# Лекция 3
# Стохастические задачи о разладке для случайных процессов. Основные статистики в задачах скорейшего обнаружения разладки

Артемов А. В.
бПМИ ФКН ВШЭ

03 февраля 2018

## 1 Обнаружение аномалий. Задачи о разладке

### 1.1 Мотивация

Как известно, многие реальные процессы (трафик, доходность акций, атмосферное давление, ЭЭГ и так далее) можно описать многомерными временными рядами. В этих временных рядах обычно выделяют различные компоненты:

- Тренды. Например, когда временной ряд на каком-то участке линейно растёт.
- Циклы, то есть наблюдается периодичность временного ряда на каком-то участке.
- Корреляции.

Для описание, оценивания и создания каких-либо разумных выводов о наблюдаемых временных рядах была разработана теория случайных процессов. В ней были разработаны такие методы, как фильтрация, сегментация, шумоподавление, анализ трендов, различные виды анализов: корреляционный, дисперсионный, регрессионный, морфологический и так далее.

Одной из компонент временного ряда является так называемая *разладка*, то есть изменение статистических свойств ряда. Вариантов разладок бывает много: например, разрывы, изломы, рост разброса и нарушения цикла. На практике это может соответствовать, например, поломке агрегата. Из этого и появилась *задача о разладке*: как обнаружить возникающее изменение?

Теперь рассмотрим несколько примеров разладок:

- Пузырь доткомов.
- Посещаемость интернет-ресурса. Тут стоит различить два варианта: краткосрочные и долгосрочные разладки. Первые могут возникнуть, например, из-за неполадок с серверами или же из-за того, что кто-то известный прорекламировал ресурс. Долгосрочные же, например, вызываются некачественным контентом.

Вообще, задача о разладке — это весьма актуальная проблема, и она возникает в самых разных областях. Перечислим лишь несколько из них:

- Обнаружение внедрений в компьютерные сети (атак, ведущих к изменению объема передаваемого трафика),
- Обнаружение аномалий в сетях передачи данных (видеопотоки в системах видеонаблюдения, сетевой трафик и так далее),
- Обнаружение и изоляция отказов узлов систем управления транспортными средствами,
- Мониторинг целостности системы геопозиционирования,
- Обнаружение изменений структуры породы при бурении скважин,
- Обнаружение начала рецессии или экономического роста,
- Обнаружение изменений волатильности индекса Dow Jones,
- Обнаружение сигнала при наблюдении подводных целей,
- Автоматическое обнаружение аномального человеческого поведения при видеонаблюдении,
- Автоматический контроль качества выпускаемой продукции,
- Мониторинг и анализ смертности и заболеваемости раком легких,
- Обнаружение возникновения эпидемий,
- Обнаружение аритмии (внезапных изменений ритма биения) сердца,
- Предсказание транзиторных ишемических атак (преходящих нарушений мозгового кровообращения),
- Диагностика задержки внутриутробного роста,
- Анализ несчастных случаев на угольных шахтах,
- Мониторинг уровня хлора в питьевой воде.

На самом деле первые работы по разладкам появились очень давно: Уолтер Шухарт написал свою статью про контроль качества выпускаемой продукции в 1931-м году. На данный же момент работ по разладкам огромное количество: поиск в системе индексации Google Scholar выдает, что на период с 2000 по 2017 год было написано статей с ключевыми словами

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

## 1.2 Математическая задача о разладке

Пусть $\{\xi_k\}_{k\geqslant 1}$ — последовательность независимых случайных величин таких, что в моменты времени $k = 1, 2, \ldots, \theta-1$ они имеют распределение $\mathsf{P}_\infty$ с плотностью $f_\infty(x)$, а в моменты $k = \theta, \theta+1, \ldots$ — распределение $\mathsf{P}_0$ с плотностью $f_0(x)$. Далее, над ними проводится наблюдение, что даёт последовательность чисел $\{x_k\}_{k\geqslant 1}$.

Параметр $\theta$ считается заранее не известным и принимающим значения $0, 1, \ldots, \infty$. Значения $x_{\theta-1}$ при $\theta = 0$ и $\theta = 1$ будем считать равными нулю. Такое искусственное введение этих моментов будет оправдано рассмотрениями переходов от дискретных схем к соответствующим непрерывным (во времени) схемам.

