1 Погрешность обнаружения разладки

Будем выбирать последовательно из всей выборки отрезки $[tnh, (t+1)nh]$, которые назовем *окнами*. Параметр h определяет ширину окна, а t – порядковый номер. Пусть $\|\theta_2^* - \theta_1^*\|$ увеличивается линейно при приближении середины окна к точке разладки. Тогда $2\triangle L$ с увеличением $\|\theta_2^* - \theta_1^*\|$ возрастает квадратично. Будем искать выступ в виде двух парабол на графике $2\triangle L(t)$ методом наименьших квадратов. График $(2\triangle L(t) - p)$ в точках t где окно содержит точку разладки, имеет вид

$$f(x) = b^2(|x-s| - nh/2)^2,$$

где s – координата точки разладки, $b^2 = 4\|\Sigma(\theta_2^* - \theta_1^*)\|^2/(nh)^2$. Хотим минимизировать функционал

$$\sum_i \left[b^2(|x_i - s| - nh/2)^2 - d_i\right]^2 \to \min_s .$$

Приравняем производную по s к нулю

$$\sum_i \left[b^2(|x_i - s| - nh/2)^2 - d_i\right](|x_i - s| - nh/2)\tau_i = 0,$$

$$\tau_i = \begin{cases} -1, & s < x_i \\ 1, & s > x_i \\ [-1,1], & s = x_i, \end{cases}$$

$$d_i = 2\triangle L(i) - p.$$

$$\sum_i \left[b^2(s-x_i)^3 + 3b^2(x_i - s)|x_i - s|(nh/2) + 3b^2(nh/2)^2(s-x_i) - d_i(s-x_i) - \tau_i b^2(nh/2)^3 + d_i(nh/2)\tau_i\right] = 0,$$

Выберем начало координат так, чтобы выполнялись равенства $\sum_i x_i^{2k+1} = 0,\ \forall k \geq 0$. Пусть $x_i \in \{-nh/2, \dots, nh/2\}$, $s \ll nh$, тогда

$$2sb^2(nh/2)^3 - 3sb^2(nh/2)^3 + 6sb^2(nh/2)^3 - s\sum_i d_i - sb^2(nh/2)^3 = -\sum_i d_i(x_i + (nh/2)\tau_i).$$

Заменим $\sum_i d_i$ на $2/3b^2(nh/2)^3$

$$s \approx \frac{-\sum_i d_i(x_i + (nh/2)\tau_i)}{3.3b^2(nh/2)^3} \approx \frac{\sum_i d_i x_i}{3.3b^2(nh/2)^3} = \frac{\sum_i d_i x_i}{3.3\|\Sigma(\theta_2^* - \theta_1^*)\|^2(nh/2)}.$$

Оценим экспериментально дисперсию $\sum_i d_i x_i$, состоящую из дисперсии $\sum_i 2\triangle\xi_{12}^T \triangle\theta_{12}^* x_i$ и дисперсии $\sum_i 2\triangle\xi_{12}^T \triangle\theta_{12}^* x_i$ и дисперсии $\sum_i \triangle\xi_{12}^2 x_i$.

$$\mathbb{D}\left\{\sum_i 2\triangle\xi_{12}^T \triangle\theta_{12}^* x_i\right\} = c_1(p)\|\theta_{12}^*\|^2(nh)^4, \quad c_1(p) \sim (\log p)^2/e^5.$$

$$\mathbb{D}\left\{\sum_i \triangle\xi_{12}^2 x_i\right\} = c_2(p)(nh)^4, \quad c_2(p) \sim p/30.$$

Аналогичный результат может быть получен из теоретических соображений, исходя из вида корреляционной функции для процесса $\triangle\xi_{12}(t)$.

Замечание. Начиная с некоторого h_0, обеспечивающем хорошую точность приближения Лапласа для функции правдоподобия, погрешность определения s (точка разладки) начинает возрастать как $\sqrt{nh}$.

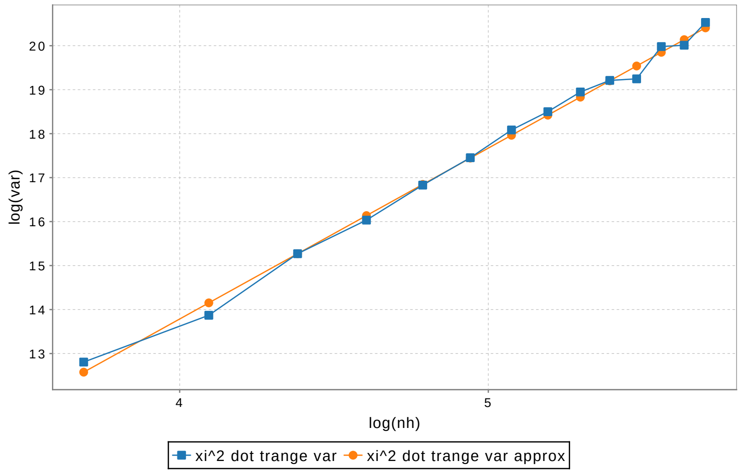

Рис. 1: Дисперсия $\sum_i \triangle\xi_{12}^2 x_i$ в зависимости от (nh) в логарифмических осях.

