ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 25 МАТЕМАТИКИ Выпуск 3
2018

Т. С. Бородина (Нижний Новгород, ННГУ). **Оценивание совместной функции распределения в зависимости «доза-эффект» методом стохастической аппроксимации.**

Пусть $\mathbf{X}_i = (X_{1i}, X_{2i}, \ldots, X_{di})$, $1 \leqslant i \leqslant n$ — независимые и одинаково распределенные d-мерные случайные величины с неизвестной совместной функцией распределения $F(\boldsymbol{x}) = F(x_1, x_2, \ldots, x_d)$ и плотностью распределения $f(\boldsymbol{x}) > 0$; $\mathbf{U}_i = (U_{1i}, U_{2i}, \ldots, U_{di})$, $1 \leqslant i \leqslant n$ — независимые и одинаково распределенные d-мерные случайные величины (независимые от $\mathbf{X}_i, i = 1, \ldots, n$) с неизвестной плотностью $q(\boldsymbol{u}) > 0$.

Мы наблюдаем последовательность одинаково распределенных $(d+1)$-мерных случайных величин $\mathcal{U}^{(n)} = \{(W_i, U_{1i}, U_{2i}, \ldots, U_{di}),\ 1 \leqslant i \leqslant n\}$, где $W_i = I(X_{1i} < U_{1i}, \ldots, X_{di} < U_{di})$ есть индикатор события $(X_{1i} < U_{1i}, \ldots, X_{di} < U_{di})$ и по выборке $\mathcal{U}^{(n)}$ хотим оценить неизвестную совместную функцию распределения $F(\boldsymbol{x})$.

Пусть $\mathbf{H} = \mathbf{H}_{d\times d}(\lambda) = \lambda \mathbf{I}_d$, где $\lambda \in \mathbf{R}$, $\mathbf{I}_d$ — единичная матрица размера $d \times d$, $\mathbf{H}_n = \mathbf{H}_{d\times d}(b_n)$. Обозначим $\mathcal{K}_{\mathbf{H}}(\boldsymbol{x}) = |\,\mathbf{H}\,|^{-1}\mathcal{K}(\mathbf{H}^{-1}\boldsymbol{x})$, где $|\,\mathbf{H}\,| = \det(\mathbf{H})$ есть определитель матрицы $\mathbf{H}$. Через $\nabla_q(\boldsymbol{x})$ и $\nabla_F(\boldsymbol{x})$ обозначим градиенты функций $q(\boldsymbol{x})$ и $F(\boldsymbol{x})$ соответственно, а через $\mathcal{H}_F(\boldsymbol{x})$ — матрицу Гессе функции $F(\boldsymbol{x})$. Пусть $\|\,\mathcal{K}\,\|_2^2 = \int_{\mathbf{R}^d} \mathcal{K}^2(\boldsymbol{u})\,d\boldsymbol{u}$.

Для оценки функции распределения $F(\boldsymbol{x})$ мы будем использовать рекурсивный метод стохастической аппроксимации [1, 2]. Преимущество рекурсивных оценок перед их нерекурсивной версией заключается в следующем: чтобы обновить оценку от выборки объема n до оценки по выборке объема $n{+}1$ требуется значительно меньше вычислений. Это свойство особенно важно в рамках оценивания функции распределения в зависимости «доза-эффект», так как количество точек, в которых оценивается функция распределения может оказаться большим. Вычисления производятся следующим образом: положим $S_{1,0}(\boldsymbol{x}) = S_{2,0}(\boldsymbol{x}) = F_0(\boldsymbol{x}) = 0$ и для $n \geqslant 1$, будем последовательно вычислять

$$S_{1,n}(\boldsymbol{x}) = (1-\gamma_n)S_{1,n-1}(\boldsymbol{x}) + \gamma_n Z_n, \quad S_{2,n}(\boldsymbol{x}) = (1-\gamma_n)\widehat{F}_{n-1}(\boldsymbol{x})S_{1,n-1}(\boldsymbol{x}) + \gamma_n W_n Z_n,$$

$$Z_n = \mathcal{K}_{\mathbf{H}_n}(\,\boldsymbol{x} - \boldsymbol{U}_n), \quad \widehat{F}_n(\boldsymbol{x}) = \frac{S_{2,n}(\boldsymbol{x})}{S_{1,n}(\boldsymbol{x})}.$$

Мы рассматриваем *правильно меняющиеся* последовательности $\{b_n\}_{n\geqslant 1}$.

Определение [3, 4]. Пусть $\alpha \in \mathbf{R}$ и $\{\nu_n\}_{n\geqslant 1}$ есть неслучайная последовательность. Мы будем говорить, что последовательность $\{\nu_n\}_{n\geqslant 1}$ является *правильно меняющейся с показателем* α (и писать $\{\nu_n\}_{n\geqslant 1} \in RV_\alpha$), если

$$\lim_{n\to\infty} n\left(1 - \frac{\nu_{n-1}}{\nu_n}\right) = \alpha.$$

Правильно меняющиеся последовательности использовались в работах [5, 6].