Несложно понять, что случаи $\theta = 0$ и $\theta = 1$ отвечают наблюдениям с плотностью $f_0(x)$. Если же $\theta = \infty$, то все наблюдения имеют плотность распределения вероятностей $f_\infty(x)$.

Пусть $p_\theta(x_1, \ldots, x_n)$ — это плотность распределения случайных величин $\xi_1, \ldots, \xi_n$ при условии, что «переход» выполняется в момент $\theta \in [0, +\infty]$. Тогда для любого $n \in \mathbb{N}$

$$p_\theta(x_1, \ldots, x_n) = \begin{cases} f_\infty(x_1) \ldots f_\infty(x_n), & \theta = 0, 1, \\ f_\infty(x_1) \ldots f_\infty(x_{\theta-1}) f_0(x_\theta) \ldots f_0(x_n), & 1 < \theta \leqslant n, \\ f_0(x_1) \ldots f_0(x_n), & \theta > n. \end{cases} \tag{1.1}$$

Далее, введём следующую статистику:

$$\gamma_n = \sup_{\theta \geqslant 0} \left( \frac{p_\theta(x_1, \ldots, x_n)}{p_\infty(x_1, \ldots, x_n)} \right).$$

Заметим, что её можно переписать следующим образом:

$$\begin{aligned} \gamma_n &= \max \left( \frac{p_0(x_1, \ldots, x_n)}{p_\infty(x_1, \ldots, x_n)}, \frac{p_1(x_1, \ldots, x_n)}{p_\infty(x_1, \ldots, x_n)}, \ldots, \frac{p_n(x_1, \ldots, x_n)}{p_\infty(x_1, \ldots, x_n)}, \frac{p_\infty(x_1, \ldots, x_n)}{p_\infty(x_1, \ldots, x_n)} \right) \\ &= \max \left( \frac{p_0(x_1, \ldots, x_n)}{p_\infty(x_1, \ldots, x_n)}, \frac{p_0(x_2, \ldots, x_n)}{p_\infty(x_2, \ldots, x_n)}, \ldots, \frac{p_0(x_n)}{p_\infty(x_n)}, 1 \right) \\ &= \max \left( 1, \max_{1 \leqslant \theta \leqslant n} \prod_{k=\theta}^{n} \frac{f_0(x_k)}{f_\infty(x_k)} \right). \end{aligned}$$

Далее, введём ещё одну статистику, пользуясь обозначением $\zeta_k = \ln f_0(x_k) - \ln f_\infty(x_k)$:

$$T_n = \ln \gamma_n = \max \left( 0, \max_{1 \leqslant \theta \leqslant n} \sum_{k=\theta}^{n} \ln \frac{f_0(x_k)}{f_\infty(x_k)} \right) = \max \left( 0, \max_{1 \leqslant \theta \leqslant n} \sum_{k=\theta}^{n} \zeta_k \right).$$

Далее, если ввести обозначение $Z_n = \zeta_1 + \ldots + \zeta_n$, то её можно переписать следующим образом:

$$T_n = \max \left( 0, \max_{1 \leqslant \theta \leqslant n} [Z_n - Z_{\theta-1}] \right) = \max \left( 0, Z_n - \min_{0 \leqslant \theta \leqslant n-1} Z_\theta \right) = Z_n - \min_{0 \leqslant \theta \leqslant n} Z_\theta.$$

Статистики $T_n$ и $\gamma_n$ играют в задачах скорейшего обнаружения исключительно важную роль. Напомним, что в задаче различения двух гипотез ($\mathrm{H}_0$ и $\mathrm{H}_\infty$) ключевую роль играет отношение правдоподобия

$$L_n = \frac{f_0(x_1) \ldots f_0(x_n)}{f_\infty(x_1) \ldots f_\infty(x_n)}, \quad L_0 = 1.$$

Заметим, что

$$\gamma_n = \max\left(1, \max_{1\leqslant\theta\leqslant n}\prod_{k=\theta}^{n}\frac{f_0(x_k)}{f_\infty(x_k)}\right) = \max\left(\frac{L_n}{L_n}, \max_{1\leqslant\theta\leqslant n}\frac{L_n}{L_{\theta-1}}\right) = \max_{0\leqslant\theta\leqslant n}\frac{L_n}{L_\theta}.$$

Для статистики $T_n$ выполняется достаточно важное рекуррентное соотношение.

