УДК 519.2:53.05

Ф.Ф. Идрисов

РЕКУРРЕНТНЫЕ МОДЕЛИ РАНДОМИЗИРОВАННЫХ ПРОЦЕССОВ В ЗАДАЧАХ ВНЕШНЕЙ РАЗВЕДКИ. Часть 1. УПРОЩЕННЫЕ МОДЕЛИ НИЗШЕГО ПОРЯДКА*

Ключевые слова: разведданные, рандомизированный процесс, тренд, сплайн, дисперсия, ковариация оценок, рекуррентные оценки, имитационное моделирование.

Введение

Во все времена внешняя разведка являлась особо ответственной зоной внимания для всех государств. Ее значение существенно возрасло в наше время, когда появились новые формы межгосударственных конфликтов (например, сетевые и гибридные войны, масштабные террористические акты и т.д.) и усложнилось пространство их проведения (к традиционным морскому, сухопутному и воздушному пространствам добавилась новая сфера — киберпространство). В связи с этим проблемы внешней разведки привлекают внимание и спецслужб, и представителей научной среды [1–3] (сюда же можно отнести и работы [4–6], где исследованы модели финансовых потоков, обеспечивающих различные криминальные и террористические группы). Но в целом представляет интерес рассматривать внешнюю разведку в более системном контексте. В частности, естественнее исследовать возможности моделей проведения разведывательных операций не только для финансовых транзакций, но и таких направлений, как передвижение людских, материальных, информационных и других ресурсов с позиций теории случайных процессов.

В данной работе не делается акцент на возможные детали этой огромной проблемы, а лишь предлагается ее концептуальное решение и исследуются его свойства.

Все дело в том, что реально на практике наблюдается очень быстрое устаревание добытых данных. И гораздо более практичным может быть подход, учитывающий быструю динамику информационной ценности этих данных. Другими словами, нередко может представлять интерес рекуррентный подход к оцениванию параметров синтезируемой модели. Но в реальности задача усложняется еще и тем, что разведываемые процессы являются не только сами по себе случайными, но и наблюдаемыми в случайные моменты времени (в работе [4] такие процессы названы рандомизированными, чтобы подчеркнуть их отличие от эквидистантных случайных процессов).

Постановка проблемы

Для внешней разведки, как правило, ставится задача выявления некоторой закономерности в наблюдаемой цепочке событий (например, в начавшейся переброске сил и средств воинских соединений). По мере появления рандомизированных данных (т.е. поступающих в случайные моменты времени) необходимо оценить возникающие тренды.

---

* Исследование выполнено за счет проекта 8.9562 2017/8.9.

Однако выделение тренда в виде полинома выше четвертого порядка [4–6] нецелесообразно в силу того, что получается большая погрешность в оценке его параметров. С другой стороны, когда длительность интервала наблюдений T достаточно велика, полином невысокого порядка может неудовлетворительно описывать искомый тренд.

Выход из создавшейся ситуации может состоять в том, что весь интервал наблюдения за объектом [0, T] разбивается на интервалы $[0, T_0], [T_0, 2T_0], \dots$ $\dots, [(n-1)T_0, nT_0]$, и на каждом таком небольшом интервале будем оценивать тренд в виде полинома невысокого порядка. На границах интервалов эти полиномы «сшиваются» так, чтобы получалась непрерывная кривая. Такая кусочно-полиномиальная кривая, как известно, носит название сплайна [7, 8]. В работе рассматривается проблема выделения тренда временного ряда в виде сплайна, при этом строятся рекуррентные процедуры оценки параметров сплайна, когда коэффициенты полиномов оцениваются не все сразу, а последовательно, один за другим.

Уточним эту постановку для сплайна первого порядка. Рассмотрим отрезок $[rT_0, (r+1)T_0]$. Если момент t_i i-го изменения наблюдаемого объекта лежит на этом отрезке, то результат наблюдения можем представить в виде

$$x_i = x(t_i) = x_r(t_i) + n_i, \tag{1}$$

где $x_r(t)$, $r = \overline{0, n-1}$, — тренд временного ряда

$$x_r(t) = a_r + b_r \frac{t - rT_0}{T_0}, \; rT_0 \le t \le (r+1)T_0. \tag{2}$$

Таким образом, на r-м отрезке тренд имеет вид полинома первого порядка.

Потребуем, чтобы значения тренда на границах отрезков были «сшиты», т.е.

$$x_r((r+1)T_0) = x_{r+1}((r+1)T_0),$$

что приводит к условию

$$a_r + b_r = a_{r+1}. \tag{3}$$

Таким образом, тренд имеет вид кусочно-ломаной линии, которая и называется сплайном первого порядка.

Выделение такого тренда сводится к нахождению оценок $\hat{a}_r, \hat{b}_r$ параметров a_r, b_r, причем значения этих оценок также должны быть сшиты, т.е.

$$\hat{a}_r + \hat{b}_r = \hat{a}_{r+1}. \tag{4}$$

Другими словами, оценка тренда также должна быть сплайном первого порядка.

Выделение тренда в виде сплайна первого порядка

Заметим, что в рассматриваемом случае достаточно оценить коэффициенты a_r, так как коэффициенты b_r определяются через a_r очень просто: $b_r = a_{r+1} - a_r$.

