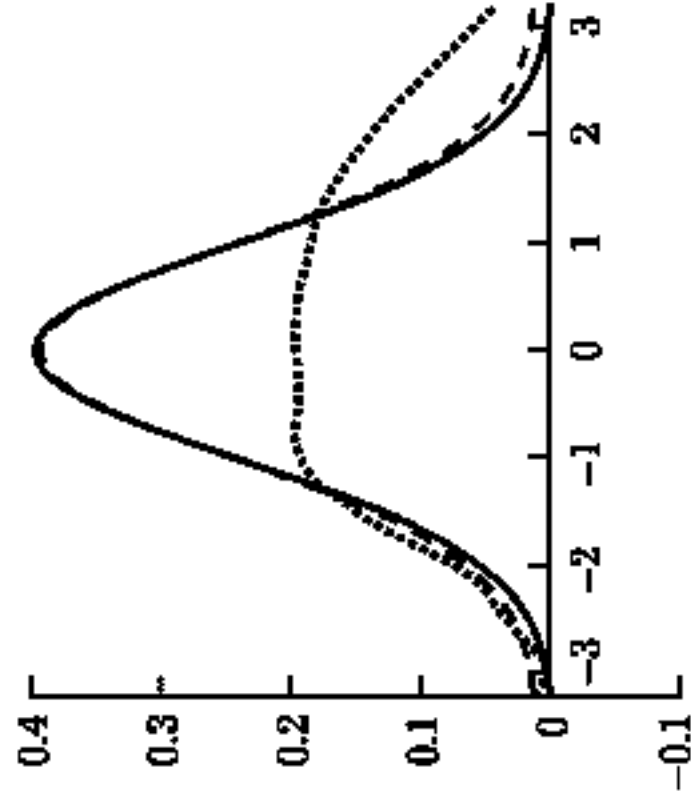
Пусть $F(v) \geq 0$:

$$\sqrt{F(v)} = F(f),$$

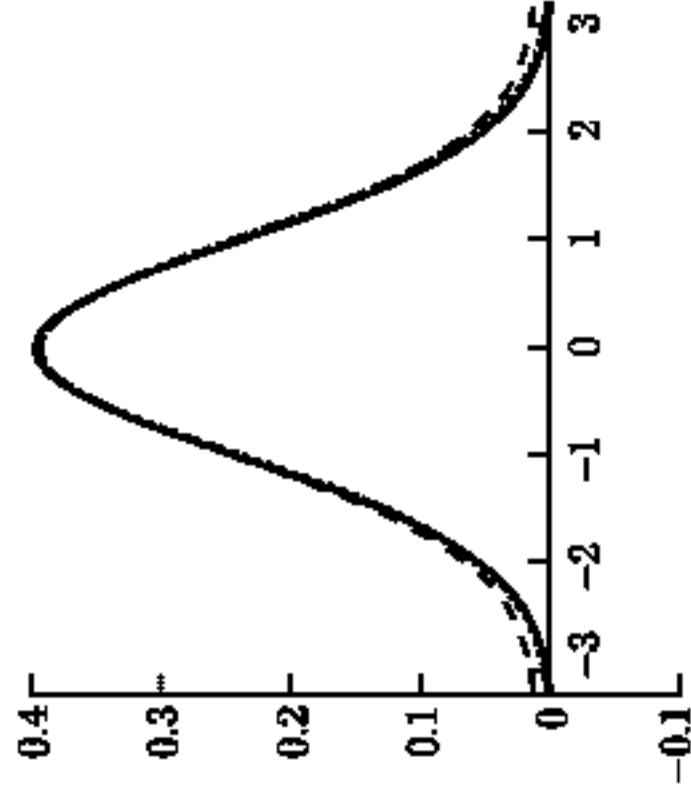
$$F^{-1}[\sqrt{F(v)}] = f.$$

Моделирование гистограмм для нормального распределения

$$n_1 = n_2 = 100$$



$$n_1 = 100, n_2 = 100000$$



МТР в распознавании образов

Теорема. Точки первого класса $\mathbf{x}_1^1, \dots, \mathbf{x}_{n_1}^1 \in Q_1 \subseteq \mathbf{R}^p$ и точки второго класса $\mathbf{x}_1^2, \dots, \mathbf{x}_{n_2}^2 \in Q_2 \subseteq \mathbf{R}^p$ имеют п.р., абсолютно непрерывные относительно меры Лебега.

Тогда МТР $r_{\alpha\beta}^x = \|\mathbf{x}_\alpha^1 - \mathbf{x}_\beta^2\|$, $\alpha = 1, \dots, n_1$, $\beta = 1, \dots, n_2$, асимптотически ($n_1, n_2 \rightarrow \infty$) попарно независимы.

Заметим, что теорема верна и для зависимых точек. Используем теорему для проверки гипотезы о разделимости двух классов.

Проверка делимости классов

$D_1(r)$ - функция распределения (ф.р.) МТР в I классе,
 $D_2(r)$ - ф.р. МТР во II классе,
 $D_x(r)$ - ф.р. МТР между классами.