$$\begin{aligned}
T_n &= Z_{n-1} + \zeta_n - \min_{0\leqslant\theta\leqslant n} Z_\theta = \zeta_n + \max_{0\leqslant\theta\leqslant n}(Z_{n-1} - Z_\theta)\\
&= \zeta_n + \max(Z_{n-1} - Z_n, \max_{0\leqslant\theta\leqslant n-1}(Z_{n-1} - Z_\theta))\\
&= \zeta_n + \max(Z_{n-1} - Z_n, Z_n - \min_{0\leqslant\theta\leqslant n-1} Z_\theta)\\
&= \zeta_n + \max(-\zeta_n, T_{n-1}) = \max(0, T_{n-1} + \zeta_n).
\end{aligned}$$

Из этого свойства, в частности, видно, что в рассматриваемом случае независимых наблюдений статистики $(T_n)_{n\geqslant 1}$, образуют марковскую цепь (по каждой из мер $\mathsf{P}_0$ и $\mathsf{P}_\infty$ в пространстве последовательностей $x = (x_1, x_2, \ldots)$). Аналогичное верно и для статистик $(L_n)_{n\geqslant 1}$, так как $L_n = L_{n-1}e^{\zeta_n}$.

Теперь можно дать названия введённым статистикам.

- Статистики $\gamma = (\gamma_n)_{n\geqslant 1}$ принято называть *обобщенными отношениями правдоподобия.* В определенных постановках задач скорейшего обнаружения на этих статистиках основаны оптимальные или асимптотически оптимальные решения.

- Статистики $T = (T_n)_{n\geqslant 1}$ называют *статистиками кумулятивных сумм* (или же CUSUM — от CUmulative SUMs). Это название объясняется тем, что там, где $T_k > 0$, $k \leqslant n$, статистика $T_n$ просто равна сумме $\zeta_k$, то есть является кумулятивной суммой. Если же $T_{n-1}+\zeta_n < 0$, то согласно рекуррентному соотношению $T_n$ просто полагается равной нулю, а не $T_{n-1} + \zeta_n$.

К слову об истории. «Первопроходец» в теории решения задач о разладке, Уолтер Шухарт, предложил следующую статистику, которую называют *контрольными картами Шухарта.* Её суть состоит в следующем: последовательность наблюдений $(x_1, x_2, \ldots)$ разбивают на группы (батчи) одинаковой длины $n$, после чего для каждого батча строится логарифм отношения правдоподобия:

$$S_m^n = \ln\frac{p_0(x_m, x_{m+1}, \ldots, x_n)}{p_\infty(x_m, x_{m+1}, \ldots, x_n)}.$$

Если все $\xi_k$ независимы и одинаково распределены, то $S_m^n = \zeta_m + \zeta_{m+1} + \ldots + \zeta_n$.

Ещё вводится так называемая *статистика Ширяева-Робертса* (SR-статистика):

$$R_n = \sum_{k=1}^{n}\frac{p_\theta(x_1, \ldots, x_n)}{p_\infty(x_1, \ldots, x_n)}.$$

В случае независимых и одинаково распределённых случайных величин SR-статистику можно записать следующим образом:

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

Теперь вспомним, что

$$\prod_{k=\theta}^{n}\frac{f_0(x_k)}{f_\infty(x_k)} = \frac{L_n}{L_{\theta-1}} \implies R_n = \sum_{\theta=1}^{n}\frac{L_n}{L_{\theta-1}} = \sum_{\theta=1}^{n}\exp\{Z_n - Z_{\theta-1}\}.$$

Для статистики $R_n$ верно следующее рекуррентное соотношение (при $R_0 = 0$):

$$R_n = \sum_{\theta=1}^{n} \exp\{Z_n - Z_{\theta-1}\} = \exp\{Z_n - Z_{n-1}\} + \sum_{\theta=1}^{n-1} \exp\{Z_n - Z_{\theta-1}\}$$
$$= e^{\zeta_n} + e^{\zeta_n} \sum_{\theta=1}^{n-1} \exp\{Z_{n-1} - Z_{\theta-1}\} = (1 + R_{n-1}) e^{\zeta_n}.$$

Допустим, что мы рассматриваем случай с нормальным распределением, когда $\mathsf{P}_0(x) = \mathcal{N}(\mu_0, \sigma^2)$, а $\mathsf{P}_\infty(x) = \mathcal{N}(\mu_\infty, \sigma^2)$. В этом случае плотности равны

$$f_0(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu_0)^2}{2\sigma^2}\right\}, \qquad f_0(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu_\infty)^2}{2\sigma^2}\right\}.$$

