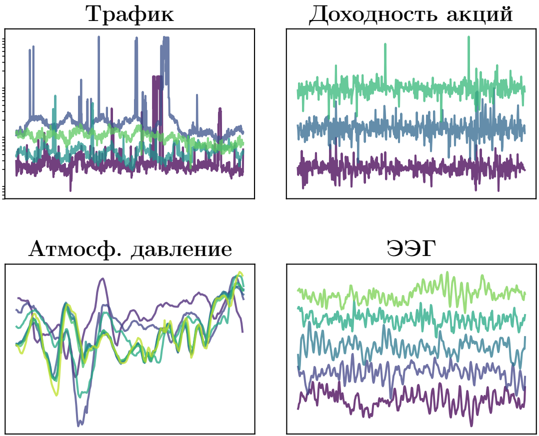
Стохастические задачи о разладке для случайных процессов. Основные статистики в задачах скорейшего обнаружения разладки

А. В. Артёмов

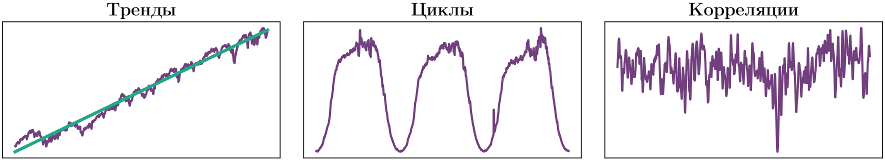
ПМИ ФКН ВШЭ, Вероятностные модели и статистика случайных процессов, весна 2017

- Многие реальные процессы описываются (многомерными) временными рядами



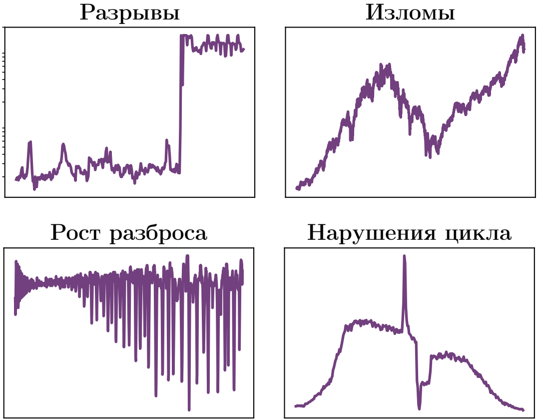
Введение

- В этих временных рядах выделяют компоненты (статистические характеристики)



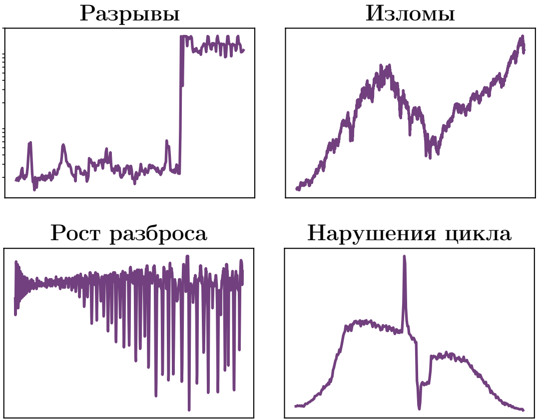
- **Описание, оценивание, выводы** о наблюдаемых временных рядах: теория и статистика случайных процессов
  - фильтрация, сегментация, шумоподавление, анализ трендов, корреляционный, дисперсионный, регрессионный, морфологический анализ, ...

- **Разладка:** изменение статистических свойств ряда



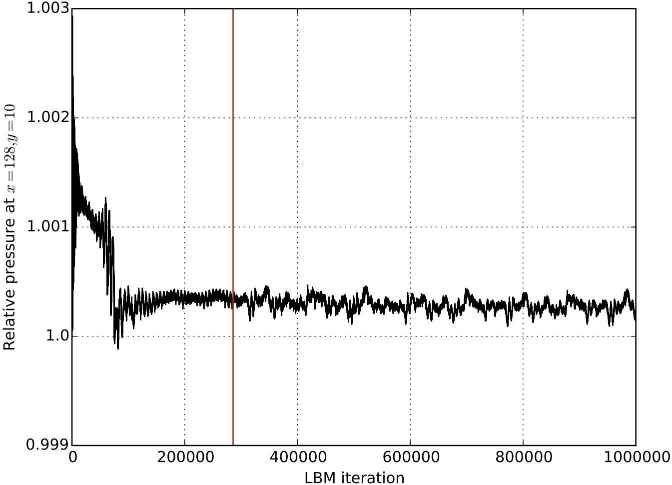
Введение

- **Разладка:** изменение статистических свойств ряда



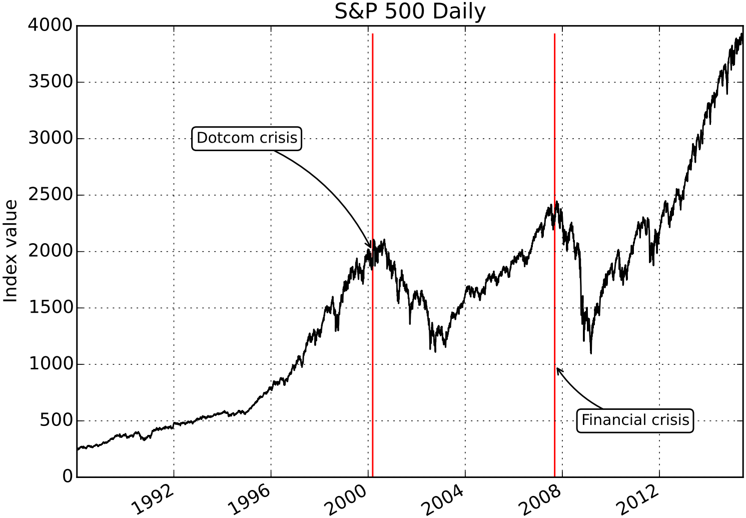
- **Задача «о разладке»:** выявить возникающее изменение

Пример: давление жидкости в гидродинамической системе (модель на основе метода Больцмана).



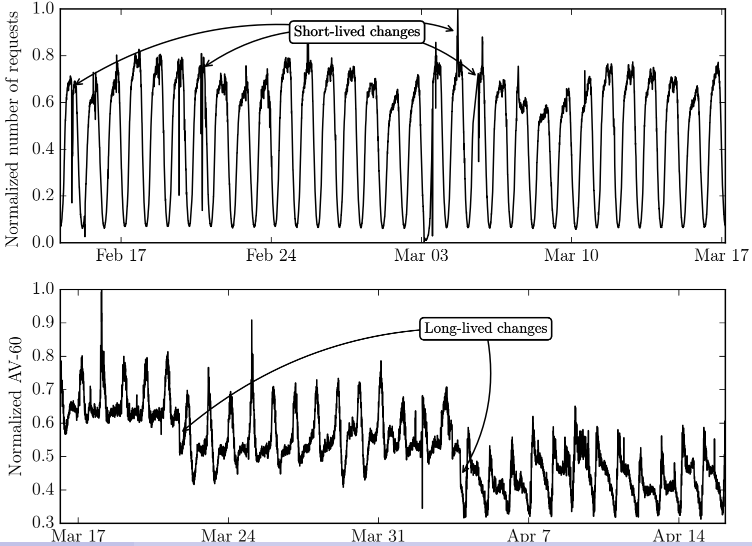
Пример разладки (реальные данные)

Пример: динамика индекса S&P 500 за 27 лет.



