Отчет по конкурсу "Молодая математика России" за 2021 год

Пучкин Никита Андреевич

1 Результаты исследований

Рассмотрена задача оценки гладкого многообразия по неточным наблюдениям $Y_1, \dots, Y_T \in \mathbb{R}^D$ временного ряда, сгенерированным из модели

$$Y_t = X_t + \varepsilon_t, \quad 1 \leqslant t \leqslant T, \tag{1}$$

где $\{X_t : 1 \leqslant t \leqslant T\}$ – марковский случайный процесс на гладком многообразии $\mathcal{M}^* \subset \mathbb{R}^D$ размерности $d < D$, $\varepsilon_1, \dots, \varepsilon_T$ – независимые от $X_1, \dots, X_T$ центрированные случайные векторы. Данная задача мотивирована практическими приложениями. Оценку многообразия $\mathcal{M}^*$ можно использовать при прогнозировании значений временного ряда. Приведем примеры известных в литературе моделей временного ряда, являющихся частным случаем модели (1).

Пример 1.1 (Модель центрального подпространства, [6]) *Пусть $g : \mathbb{R}^d \to \mathbb{R}^p$ – гладкая функция, Φ – матрица размера $(p \times d)$, $d < p$. Модель центрального подпространства предполагает, что наблюдения временного ряда $Z_1, \dots, Z_T$ сгенерированы из модели*

$$Z_t = g(\Phi^\top Z_{t-1}) + \xi_t, \quad 1 \leqslant t \leqslant T,$$

где $\xi_1, \dots, \xi_T$ – независимые центрированные случайные векторы с конечной дисперсией. Заметим, что векторы $X_t = (Z_{t-1}, g(\Phi^\top Z_{t-1})) \in \mathbb{R}^D$, где $D = 2p$, $2 \leqslant t \leqslant T$, образуют марковский процесс на гладком многообразии размерности d. Взяв $Y_t = (Z_{t-1}, Z_t)$, $\varepsilon_t = (0, \xi_t)$, $2 \leqslant t \leqslant T$, приходим к модели (1).

Пример 1.2 (Одномерная авторегрессионная модель) *В стандартной одномерной авторегрессионной модели порядка $\tau \in \mathbb{N}$ наблюдения $Z_1, \dots, Z_T$ генерируются из модели*

$$Z_t = \sum_{i=1}^{\tau} a_i Z_{t-i} + \xi_t, \quad \tau + 1 \leqslant t \leqslant T.$$

Зафиксируем $D > \tau$ и рассмотрим векторы $Y_t = (Z_t, \dots, Z_{t-D+1})^\top \in \mathbb{R}^D$. Тогда авторегрессионная модель может быть переписана в виде

$$Y_t = AY_{t-1} + \varepsilon_t, \quad D + 1 \leqslant t \leqslant T,$$

где $\varepsilon_t = (\xi_t, 0, \ldots, 0)^\top \in \mathbb{R}^D$ *и*

$$A = \begin{pmatrix} a_1 & a_2 & \ldots & a_\tau & 0 & \ldots & 0 & 0 \\ 1 & 0 & \ldots & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 1 & 0 \end{pmatrix} \in \mathbb{R}^{D \times D}.$$

В этом случае очевидно, что векторы $X_t = AY_{t-1}$, $D + 1 \leqslant t \leqslant T$ *образуют марковский процесс в* $Im(A)$. *Так как* $rank(A) < D$, $Im(A)$ – *линейное собственное подпространство* $\mathbb{R}^D$ *размерности* $rank(A)$.

Были сформулированы следующие предположения относительно многообразия $\mathcal{M}^*$, марковского процесса $\{X_t : 1 \leqslant t \leqslant T\}$ и шума $\varepsilon_1, \ldots, \varepsilon_T$. Во-первых, предполагается, что $\mathcal{M}^*$ принадлежит классу $\mathscr{M}^d_\varkappa$ компактных, линейно связных $\mathcal{C}^2$ многообразий без края, которые содержатся в шаре $\mathcal{B}(0, R)$, имеют рич не менее $\varkappa$ и размерность d:

$$\begin{aligned} \mathcal{M}^* \in \mathscr{M}^d_\varkappa = \big\{ & \mathcal{M} \subset \mathbb{R}^D : \mathcal{M} \text{ – компактное линейно-связное} \\ & \text{многообразие без края}, \mathcal{M} \in \mathcal{C}^2, \mathcal{M} \subseteq \mathcal{B}(0, R), \\ & \operatorname{reach}(\mathcal{M}) \geqslant \varkappa, \dim(\mathcal{M}) = d < D \big\}. \end{aligned} \tag{A1}$$

Рич многообразия $\mathcal{M}$ определяется как супремум таких $r > 0$, что любая точка в $\mathcal{M} + \mathcal{B}(0, r)$ имеет единственную проекцию на $\mathcal{M}$. Здесь и далее $A + B = \{a + b : a \in A, b \in B\}$ — сумма Минковского множеств A и B. Условие $\operatorname{reach}(\mathcal{M}^*) \geqslant \varkappa$ играет ключевую роль при оценке многообразия $\mathcal{M}^*$ и часто используется в литературе (см., например, [5, 4, 3, 2, 1]).