Следовательно, $\zeta_n$ равна

$$\zeta_n = \ln \frac{f_0(x_n)}{f_\infty(x_n)} = \frac{(x-\mu_\infty)^2}{2\sigma^2} - \frac{(x-\mu_0)^2}{2\sigma^2} = \frac{\mu_0 - \mu_\infty}{\sigma^2} \left( x_n - \frac{\mu_0 + \mu_\infty}{2} \right).$$

Теперь несложно выписать рекуррентные формулы для $T_n$ и $R_n$:

$$T_n = \max \left( 0, T_{n-1} + \frac{\mu_0 - \mu_\infty}{\sigma^2} \left( x_n - \frac{\mu_0 + \mu_\infty}{2} \right) \right)$$
$$R_n = (1 + R_{n-1}) \exp \left\{ \frac{\mu_0 - \mu_\infty}{\sigma^2} \left( x_n - \frac{\mu_0 + \mu_\infty}{2} \right) \right\}.$$

Теперь заменим измерение случайной величины $x_n$ на саму случайную величину $\xi_n$. Тогда можно посчитать матожидания $\zeta_n$ по мерам $\mathsf{P}_0$ и $\mathsf{P}_\infty$:

$$\mathsf{E}_0[\zeta_n] = \frac{\mu_0 - \mu_\infty}{\sigma^2} \left( \mu_0 - \frac{\mu_0 + \mu_\infty}{2} \right) = \frac{(\mu_0 - \mu_\infty)^2}{\sigma^2},$$
$$\mathsf{E}_\infty[\zeta_n] = \frac{\mu_0 - \mu_\infty}{\sigma^2} \left( \mu_\infty - \frac{\mu_0 + \mu_\infty}{2} \right) = -\frac{(\mu_0 - \mu_\infty)^2}{\sigma^2}.$$

Отсюда видно, что при $n < \theta$ статистика $T_n$ близка к нулю, однако при $n \geqslant \theta$ она начинает (в среднем) возрастать. Это рассуждение наглядно показывает, что статистики $T_n$ и $R_n$ позволяют «ухватывать» момент появления разладки, то есть момент $\theta$. Именно эти наблюдения послужили основанием применения этих статистик к реальным данным. Разумеется, для этой цели надо знать или уметь «хорошо» оценивать величины $\zeta_n$, когда плотности $f_0(x)$ и $f_\infty(x)$ точно не известны.

Среди других методов, в которых не требуется знание плотностей, укажем на “метод экспоненциального сглаживания”. Он состоит в следующем: по наблюдениям $(x_1, x_2, \ldots)$ строятся статистики $(Y_n)_{n\geqslant 0}$ по следующему правилу:

$$Y_n = (1 - \beta) Y_{n-1} + \beta x_n,$$

где $Y_0 = 0$ и $\beta \in (0, 1)$ — некоторый (подбираемый) параметр. Здесь несложно развернуть рекурсифную формулу, которая объясняет название метода:

$$Y_n = (1-\beta) Y_{n-1} + \beta x_n = (1-\beta)^2 Y_{n-2} + \beta(1-\beta) x_{n-1} + \beta x_n$$
$$= \ldots = \beta \sum_{k=1}^{n} (1-\beta)^{n-k} x_k.$$

В случае различения двух гипотез мы рассмотрели две постановки задач. Одна (формулировка Неймана–Пирсона) носила, так сказать, ретроспективный характер, поскольку решение выносилось по заданным наблюдениям $(x_1, x_2, \ldots, x_n)$. Вторая постановка (формулировка А. Вальда) исходила из того, что наблюдения поступают последовательно, шаг за шагом, и решение принимается по поведению текущих данных.

В задаче о разладке также принято различать две постановки — ретроспективная и последовательная. В первом случае наблюдения $(x_1, x_2, \ldots, x_n)$ заданы *полностью* и считается, что $\theta \in \{1, 2, \ldots, n, \infty\}$. Во втором случае наблюдения поступаюь последовательно и, вообще говоря, $\theta \in \mathbb{N}$. Введенные статистики кумулятивных сумм $T_n$ и статистики Ширяева-Робертса $R_n$ пригождаются в обоих сформулированных случаях.