Соответствующие количества МТР обозначим через $m_1, m_2, m_x, m_+ := m_1 + m_2$, а через $D_+(r)$ ф.р. m_+ внутри классовых МТР. Через D_* будем обозначать эмпирические ф.р. Если распределения классов совпадают, то совпадают все ф.р. Если не совпадают, то наибольшая разница будет между D_+ и D_x .

Многомерная задача свелась к одномерной, можно пользоваться критериями согласия, их много.

Критерий Колмогорова-Смирнова

При совпадении распределений величина

$$A_{m_+, m_x} = \sqrt{\frac{m_+ m_x}{m_+ + m_x} \max_r |D_+^* - D_x^*|}$$

имеет асимптотически ($n \rightarrow \infty$) табулированное распределение Колмогорова.

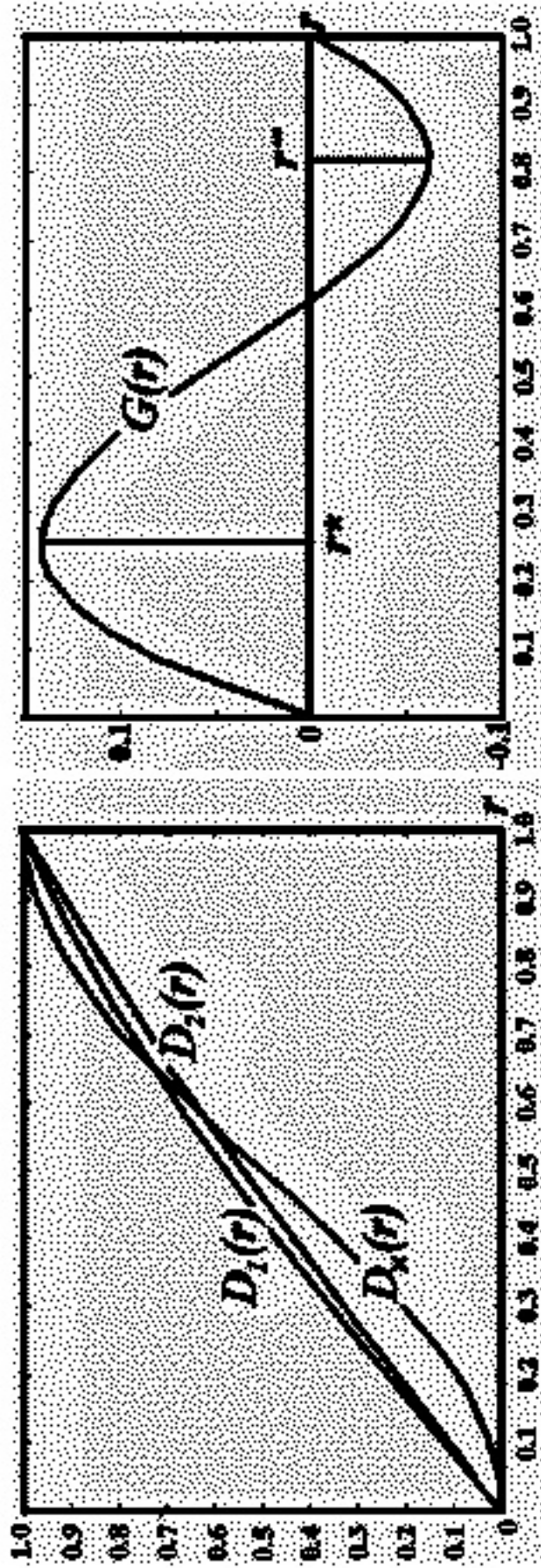
Выбор оптимального радиуса

Классифицируемую точку \mathbf{x} ? сделаем центром шара радиуса r и отнесём её к классу, преобладающему в шаре (прямая классификация) или к другому классу (обратная классификация для чередующихся точек).

Пусть $\mathbf{x}^2 = \mathbf{x}^1$. Тогда вероятность того, что следующая точка тоже \mathbf{x}^1 , равна $D_1(r)$, а вероятность того, что она \mathbf{x}^2 , равна $D_x(r)$. Осреднив эти разности по классам, получим функцию

$$G(r) = [D_1(r) + D_2(r)] / 2 - D_x(r).$$

Радиус $r^* = \arg \max_r G(r)$ наилучший для деления «компактных» классов, а величину $d = G(r^*)$ можно назвать компактностью классов. Если $|G(r^-)| > G(r^*)$, где $r^- = \arg \min_r G(r)$, то обратная классификация будет более эффективной, чем прямая, она может быть полезной в задачах микробиологии. Величину $d^- = |G(r^-)|$ можно назвать смешанностью классов. Для определённости будем говорить о прямой классификации.



Отбор информативных признаков

Если гипотеза о совпадении распределений отвергается, то полезно избавиться от «шумовых» признаков. Изложенное даёт для этого два способа.

- 1) Максимизировать разобщённость d .
- 2) Максимизировать статистику Колмогорова-Смирнова или другой критерий согласия.

Оценка качества решения

Оценка погрешности $D_x(r^*)$ имеет преимущества перед традиционными методами.