Поэтому естественно с самого начала записать тренд так, чтобы он содержал только коэффициенты a_r: на участке $[rT_0, (r+1)T_0]$ тренд $x_r(t)$ может быть записан в виде

$$x_r(t) = a_r\left(r + 1 - \frac{t}{T_0}\right) + a_{r+1}\left(\frac{t}{T_0} - r\right), \quad rT_0 \le t \le (r+1)T_0, \tag{5}$$

так что если момент t_i попадает на интервал $[rT_0, (r+1)T_0]$, то $x_i = x(t_i)$ имеет вид

$$x_i = x(t_i) = a_r\left(r+1-\frac{t_i}{T_0}\right) + a_{r+1}\left(\frac{t_i}{T_0} - r\right) + n_i, \tag{6}$$

где n_i — независимые случайные величины с $M\{n_i\} = 0$ и $D\{n_i\} = \sigma^2$.

А. Моменты изменений наблюдаемого объекта известны точно. Пусть моменты t_i известны точно. Образуем множества M_r индексов *i* следующим образом:

$$M_r = \{i : rT_0 \le t_i \le (r+1)T_0\}, \quad r = \overline{0, n-1}. \tag{7}$$

Другими словами, M_r — это множество индексов у тех t_i, которые попали на *r*-й интервал. Для оценки $\hat{a}_r$ параметров a_r рассмотрим статистики вида

$$S_r = \frac{1}{\lambda T_0}\sum_{i\in M_r}\left(\alpha + \beta\frac{t_i - rT_0}{T_0}\right)x_i. \tag{8}$$

Вычислим математическое ожидание этой статистики. Учитывая представление для x_i (6) и усредняя сначала по n_i с учетом того, что $M\{n_i\} = 0$, а затем по $\{t_i\}$, получим

$$M\{S_r\} = \frac{1}{T_0}\int_{rT_0}^{(r+1)T_0}\left(\alpha + \beta\frac{t - rT_0}{T_0}\right)\left(a_r\left(r+1-\frac{t}{T_0}\right) + a_{r+1}\left(\frac{t}{T_0} - r\right)\right)dt =$$

$$= \int_0^1 (\alpha + \beta x)(a_{r+1}x + a_r(1-x))\,dx = a_r\left(\frac{\alpha}{2} + \frac{\beta}{6}\right) + a_{r+1}\left(\frac{\alpha}{2} + \frac{\beta}{3}\right). \tag{9}$$

Отсюда следует, что если взять α = 4, β = – 6, то $M\{S_r\} = a_r$, а если взять α = – 2, β = 6, то $M\{S_r\} = a_{r+1}$.

Это позволяет предложить следующий рекуррентный алгоритм оценок параметров a_r начиная с a_0.

Оценку $\hat{a}_0$ параметра a_0 следует брать в виде

$$\hat{a}_0 = \frac{1}{\lambda T_0}\sum_{i\in M_0}\left(4 - 6\frac{t_i}{T_0}\right)x_i. \tag{10}$$

Оценку $\hat{a}_1$ параметра a_1 можно получить как из отрезка $[0, T_0]$, так и $[T_0, 2T_0]$. В общем случае можно рассматривать некоторое взвешенное среднее этих оценок, т.е. оценку вида

$$\hat{a}_1 = \frac{1}{\lambda T_0}\left[\alpha\sum_{i\in M_0}\left(6\frac{t_i}{T_0} - 2\right)x_i + (1-\alpha)\sum_{i\in M_1}\left(4 - 6\frac{t_i - T_0}{T_0}\right)x_i\right]$$

с некоторым α ∈ [0, 1].

Аналогично можно построить оценки для $a_2, a_3, ..., a_{n-1}$ в виде

$$\hat{a}_r = \frac{1}{\lambda T_0}\left[\alpha\sum_{i\in M_{r-1}}\left(6\frac{t_i - (r-1)T_0}{T_0} - 2\right)x_i + (1-\alpha)\sum_{i\in M_r}\left(4 - 6\frac{t_i - rT_0}{T_0}\right)x_i\right], \tag{11}$$

и лишь оценка $\hat{a}_n$ последнего параметра $a_n = a_{n-1} + b_{n-1}$ имеет особый вид

$$\hat{a}_n = \frac{1}{\lambda T_0} \sum_{i \in M_{n-1}} \left(6 \frac{t_i - (n-1)T_0}{T_0} - 2 \right) x_i . \qquad (12)$$

По-видимому, достаточно разумным выбором для α является значение $\alpha = 0$, когда

$$\hat{a}_r = \frac{1}{\lambda T_0} \sum_{i \in M_r} \left(4 - 6 \frac{t_i - rT_0}{T_0} \right) x_i . \qquad (13)$$

В этом случае оценки $a_0, a_1, a_2, ..., a_{n-1}$ будут некоррелированы, так как в них фигурируют n_i и t_i из непересекающихся множеств, и лишь $\hat{a}_{n-1}$ и $\hat{a}_n$ будут зависимы.