Пример разладки (реальные данные)

Пример: посещаемость интернет-сервиса.



Актуальность проблематики. Примеры приложений I

- Обнаружение внедрений в компьютерные сети (атак, ведущих к изменению объема передаваемого трафика) [Kim и др. 2004; Alexander G Tartakovsky 2003; Alexander G. Tartakovsky и др. 2006]
- Обнаружение аномалий в сетях передачи данных (видеопотоки в системах видеонаблюдения, сетевой трафик и др.) [Casas и др. 2010; Lakhina, Crovella и Christiphe Diot 2004; Lakhina, Crovella и Christophe Diot 2004; Pham и др. 2014; A. Tartakovsky 2013]
- Обнаружение и изоляция отказов узлов систем управления транспортными средствами [Malladi и др. 1999; Willsky 1976]
- Мониторинг целостности системы геопозиционирования [M Basseville и др. 2002]
- Обнаружение изменений структуры породы при бурении скважин [Adams и др. 2007]

Актуальность проблематики. Примеры приложений II

- Обнаружение начала рецессии или экономического роста [Andersson и др. 2002]
- Обнаружение изменений волатильности индекса Dow Jones [Adams и др. 2007]
- Обнаружение сигнала при наблюдении подводных целей [Streit и др. 1999]
- Автоматическое обнаружение аномального человеческого поведения при видеонаблюдении [Pham и др. 2014]
- Автоматический контроль качества выпускаемой продукции [Ben-Gal и др. 2003; Girshick, Meyer A and Rubin 1952; Shewhart 1931]
- Мониторинг и анализ смертности и заболеваемости раком легких [Dass 2009; Dass и др. 2011; Taweab и др. 2015]
- Обнаружение возникновения эпидемий [MacNeill и др. 1995]

Актуальность проблематики. Примеры приложений III

- Обнаружение аритмии (внезапных изменений ритма биения) сердца [Willsky 1976]
- Предсказание транзиторных ишемических атак (преходящих нарушений мозгового кровообращения) [Cerutti S. и др. 1993]
- Диагностика задержки внутриутробного роста [Petzold и др. 2004]
- Анализ несчастных случаев на угольных шахтах [Adams и др. 2007]
- Мониторинг уровня хлора в питьевой воде [Guépié и др. 2012]

Актуальность проблематики. Объем публикуемой литературы

Поиск в системе индексации Google Scholar выдает, **начиная с 2000 года:**

- `change point detection` — **10 200** статей
- `anomaly detection` — **53 900** статей
- `break detection` — **3 980** статей
- `обнаружение разладок, обнаружение аномалий, обнаружение изменений` — **765** статей

Первые работы по разладкам: 1931 год, W. A. Shewhart (цель — контроль качества выпускаемой продукции).

Математическая задача «о разладке»

- Наблюдаемый случайный процесс $\xi = (\xi_t)_{t \geqslant 0}$ задан на пространстве $(\Omega, \mathcal{F}, \mathrm{P})$

Математическая задача «о разладке»

- Наблюдаемый случайный процесс $\xi = (\xi_t)_{t \geqslant 0}$ задан на пространстве $(\Omega, \mathcal{F}, \mathrm{P})$
- Наблюдаемый процесс имеет структуру

$$\xi_1, \xi_2, \ldots, \xi_{\theta-1}, \xi_\theta, \xi_{\theta+1}, \ldots$$

и описывается своей реализацией $X = (X_t)_{t \geqslant 0}$

$$X_1, X_2, \ldots, X_{\theta-1}, X_\theta, X_{\theta+1}, \ldots$$

Математическая задача «о разладке»

- Наблюдаемый случайный процесс $\xi = (\xi_t)_{t\geqslant 0}$ задан на пространстве $(\Omega, \mathcal{F}, \mathrm{P})$
- Наблюдаемый процесс имеет структуру

$$\underbrace{\xi_1, \xi_2, \ldots, \xi_{\theta-1},}_{\text{до разладки } \mathrm{P}=\mathrm{P}_\infty} \quad \underbrace{\xi_\theta, \xi_{\theta+1}, \ldots}_{\text{разладка: } \mathrm{P}=\mathrm{P}_0}$$

и описывается своей реализацией $X = (X_t)_{t\geqslant 0}$

$$X_1, X_2, \ldots, X_{\theta-1}, \quad X_\theta, X_{\theta+1}, \ldots$$

Математическая задача «о разладке»

- Наблюдаемый случайный процесс $\xi = (\xi_t)_{t \geqslant 0}$ задан на пространстве $(\Omega, \mathcal{F}, \mathrm{P})$
- Наблюдаемый процесс имеет структуру

$$\underbrace{\xi_1, \xi_2, \ldots, \xi_{\theta-1},}_{\text{до разладки } \mathrm{P}=\mathrm{P}_\infty} \quad \underbrace{\xi_\theta, \xi_{\theta+1}, \ldots}_{\text{разладка: } \mathrm{P}=\mathrm{P}_0}$$

и описывается своей реализацией $X = (X_t)_{t \geqslant 0}$

$$X_1, X_2, \ldots, X_{\theta-1}, \quad X_\theta, X_{\theta+1}, \ldots$$

- θ — момент появления **разладки** (заранее неизвестен)

Математическая задача «о разладке»

- Наблюдаемый процесс $\xi = (\xi_t)_{t \geqslant 0}$ имеет структуру

$$\xi_1, \xi_2, \ldots, \xi_{\theta-1}, \quad \xi_\theta, \xi_{\theta+1}, \ldots$$

Математическая задача «о разладке»

- Наблюдаемый процесс $\xi = (\xi_t)_{t\geqslant 0}$ имеет структуру

$$\xi_1, \xi_2, \ldots, \xi_{\theta-1}, \quad \xi_\theta, \xi_{\theta+1}, \ldots$$

- Пример из классики:

$$\xi_t = \mu \mathbb{1}_{\{t\geqslant\theta\}}(t) + W_t,$$

где $W = (W_t)_{t\geqslant 0}$ — нормальные i.i.d.r.v.

Математическая задача «о разладке»

- Наблюдаемый процесс $\xi = (\xi_t)_{t \geqslant 0}$ имеет структуру

$$\xi_1, \xi_2, \dots, \xi_{\theta-1}, \quad \xi_\theta, \xi_{\theta+1}, \dots$$

- P_∞: распределение ξ в предположении, что разладка не появляется никогда ($\theta = \infty$)

$$\xi_1, \xi_2, \dots, \xi_{\theta-1}, \xi_\theta, \xi_{\theta+1}, \dots \quad \sim \quad \mathrm{P}_\infty$$

Математическая задача «о разладке»

- Наблюдаемый процесс $\xi = (\xi_t)_{t \geqslant 0}$ имеет структуру

$$\xi_1, \xi_2, \ldots, \xi_{\theta-1}, \quad \xi_\theta, \xi_{\theta+1}, \ldots$$

- P_∞: распределение ξ в предположении, что разладка не появляется никогда ($\theta = \infty$)

$$\xi_1, \xi_2, \ldots, \xi_{\theta-1}, \xi_\theta, \xi_{\theta+1}, \ldots \quad \sim \quad \mathrm{P}_\infty$$

- P_0: распределение ξ в предположении, что разладка произошла в момент старта наблюдений ($\theta = 0$)

$$\xi_1, \xi_2, \ldots, \xi_{\theta-1}, \xi_\theta, \xi_{\theta+1}, \ldots \quad \sim \quad \mathrm{P}_0$$