Стоит заметить, что введённые выше статистики предназначены для тех ситуаций, когда относительно параметра $\theta$ не делается никаких априорных предположений. А как можно улучшить ситуацию, если такое предположение есть? Рассмотрим на следующем примере. Пусть $\theta$ — это случайная величина, принимающая значения из $\mathbb{Z}_+$ со следующими вероятностями:

$$\begin{gathered}\mathsf{P}(\theta = 0) = \pi,\\ \mathsf{P}(\theta = n \mid \theta > 0) = pq^{n-1}, \quad n \in \mathbb{N},\end{gathered}$$

где $\pi \in [0, 1]$, $q \equiv 1 - p$. Сразу же скажем, что $\mathsf{P}(\theta = n) = (1 - \pi)pq^{n-1}$ для любого $n \in \mathbb{N}$.

Как и раньше, считаем, что случайные величины независимы и одинаково распределены с плотностями $f_\infty(x)$ (до момента разладки) и $f_0(x)$ (после). Тогда совместная плотность $p_\theta(x_1, \ldots, x_n)$ задаётся формулой (1.1).

Пусть $p^\pi(x_1, \ldots, x_n)$ — это плотность набора случайных величин $(\xi_1, \ldots, \xi_n)$ в точках $(x_1, \ldots, x_n)$ при условии, что $\theta$ есть случайная величина с распределением, заданным выше. Тогда по формуле полной вероятности

$$\begin{aligned}p^\pi(x_1, \ldots, x_n) &= \sum_{k=0}^{\infty} p_k(x_1, \ldots, x_n)\,\mathsf{P}(\theta = k)\\ &= (\pi + p(1 - \pi))f_0(x_1) \ldots f_0(x_n)\\ &\quad + (1 - \pi)p\sum_{k=1}^{n} q^k f_\infty(x_1) \ldots f_\infty(x_{k-1})f_0(x_k) \ldots f_0(x_n)\\ &\quad + (1 - \pi)q^n f_\infty(x_1) \ldots f_\infty(x_n).\end{aligned}$$

Далее, введём статистику $(\pi_n)_{n\geqslant 1}$ по следующему правилу:

$$\pi_n = \mathsf{P}(\theta \leqslant n \mid x_1, \ldots, x_n).$$

Она будет играть исключительно важную роль во многих постановках задач скорейшего обнаружения.

Можно ли считать её рекурсивно? Можно. Для этого введём следующее обозначение:

$$\varphi_n \equiv \frac{\pi_n}{1 - \pi_n}.$$

Теперь докажем одну лемму, связанную с $\varphi_n$:

**Лемма.** *Для любого натурального $n$*

$$\varphi_n = (p + \varphi_{n-1})\frac{f_0(x_n)}{qf_\infty(x_n)}.$$

*Доказательство.* Для начала вспомним формулу Байеса:

$$\mathsf{P}(A \mid B) = \frac{\mathsf{P}(A)}{\mathsf{P}(B)} \mathsf{P}(B \mid A).$$

Отсюда получаем, что

$$\mathsf{P}(\theta = k \mid x_1, \ldots, x_n) = \frac{\mathsf{P}(\theta = k)}{p^\pi(x_1, \ldots, x_n)} p_k(x_1, \ldots, x_n).$$

Воспользуемся этим и распишем $\varphi_n$:

$$\begin{aligned}\varphi_n &= \frac{\mathsf{P}(\theta \leqslant n \mid x_1, \ldots, x_n)}{\mathsf{P}(\theta > n \mid x_1, \ldots, x_n)} = \sum_{k=0}^{n} \frac{p_k(x_1, \ldots, x_n)\,\mathsf{P}(\theta = k)}{p_\infty(x_1, \ldots, x_n)\,\mathsf{P}(\theta > n)} \\ &= \frac{p_0(x_1, \ldots, x_n)\,\mathsf{P}(\theta = 0)}{p_\infty(x_1, \ldots, x_n)\,\mathsf{P}(\theta > n)} + \sum_{k=1}^{n} \frac{p_k(x_1, \ldots, x_n)\,\mathsf{P}(\theta = k)}{p_\infty(x_1, \ldots, x_n)\,\mathsf{P}(\theta > n)}.\end{aligned}$$

Теперь вспомним, что $\mathsf{P}(\theta = 0) = \pi$ и $\mathsf{P}(\theta = k) = (1-\pi)pq^{k-1}$ для любого $k \in \mathbb{N}$. Тогда

$$\mathsf{P}(\theta > n) = \sum_{k=n+1}^{\infty} \mathsf{P}(\theta = k) = (1-\pi)pq^n \sum_{k=0}^{\infty} q^k = (1-\pi)q^n.$$