Найдем теперь дисперсию построенных оценок, считая поток моментов измерений $\{t_i\}$ пуассоновским. Учитывая выражение (6) для x_i, получаем

$$\hat{a}_r^2 = \frac{1}{(\lambda T_0)^2} \sum_{i,\, j \in M_r} \left(4 - 6 \frac{t_i - rT_0}{T_0} \right) \left(4 - 6 \frac{t_j - rT_0}{T_0} \right) \times$$

$$\times \left[a_r \left((r+1) - \frac{t_i}{T_0} \right) + a_{r+1} \left(\frac{t_i}{T_0} - r \right) + n_i \right] \left[a_r \left((r+1) - \frac{t_j}{T_0} \right) + a_{r+1} \left(\frac{t_j}{T_0} - r \right) + n_j \right].$$

Усредняя по n_i с учетом их независимости, запишем следующее выражение:

$$M\{\hat{a}_r^2 \big| \{t_i\}\} = \frac{\sigma^2}{(\lambda T_0)^2} \sum_{i \in M_r} \left(4 - 6 \frac{t_i - rT_0}{T_0} \right)^2 + \frac{1}{(\lambda T_0)^2} \sum_{i,\, j \in M_r} \left(4 - 6 \frac{t_i - rT_0}{T_0} \right) \times$$

$$\times \left(4 - 6 \frac{t_j - rT_0}{T_0} \right) \left(a_r \left((r+1) - \frac{t_i}{T_0} \right) + a_{r+1} \left(\frac{t_i}{T_0} - r \right) \right) \cdot \left(a_r \left(r + 1 - \frac{t_j}{T_0} \right) + a_{r+1} \left(\frac{t_j}{T_0} - r \right) \right).$$

Усредняя далее по моментам измерений $\{t_i\}$ с учетом того, что поток этих моментов пуассоновский, после замены переменных получим

$$M\{\hat{a}_r^2\} = \frac{\sigma^2}{\lambda T_0} \int_0^1 (4 - 6x)^2 dx + \left(\int_0^1 (4 - 6x)(a_r(1-x) + a_{r+1} x) dx \right)^2 +$$

$$+ \frac{1}{\lambda T_0} \int_0^1 (4 - 6x)^2 (a_r(1-x) + a_{r+1} x)^2 dx.$$

Вычисляя интегралы, получаем

$$M\{\hat{a}_r^2\} = a_r^2 + 4 \frac{\sigma^2}{\lambda T_0} + \frac{1}{\lambda T_0} \left(\frac{38}{15} a_r^2 + \frac{14}{15} a_r a_{r+1} + \frac{8}{15} a_{r+1}^2 \right),$$

отсюда окончательно

$$D\{\hat{a}_r\} = 4 \frac{\sigma^2}{\lambda T_0} + \frac{1}{15 \lambda T_0} (38 a_r^2 + 14 a_r a_{r+1} + 8 a_{r+1}^2) . \qquad (14)$$

Построим теперь несмещенную оценку величины $D\{\hat{a}_r\}$. Для этого рассмотрим статистики вида

$$S = \frac{1}{\lambda T_0} \sum_{i \in M_r} \left(4 - 6\frac{t_i - rT_0}{T_0}\right)^2 x_i^2. \tag{15}$$

Вычисляя математическое ожидание этой статистики аналогично тому, как это сделано выше при вычислении $M\{\hat{a}_r^2\}$, получим

$$M\{S\} = 4\sigma^2 + \frac{1}{15}(38a_r^2 + 14a_r a_{r+1} + 8a_{r+1}^2),$$

так что для пуассоновского потока моментов измерений несмещенная оценка $\hat{D}\{\hat{a}_r\}$ дисперсии $D\{a_r\}$ имеет вид

$$\hat{D}\{\hat{a}_r\} = \frac{1}{(\lambda T_0)^2} \sum_{i \in M_r} \left(4 - 6\frac{t_i - rT_0}{T_0}\right)^2 x_i^2. \tag{16}$$

Заметим, что относительно асимптотических свойств рассмотренных выше статистик верны все теоремы, доказанные в [4–6].

В случае, если поток измерений является рекуррентным, изменяются лишь выражения для $D\{\hat{a}_r\}$ и ее оценки. Выражение для $D\{\hat{a}_r\}$ примет вид

$$D\{\hat{a}_r\} = 4\frac{\sigma^2}{\lambda T_0} + \frac{c^2}{15\lambda T_0}(38a_r^2 + 14a_r a_{r+1} + 8a_{r+1}^2), \tag{17}$$

и в качестве ее оценки можно взять статистику

$$\hat{D}\{\hat{a}_r\} = \frac{1}{(\lambda T_0)^2} \sum_{i \in M_r} \left(4 - 6\frac{t_i - rT_0}{T_0}\right)^2 x_i^2 + \frac{c^2 - 1}{15\lambda T_0}(38\hat{a}_r^2 + 14\hat{a}_r \hat{a}_{r+1} + 8\hat{a}_{r+1}^2). \tag{18}$$

Б. Моменты изменений в наблюдаемом объекте неизвестны. Рассмотрим теперь ситуацию, когда моменты изменений наблюдаемого объекта на интервале [0, *T*] неизвестны, но предположим, что их было *N* и $0 < t_1 < t_2 < \ldots < t_N < T$. При исследовании данной ситуации воспользуемся результатами [5] относительно свойств моментов t_i, а именно тем, что

$$\begin{gathered} M\{t_i | N\} = T\frac{i}{N+1} \approx T\frac{i}{N}, \\ D\{t_i | N\} = T^2 c^2 \frac{i(N+1-i)}{(N+1)^2 (N+2)} \approx T^2 c^2 \frac{i(N-i)}{N^3}. \end{gathered} \tag{19}$$

Разобьем все измеренные *N* значений на *n* равных групп по N_0 измерений в каждой, так что $N = nN_0$. Тогда в *k*-й группе ($k = \overline{0, n-1}$) индекс измерения *i* меняется в пределах $kN_0 \le i \le (k+1)N_0$. Обозначая $i = kN_0 + j$, представим измеренные значения в *k*-й группе в виде

$$x_{kN_0+j} = a_{k+1}\left(\frac{t_{kN_0+j}}{T_0} - k\right) + a_k\left(k + 1 - \frac{t_{kN_0+j}}{T_0}\right) + n_{kN_0+j}, \tag{20}$$

где пренебрегли возможностью «перескока» измерения из одной группы в другую на стыках групп. При достаточно большом N_0 этот эффект играет незначительную роль.