Цепь разнородных импульсов

На окружности импульсы k родов a_1, \dots, a_k в количестве n_1, \dots, n_k , $\sum n_i = n$, $n_i/n = p_i$. Количество переходов от импульса a_i к импульсу a_j обозначим n_{ij} .

В тривиальной цепи разнородных импульсов (ТЦРИ) импульсы независимы и равномерно распределены на окружности.

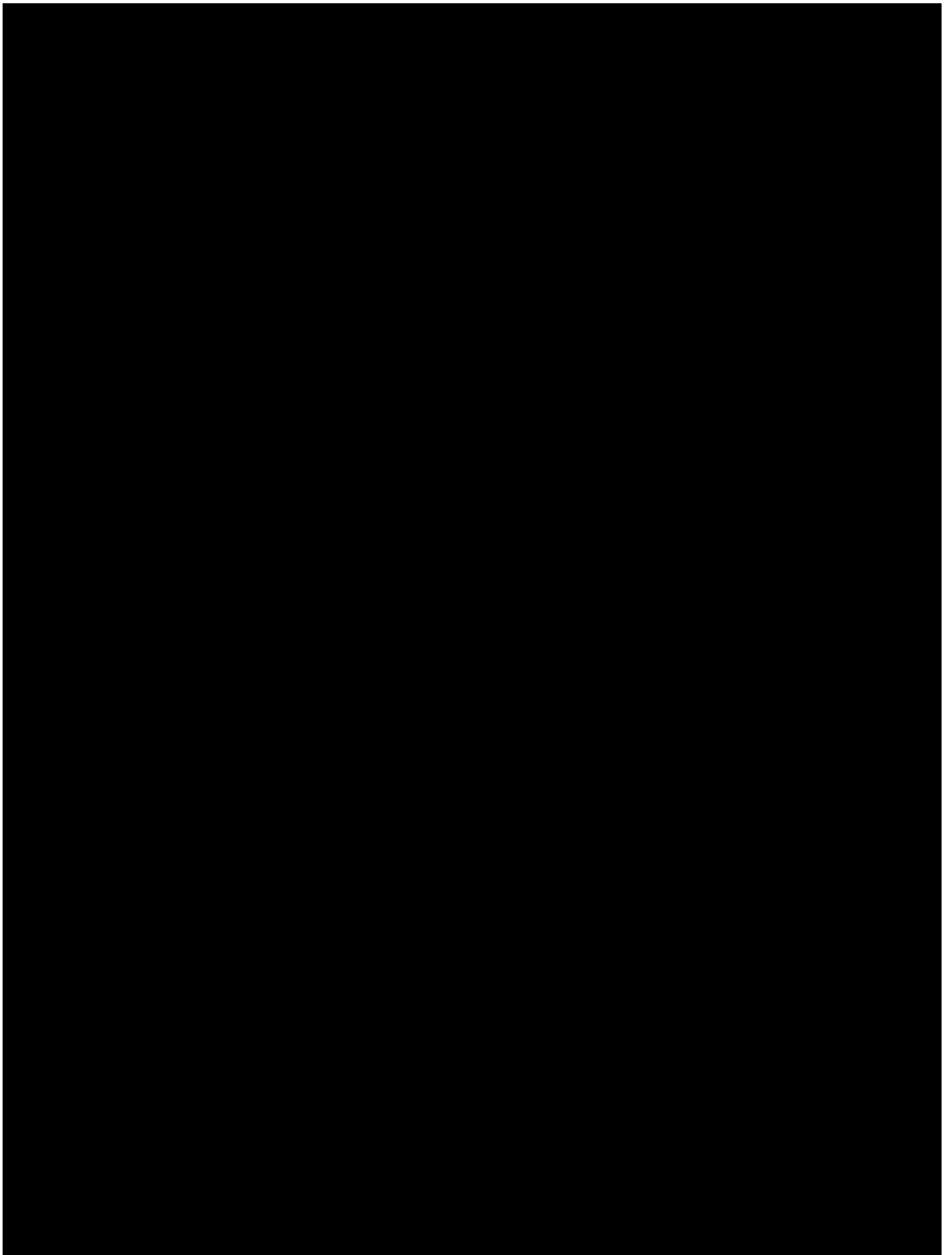
Теорема. В ТЦРИ $E n_{ij} = n_i p_j$,

$$D n_{ij} = n_i p_j (1 - p_j),$$

$$\text{Cov}(n_{ij}, n_{il}) = -n_i p_j p_l \quad \text{при } j \neq l.$$

Если импульсы i -го рода не зависят от остальных, то по критерию Пирсона

$$\sum_{j=1}^k \frac{(n_{ij} - n_i p_j)^2}{n_i p_j} = \chi^2_{k-1}$$



Литература

- Шурыгин А. М. Статистика при подсчёте запасов месторождений. Изд. МГУ, 1978
- Шурыгин А. М. Прикладная стохастика: робастность, оценивание, прогноз. «Финансы и статистика», 2000
- Стригунова М. С., Шурыгин А. М. Скалярное умножение матриц со скользящим окном и прогноз землетрясений. //Автоматика и телемеханика, №7, 2004, с.27-34
- <http://www.ccas.ru/cito/eq>