Математическая задача «о разладке»

- Наблюдаемый процесс $\xi = (\xi_t)_{t\geqslant 0}$ имеет структуру

$$\xi_1, \xi_2, \ldots, \xi_{\theta-1}, \quad \xi_\theta, \xi_{\theta+1}, \ldots$$

- P_∞: распределение ξ в предположении, что разладка не появляется никогда ($\theta = \infty$)

$$\xi_1, \xi_2, \ldots, \xi_{\theta-1}, \xi_\theta, \xi_{\theta+1}, \ldots \quad \sim \quad \mathrm{P}_\infty$$

- P_0: распределение ξ в предположении, что разладка произошла в момент старта наблюдений ($\theta = 0$)

$$\xi_1, \xi_2, \ldots, \xi_{\theta-1}, \xi_\theta, \xi_{\theta+1}, \ldots \quad \sim \quad \mathrm{P}_0$$

- У мер P_∞ и P_0 есть плотности $f_\infty(x)$ и $f_0(x)$ и соответствующие матожидания E_∞ и E_0

Математическая задача «о разладке»

- Наблюдаемый процесс $\xi = (\xi_t)_{t \geqslant 0}$ имеет структуру

$$\xi_1, \xi_2, \ldots, \xi_{\theta-1}, \quad \xi_\theta, \xi_{\theta+1}, \ldots$$

- P_θ: распределение ξ в предположении, что разладка произошла в момент θ
- Плотность $f_\theta(x), x \in \mathbb{R}^n$ меры P_θ имеет специальный вид

$$f_\theta(X_1, \ldots, X_n) = f_\infty(X_1) f_\infty(X_2) \cdot \ldots \cdot f_0(X_\theta) f_0(X_{\theta+1})$$

Пример: бракованные изделия

- $\xi_1, \xi_2, \ldots$: размеры выпускаемых изделий

Пример: бракованные изделия

- $\xi_1, \xi_2, \ldots$: размеры выпускаемых изделий
- *Нормальный* ход индустриального процесса: $\xi_1, \xi_2, \ldots$ — i.i.d., $\xi_i \sim f_\infty(x)$, $\theta = \infty$

Пример: бракованные изделия

- $\xi_1, \xi_2, \ldots$: размеры выпускаемых изделий
- *Нормальный* ход индустриального процесса: $\xi_1, \xi_2, \ldots$ — i.i.d., $\xi_i \sim f_\infty(x)$, $\theta = \infty$
- Изначально производятся *бракованные* изделия: $\xi_1, \xi_2, \ldots$ — i.i.d., $\xi_i \sim f_0(x)$, $(\theta = 0)$

Пример: бракованные изделия

- $\xi_1, \xi_2, \ldots$: размеры выпускаемых изделий
- *Нормальный* ход индустриального процесса: $\xi_1, \xi_2, \ldots$ — i.i.d., $\xi_i \sim f_\infty(x)$, $\theta = \infty$
- Изначально производятся *бракованные* изделия: $\xi_1, \xi_2, \ldots$ — i.i.d., $\xi_i \sim f_0(x)$, ($\theta = 0$)
- Типичный случай: сначала имеет место нормальный ход, но в момент θ наступает «сбой» («разладка»):

$$\underbrace{\xi_1, \xi_2, \ldots, \xi_{\theta-1},}_{f_\infty(x)} \quad \underbrace{\xi_\theta, \xi_{\theta+1}, \ldots}_{f_0(x)}$$

- Пусть до момента времени n доступны наблюдения

$$\mathbf{X}_n = (X_1, \ldots, X_n)$$

Математическая задача «о разладке»

- Пусть до момента времени n доступны наблюдения
$$\mathbf{X}_n = (X_1, \ldots, X_n)$$
- Требуется подать *сигнал тревоги*, если к моменту n есть доказательства появления разладки

Математическая задача «о разладке»

- Пусть до момента времени n доступны наблюдения
$$\mathbf{X}_n = (X_1, \dots, X_n)$$
- Требуется подать *сигнал тревоги*, если к моменту n есть доказательства появления разладки
- **Момент остановки:** статистика $\tau = \tau(\mathbf{X}_n)$
$$\tau \in \{0, 1 \dots, \infty\}, \qquad \{\mathbf{X}_n : \tau(\mathbf{X}_n) = n\} \in \sigma(\mathbf{X}_n)$$

Математическая задача «о разладке»

- Пусть до момента времени n доступны наблюдения
$$\mathbf{X}_n = (X_1, \ldots, X_n)$$
- Требуется подать *сигнал тревоги*, если к моменту n есть доказательства появления разладки
- **Момент остановки:** статистика $\tau = \tau(\mathbf{X}_n)$
$$\tau \in \{0, 1 \ldots, \infty\}, \qquad \{\mathbf{X}_n : \tau(\mathbf{X}_n) = n\} \in \sigma(\mathbf{X}_n)$$
- Используется лишь накопленная к *настоящему времени* информация (не используется будущее)

Математическая задача «о разладке»

- Момент остановки: статистика $\tau = \tau(\mathbf{X}_n)$

Математическая задача «о разладке»

- Момент остановки: статистика $\tau = \tau(\mathbf{X}_n)$
- **Хорошо:** $\mathrm{E}_\infty \tau \to \infty$ (редкие ложные тревоги)
- $\mathrm{E}_\infty \tau$: среднее время до ложной тревоги (false detection delay, $\mathrm{FDD}(\tau)$)

Математическая задача «о разладке»

- Момент остановки: статистика $\tau = \tau(\mathbf{X}_n)$
- **Хорошо:** $\mathrm{E}_\infty \tau \to \infty$ (редкие ложные тревоги)
- $\mathrm{E}_\infty \tau$: среднее время до ложной тревоги (false detection delay, $\mathrm{FDD}(\tau)$)
- **Хорошо:** $\mathrm{E}_0 \tau \to 0$ (быстрое обнаружение)
- $\mathrm{E}_0 \tau$ или $\mathrm{E}_\theta[\tau - \theta | \tau > \theta]$: средняя задержка в обнаружении (average detection delay, $\mathrm{ADD}(\tau)$)

Математическая задача «о разладке»

- Момент остановки: статистика $\tau = \tau(\mathbf{X}_n)$
- **Хорошо:** $\mathrm{E}_\infty \tau \to \infty$ (редкие ложные тревоги)
- $\mathrm{E}_\infty \tau$: среднее время до ложной тревоги (false detection delay, $\mathrm{FDD}(\tau)$)
- **Хорошо:** $\mathrm{E}_0 \tau \to 0$ (быстрое обнаружение)
- $\mathrm{E}_0 \tau$ или $\mathrm{E}_\theta[\tau - \theta | \tau > \theta]$: средняя задержка в обнаружении (average detection delay, $\mathrm{ADD}(\tau)$)
- **Плохо:** $\mathrm{P}_\theta(\tau < \theta) \to 1$ (частые ложные тревоги)
- $\mathrm{P}_\theta(\tau < \theta)$: вероятность ложной тревоги (probability of false alarm, $\mathrm{PFA}(\tau)$)

Математическая задача «о разладке»