Отсюда

$$M\{x_{kN_0+j}|N\} = a_{k+1}\left(T\frac{kN_0+j}{NT_0} - k\right) + a_k\left(k+1-T\frac{kN_0+j}{NT_0}\right). \quad (21)$$

Однако

$$T\frac{kN_0+j}{NT_0} = \frac{nT_0}{nN_0T_0}(kN_0+j) = k + \frac{j}{N_0},$$

так что

$$M\{x_{kN_0+j}|N\} = a_{k+1}\frac{j}{N_0} + a_k\left(1-\frac{j}{N_0}\right). \quad (22)$$

Рассмотрим теперь статистики вида

$$S_k = \frac{1}{N_0}\sum_{j=1}^{N_0}\left(\alpha+\beta\frac{j}{N_0}\right)x_{kN_0+j}. \quad (23)$$

Тогда, принимая во внимание (22), получаем

$$M\{S_k|N\} = a_{k+1}\left(\alpha\frac{1}{N_0^2}\sum_{j=1}^{N_0} j + \beta\frac{1}{N_0^3}\sum_{j=1}^{N_0} j^2\right) +$$

$$+ a_k\left(\alpha\frac{1}{N_0}\sum_{j=1}^{N_0}\left(1-\frac{j}{N_0}\right) + \beta\frac{1}{N_0^2}\sum_{j=1}^{N_0} j\left(1-\frac{j}{N_0}\right)\right).$$

Входящие сюда суммы легко вычисляются [9]

$$\frac{1}{N_0}\sum_{j=1}^{N_0} j = \frac{N_0+1}{2N_0}; \quad \frac{1}{N_0}\sum_{j=1}^{N_0}\left(1-\frac{j}{N_0}\right) = \frac{N_0-1}{2N_0};$$

$$\frac{1}{N_0^3}\sum_{j=1}^{N_0} j^2 = \frac{(N_0+1)(2N_0+1)}{6N_0^2}; \quad \frac{1}{N_0^2}\sum_{j=1}^{N_0} j\left(1-\frac{j}{N_0}\right) = \frac{N_0^2-1}{6N_0^2}$$

При больших N_0 приближенно

$$\frac{1}{N_0}\sum_{j=1}^{N_0} j \approx \frac{1}{2}; \quad \frac{1}{N_0}\sum_{j=1}^{N_0}\left(1-\frac{j}{N_0}\right) \approx \frac{1}{2}; \quad \frac{1}{N_0^3}\sum_{j=1}^{N_0} j^2 \approx \frac{1}{3}; \quad \frac{1}{N_0^2}\sum_{j=1}^{N_0} j\left(1-\frac{j}{N_0}\right) \approx \frac{1}{6},$$

так что при $N_0 >> 1$

$$M\{S_k|N\} = a_{k+1}\left(\frac{\alpha}{2}+\frac{\beta}{3}\right) + a_k\left(\frac{\alpha}{2}+\frac{\beta}{6}\right).$$

Если взять β = – 6, α = + 4, то $M\{S_k|N\} = a_k$. Поэтому в качестве оценки $\hat{a}_k$ параметра a_k можно брать статистику

$$\hat{a}_k = \frac{1}{N_0}\sum_{j=1}^{N_0}\left(-6\frac{j}{N_0}+4\right)x_{kN_0+j}, \quad (24)$$

которая дает асимптотически несмещенную оценку параметра a_k.

ISSN 0572-2691

Вычислим теперь асимптотическую (при $N_0 \to \infty$) дисперсию оценки $\hat{a}_k$. Так как $M\{t_{kN_0+j} | N\} = T(kN_0 + j)/N$, то представим t_{kN_0+j} в виде

$$t_{kN_0+j} = T\frac{kN_0+j}{N} + \Delta t_{kN_0+j}. \tag{25}$$

Полагая $T = nT_0,\ N = nN_0,$ приведем это выражение к виду

$$t_{kN_0+j} = T_0\left(k + \frac{j}{N_0}\right) + \Delta t_{kN_0+j},$$

так что

$$\frac{t_{kN_0+j}}{T_0} = k + \frac{j}{N_0} + \frac{\Delta t_{kN_0+j}}{T_0}. \tag{26}$$

Поэтому x_{kN_0+j} можно представить в виде

$$x_{kN_0+j} = a_{k+1}\left(\frac{j}{N_0} + \frac{\Delta t_{kN_0+j}}{T_0}\right) + a_k\left(1 - \frac{j}{N_0} - \frac{\Delta t_{kN_0+j}}{T_0}\right) +$$

$$+ n_{kN_0+j} = a_{k+1}\frac{j}{N_0} + a_k\left(1 - \frac{j}{N_0}\right) + \frac{\Delta t_{kN_0+j}}{T_0}\left(a_{k+1} - a_k\right) + n_{kN_0+j}. \tag{27}$$