Условия.

(A1) $\mathcal{K} : \mathbf{R}^d \to \mathbf{R}^+$ есть неотрицательные, непрерывные, ограниченные функции, удовлетворяющие условию $\int_{\mathbf{R}^d} \mathcal{K}(\boldsymbol{x})\, d\boldsymbol{x} = 1$; $\int_{\mathbf{R}^d} x_j\, \mathcal{K}(\boldsymbol{x})\, d\boldsymbol{x} = 0$, $j = 1, \ldots, d$ и пусть $\mu_2(\mathcal{K}) = (\mu_{ij})_{d\times d}$, $\mu_{ij} = \int_{\mathbf{R}^d} x_i x_j\, \mathcal{K}(\boldsymbol{x})\, d\boldsymbol{x} < \infty$.

(A2) $(b_n) \in RV_a$, $a = -1/(d+4)$.

(A3) $0 < \xi = \lim\limits_{n\to\infty} (n\gamma_n)^{-1} < \infty$.

(A4) f, q есть ограниченные, трижды непрерывно дифференцируемые функции и для любых $i, j \in \{1, \ldots, d\}$, $q_{ij} = \partial^2 q/\partial x_i \partial x_j$, $(Fq)_{ij} = \partial^2 (Fq)/\partial x_i \partial x_j$, $\mathrm{F}_{ij} = \partial^2 F/\partial x_i \partial x_j$ есть ограниченные функции, $\mathcal{H}_F = (F_{ij})_{d\times d}$.

Теорема. *Пусть выполнены предположения* (A1)–(A4) *Тогда при* $n \to \infty$,

(i) $$\mathbf{E}(\widehat{F}_n(\boldsymbol{x})) - F(\boldsymbol{x}) = \frac{\mu_2(\mathcal{K})(\nabla_F(\boldsymbol{x})\, \mathbf{H}_n^T \mathbf{H}_n \nabla_q(\boldsymbol{x}) + (1/2)\mathrm{tr}(\mathbf{H}_n^T \mathcal{H}_F(\boldsymbol{x}) \mathbf{H}_n))}{(1 + 2a\xi)\, q(\boldsymbol{x})}\, (1 + o\,(1)),$$

(ii) $$\mathbf{Var}(\widehat{F}_n(\boldsymbol{x})) = \frac{\gamma_n \|\, \mathcal{K}\, \|_2^2 F(\boldsymbol{x})(1 - F(\boldsymbol{x}))}{(2 - (1 + ad)\xi)\, |\, \mathbf{H}_n\, |\, q(\boldsymbol{x})}\, (1 + o\,(1)).$$

Заметим, что если матрица $\mathbf{H}$ симметричная, то $\mathbf{H}^T \mathbf{H} = \mathbf{H}^2$. В качестве ядра обычно берут либо произведение одномерных ядер, либо эллиптическое ядро.

Было проведено численное моделирование, которое показало эффективность предложенного подхода. Рассмотрен также случай, произвольной матрицы $\mathbf{H}$.

Что касается применений предложенного подхода, то заметим, что практически вся лекарственная терапия основана на принципах одновременного действия нескольких разных препаратов для достижения лечебного эффекта. Кроме того, зависимость «доза-эффект» — это условное название. Данная модель может быть применена, например, к оценке послепожарного состояния лесов, где в качестве показателя W выступает *выживаемость после пожара*, а в качестве дозы: 1) U_1 — *высота нагара на стволах*, 2) U_2 — *доля поврежденных корней* и др. (см. [7]).

СПИСОК ЛИТЕРАТУРЫ

1. Robbins H., Monro S. A Stochastic Approximation Method. — Ann. Math. Statist., 1951, v. 22, № 3, p. 400–407.
2. Вазан М. Стохастическая аппроксимация, М.: Мир, 1972, 296 с.
3. Galambos J., Seneta E. Regularly varying sequences. — Proc. Amer. Math. Soc., 1973, v. 41, № 1, p. 110–116.
4. Seneta E. Variants of Karamata's Iteration Theorem. — Pub. Inst. Math., 2006, v. 80(94), p. 241–251.
5. Mokkadem A., Pelletier M., Slaoui Y. The stochastic approximation method for the estimation of a multivariate probability density. — J. Stat. Plann. and Infer., 2009, v. 139, p. 2459–2478.
6. Slaoui Y. The stochstic approximation method for estimation of a distribution function. — Math. Methods Statist., 2014, v. 23, Is. 4, p. 306–325.
7. Абаимов А. П., Прокушкин С. Г., Суховольский В. Г., Овчинникова Т. М. Оценка и прогноз послепожарного состояния лиственницы Гмелина на мерзлотных почвах средней Сибири. — Лесоведение, 2004, № 2, с. 3–11.