

НЕПАРАМЕТРИЧЕСКИЕ ДАТЧИКИ ДЛЯ СЛУЧАЙНЫХ СТАЦИОНАРНЫХ ПРОЦЕССОВ

Рассмотрены алгоритмы непараметрических датчиков случайных стационарных процессов. На основе непараметрических алгоритмов прогноза находится интервал зависимости случайного процесса.

Ключевые слова: датчик, непараметрический, прогноз, процесс, моделирование.

Методы статистического имитационного моделирования нашли широкое распространение как в научных исследованиях, так и в технических приложениях. Именно с их помощью удается получить выводы при исследовании сложных систем, зависящих от большого числа случайных факторов (переменных, показателей, элементов). Особая роль при таких исследованиях отводится датчикам случайных величин, которые должны адекватно отражать суть физических явлений, характерных для случайных процессов, участвующих в исследовании. Определение функции распределения случайного процесса фактически определяет задание математической модели физического явления и тем самым определяет адекватность получаемых решений. Обычно, исходя из физических или интуитивных соображений, выбирают некоторое стандартное распределение $G(x(t), \bar{\theta})$, а затем по обучающей выборке x_1, \dots, x_N оценивают неизвестный параметр $\bar{\theta}$ и проверяют по некоторым критериям адекватность параметрической модели, например, линейной авторегрессии (АР), скользящего среднего (МА), АРМА, ARCH, GARCH и т. д. – параметрический подход. Понятно, что при этом задается вид параметрической модели и распределение случайных помех. Построить датчик случайного процесса в данном случае не представляет трудностей. Но при параметрическом подходе имеется ряд трудноформализуемых моментов (например, выбор вида параметрической модели процесса и распределения $G(x(t), \bar{\theta})$, критерия адекватности, интервала зависимости и т. д.). Но необходимо понимать, что любая ошибка в задании априорной информации о виде модели и распределения $G(x(t), \bar{\theta})$ приводит к неизвестному смещению при окончательных выводах. Именно этот момент и вызывает значительный интерес к непараметрическому подходу, когда вид распределения $F(x(t))$ случайного процесса $X(t)$ априори считается неизвестным. Действительно, выборка x_1, \dots, x_N в виде эмпирических функций распределения $F_N(x(t))$ несет информацию о неизвестном распределении $F(x(t))$ случайного процесса $X(t)$. Следовательно, возможно построение датчиков на основе непараметрических оценок $F_N(x(t))$. Такие датчики назовем непараметрическими. Интерес к непараметрическим датчикам случай-

ных величин в математической статистике особенно вырос с появлением такого статистического метода, как бутстреп [1], который получил широкое распространение [2]. Как известно, в математической статистике большинство результатов исследований носит асимптотический характер. Желание узнать свойства статистических процедур при конечных объемах выборки N с помощью метода статистических испытаний и породило бутстреп-метод [1]. В последние десятилетия исследования в области создания бутстреп-процедур сосредоточены на создании непараметрических датчиков случайных процессов [3; 4]. В рамках этого направления был предложен ряд непараметрических датчиков случайных процессов: блочный бутстреп, решетчатый бутстреп, локальный бутстреп. Но то, что хорошо для исследования статистических процедур в задачах математической статистики, в большинстве случаев не совсем приемлемо в задачах статистического моделирования, где требуются большие массивы случайных неповторяющихся чисел. При блочном бутстрепе нарушается структура статистической зависимости, решетчатый бутстреп – это фактически параметрический датчик, локальный бутстреп основан на непараметрической регрессии и создает датчик, имитирующий исходную выборку [3].

В данной работе рассмотрены основные принципы построения непараметрических датчиков случайных процессов на основе непараметрических оценок условных функций распределения.

Постановка задачи и алгоритм датчика. Пусть $(x)t$ – стационарный в узком смысле случайный процесс, удовлетворяющий условию слабой зависимости [5] с функцией распределения $F(x(t))$. В реальных задачах бесконечномерная функция распределения $F(x(t))$ представляется в виде k -мерной $F_k(\vec{t}_k)$ с конечным интервалом зависимости.