Подставляя это разложение в выражение (24) для оценки $\hat{a}_k$ параметра a_k, получим

$$\hat{a}_k = a_k + \frac{1}{N_0}\sum_{j=1}^{N_0}\left(4 - 6\frac{j}{N_0}\right)\left[n_{kN_0+j} + \frac{\Delta t_{kN_0+j}}{T_0}(a_{k+1} - a_k)\right], \tag{28}$$

отсюда следует, что

$$D\{\hat{a}_k\} = \sigma^2\frac{1}{N_0^2}\sum_{j=1}^{N_0}\left(4 - 6\frac{j}{N_0}\right)^2 + (a_{k+1} - a_k)^2 \times$$

$$\times\frac{1}{N_0^2 T_0^2}\sum_{j=1}^{N_0}\sum_{l=1}^{N_0}\left(4 - 6\frac{j}{N_0}\right)\left(4 - 6\frac{l}{N_0}\right)\mathrm{cov}(\Delta t_{kN_0+j}, \Delta t_{kN_0+l}). \tag{29}$$

При $N_0 >> 1$ суммы приближенно можно заменить интегралами, тогда

$$\frac{1}{N_0}\sum_{j=1}^{N}\left(4 - 6\frac{j}{N_0}\right)^2 \approx \int_0^1 (4 - 6x)^2 dx = 4. \tag{30}$$

Для $\mathrm{cov}(\Delta t_{kN_0+j}, \Delta t_{kN_0+l})$ нетрудно показать, что

$$\mathrm{cov}(\Delta t_{kN_0+j}, \Delta t_{kN_0+l}) \approx \frac{T^2}{N^3}[N \cdot \min(kN_0 + j, kN_0 + l) - (kN_0 + j)(kN_0 + l)],$$

для второй суммы в (29) получим выражение

$$\frac{1}{N_0^2 T_0^2}\cdot\frac{T^2}{N^3}\sum_{j=1}^{N_0}\sum_{l=1}^{N_0}\left(4 - 6\frac{j}{N_0}\right)\left(4 - 6\frac{l}{N_0}\right)\times$$

$$\times[N \cdot \min(kN_0 + j, kN_0 + l) - (kN_0 + j)(kN_0 + l)].$$

Снова полагая $T = nT_0$, $N = nN_0$, приведем это выражение к виду

$$\frac{1}{N_0} \cdot \frac{1}{N_0^2} \sum_{j=1}^{N_0} \sum_{l=1}^{N_0} \left(4-6\frac{j}{N_0}\right)\left(4-6\frac{l}{N_0}\right) \times$$

$$\times \left[\min\left(k+\frac{j}{N_0}, k+\frac{l}{N_0}\right) - \frac{1}{n}\left(k+\frac{j}{N_0}\right)\left(k+\frac{l}{N_0}\right)\right].$$

Считая $N_0 >> 1$ и заменяя суммы интегралами, получим, что

$$\frac{1}{N_0^2} \sum_{j=1}^{N_0} \sum_{l=1}^{N_0} \left(4-6\frac{j}{N_0}\right)\left(4-6\frac{l}{N_0}\right) \cdot \left[\min\left(k+\frac{j}{N_0}, k+\frac{l}{N_0}\right) - \frac{1}{n}\left(k+\frac{j}{N_0}\right)\left(k+\frac{l}{N_0}\right)\right] \approx$$

$$\approx \int_0^1 \int_0^1 (4-6x)(4-6y)\left[\min(k+x, k+y) - \frac{1}{n}(k+x, k+y) \right] dxdy.$$

Наконец, вычисляя полученный интеграл, получим, что он равен

$$\frac{2}{15} + k - \frac{k^2}{n} = \frac{2}{15} + k\left(1-\frac{k}{n}\right),$$

так что окончательное выражение для дисперсии $D\{\hat{a}_k\}$ оценки $\hat{a}_k$ принимает вид

$$D\{\hat{a}_k\} = 4\frac{\sigma^2}{N_0} + \frac{(a_{k+1}-a_k)^2}{N_0}\left(\frac{2}{15} + k\left(1-\frac{k}{n}\right)\right). \tag{31}$$

В этом выражении неизвестны σ^2 и $(a_{k+1}-a_k)^2$. Поэтому построим оценку $\hat{D}\{\hat{a}_k\}$ этой дисперсии. Удобнее строить оценку величины $N_0 D\{\hat{a}_k\}$, так как при $N_0 \to \infty$ эта величина стремится к константе.