- Момент остановки: статистика $\tau = \tau(\mathbf{X}_n)$
- **Хорошо:** $\mathrm{E}_\infty \tau \to \infty$ (редкие ложные тревоги)
- $\mathrm{E}_\infty \tau$: среднее время до ложной тревоги (false detection delay, $\mathrm{FDD}(\tau)$)
- **Хорошо:** $\mathrm{E}_0 \tau \to 0$ (быстрое обнаружение)
- $\mathrm{E}_0 \tau$ или $\mathrm{E}_\theta[\tau - \theta | \tau > \theta]$: средняя задержка в обнаружении (average detection delay, $\mathrm{ADD}(\tau)$)
- **Плохо:** $\mathrm{P}_\theta(\tau < \theta) \to 1$ (частые ложные тревоги)
- $\mathrm{P}_\theta(\tau < \theta)$: вероятность ложной тревоги (probability of false alarm, $\mathrm{PFA}(\tau)$)
- **Плохо:** $\mathrm{E}_\theta[\tau - \theta | \tau > \theta] \to \infty$ (медленное обнаружение)

Общие обозначения

- Пусть $X_1, \ldots, X_n$ — наблюдения, доступные до момента времени n
- Основная статистика — отношение правдоподобия

$$L_n = \frac{f_0(X_1, \ldots, X_n)}{f_\infty(X_1, \ldots, X_n)}$$

- Удобно также использовать статистику $Z_n = \log L_n$
- Если наблюдения $X_1, \ldots, X_n$ независимы:

$$L_n = \prod_{k=1}^{n} \frac{f_0(X_k)}{f_\infty(X_k)} = \prod_{k=1}^{n} l_k, \quad Z_n = \sum_{k=1}^{n} \log \frac{f_0(X_k)}{f_\infty(X_k)} = \sum_{k=1}^{n} \zeta_k$$

Простой пример

- Пусть $\xi_1, \ldots, \xi_n$ — нормальные i.i.d.r.v., причем
$$f_0(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-r_0)^2}{2\sigma^2}}, \qquad f_\infty(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-r_\infty)^2}{2\sigma^2}}$$
Тогда правдоподобие выборки $X_1, \ldots, X_n$
$$L_n = \prod_{k=1}^{n} \exp\left\{ \frac{r_\infty - r_0}{\sigma^2} \left[X_k - \frac{r_0 + r_\infty}{2} \right] \right\},$$
а его логарифм —
$$Z_n = \frac{r_\infty - r_0}{\sigma^2} \left[\overline{X_n} - \frac{r_0 + r_\infty}{2} n \right]$$

Контрольные карты Шухарта

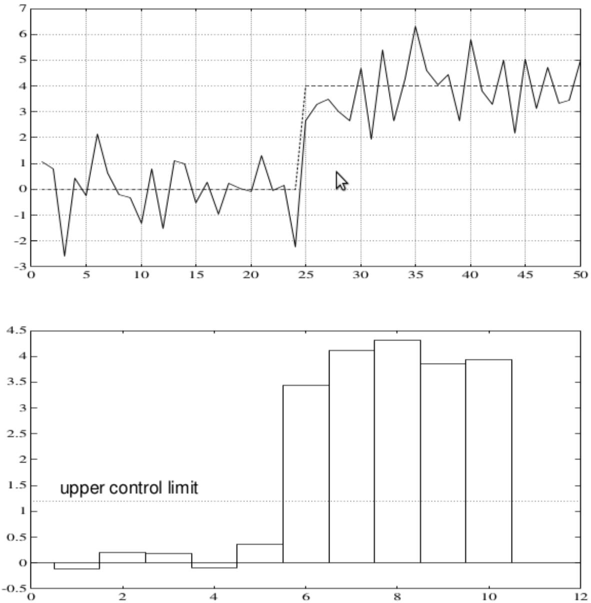
- Наблюдения $X_1, X_2, \ldots$ разбиваются на группы (батчи) размера N (N — параметр алгоритма)
- Для каждой группы $\mathbf{X}_K = (X_{N\cdot(K-1)+1}, \ldots, X_{N\cdot K})$, $K = 1, 2, \ldots$ подсчитывается логарифм правдоподобия:

$$S_i^k = \sum_{j=i}^{k} \zeta_j$$

- Момент остановки — первый момент выхода статистики $S_{N\cdot(K-1)+1}^{N\cdot K}$ на заданный уровень h:

$$\tau_{\mathrm{SH}} = N \cdot \min\{K : S_{N\cdot(K-1)+1}^{N\cdot K} \geqslant h\}$$

Пример [Michèle Basseville и др. 1993]



- Задается рекурсивная оценка среднего значения

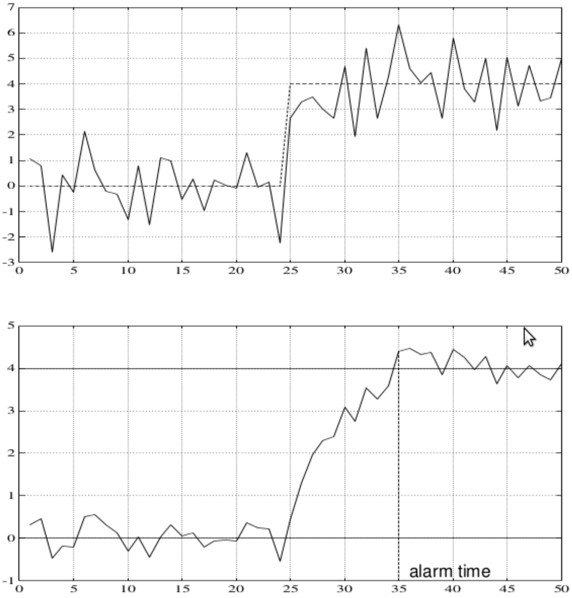
$$\hat{m}_k = (1-\lambda)\hat{m}_{k-1} + \lambda X_k, \qquad k = 1, 2, \dots,$$

где λ («вес» новых данных) — параметр алгоритма

- Момент остановки — момент первого выхода статистики $\hat{m}_k$ на заданный уровень h:

$$\tau_{\mathrm{EWMA}} = \min\{k \geqslant 1 : \hat{m}_k \geqslant h\}$$

Пример [Michèle Basseville и др. 1993]



Кумулятивные суммы

- О параметре θ не делается никаких предположений

- О параметре θ не делается никаких предположений
- Фиксируется $T > 0$ и задается класс

$$\mathcal{M}_T = \{\tau : \mathrm{E}_\infty \tau \geqslant T\}$$

тех моментов остановки, для которых среднее время до ложной тревоги не меньше T

Кумулятивные суммы

- О параметре θ не делается никаких предположений
- Фиксируется $T > 0$ и задается класс
$$\mathcal{M}_T = \{\tau : \mathrm{E}_\infty \tau \geqslant T\}$$
тех моментов остановки, для которых среднее время до ложной тревоги не меньше T
- Качество алгоритма задается величиной
$$\mathbf{D}(T) = \sup_{\theta \geqslant 0} \operatorname*{ess\,sup}_{\omega} \mathrm{E}_\theta\Big((\tau - \theta)^+ \Big| \mathcal{F}_\theta\Big)(\omega) \quad \sim \quad \inf_{\tau \in \mathcal{M}_T}$$

- О параметре θ не делается никаких предположений
- Фиксируется $T > 0$ и задается класс

$$\mathcal{M}_T = \{\tau : \mathrm{E}_\infty \tau \geqslant T\}$$

  тех моментов остановки, для которых среднее время до ложной тревоги не меньше T
- Качество алгоритма задается величиной

$$\mathbf{D}(T) = \sup_{\theta \geqslant 0} \operatorname{ess\,sup}_{\omega} \underbrace{\mathrm{E}_\theta\Big((\tau - \theta)^+ \Big| \mathcal{F}_\theta\Big)(\omega)}_{\text{среднее время обнаружения разладки}} \quad \sim \quad \inf_{\tau \in \mathcal{M}_T}$$