Пусть x_1, \dots, x_N – выборка из $X(t)$ с неизвестной функцией распределения $F(x(t))$, $\vec{X} = (X_1, \dots, X_k)^T$ – k -мерная случайная величина с функцией распределения $F_k(\vec{t}_k)$. Обозначим через $F_j(t_j / \vec{t}_{j-1})$, $j = \overline{1, k}$ условные функции распределения случайной величины \vec{X} , $F_1(t_1) = F_j(t_j / \vec{t}_{j-1})$, $j = 1$, $\vec{t}_{j-1} = (t_1, \dots, t_{j-1})^T$.

Для построения непараметрических датчиков воспользуемся стандартным методом генерирования случайных векторов с помощью преобразования [6]

$$\begin{cases} U_1 = F_1(X_1); \\ U_2 = F_2(X_2 / X_1); \\ \dots \\ U_k = F_k(X_k / \bar{X}_{k-1}). \end{cases} \quad (1)$$

Как известно [6], случайные величины U_1, \dots, U_k будут независимы и равномерно распределены на $[0; 1]$. С помощью обратного преобразования системы (1) и получают датчики случайных векторов. Метод получения непараметрических датчиков заключается в следующем. Вместо неизвестных $F_1(X_1), \dots, F_k(X_k / \bar{X}_{k-1})$ в систему (1) подставляются их непараметрические оценки $F_{1N}(X_1), \dots, F_{kN}(X_k / \bar{X}_{k-1})$ и затем решается система типа (1). Для реализации данного метода необходимо выполнение следующих условий:

– оценки $F_{1N}(X_1), \dots, F_{kN}(X_k / \bar{X}_{k-1})$ должны быть из класса непрерывных непараметрических оценок условных функций распределения;

– преобразование типа (1), где вместо неизвестных $F_1(X_1), \dots, F_k(X_k / \bar{X}_{k-1})$ стоят их оценки $F_{1N}(X_1), \dots, F_{kN}(X_k / \bar{X}_{k-1})$, должно также асимптотически приводить к независимым и равномерным в $[0; 1]$ случайным величинам U_1, \dots, U_k ;

– так как решение системы типа (1) сводится к нахождению непараметрических оценок условных квантилей распределения $F_{jN}(X_j / \bar{X}_{j-1})$, то необходимо, чтобы эти оценки сходились по вероятности к квантилям распределения $F_j(X_j / \bar{X}_{j-1})$.

Для этого воспользуемся непараметрическими оценками условных функций распределения [5]:

$$F_{jN}(t_j / \bar{t}_{j-1}) = \frac{A_{jN}}{B_{jN}}; \quad (2)$$

$$A_{jN} = \frac{1}{Nh_N^{j-1}} \sum_{i=1}^{N-j} G\left(\frac{t_j - x_j}{h_N}\right) \prod_{\lambda=1}^{j-1} K_\lambda\left(\frac{t_\lambda - x_{i-\lambda}}{h_N}\right);$$

$$B_{jN} = \frac{1}{Nh_N^{j-1}} \sum_{i=1}^{N-j} \prod_{\lambda=1}^{j-1} K_\lambda\left(\frac{t_\lambda - x_{i-\lambda}}{h_N}\right),$$

где K_λ – ядерные функции [6]; $G(u)$ – симметричная функция распределения; $h_N \rightarrow 0$, $N \cdot h_N^k \rightarrow \infty$, $N \rightarrow \infty$.

При некоторых условиях регулярности на $F_k(\bar{x}_k)$, $K_\lambda(u)$, h_N , $G(u)$ можно доказать [5]:

$$1) F_{jN}(t_j / \bar{t}_{j-1}) \xrightarrow{\text{Pr}} F_j(t_j / \bar{t}_{j-1});$$

2) вектор $\sqrt{N}(T_{1N}(t_1), \dots, T_{kN}(t_k / \bar{t}_{k-1}))^T$ имеет асимптотически нормальное распределение с нулевым вектором средних и диагональной корреляционной матрицей

$$h_N^j \sigma_j^2 = Q^j \cdot f^{-1}(\bar{t}_{j-1}) F(t_j / \bar{t}_{j-1}) [1 - F(t_j / \bar{t}_{j-1})],$$

