УДК 519

А. В. Овсянников, доцент

ИНТЕРВАЛЬНОЕ ПРОГНОЗИРОВАНИЕ СТОХАСТИЧЕСКИХ ПРОЦЕССОВ

In article two approaches to the decision of a problem of interval prediction stochastic processes are considered. The first approach assumes designing non-stationary density of distribution of probability of casual process and further the sanction of the integrated equation concerning a confidential limit of the prognosis. In the second approach modification of a classical technique interval estimation of random parameters is used. Thus, feature of the received interval estimation of the prognosis is that it is applied to any laws of distribution, and also there is probable an obtaining "coarsened", robust interval estimations of the prognosis.

Введение. Задача прогнозирования фазовых траекторий является одним из актуальных направлений современных исследований в области анализа поведения сложных гибридных технологических и технических систем и их управления.

Известны и достаточно хорошо изучены методы и алгоритмы точечного оценивания стационарных случайных процессов с известным законом распределения – плотностью распределения вероятности (ПРВ) [1, 2].

Однако ряд задач требует иного подхода, основанного на интервальном оценивании и прогнозировании. К таким задачам, например, можно отнести технические задачи: допускового контроля и технической диагностики, интервального управления объектами с переменной структурой, прогнозирования нестационарных процессов; в финансово-экономической и социальной сферах – прогнозирование микро- и макроэкономических показателей развития и т. д.

С формальной математической точки зрения, если стохастический процесс стационарен в узком и широком смысле, то, построив, например, уравнение тренда, можно получить прогнозные значения. Прогноз, полученный подстановкой в уравнение тренда ожидаемого значения параметра, является точечным. Вероятность реализации такого прогноза мала. Целесообразно определить доверительный интервал прогноза. Для индивидуальных значений показателя интервал должен учитывать ошибки в положении линии тренда и отклонения индивидуальных значений от этой линии. В литературе приводятся формулы для расчета доверительных интервалов прогноза для случая линейной модели регрессии. В то же время доверительный интервал остается постоянным.

Однако фактически неопределенность прогноза (энтропия) с ростом количества шагов прогноза (интервала, глубины прогноза) также должна увеличиваться.

Особо можно выделить задачу прогнозирования нестационарных случайных процессов в связи с возросшим в последнее время интересом к исследованию систем переменной структуры, гибридных систем. В этом случае постановка задачи усложняется за счет наличия дополнительного дискретного параметра, характеризующего состояние системы или процесса.

Основная часть. Будем рассматривать задачу прогнозирования скалярного параметра на k шагов в условиях регулярного статистического эксперимента. Пусть имеется однородная независимая выборка $R=\{r_1, ..., r_n\}$ наблюдаемых данных, принадлежащая распределению P_ϑ с плотностью $P_\vartheta = P(r \mid \vartheta)$, где $\vartheta \in \Theta$ и Θ – интервал на действительной оси. При фиксированном объеме выборки R выборочным пространством является n – мерное евклидово пространство $R = R^n$, на котором задана плотность P_ϑ, $r \in R$.

Графическая интерпретация задачи интервального прогнозирования представлена на рис. 1.

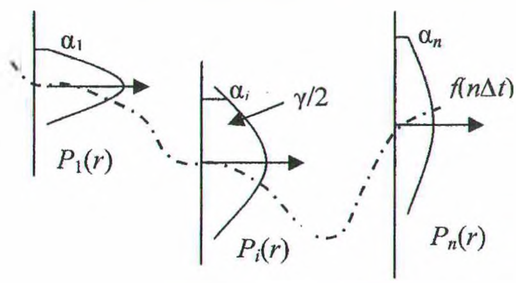

Рис. 1. Нестационарная ПРВ

Если задаться вероятностью γ (коэффициентом доверия), то, например, верхний доверительный предел α определится выражением $P(0 < r \leq \alpha_i) = \gamma / 2$ и, следовательно, при известной нестационарной ПРВ $\int_0^{\alpha_i} P_i(r)dr = \gamma / 2$.

Рассмотрим два подхода к решению задачи прогнозирования.

I подход. Исчерпывающей характеристикой стохастического процесса является его нестационарная ПРВ. Поскольку такая ПРВ неизвестна, можно попытаться ее сконструировать. Одним из возможных подходов в этом случае является использование ядерных оценок плотности типа Парзена – Розенблатта [3]. Непараметрическая оценка нестационарной плотности Парзена – Розенблатта имеет вид следующей функции:

$$P_n(r) = (n\sigma^d)^{-1} \sum_{i=1}^{n} K\left[\frac{r - \vartheta_i}{\sigma}\right], \qquad (1)$$

где $\sigma = \sigma_n$ – последовательность положительных чисел ($\lim\limits_{n\to\infty} \sigma = 0$, $\lim\limits_{n\to\infty} n\sigma^d = \infty$); $K\left[\frac{r-\vartheta_i}{\sigma}\right]$ – борелевская функция (ядро) в пространстве $L_1 : J_n = \int |P_n - P| dx$, удовлетворяющая условиям $K \geq 0$, $\int K dx = 1$.