Кумулятивные суммы

- О параметре θ не делается никаких предположений
- Фиксируется $T > 0$ и задается класс
$$\mathcal{M}_T = \{\tau : \mathrm{E}_\infty \tau \geqslant T\}$$
тех моментов остановки, для которых среднее время до ложной тревоги не меньше T
- Качество алгоритма задается величиной
$$\mathbf{D}(T) = \sup_{\theta \geqslant 0} \underbrace{\operatorname*{ess\,sup}_{\omega} \underbrace{\mathrm{E}_\theta\Big((\tau-\theta)^+ \Big| \mathcal{F}_\theta\Big)(\omega)}_{\text{среднее время обнаружения разладки}}}_{\text{наихудшее среди всех траекторий}} \quad \sim \quad \inf_{\tau \in \mathcal{M}_T}$$

- О параметре θ не делается никаких предположений
- Фиксируется $T > 0$ и задается класс

$$\mathcal{M}_T = \{\tau : \mathrm{E}_\infty \tau \geqslant T\}$$

тех моментов остановки, для которых среднее время до ложной тревоги не меньше T

- Качество алгоритма задается величиной

$$\mathbf{D}(T) = \underbrace{\sup_{\theta \geqslant 0} \underbrace{\operatorname{ess\,sup}_{\omega} \underbrace{\mathrm{E}_\theta\Big((\tau-\theta)^+ \Big| \mathcal{F}_\theta\Big)(\omega)}_{\text{среднее время обнаружения разладки}}}_{\text{наихудшее среди всех траекторий}}}_{\text{и всех моментов } \theta \text{ появления разладки}} \quad \sim \quad \inf_{\tau \in \mathcal{M}_T}$$

Кумулятивные суммы

- Вводятся статистики

$$\gamma_n = \sup_{\theta \geqslant 0} \frac{f_\theta(X_1, \ldots, X_n)}{f_\infty(X_1, \ldots, X_n)} \qquad \text{и} \qquad T_n = \log \gamma_n$$

Кумулятивные суммы

- Вводятся статистики

$$\gamma_n = \sup_{\theta \geqslant 0} \frac{f_\theta(X_1, \ldots, X_n)}{f_\infty(X_1, \ldots, X_n)} \qquad \text{и} \qquad T_n = \log \gamma_n$$

- Если случайные величины $\xi_1, \ldots, \xi_n$ независимы, то

$$\gamma_n = \max\left\{1, \max_{1 \leqslant \theta \leqslant n} \prod_{k=\theta}^{n} \frac{f_0(X_k)}{f_\infty(X_k)}\right\},$$

$$T_n = \max\left\{0, \max_{1 \leqslant \theta \leqslant n} \sum_{k=\theta}^{n} \log \frac{f_0(X_k)}{f_\infty(X_k)}\right\} =$$

$$= \max\left\{0, \max_{1 \leqslant \theta \leqslant n} \sum_{k=\theta}^{n} \zeta_k\right\}$$

Кумулятивные суммы

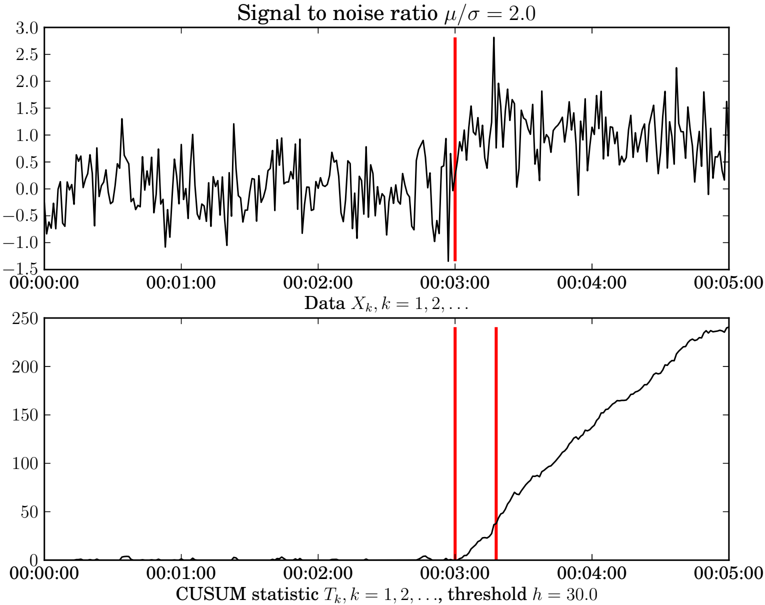
- Статистика T_n обладает свойством $T_n = \max(0, T_{n-1} + \zeta_n)$ и называется статистикой кумулятивных сумм (CUmulative SUMs, CUSUM).

Кумулятивные суммы

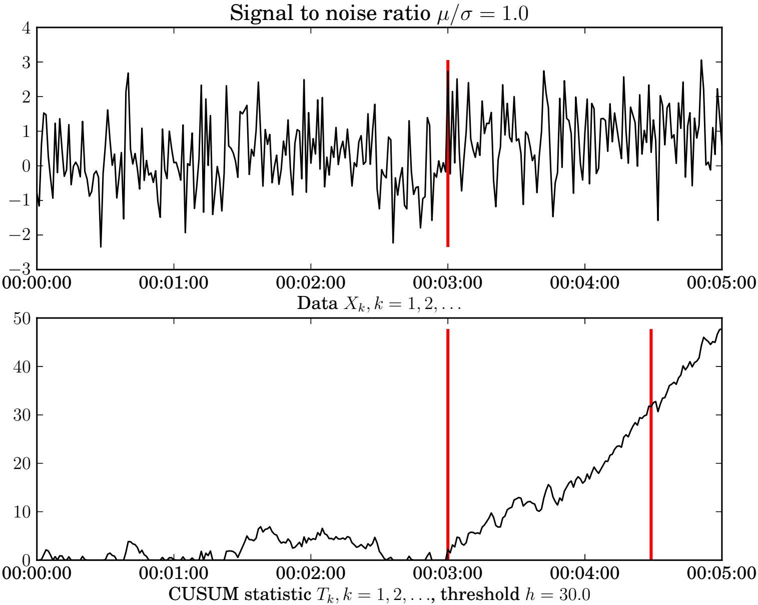
- Статистика T_n обладает свойством $T_n = \max(0, T_{n-1} + \zeta_n)$ и называется статистикой кумулятивных сумм (CUmulative SUMs, CUSUM).
- Остановка в момент τ_{CUSUM} минимизирует величину $\mathbf{D}(T)$

$$\tau_{\text{CUSUM}} = \inf\{n \geqslant 0 : T_n \geqslant h\}$$

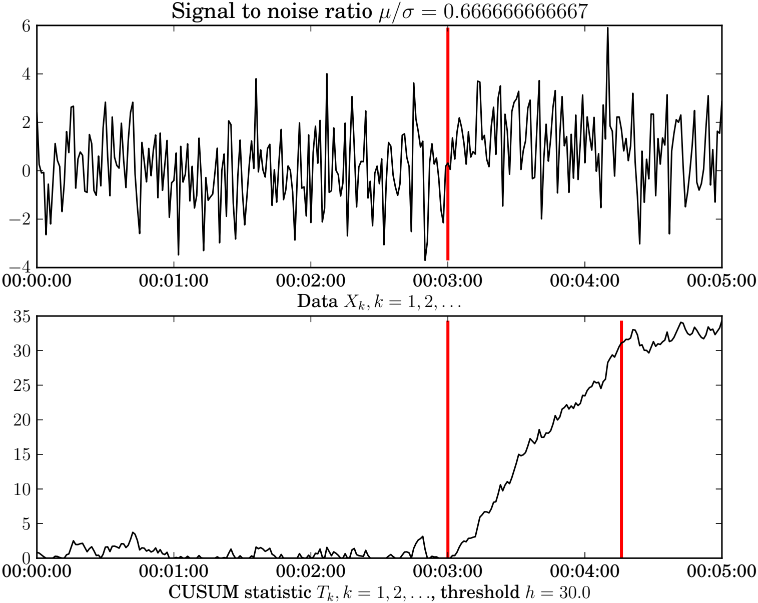
Пример 8.1 [Razladki]