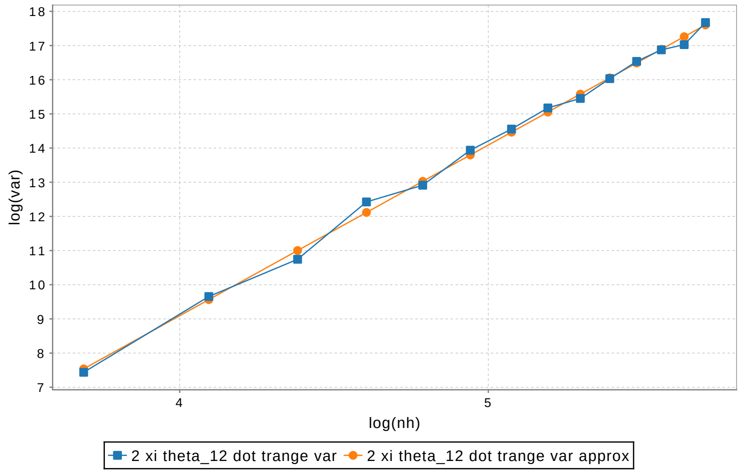

Рис. 2: Дисперсия $\sum_i 2\triangle\xi_{12}^T \triangle\theta_{12}^* x_i$ в зависимости от (nh) в логарифмических осях.

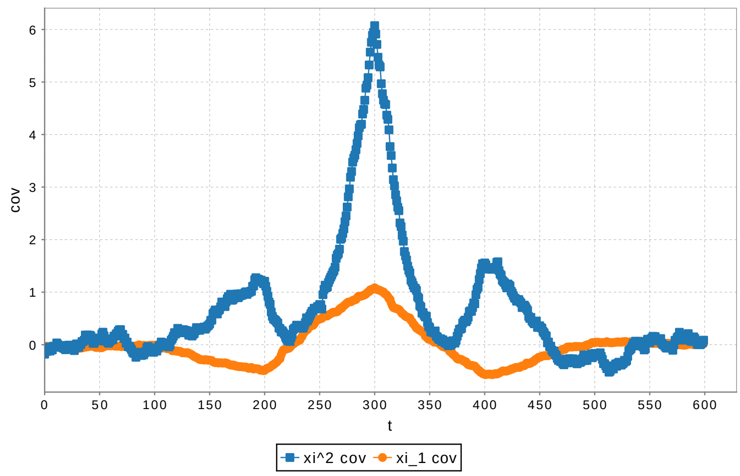

Рис. 3: Корреляция $\triangle\xi_{12}(t)^2$ и $\triangle\xi_{12}(t)(1)$.

2 Пример

$$Y_i = X\theta + \varepsilon_i.$$

Приведены графики статистик при различных величинах разности параметра θ, взятых до и после разладки

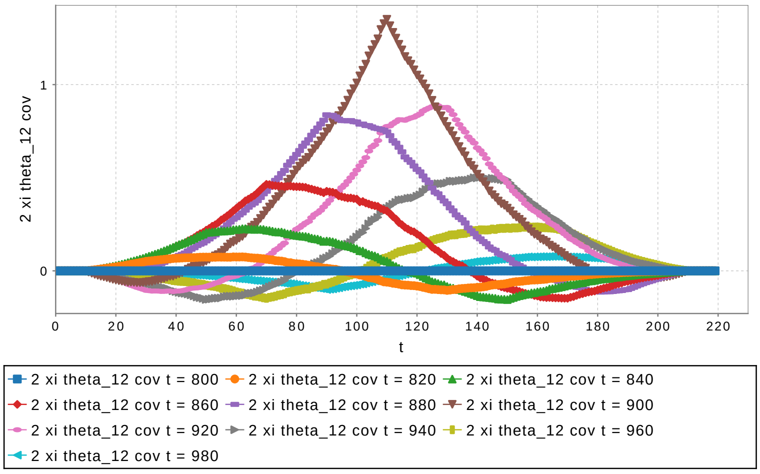

Рис. 4: Корреляция $2\triangle\xi_{12}(t)^T\triangle\theta^*_{12}$.

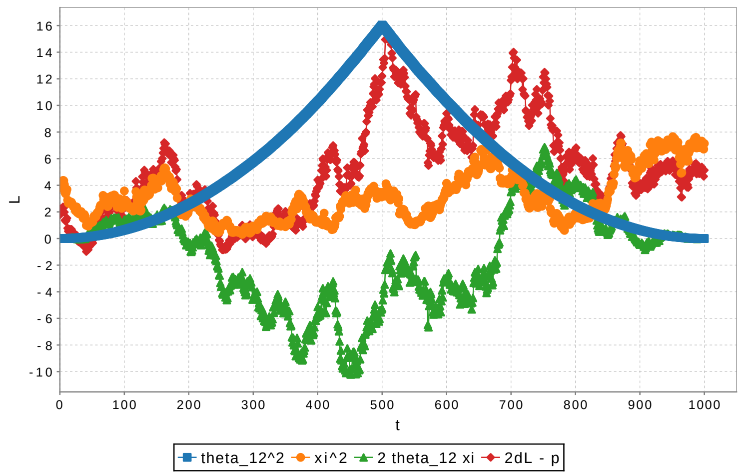

Рис. 5: Вид статистик $2\triangle L(t)$, $\triangle\xi_{12}(t)^2$, $2\triangle\xi_{12}(t)^T\triangle\theta^*_{12}(t)$ и $\triangle\theta^{*2}_{12}(t)$.