Далее, введём следующее обозначение:

$$L_m^n \equiv \frac{f_0(x_m) f_0(x_{m+1}) \ldots f_0(x_n)}{f_\infty(x_m) f_\infty(x_{m+1}) \ldots f_\infty(x_n)}.$$

Пользуясь вышесказанным и (1.1), преобразуем $\varphi_n$:

$$\varphi_n = \frac{1}{q^n} \left( \frac{\pi}{1-\pi} L_1^n + pL_1^n + pqL_2^n + \ldots + pq^{n-1} L_n^n \right).$$

Далее заметим, что

$$\varphi_n = \frac{f_0(x_n)}{q f_\infty(x_n)} \left( p + \frac{1}{q^{n-1}} \left( \frac{\pi}{1-\pi} L_1^{n-1} + pL_1^{n-1} + pqL_2^{n-1} + \ldots + pq^{n-2} L_{n-1}^{n-1} \right) \right).$$

А это и есть желаемое:

$$\varphi_n = (p + \varphi_{n-1}) \frac{f_0(x_n)}{q f_\infty(x_n)}.$$ □

Отсюда несложно получить рекурсивное выражение для $\pi_{n+1}$:

$$\pi_{n+1} = \frac{f_0(x_{n+1})(\pi_n + p(1-\pi_n))}{f_0(x_{n+1})(\pi_n + p(1-\pi_n)) + f_\infty(x_{n+1}) q (1-\pi_n)}.$$

*Упражнение.* Оказывается, что получить рекуррентную формулу для $\pi_{n+1}$ можно напрямую, пользуясь формулой Байеса. Покажите это.

Пусть

$$\zeta_n(p) = \ln \frac{f_0(x_n)}{(1-p) f_\infty(x_n)}.$$

Следовательно,

$$\varphi_n = (p + \varphi_{n-1}) e^{\zeta_n(p)} \implies \frac{\varphi_n}{p} = \left(1 + \frac{\varphi_n}{p}\right) e^{\zeta_n(p)}.$$

Тем самым, статистика $R_n(p) \equiv \varphi_n/p$ достаточно похожа на статистику Ширяева-Робертса $R_n$ (в том плане, что их рекуррентные соотношения похожи). Более того, при $p \to 0$ $\zeta_n(p) \to \zeta_n$ и $R_n(p) \to R_n$.

Это показывает, что введенная выше ad-hoc статистика $R_n$ возникает весьма естественным образом из статистики $R_n(p)$ при $p \to 0$. Стоит заметить, что

$$\mathsf{E}[\theta] = \frac{1-\pi}{p}.$$

Таким образом, предельный переход $p \to 0$ означает, что делается предположение, что появление разладки следует ожидать нескоро.

## 1.3 Основные постановки задачи скорейшего обнаружения для броуновского движения

В статистике случайных процессов хорошо развита та ее часть, в которой исходными предположениями является то, что рассматриваемые процессы являются *стационарными*. В основе теории таких процессов лежат ковариационно-спектральные характеристики (то есть ковариационная функция, так называемая спектральная плотность и иже с ними). Типичной задачей статистики таких процессов является задача определения корреляционной функции и среднего значения наблюдаемого процесса.

Сейчас же мы будем рассматривать сугубо *нестационарные* процессы. Это обстоятельство приводит и к новым постановкам задач и к необходимости развития соответствующей теории их решения. Описанный ниже материал будет посвящён моделям, основанным на броуновском движении, что объясняется и тем, что такие модели представляют практический интерес, и тем, что для них во многих случаях удается получить прозрачные и точные результаты.[1]

Пусть $B = (B_t)_{t\geqslant 0}$ — броуновское движение, заданное на некотором вероятностном пространстве $(\Omega, \mathcal{F}, \mathsf{P})$. Построим по нему процесс $X = (X_t)_{t\geqslant 0}$ по следующему правилу:

$$X_t = \mu\, \mathrm{I}_{\{t\geqslant\theta\}} + \sigma B_t \implies X_t = \begin{cases} \sigma B_t, & t < \theta, \\ \mu + \sigma B_t, & t \geqslant \theta, \end{cases}$$

где $\sigma > 0$, $\mu \neq 0$ и $\theta \in [0, \infty]$.