Пример 8.2 [Razladki]



Пример 8.3 [Razladki]



Статистика Ширяева-Робертса

- Условия для моментов остановки и параметра θ в точности совпадают с условиями для процедуры кумулятивных сумм

Статистика Ширяева-Робертса

- Условия для моментов остановки и параметра θ в точности совпадают с условиями для процедуры кумулятивных сумм
- Качество алгоритма задается величиной

$$\mathbf{C}(T) = \sup_{\theta \geqslant 0} \mathbb{E}_\theta\Big((\tau - \theta)^+ \Big| \tau \geqslant \theta\Big) \quad \sim \quad \inf_{\tau \in \mathcal{M}_T}$$

Статистика Ширяева-Робертса

- Условия для моментов остановки и параметра θ в точности совпадают с условиями для процедуры кумулятивных сумм
- Качество алгоритма задается величиной

$$\mathbf{C}(T) = \sup_{\theta \geqslant 0} \underbrace{\mathbb{E}_\theta\Big((\tau - \theta)^+ \Big| \tau \geqslant \theta\Big)}_{\text{(условное) среднее время обнаружения разладки}} \quad \sim \quad \inf_{\tau \in \mathcal{M}_T}$$

Статистика Ширяева-Робертса

- Условия для моментов остановки и параметра θ в точности совпадают с условиями для процедуры кумулятивных сумм
- Качество алгоритма задается величиной

$$\mathbf{C}(T) = \underbrace{\sup_{\theta \geqslant 0} \underbrace{\mathbb{E}_\theta\Big((\tau-\theta)^+ \Big| \tau \geqslant \theta\Big)}_{\text{(условное) среднее время обнаружения разладки}}}_{\text{наихудшее среди всех моментов } \theta \text{ появления разладки}} \quad \sim \quad \inf_{\tau \in \mathcal{M}_T}$$

Статистика Ширяева-Робертса

- Вводится статистика $R_n = \sum\limits_{\theta=1}^{n} \frac{f_\theta(X_1,...,X_n)}{f_\infty(X_1,...,X_n)}$

Статистика Ширяева-Робертса

- Вводится статистика $R_n = \sum\limits_{\theta=1}^{n} \frac{f_\theta(X_1,\ldots,X_n)}{f_\infty(X_1,\ldots,X_n)}$
- Если случайные величины $\xi_1, \ldots, \xi_n$ независимы, то

$$R_n = \sum_{\theta=1}^{n} \prod_{k=\theta}^{n} \frac{f_0(X_k)}{f_\infty(X_k)} = \sum_{\theta=1}^{n} \prod_{k=\theta}^{n} l_k.$$

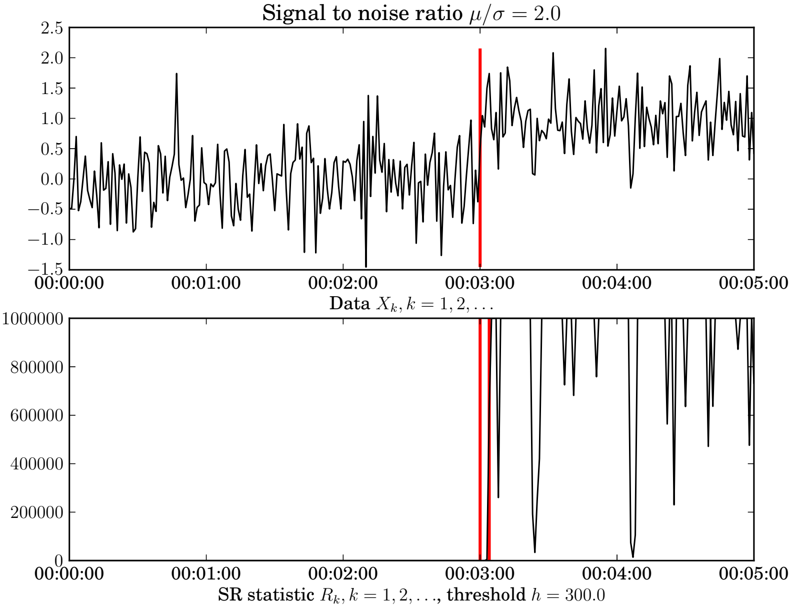
Статистика Ширяева-Робертса

- Вводится статистика $R_n = \sum_{\theta=1}^{n} \frac{f_\theta(X_1,\ldots,X_n)}{f_\infty(X_1,\ldots,X_n)}$
- Если случайные величины $\xi_1, \ldots, \xi_n$ независимы, то
$$R_n = \sum_{\theta=1}^{n} \prod_{k=\theta}^{n} \frac{f_0(X_k)}{f_\infty(X_k)} = \sum_{\theta=1}^{n} \prod_{k=\theta}^{n} l_k.$$
- Статистика R_n обладает свойством $R_n = (1 + R_{n-1}) l_k$ и называется статистикой Ширяева-Робертса (Shiryaev-Roberts, SR).

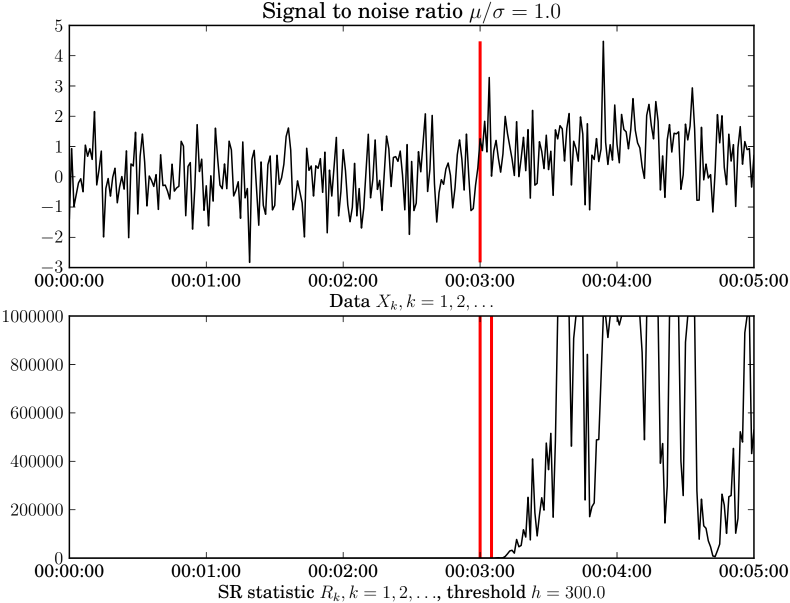
Статистика Ширяева-Робертса

- Вводится статистика $R_n = \sum\limits_{\theta=1}^{n} \frac{f_\theta(X_1,\ldots,X_n)}{f_\infty(X_1,\ldots,X_n)}$
- Если случайные величины $\xi_1, \ldots, \xi_n$ независимы, то
$$R_n = \sum_{\theta=1}^{n} \prod_{k=\theta}^{n} \frac{f_0(X_k)}{f_\infty(X_k)} = \sum_{\theta=1}^{n} \prod_{k=\theta}^{n} l_k.$$
- Статистика R_n обладает свойством $R_n = (1 + R_{n-1}) l_k$ и называется статистикой Ширяева-Робертса (Shiryaev-Roberts, SR).
- Остановка в момент τ_{SR} минимизирует $\mathbf{C}(T)$
$$\tau_{\mathrm{SR}} = \inf\{n \geqslant 0 : R_n \geqslant h\}$$