При наличии тренда $f(i\Delta t)$ (рис. 1) функцию (1) можно записать так

$$P_n(r) = \left(n\sigma^d\right)^{-1} \sum_{i=1}^{n} K\left[\frac{r - f(\vartheta_i)}{\sigma}\right]. \qquad (2)$$

С вычислительной точки зрения при достаточно большой априорной информации ($n >> 1$) использование (1) или (2) неудобно, поэтому можно предложить модификацию дискретной оценки Парзена – Розенблатта, эквивалентную непрерывному функционалу

$$P_n(r) = \left(n\sigma^d\right)^{-1} \exp\left\{-\sum_{i=1}^{n} B\left[\frac{r - f(\vartheta_i)}{\sigma}\right]\right\}, \qquad (3)$$

где $B(r)$ – семейство однопараметрических функций.

Так, например, для функционала гауссовского случайного процесса из (3) получаем

$$P_n(r) = \left(\sqrt{2\pi\sigma^2}\right)^{-n} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left[\frac{r - f(\vartheta_i)}{\sigma}\right]^2\right\}. \qquad (4)$$

С учетом предельного перехода $n\Delta t \to i$ из дискретных оценок плотности можно получить оценку нестационарной ПРВ стохастического процесса $P_n(r) \to P(r, n\Delta t) \to P(r, t)$ и, следовательно, $\int\limits_0^{\alpha_n} P_n(r)dr = \frac{\gamma}{2} \to \int\limits_0^{\alpha_t} P(r,t)dr = \frac{\gamma}{2}$.

Как видно из выражений (2)–(4), при осуществлении прогноза необходимо знание функции тренда – регрессии $f(i\Delta t)$. Получить оценку, восстановить функцию тренда можно одним из известных способов, например, регрессионного, факторного, спектрального анализа и т. п.

Для диффузионного процесса с односторонней спектральной плотностью N из (3), (4) следует, что

$$\alpha_k = f^*(k\Delta t) + \sqrt{k\Delta t N}\ \mathrm{Erf}^{-1}\left\{\gamma - \mathrm{Erf}\left[\frac{f^*(k\Delta t)}{\sqrt{k\Delta t N}}\right]\right\}, \qquad (5)$$

где $f^*(k\Delta t)$ – восстановленная по n предыдущим значениям функция тренда; $\mathrm{Erf}(z)$ – функция ошибки (интеграл вероятности).

Для функционала, описывающего случайный процесс с мощными импульсными выбросами, из (3) получаем

$$P_n(r) = \left(2n\Delta t N\right)^{-1} \exp\left\{-\sum_{i=1}^{n}\frac{|r - f(\vartheta_i)|}{n\Delta t N}\right\}. \qquad (6)$$

В этом случае рекуррентная оценка доверительного предела определиться зависимостью

$$\alpha_k = f^*(k\Delta t) - k\Delta t N \ln\left[e^{\frac{f^*(k\Delta t)}{k\Delta t N}} - \gamma\right]. \qquad (7)$$

Проиллюстрируем полученные выражения для доверительного верхнего α_k и нижнего β_k пределов по формулам (5) и (7). Заметим, что нижний доверительный предел β_k находится из тех же соображений, что и верхний. Зададимся доверительной вероятностью $\gamma = 0.9$ и пусть восстановленная функция тренда имеет следующий вид: $f^*(i\Delta t) = a + b\ i\Delta t$, тогда результаты расчетов можно отобразить в виде зависимостей, представленных на рис. 2.

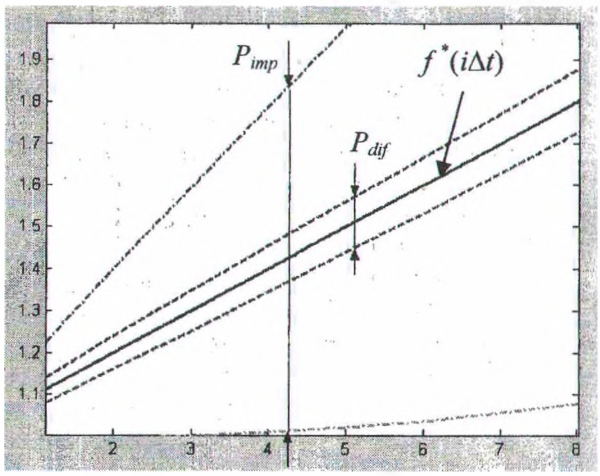

Рис. 2. Доверительные интервалы прогноза

Величина размаха доверительного интервала нестационарного процесса с односторонней спектральной плотностью N на рис. 2 обозначена P_{dif}, а случайного нестационарного процесса, определенная по формуле (6), – P_{imp}. Видим, что даже при относительно небольшой глубине прогноза размах P_{imp} значительно превосходит интервальную оценку прогноза P_{dif}.