Относительно природы момента $\theta$, то есть того, является ли $\theta$ случайной величиной или же просто константным параметром, можно будет делать вполне конкретные предположения. Сейчас же важно лишь то, что момент $\theta$ — это тот момент, когда у наблюдаемого процесса $X$ меняется снос у броуновского движения (с 0 на $\mu$). Момент $\theta$ называется моментом появления *разладки*. При этом «нормальный» ход процесса $X$, характеризующийся тем, что $X_t = \sigma B_t$, переходит в «разлаженный»: $X_t = \mu + \sigma B_t$, где снос $\mu \neq 0$.

Несложно понять, что означают крайние случаи $\theta = 0$ и $\theta = \infty$. Случай, когда $\theta = 0$, соответствует тому, что с самого начала (т. е. с момента $t = 0$) идет «разлаженный» процесс. Случай $\theta = \infty$ соответствует тому, что «разладка» не появляется и вовсе, следовательно, у наблюдаемого процесса $X$ все время $X_t = \sigma B_t$.

Пусть $\mathsf{P}_\theta = \mathrm{Law}(X \mid \theta)$ — распределение вероятностей (по сути, семейство конечномерных функций распределения) процесса $X = (X_t)_{t\geqslant 0}$ при условии, что разладка происходит в момент времени $\theta$. В частности, $\mathsf{P}_\infty$ есть распределение вероятностей процесса $X$ при условии, что разладка вообще не происходит.

[1] Действительно, для броуновского движения задача различения двух гипотез относительно сноса может быть решена точно, но в общем случае получить что-то лучше приближенных результатов не так уж и просто.

Для дальнейшего повествования понадобится понятие момента остановки, или же марковского момента. Наглядный смысл этого поняти состоит в том, что для каждого $t \geqslant 0$ решение вопроса о том, прекратить ли наблюдения или же продолжать их, зависит лишь от информации о процессе $X$, полученной на интервале времени $[0, t]$ и не зависит от «будущего». В данный момент марковские моменты удобно рассматривать, как моменты подачи сигнала тревоги о появлении разладки.

Рассмотрим два события: $\{\tau < \theta\}$ и $\{\tau \geqslant \theta\}$. Первое событие означает ложную тревогу (так как сигнал подаётся до момента разладки). Во втором же событии нас обычно интересует то, насколько велико запаздывание $\tau - \theta$ при правильной подаче сигнала.

Теперь сформулируем четыре варианта задачи о разладке.

(**A**) Пусть $\theta$ — это случайная величина со следующим распределением:

$$\mathsf{P}(\theta = 0) = \pi$$
$$\mathsf{P}(\theta > t \mid \theta > 0) = e^{-\lambda t},$$

где $\pi \in [0, 1)$ и $\lambda > 0$ — какие-то известные константы.

Далее, зафиксируем какую-то константу $\alpha > 0$ и построим для неё следующий класс:

$$\mathcal{M}_\alpha = \{\tau \colon \mathsf{P}(\tau < \theta) \leqslant \alpha\}.$$

Другими словами, $\mathcal{M}_\alpha$ — это класс тех моментов остановки $\tau$ (относительно естественной фильтрации $\mathbb{F}^X$), для которых вероятность ложной тревоги не превосходит $\alpha$.

Нужно найти такой момент остановки $\tau^*_\alpha \in \mathcal{M}_\alpha$, для которого время срабатывания минимально:

$$\mathsf{E}[\tau^*_\alpha - \theta \mid \tau^*_\alpha \geqslant \theta] = \inf_{\tau \in \mathcal{M}_\alpha} \mathsf{E}[\tau - \theta \mid \tau \geqslant \theta].$$

Для решения этой *условно-вариационной* задачи посмотрим на следующую байесовскую постановку задачи о разладке. Пусть

$$\mathbf{A}(c) = \inf_{\tau \in \mathcal{M}} \big(\mathsf{P}(\tau < \theta) + c\, \mathsf{E}\big[(\tau - \theta)^+\big]\big),$$

где $\mathcal{M}$ — класс всех конечных моментов остановки, $c > 0$ — некоторая константа, а $\xi^+ = \max(0, \xi)$. Тогда

$$\mathsf{E}\big[(\tau - \theta)^+\big] = \mathsf{E}\big[(\tau - \theta)\, \mathrm{I}_{\{\tau \geqslant \theta\}}\big] = \mathsf{E}[\tau - \theta \mid \tau \geqslant \theta]\, \mathsf{P}(\tau \geqslant \theta).$$

Момент $\tau^*_c$ будем называть *оптимальным*, если на нём достигается инфинум:

$$\underbrace{\mathsf{P}(\tau^*_c < \theta)}_{\substack{\text{вероятность}\\ \text{ложного срабатывания}}} + \underbrace{c\, \mathsf{E}\big[(\tau^*_c - \theta)^+\big]}_{\substack{\text{(условное) среднее}\\ \text{время срабатывания}}} = \mathbf{A}(c).$$

В трёх следующих вариантах — **B**, **C** и **D** — параметр $\theta$ будем считать постоянным и принимающим значения в $[0, \infty]$.