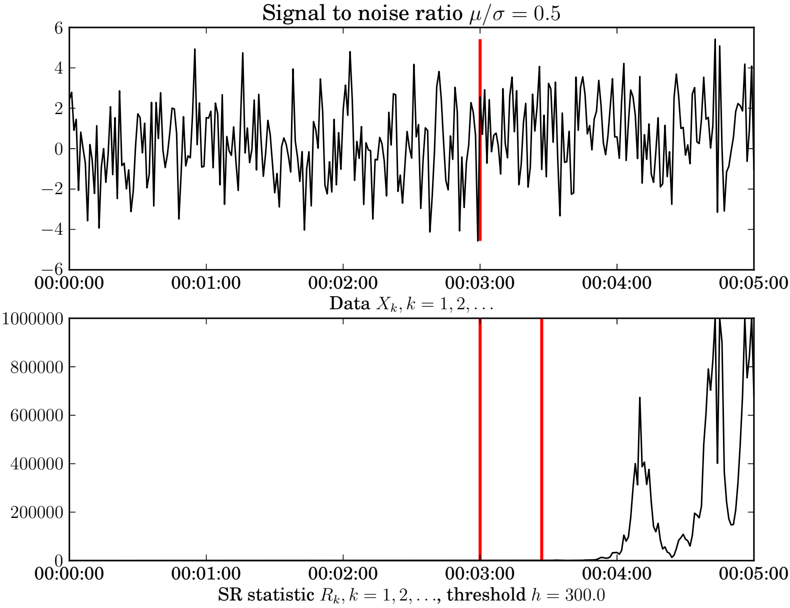
Пример 9.1 [Razladki]



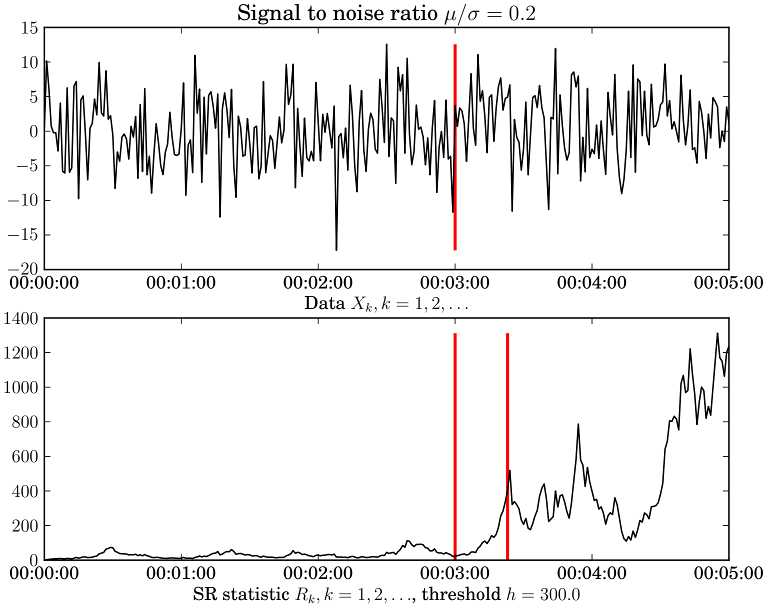
Пример 9.2 [Razladki]



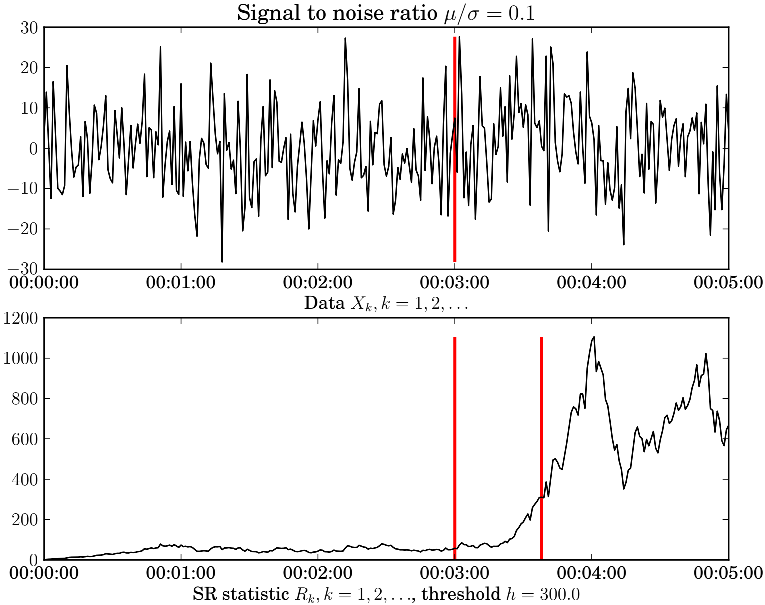
Пример 9.3 [Razladki]



Пример 9.4 [Razladki]



Пример 9.5 [Razladki]



Статистика апостериорной вероятности π_t I

- Пусть $\theta = \theta(\omega)$ — случайная величина, $\theta \perp Z_t$, имеющая (геометрическое) распределение
$$\mathrm{P}(\theta = 0) = \pi, \quad \mathrm{P}(\theta = n \big| \theta > 0) = pq^{n-1},$$
причем $\pi \in [0, 1)$ и $p \in (0, 1)$ известны, $q = 1 - p$.

Статистика апостериорной вероятности π_t II

- Качество алгоритма задается величиной

$$\mathrm{E}(\tau - \theta \big| \tau > \theta) \quad \sim \quad \inf_{\tau \in \mathcal{M}_\alpha}$$

- Этот критерий эквивалентен безусловному

$$\mathbf{A}(c) = \underbrace{\mathrm{P}(\tau < \theta)}_{\substack{\text{вероятность}\\ \text{ложной тревоги}}} + \underbrace{c\,\mathrm{E}(\tau - \theta \big| \tau > \theta)}_{\substack{\text{(условное) среднее время}\\ \text{обнаружения разладки}}} \quad \sim \quad \inf_{\tau}$$

Статистика апостериорной вероятности

- Вводится статистика

$$\pi_n = \mathrm{P}(\theta \leqslant n \big| X_1, \ldots, X_n),$$

представимая в виде

$$\pi_n = \frac{\varphi_n}{1+\varphi_n}, \qquad \varphi_{n+1} = \frac{(p+\varphi_n)}{q} l_k$$