II подход. Рассмотрим решение задачи интервального прогнозирования в рамках классического подхода. Предполагается известной некоторая априорная выборка данных наблюдений $R = \{r_1, ..., r_n\} = r_1^n$, $r_i = f(\vartheta_i) + \xi_i$, где ξ_i – случайная величина с ПРВ $P(\xi)$. Выбрав некоторое положительное число ε, уравнение для доверительной вероятности можем записать в виде

$$P(f_n^0(1-\varepsilon) \leq f_{r+k}^* \leq f_n^0(1+\varepsilon)) = \int\limits_{f_n^0(1-\varepsilon)}^{f_n^0(1+\varepsilon)} P_\xi(r_1^n | f_{n+k}^0) df_{n+k}^0 = \gamma,$$

где $f^*_{n+k} = f^*((n+k)\Delta t)$ – прогноз на k шагов вперед по n априорным данным; $f^0_n = f^0(n\Delta t)$ – оценка функции тренда по n априорным данным; $P_\xi(r_1^n \mid f^0_{n+k})$ – функция правдоподобия.

Сформулируем следующее утверждение. В условиях регулярного статистического эксперимента и

$$\Delta f = \frac{f^*_{n+k} - f^0_n}{\sqrt{n+k}} \to 0 \qquad (8)$$

справедливо следующее неравенство:

$$\int\limits_{f^0_m(1-\varepsilon)}^{f^0_m(1+\varepsilon)} P_\xi(y_n)dy_n \geq \gamma - S_{\Delta f}\sqrt{I_\Phi} = \gamma_1, \qquad (9)$$

где

$$S^2_{\Delta f} = \sum \Delta f^2 P(y), \qquad (10)$$

$$I_\Phi = \sum \left[\frac{\partial \ln P(y)}{\partial y}\right]^2 P(y). \qquad (11)$$

Величина $S^2_{\Delta f}$ представляет собой выборочную дисперсию (8), а I_Φ – информацию Фишера [1] относительно ПРВ стационарного распределения $P_\xi(r_k \mid f^0_{n+k}) = P_\xi(r_k - f^0_{n+k})$. Утверждение (9) может быть получено путем разложения функции правдоподобия $P_\xi(r_k - f^0_{n+k})$ по степеням Δf. Исходя из неравенства Коши – Буняковского – Шварца как частного случая общего неравенства Гельдера [4]

$$\left[\int\limits_\beta^\alpha \frac{\partial P(y)}{\partial y}\Delta f dy\right]^2 \leq$$

$$\leq \int\limits_\beta^\alpha \left[\frac{\partial \ln P(y)}{\partial y}\right]^2 P(y)dy \int\limits_\beta^\alpha (\Delta f)^2 P(y)dy,$$

с учетом перехода к дискретным отсчетам получим

$$\gamma - \int\limits_\beta^\alpha P_\xi(y_n)dy_n \leq S_{\Delta f}\sqrt{I_\Phi},$$

из чего непосредственно следует (9).

В качестве нормированной ошибки прогноза на k шагов вперед выберем величину

$$t_{\gamma_1} = \Delta f\sqrt{I_\Phi} = \frac{f^*_{n+k} - f^0_n}{\sqrt{n+k}}\sqrt{I_\Phi},$$

которая, как можно показать, приближенно имеет распределение Стьюдента с $(n + k - 1)$ степенями свободы.

Учитывая симметрию распределения Стьюдента [5], получим

$$P(|t| \leq t_{\gamma_1}) = 2\int\limits_0^{t_{\gamma_1}} S_{n+k-1}(y)dy = \gamma_1,$$

где $S_{n+k-1}(y)$ – распределение Стьюдента.

Доверительный интервал прогноза f^*_{n+k} теперь можно представить в виде неравенств

$$f^0_n - t_{\gamma_1}\sqrt{k\left(1+\frac{k}{n}\right)\frac{1}{I_\Phi}} \leq f^*_{n+k} \leq$$

$$\leq f^0_n + t_{\gamma_1}\sqrt{k\left(1+\frac{k}{n}\right)\frac{1}{I_\Phi}}.$$

Заключение. Особенностью полученной с помощью второго подхода интервальной оценки прогноза является то, что она применима к произвольным законам распределения $P(\xi)$. Кроме того, существенным является и тот факт, что на ее основе могут быть получены «огрубленные», робастные интервальные оценки прогноза. При этом используется понятие «наихудшей» ПРВ в классе минимизирующей фишеровскую информацию.

Литература

1. Ибрагимов, И. А. Асимптотическая теория оценивания / И. А. Ибрагимов, Р. З. Хасьминский. – М.: Наука, 1979. – 528 с.
2. Гранберг, А. Г. Статистическое моделирование и прогнозирование / А. Г. Гранберг. – М.: Финансы и статистика, 1990. – 446 с.
3. Деврой, Л. Непараметрическое оценивание плотности. L_1 – подход / Л. Деврой, Л. Дьёрфи; пер. с англ. – М.: Мир, 1988. – 408 с.
4. Маршалл, А. Неравенства: Теория мажоризации и ее приложения / А. Маршалл, И. Олкин; пер. с англ. – М.: Мир, 1983. – 576 с.
5. Корн, Г. Справочник по математике (для научных работников и инженеров) / Г. Корн, Т. Корн; пер. с англ. – М.: Наука, 1973. – 831 с.