Для этого рассмотрим статистику вида

$$S = \frac{1}{N_0} \sum_{j=1}^{N_0} \left(4-6\frac{j}{N_0}\right)^2 x_{kN_0+j}^2 . \tag{32}$$

Подставляя сюда выражение (27), получим

$$S = \frac{1}{N_0} \sum_{j=1}^{N_0} \left(4-6\frac{j}{N_0}\right)^2 \left[a_{k+1}\frac{j}{N_0} + a_k\left(1-\frac{j}{N_0}\right) + \frac{\Delta t_{kN_0+j}}{T_0}\left(a_{k+1}-a_k\right) + n_{kN_0+j} \right]^2 .$$

С учетом того, что $M\{n_{kN_0+j}\} = 0$, $M\{\Delta t_{kN_0+j}\} = 0$ и $n_{kN_0+j}, \Delta t_{kN_0+j}$ независимы, получаем

$$M\{S\} = \sigma^2 \frac{1}{N_0} \sum_{j=1}^{N_0} \left(4-6\frac{j}{N_0}\right)^2 + \frac{1}{N_0} \sum_{j=1}^{N_0} \left(4-6\frac{j}{N_0}\right)^2 \left(a_{k+1}\frac{j}{N_0} + a_k\left(1-\frac{j}{N_0}\right)\right)^2 +$$

$$+ \frac{(a_{k+1}-a_k)^2}{N_0 T_0^2} \sum_{j=1}^{N_0} \left(4-6\frac{j}{N_0}\right)^2 \cdot M\{\Delta t_{kN_0+j}^2\}. \tag{33}$$

Если заменить суммы соответствующими интегралами, то в (33) первое слагаемое равно $4\sigma^2$**,** второе слагаемое — $\frac{1}{15}(38a_k^2+14a_k a_{k+1}+8a_{k+1}^2)$. Что касается последнего слагаемого, то учтем, что

$$M\{\Delta t^2_{kN_0+j}\}=D\{\Delta t_{kN_0+j}\}\approx T^2\,\frac{(kN_0+j)(N-kN_0-j)}{N^3},$$

так что коэффициент при $(a_{k+1}-a_k)^2$ равен

$$\frac{1}{nN_0^2}\sum_{j=1}^{N_0}\left(4-6\frac{j}{N_0}\right)^2\left(k+\frac{j}{N_0}\right)\left(n-k-\frac{j}{N_0}\right). \tag{34}$$

Однако выражение

$$\frac{1}{nN_0}\sum_{j=1}^{N_0}\left(4-6\frac{j}{N_0}\right)^2\left(k+\frac{j}{N_0}\right)\left(n-k-\frac{j}{N_0}\right)$$

при $N_0\to\infty$ стремится к конечному пределу, равному

$$\frac{1}{n}\int_0^1(4-6x)^2(k+x)(n-k-x)dx=1-2\frac{k}{n}+4k-4\frac{k^2}{n}-\frac{8}{15n},$$

и поэтому все выражение (34) при $N_0\to\infty$ стремится к нулю.

Таким образом, при $N_0\to\infty$ имеем асимптотическое равенство

$$M\{S\}\cong 4\sigma^2+\frac{1}{15}(38a_k^2+14a_k a_{k+1}+8a_{k+1}^2), \tag{35}$$

и поэтому с учетом сходимости $\hat{a}_k$ к a_k, по крайней мере в среднеквадратичном смысле, оценкой величины $N_0D\{\hat{a}_k\}$ является величина

$$\frac{1}{N_0}\sum_{j=1}^{N_0}\left(4-6\frac{j}{N_0}\right)x_i^2-\frac{1}{15}(38\hat{a}_k^2+14\hat{a}_k\hat{a}_{k+1}+8\hat{a}_{k+1}^2)+(\hat{a}_{k+1}-\hat{a}_k)^2\times$$

$$\times\left[\frac{2}{15}+k\left(1-\frac{k}{n}\right)\right]=\frac{1}{N_0}\sum_{j=1}^{N_0}\left(4-6\frac{j}{N_0}\right)x_j^2+(\hat{a}_{k+1}-\hat{a}_k)^2k\left(1-\frac{k}{n}\right)-$$

$$-\frac{1}{15}(36\hat{a}_k^2+18\hat{a}_k\hat{a}_{k+1}+10\hat{a}_{k+1}^2), \tag{36}$$

что и дает асимптотически несмещенную оценку $\hat{D}\{\hat{a}_k\}$ дисперсии $D\{\hat{a}_k\}$:

$$\hat{D}\{\hat{a}_k\}=\frac{1}{N_0^2}\sum_{j=1}^{N_0}\left(4-6\frac{j}{N_0}\right)^2x_i^2+$$

$$+\frac{1}{N_0}\left[(\hat{a}_{k+1}-\hat{a}_k)^2k\left(1-\frac{k}{n}\right)-\frac{1}{15}(36\hat{a}_k^2+18\hat{a}_k\hat{a}_{k+1}+10\hat{a}_{k+1}^2)\right]. \tag{37}$$

Знание $\hat{D}\{\hat{a}_k\}$ позволяет строить доверительные интервалы для $\hat{a}_k$ и самого тренда.

Оценка сплайном первого порядка тренда произвольного вида

Выше изучен случай, когда истинный тренд — сплайн первого порядка и его оценка — также сплайн. Теперь пусть тренд является произвольной функцией от времени, т.е.

$$x_i = x(t_i) = f(t_i) + n_i. \tag{38}$$

В этом случае основная идея заключается в том, чтобы для произвольной функции тренда $f(t)$ все-таки сохранить те оценки $\hat{a}_k$, которые были получены для случая, когда $f(t)$ — сплайн первого порядка, и выделить тренд в форме сплайна.

Однако в данном случае возникает неустранимая ошибка (которая не стремится к нулю при $N_0 \to \infty$), связанная с тем, что $f(t)$ не является сплайном. Ниже вычисляется эта неустранимая ошибка.

Данная ошибка проявляется при $\lambda \to \infty$, т.е. когда все измерения выполняются через бесконечно малые интервалы времени, и поэтому в выражениях для оценок параметров все суммы превращаются в интегралы.