Статистика апостериорной вероятности

- Вводится статистика

$$\pi_n = \mathrm{P}(\theta \leqslant n \big| X_1, \ldots, X_n),$$

представимая в виде

$$\pi_n = \frac{\varphi_n}{1+\varphi_n}, \qquad \varphi_{n+1} = \frac{(p+\varphi_n)}{q} l_k$$

- Остановка в момент τ_π минимизирует $\mathbf{A}(c)$:

$$\tau_\pi = \inf\{n \geqslant 0 : \pi_n \geqslant h\}$$

Статистика апостериорной вероятности

- Вводится статистика

$$\pi_n = \mathrm{P}(\theta \leqslant n \big| X_1, \ldots, X_n),$$

представимая в виде

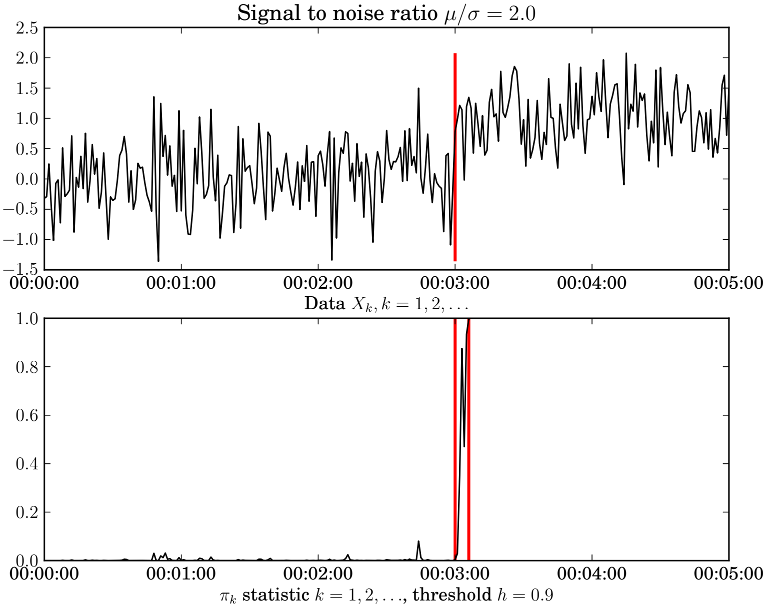
$$\pi_n = \frac{\varphi_n}{1+\varphi_n}, \qquad \varphi_{n+1} = \frac{(p+\varphi_n)}{q} l_k$$

- Остановка в момент τ_π минимизирует $\mathbf{A}(c)$:

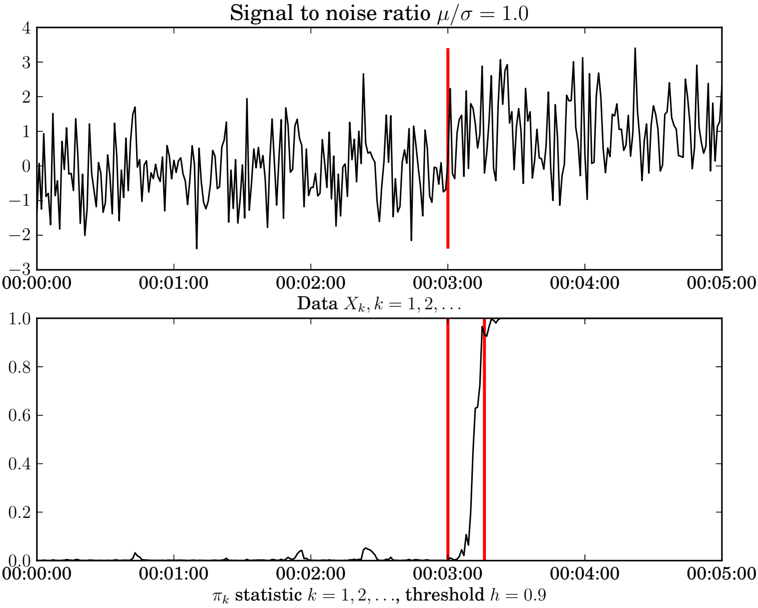
$$\tau_\pi = \inf\{n \geqslant 0 : \pi_n \geqslant h\}$$

- π_n — апостериорная вероятность появления разладки до момента времени n в предположении, что получены наблюдения $X_1, \ldots, X_n$

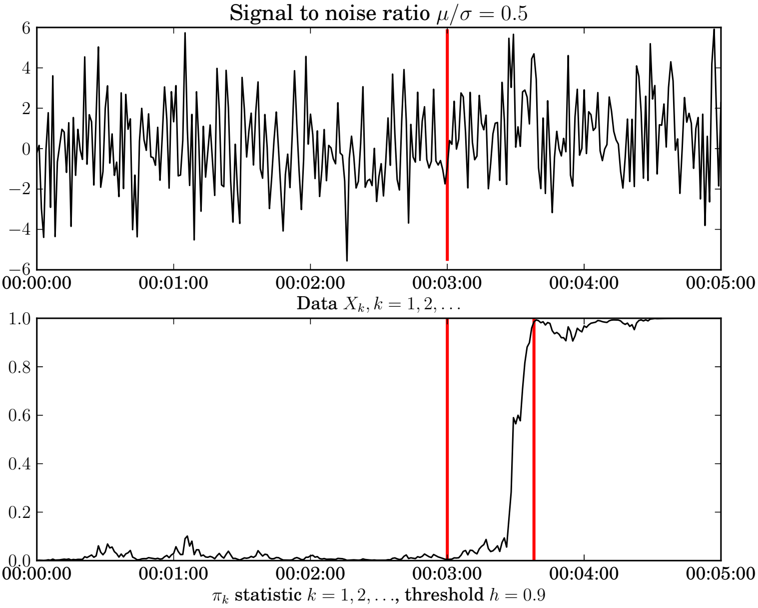
Пример 10.1 [Razladki]



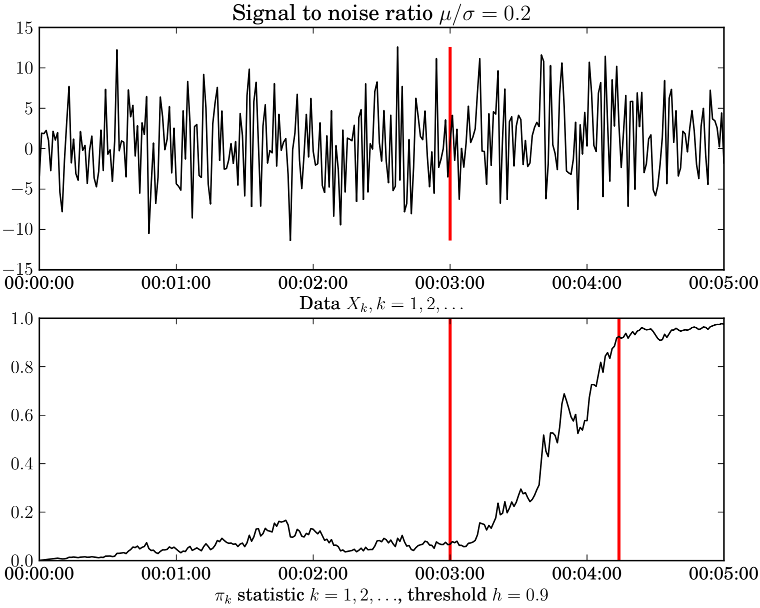
Пример 10.2 [Razladki]



Пример 10.3 [Razladki]



Пример 10.4 [Razladki]



Пример 10.5 [Razladki]