где $T_{jN}(t_j / \bar{t}_{j-1}) = \sqrt{h_N^j} (F_{jN}(t_j / \bar{t}_{j-1}) - F_j(t_j / \bar{t}_{j-1}))$, $\int K^2(u) du = Q < \infty$, (т. е. асимптотически $F_{1N}(t_1), \dots, F_{kN}(t_k / \bar{t}_{k-1})$ – независимы);

3) пусть ξ_j квантиль уровня p условной функции распределения $F_j(t_j / \bar{t}_{j-1})$, тогда $\xi_{jN} \xrightarrow{\text{Pr}} \xi_j$, где ξ_{jN} – оценка условного квантиля, удовлетворяющего уравнению $F_{jN}(\xi_{jN} / \bar{t}_{j-1}) = p$, $0 < p < 1$.

Данные теоретические результаты и позволяют практически создавать алгоритмы непараметрических датчиков стационарных случайных процессов. Для этого необходимо использовать алгоритм стохастической аппроксимации нахождения корня уравнения $F_{jN}(t_j / \bar{t}_{j-1}) = p$ одной переменной t_j при фиксированном векторе \bar{t}_{j-1} . Рассмотрим алгоритм получения одного многомерного случайного числа $\bar{x}^* = (x_1^*, \dots, x_k^*)^T$. Пусть u_1, \dots, u_k – равномерные на $[0; 1]$ числа. Генерируем u_1 и решаем уравнение $F_{1N}(x_1^*) = u_1 \Rightarrow x_1^* = F_{1N}^{-1}(u_1)$. В результате получаем x_1^* . Данный алгоритм давно используется для получения одномерных случайных чисел. Генерируем u_2 и решаем уравнение $F_{2N}(x_2^* / x_1^*) = u_2 \Rightarrow x_2^* = F_{2N}^{-1}(u_2 / x_1^*)$, генерируем u_k и решаем уравнение $F_{kN}(x_k^* / \bar{x}_{k-1}^*) = u_k \Rightarrow x_k^* = F_{kN}^{-1}(u_k / \bar{x}_{k-1}^*)$. На j такте необходимо решить уравнение $F_{jN}(x_j^* / \bar{x}_{j-1}^*) = u_j$ относительно x_k^* , при условии, что известно число $\bar{x}_{j-1}^* = (x_1^*, \dots, x_{j-1}^*)^T$ вида

$$\frac{\sum_{i=1}^N G\left(\frac{x_j^* - x_i}{h_N}\right) \prod_{\lambda=1}^{j-1} K_\lambda\left(\frac{x_\lambda^* - x_{i-\lambda}}{h_N}\right)}{\sum_{i=1}^N \prod_{\lambda=1}^{j-1} K_\lambda\left(\frac{x_\lambda^* - x_{i-\lambda}}{h_N}\right)} = u_j, \quad j = \overline{2, k}. \quad (3)$$

В результате, при заданном интервале зависимости k , мы получаем начальные значения x_1^*, \dots, x_k^* случайного временного ряда. Для получения следующего значения x_{k+1}^* при заданных x_1^*, \dots, x_k^* генерируем u_{k+1} и решаем уравнение

$$\frac{\sum_{i=k+1}^N G\left(\frac{x_{k+1}^* - x_i}{h_N}\right) \prod_{\lambda=1}^k K_\lambda\left(\frac{x_\lambda^* - x_{i-\lambda}}{h_N}\right)}{\sum_{i=k+1}^N \prod_{\lambda=1}^k K_\lambda\left(\frac{x_\lambda^* - x_{i-\lambda}}{h_N}\right)} = u_{k+1}. \quad (4)$$

Алгоритмы (3), (4) зависят от выбора ядерных функций $G(u)$, K_λ , параметра размытости h_N , интервала зависимости k . Выбор ядерных функций $G(u)$, K_λ и параметра размытости h_N хорошо исследован в рамках непараметрической статистики (на-