Пусть $M\{x(t)\} = f(t)$. Тогда параметры сплайна a_k определяются следующим образом:

$$a_k = \frac{1}{T_0} \int\limits_{kT_0}^{(k+1)T_0} f(t)\left(4 - 6\frac{t - kT_0}{T_0}\right)dt = \int\limits_0^1 f(xT_0 + kT_0)(4 - 6x)dx \tag{39}$$

для k = 0, 1, 2,... .

На участке $[kT_0, (k+1)T_0]$ $f(t)$ аппроксимируется сплайном

$$a_{k+1}\left(\frac{t}{T_0} - k\right) + a_k\left(k + 1 - \frac{t}{T_0}\right), \tag{40}$$

и поэтому интеграл от квадрата разности $f(t)$ и аппроксимирующего его сплайна равен

$$\int\limits_{kT_0}^{(k+1)T_0} \left[f(t) - a_{k+1}\left(\frac{t}{T_0} - k\right) - a_k\left(k + 1 - \frac{t}{T_0}\right)\right]^2 dt =$$

$$= T_0 \int\limits_0^1 [f(yT_0 + kT_0) - a_{k+1}y - a_k(1 - y)]^2 dy, \tag{41}$$

тогда среднеквадратичная погрешность ε аппроксимации тренда $f(t)$ таким сплайном равна

$$\varepsilon^2 = \frac{1}{nT_0} \sum_{k=0}^{n-1} \int\limits_{kT_0}^{(k+1)T_0} \left[f(t) - a_{k+1}\left(\frac{t}{T_0} - k\right) - a_k\left(k + 1 - \frac{t}{T_0}\right)\right]^2 dt =$$

$$= \frac{1}{n} \sum_{k=0}^{n-1} \int\limits_0^1 [f(yT_0 + kT_0) - a_{k+1}y - a_k(1 - y)]^2 dy, \tag{42}$$

где коэффициенты a_k определяются формулой (39).

На одном примере покажем, чему равна среднеквадратичная погрешность ε. Пусть $f(t) = e^{-\alpha t}$. Тогда (все вычисления проведены с помощью программы Mathcad Plus 6.0 Pro)

 ISSN 0572-2691

$$a_k = \int_0^1 (4-6x)e^{-a(x+k)}dx = a_0 \cdot e^{-ak},$$

где $a = \alpha T_0$ и

$$a_0 = \int_0^1 (4-6x)e^{-ax}dx = \frac{2}{a}[2a-3+(a+3)\cdot e^{-a}]\,. \qquad (43)$$

Поэтому k-е слагаемое в (42) имеет вид

$$\int_0^1 [e^{-ay-ak} - a_0 e^{-ak} e^{-a} y - a_0 e^{-ak}(1-y)]^2 dy = e^{-2ak} \cdot \varphi(a),$$

$$\varphi(a) = \int_0^1 [e^{-ay} - a_0 e^{-a} y - a_0(1-y)]^2 dy,$$

и тогда

$$\varepsilon^2 = \frac{1}{n}\sum_{k=0}^{n-1} \varphi(a)e^{-2ak} = \varphi(a)\frac{1-e^{-2an}}{n(1-e^{-2a})}\,. \qquad (44)$$

Основным здесь является сомножитель $\varphi(a)$, так как второй сомножитель всегда меньше единицы. Что касается $\varphi(a)$, то его явный вид следующий:

$$\varphi(a) = \frac{1}{6a^4}[(24a-16a^2+3a^3)+e^{-a}(144-192a+40a^2)+e^{-2a}(-216-72a+$$

$$+120a^2-3a^3)+e^{-3a}(192a+64a^2)+e^{-4a}(72+48a+8a^2)]. \qquad (45)$$

Разложение $\varphi(a)$ в ряд Тейлора по a имеет вид

$$\varphi(a) = \frac{a^4}{720}\left(1 - a + \frac{169}{105}a^2 - \frac{82}{35}a^3 + \frac{1931}{8340}a^4 + ...\right), \qquad (46)$$

так что при малых a приближенно

$$\varepsilon \approx \sqrt{\varphi(a)} \approx \frac{a^2}{12\sqrt{5}} \approx 0{,}037a^2, \qquad (47)$$

т.е. при малых a среднеквадратичная погрешность аппроксимации $e^{-\alpha t}$ сплайном первого порядка мала.

Для проверки работоспособности исследованных алгоритмов проведено имитационное моделирование. Результаты моделирования, как и общее заключение, приведены в ч. 2 данной работы.