(**B**) Зафиксируем некоторое число $T > 0$ и рассмотрим следующий класс:

$$\mathcal{M}_\tau = \{\tau \colon \mathsf{E}_\infty[\tau] \geqslant T\}.$$

Этот класс содержит такие моменты остановки $\tau$, для которых среднее время до ложной тревоги не меньше $T$. Далее, пусть

$$\mathbf{B}(T) = \inf_{\tau \in \mathcal{M}_T} \frac{1}{T} \int\limits_0^\infty \mathsf{E}_\theta[(\tau - \theta)^+]\, d\theta.$$

В этом варинате момент остановки $\tau_T^*$ будет оптимальным, если на нём достигается инфинум:

$$\inf_{\tau \in \mathcal{M}_T} \frac{1}{T} \int\limits_0^\infty \mathsf{E}_\theta[(\tau_T^* - \theta)^+]\, d\theta = \mathbf{B}(T).$$

**(C)** Этот вариант ещё называют *первой минимаксной задачей.* Пусть

$$\mathbf{C}(T) = \inf_{\tau \in \mathcal{M}_T} \sup_{\theta \geqslant 0} \mathsf{E}_\theta[\tau - \theta \,|\, \tau \geqslant \theta].$$

Будем говорить, что момент $\tau_T^* \in \mathcal{M}_T$ является оптимальным (если он существует) в варианте **C**, если на нём достигается инфинум:

$$\underbrace{\sup_{\theta \geqslant 0} \underbrace{\mathsf{E}_\theta[\tau_T^* - \theta \,|\, \tau_T^* \geqslant \theta]}_{\substack{\text{(условное) среднее}\\ \text{время срабатывания}}}}_{\text{наихудшее по всем } \theta} = \mathbf{C}(T).$$

Для введения варианта **D** нужно ввести понятие вещественного супремума.

**Определение 1.** Будем говорить, что $a$ является *вещественным супремумом* функции $f$, если $f(x) \leqslant a$ почти везде, то есть мера множества $\{x \colon f(x) > a\}$ равна нулю.
**Обозначение**: $a = \operatorname{ess\,sup} f(x)$.

**(D)** Этот вариант называют *второй минимаксной задачей.* Пусть

$$\mathbf{D}(T) = \inf_{\tau \in \mathcal{M}_T} \sup_{\theta \geqslant 0} \operatorname*{ess\,sup}_{\omega} \mathsf{E}_\theta[(\tau - \theta)^+ \,|\, \mathcal{F}_\theta^X](\omega),$$

где $\mathcal{F}_\theta^X = \sigma(X_t, t \leqslant \theta)$. Момент $\tau_T^* \in \mathcal{M}_T$ называется оптимальным (если он существует) в варианте **D**, если

$$\underbrace{\sup_{\theta \geqslant 0} \underbrace{\operatorname*{ess\,sup}_{\omega} \underbrace{\mathsf{E}_\theta[(\tau - \theta)^+ \,|\, \mathcal{F}_\theta^X](\omega)}_{\text{среднее время срабатывания}}}_{\text{наихудшее по всем траекториям}}}_{\text{наихудшее по всем } \theta} = \mathbf{D}(T).$$

На самом деле есть ещё один вариант, который обычно называют вариантом **E**. Он кардинально отличается от остальных вариантов тем, что в них мы предполагаем, что после объявления сигнала тревоги наблюдение останавливается. На самом же деле в реальной практике дело обстоит так, что после объявления тревоги системы обнаружения продолжают свое функционирование, ожидая, скажем, появления следующей разладки. Этот же вариант задачи о разладке носит *многократный характер.* При этом целесообразно считать, что сам процесс наблюдения начался «давно», он мог прерываться ложными тревогами и «разладка» появляется на «фоне установившегося режима наблюдения».