пример, в [5]) и в рамках данной статьи не рассматривается. Выбор интервала зависимости k является достаточно критичным во всех бутстреп-процедурах генерации re -выборок случайных процессов (длина блока в блочном бутстреппе, порядок АР-модели) и, несмотря на многочисленные исследования, остается достаточно субъективным. В данной работе предлагается алгоритм нахождения параметра k , основанный на непараметрических алгоритмах прогноза [7], который может быть использован в разных бутстреп-процедурах. Возьмем простой непараметрический алгоритм прогноза регрессионного типа [5; 7]. Алгоритм нахождения параметра k заключается в следующем. На основании первых значений ряда $x_1, \dots, x_j, 1 \leq j \leq k$ делаем прогноз следующего значения X_{j+1} . Образует квадрат невязок прогноза

$$Q(j) = \sum_{i=j+1}^N (X_{j+1} - x_i)^2.$$

Значение параметра k находим из условия минимума $Q(j)$ (для циклических гетероскедастических случайных процессов наблюдается минимум) или $Q(j) \leq l$, где l задается пользователем.

Моделирование. Для теста возьмем случайный стационарный процесс

$$X(t) = A \sin(\omega t + \varphi) + b \cdot \varepsilon,$$

где φ равномерная на $[0; 2\pi]$ случайная величина; b – константа уровня помех; ε – стандартная нормальная величина. Вначале определялся интервал зависимости k . На рис. 1, 2 приводятся зависимости $Q(j)$ (рис. 1. – без помех; рис. 2. – с помехой $b = 0,1$; $N = 100$), из которых следует, что интервал зависимости процесса без помех равен 2, с помехами – 5.

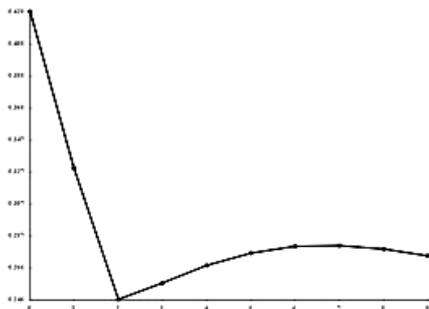


Рис. 1

Без помех: $F_{1N}(t)$ получается равномерным на $[-1; 1]$; $F_{2N}(t/t_1)$ получается двухточечным; $F_{3N}(t/t_1, t_2)$ выродается в сингулярное. Моделировались и другие процессы, но наглядность, как в тестовом примере, существенно снижается.

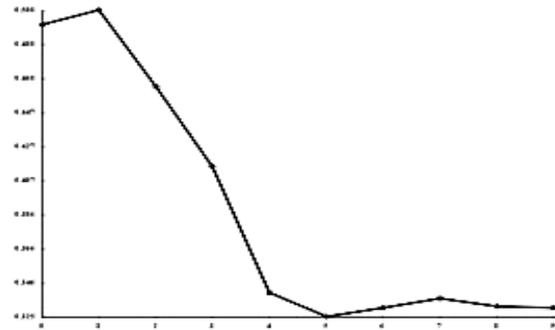


Рис. 2

Таким образом, в работе предложен метод генерации случайных процессов, основанный на классической схеме генерации (3) с использованием непараметрических оценок условных функций распределения.

Предложен метод нахождения интервала зависимости на основе непараметрических оценок прогноза. Данный метод может быть использован в качестве блока адаптивной настройки в классических бутстреп-процедурах (блочный бутстреп, локальный бутстреп).

К недостаткам предложенного метода следует отнести большую вычислительную емкость. Хотя моделирование производилось на кластерах, требуется большая вычислительная производительность. Сейчас рассматривается реализация таких алгоритмов в нейросетевом базисе.

Библиографические ссылки

1. Эфрон Б. Нетрадиционные методы многомерного статистического анализа. М. : Финансы и статистика, 1998.
2. Davison A. C., Hinkley D. V. Bootstrap Methods and Their Application. Cambridge : Cambridge University Press, 1997.
3. Бюльман П. Бутстреп-схемы для временных рядов // Квантиль. 2007. № 3. С. 37–56.
4. Hardle W., Horovitz J., Kreiss J-P. Bootstrap methods for time series // International Statist. Review. 2003. № 71. P. 435–459.
5. Васильев В. А., Добровидов А. В., Кошкин Г. М. Непараметрическое оценивание функционалов от распределений стационарных последовательностей. М. : Наука, 2004.
6. Ермаков С. М., Михайлов Г. А. Курс статистического моделирования. М. : Наука, 1976.
7. Simakhin V. A. Nonparametric Robust Prediction Algorithms // Book of Abstracts of the International Symposium on Stochastic Models in Reliability Engineering, Life Science and Operations management. Beer Sheva, 2010. P. 204–218.