Ф.Ф. Ідрісов

РЕКУРЕНТНІ МОДЕЛІ РАНДОМІЗОВАНИХ ПРОЦЕСІВ В ЗАДАЧАХ ЗОВНІШНЬОЇ РОЗВІДКИ. Частина 1. СПРОЩЕНІ МОДЕЛІ НИЖЧОГО ПОРЯДКУ

Для держави зовнішня розвідка в усі часи була сферою особливої уваги. Роль зовнішньої розвідки значно зросла в наш час, коли з’явилися нові форми міждержавних конфліктів (мережеві і гібридні війни, масштабні терористичні акти, системне втручання у внутрішні справи іншої держави і т.д.) і ускладнився простір їх проведення (до традиційних просторів, морського, сухопутного та повітряного, додалася нова сфера — кіберпростір). У зв'язку з цим ускладнилися умови отримання розвідданих і вимоги до їх обробки. Найбільший інтерес становить контекст проведення розвідувальних операцій, тобто дослідження потенційних можливостей моделей обробки даних про фінансові транзакції, про пересування людських, матеріальних, інформаційних та інших ресурсів з позицій теорії випадкових процесів. Однак класична теорія випадкових процесів розроблялася для задач, коли спостереження за об'єктом проводилися через рівні інтервали часу (наприклад, врожайність, народжуваність і т.д.). Еквідистантність малася на увазі за замовчуванням. Але для задач розвідки таке уявлення об'єкта, що спостерігається, неприйнятно, оскільки інформація про нього, будучи випадковою, надходить у випадкові моменти часу. Назвемо такі процеси рандомізованими, підкреслюючи цим самим рандомізованість моментів надходження розвідданих. Але є ще одна важлива обставина — це висока динаміка старіння даних і обмеженість їх обсягу. Природно, в цьому випадку необхідні рекурентні моделі, що враховують як недостатність інформації, так і її «старіння» в умовах рандомізованих спостережень. Всі ці моменти враховано у запропонованих моделях, заснованих на модифікованих сплайнах. Розглянуто моделі, досліджені для випадків, коли моменти появи даних відомі точно або з тих чи інших причин невідомі. Побудовано відповідні алгоритми виділення трендів спостережуваних даних, а також проаналізовано статистичні властивості їх параметрів. Наведено результати імітаційного моделювання. У цій частині роботи досліджено досить прості моделі порівняно невисокого порядку, тоді як у ч. 2 завдання вирішується шляхом побудови моделі більш високого порядку.

Ключові слова: розвіддані, рандомізований процес, тренд, сплайн, дисперсія, коваріація оцінок, рекурентні оцінки, імітаційне моделювання.

F.F. Idrisov

RECURSIVE MODELS OF RANDOMIZED PROCESSES IN FOREIGN INTELLIGENCE TASKS. Part I. SIMPLIFIED MODELS OF LOWER ORDER

For the state, foreign intelligence has always been a sphere of special attention. The role of foreign intelligence has grown significantly in our time, when new forms of interstate conflicts appeared (network and hybrid wars, large-scale terrorist acts, systemic interference in the internal affairs of another state, etc.) and the space for their conduct became more complicated (for traditional maritime, land and airspace added a new sphere — cyberspace). In this regard, the conditions for obtaining intelligence and requirements for their processing have become more complicated. Of most interest is the context of intelligence operations, i.e. study of the potential possibilities of data processing models for financial transactions, the movement of human material, information and other resources from the standpoint of the theory of random processes. However, the classical theory of random processes was developed for tasks when observations of an object were made at regular time intervals (for example, yield, fertility, etc.). Equidistance was implied by default. But for reconnaissance tasks, such representation of the observed object is unacceptable, since information about it, being random, arrives at random times. We call such processes randomized, emphasizing the very randomization of the moments of receipt of intelligence. But there is another important circumstance — this is the high dynamics of data obsolescence and the limitedness of their volume. Naturally, in this case, recurrence models are necessary that take into account both the lack of information and its «obsolescence» in the context of randomized observations. All these points are taken into account in the proposed models based on modified splines. The models presented in this work are investigated for cases when the moments of the appearance of the data are known exactly, or we cannot know these moments for one reason or another. Corresponding algorithms for distinguishing trends of the observed data are constructed, as well as the analyzed statistical properties of their parameters. The results of simulation are presented. In the first part of this work, fairly simple models of a relatively low order are investigated, while in the second part of the floor the problem is solved by constructing a model of a higher order.

Keywords: Intelligence, randomized process, trend, spline, variance, covariance of estimates, recurrence estimates, simulation.

1. Шаваев А.Г., Лекарев С.В. Разведка и контрразведка. Фрагменты мирового опыта и теории. М. : Издательская группа «БДЦ–Пресс», 2003. 544 с.
2. Соколов Г.Е. Шпионаж и политика. Тайная христоматия. М. : Изд-во «Алгоритм», 2017. 430 с.
3. Бобылов Ю.А. Специальные операции и технологическая модернизация России. Berlin : Lambert Academic Publishing, 2016. 684 с.
4. Идрисов Ф.Ф. Приближенные алгоритмы выделения трендов в задачах финансовой разведки. Часть 1. Моменты появления элементов финансового потока известны точно. *Международный научно-технический журнал «Проблемы управления и информатики»*. 2017. № 6. С. 7–18.
5. Идрисов Ф.Ф. Приближенные алгоритмы выделения трендов в задачах финансовой разведки. Часть 2. Моменты появления элементов финансового потока неизвестны. *Международный научно-технический журнал «Проблемы управления и информатики»*. 2018. № 1. С. 146–155.
6. Идрисов Ф.Ф. Приближенные модели финансовой разведки при неточно заданных моментах времени осуществления финансовых транзакций *Международный научно-технический журнал «Проблемы управления и информатики»*. 2018. № 6. С. 119–131.
7. Завьялов Ю.С., Леус В.А., Скоропостелов В.А. Сплайны в инженерной геометрии. М. : Машиностроение, 1985. 224 с.
8. Лифшиц К.И. Сглаживание экспериментальных данных сплайнами. Томск : Изд-во Том. ун-та, 1991. 180 с.

Получено 29.03.2019