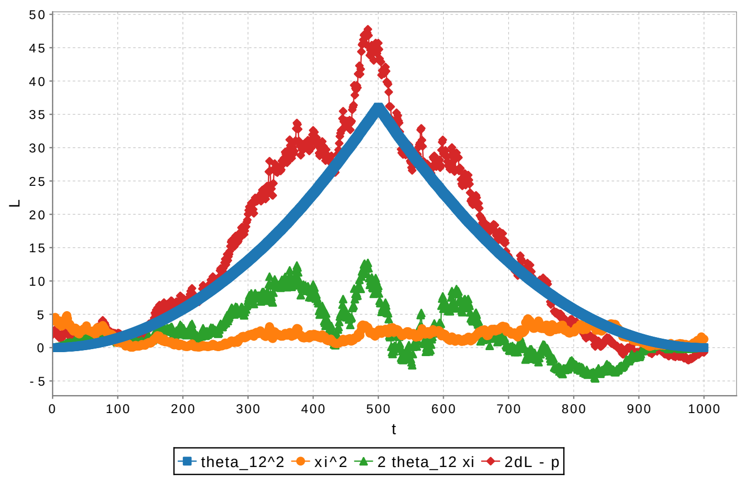

Рис. 6: Вид статистик $2\triangle L(t)$, $\triangle\xi_{12}(t)^2$, $2\triangle\xi_{12}(t)^T\triangle\theta^*_{12}(t)$ и $\triangle\theta^{*2}_{12}(t)$.

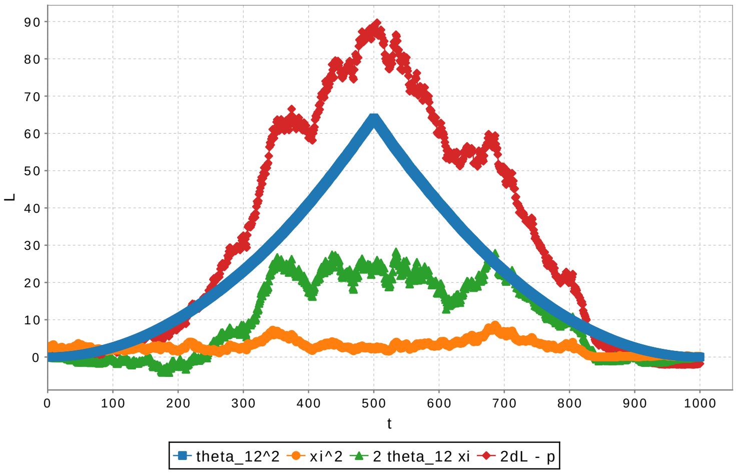

Рис. 7: Вид статистик $2\triangle L(t)$, $\triangle\xi_{12}(t)^2$, $2\triangle\xi_{12}(t)^T\triangle\theta^*_{12}(t)$ и $\triangle\theta^{*2}_{12}(t)$.

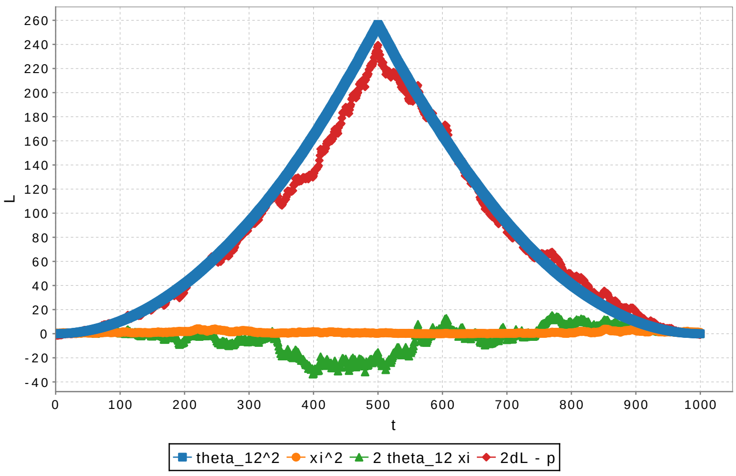

Рис. 8: Вид статистик $2\triangle L(t)$, $\triangle\xi_{12}(t)^2$, $2\triangle\xi_{12}(t)^T\triangle\theta^*_{12}(t)$ и $\triangle\theta^{*2}_{12}(t)$.