NONPARAMETRIC SENSORS FOR STOCHASTIC STATIONARY PROCESS

In the article we consider an algorithm of nonparametric generator's building for stochastic stationary process. Dependency interval of stochastic process is determined with the help of nonparametric algorithms of prognosis.

Keywords: generator, nonparametric, prognosis, process, modeling.

© Маер А. В., Симахин В. А., 2010

УДК 519.224:330.46

А. А. Новоселов

ПОСТРОЕНИЕ МНОГОМЕРНЫХ ДИСКРЕТНЫХ РАСПРЕДЕЛЕНИЙ С ЗАДАННОЙ КОРРЕЛЯЦИОННОЙ СТРУКТУРОЙ

Рассмотрены методы воспроизведения многомерного дискретного распределения с заданной корреляционной структурой и маргинальными распределениями. Для воспроизведения используются смеси базовых распределений и решение некоторых оптимизационных задач.

Ключевые слова: дискретное распределение, корреляция, копула, смесь.

Пусть заданы нормальные распределения со средними значениями μ_1, \dots, μ_d и стандартными отклонениями $\sigma_1, \dots, \sigma_d$. Для произвольной корреляционной матрицы R существует единственное многомерное нормальное распределение, обладающее такими маргинальными распределениями и корреляционной матрицей. Хорошо известный алгоритм воспроизведения соответствующего случайного вектора $X = (X_1, \dots, X_d)$ основан на факторизации ковариационной матрицы.

Обозначим

$$\Lambda = \begin{pmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_d \end{pmatrix}$$

диагональную матрицу со стандартными отклонениями на диагонали, тогда $C = \Lambda R \Lambda$ является ковариационной матрицей распределения вектора X . Будучи неотрицательно определенной и симметричной, ковариационная матрица C может быть представлена в виде

$$C = A'A \tag{1}$$

с некоторой матрицей A , причем последняя определяется не одним образом. Примерами такого представления являются разложение Холецкого и ортогональное разложение.

При наличии разложения (1) вектор X воспроизводится из стандартного нормального случайного вектора Z по формуле

$$X = A'Z. \tag{2}$$

Действительно, для Z справедливо $EZZ' = I$, где I – единичная матрица соответствующего размера, поэтому $EXX' = E(A'ZZ'A) = A'(EZZ')A = A'A = C$, так

что X обладает требуемой ковариационной структурой.

В случае, когда компоненты X имеют фиксированные дискретные распределения, аналогичный метод оказывается неприменимым. Во-первых, заданным маргинальным распределениям и ковариационной матрице соответствуют, вообще говоря, многие многомерные дискретные распределения. Может оказаться и так, что подходящее многомерное распределение не существует.

Во-вторых, алгоритм вращения (2) не сохраняет дискретную решетку значений, на которой задано распределение.

В работах [1; 2] анонсированы методы воспроизведения двумерного дискретного распределения с заданными маргинальными распределениями и корреляцией, основанные на смесях некоторых базовых распределений и минимизации отклонения от независимого распределения. В настоящей работе предлагается обоснование этих методов.

Описание двумерного дискретного распределения. Пусть размерность $d = 2$. Обозначим $K = \{1, \dots, m\} \times \{1, \dots, n\}$. Дискретное распределение вектора $X = (X_1, X_2)'$ задается на прямоугольной сетке значений $\{x_{11}, \dots, x_{1m}\} \times \{x_{21}, \dots, x_{2n}\}$ в виде $P(X_1 = x_{1i}; X_2 = x_{2j}) = r_{ij}, (i, j) \in K$. Обозначим совместное распределение компонент вектора X :

$$r = \{r_{ij}, (i, j) \in K\}. \tag{3}$$

Здесь $P(X_1 = x_{1i}) = p_i, i = 1, \dots, m$ и $P(X_2 = x_{2j}) = q_j, j = 1, \dots, n$, так что векторы

$$p = (p_1, \dots, p_m), q = (q_1, \dots, q_n) \tag{4}$$

описывают маргинальные распределения компонент.