Перейдем к описанию свойств марковского процесса $\{X_t : 1 \leqslant t \leqslant T\}$. Были рассмотрены случаи эргодического и неэргодического процесса $\{X_t : 1 \leqslant t \leqslant T\}$. Пусть $\mathbb{P}_t$ – маргинальное распределение X_t. В эргодическом случае предполагается существование у $\{X_t\}$ стационарного распределения π и констант $A > 0$, $\rho \in (0, 1]$, таких что для любого $t \in \{1, \ldots, T\}$ мера $\mathbb{P}_t$ удовлетворяет неравенству

$$\|\mathbb{P}_t - \pi\|_{\mathrm{TV}} \leqslant A(1 - \rho)^t, \tag{A2}$$

где $\|\cdot\|_{\mathrm{TV}}$ – расстояние полной вариации. Стоит отметить, что в работах, посвященных оценке гладких многообразий на основе простой выборки (см., например, [4, 3, 1]), часто предполагается, что для любого t маргинальная плотность X_t отделена от нуля, что в нашем случае автоматически влекло бы эргодичность марковского процесса. В неэргодическом случае условие (A2) значительно ослабляется. Для каждого $t \in \{1, \ldots, T\}$ обозначим сигма-алгебру, порожденную $X_1, \ldots, X_t$, через $\mathcal{F}_t$, и пусть $\mathcal{F}_0$ – тривиальная сигма-алгебра. Мы предполагаем, что существуют $k \in \mathbb{N}, h_0 > 0$ и $p_1 \geqslant p_0 > 0$, такие что для всех $t \in \{0, \ldots, T - k\}$, $h \in (0, h_0)$ и $x \in \mathcal{M}^*$ выполнено неравенство

$$p_0 h^d \leqslant \frac{1}{k} \sum_{j=1}^{k} \mathbb{P}\left(X_{t+j} \in \mathcal{B}(x, h) \,|\, \mathcal{F}_t\right) \leqslant p_1 h^d. \tag{A3}$$

Условие (A3) аналогично предположению, введенному в [7].

Наконец, опишем требования, предъявляемые к векторам $\varepsilon_1, \ldots, \varepsilon_T$. Случайный вектор $\xi \in \mathbb{R}^D$ называется субгауссовским с параметром σ^2, если

$$\sup_{\|u\|=1} \mathbb{E} e^{\lambda u^T (\xi - \mathbb{E}\xi)} \leqslant e^{\lambda^2 \sigma^2 / 2}, \quad \forall \lambda \in \mathbb{R}.$$

Предполагаем, что $\varepsilon_1, \ldots, \varepsilon_T$ – независимые центрированные субгауссовские случайные векторы:

$$\mathbb{E}\varepsilon_t = 0, \quad \varepsilon_t \in \mathrm{SG}(\sigma_t^2), \quad \forall\, t \in \{1, \ldots, T\}. \tag{A4}$$

Одной из естественных оценок $\mathcal{M}^*$ является минимизатор эмпирического риска:

$$\widehat{\mathcal{M}} \in \operatorname*{argmin}_{\mathcal{M} \in \mathscr{M}_\varkappa^d} \frac{1}{T} \sum_{t=1}^{T} d^2(Y_t, \mathcal{M}). \tag{2}$$

Грубо говоря, $\widehat{\mathcal{M}}$ – гладкое многообразие, которое лучше всех подстраивается под наблюдения $Y_1, \ldots, Y_T$. Нас интересует вопрос об обобщающей способности оценки (2), то есть насколько хорошо оценка $\widehat{\mathcal{M}}$, полученная по одной траектории $Y_1, \ldots, Y_T$, подстраивается под другие траектории процесса. Пусть случайные векторы $Y_1', \ldots, Y_T'$ имеют то же совместное распределение, что и $Y_1, \ldots, Y_T$, но при этом независимы от них. Для каждого $\mathcal{M} \in \mathscr{M}_\varkappa^d$ можно ввести следующую метрику качества, которую будем называть риском:

$$\frac{1}{T} \sum_{t=1}^{T} \mathbb{E} d^2(Y_t', \mathcal{M}).$$

Тогда качество оценки $\widehat{\mathcal{M}}$ можно охарактеризовать величиной избыточного риска по сравнению с $\mathcal{M}^*$:

$$\frac{1}{T} \sum_{t=1}^{T} \mathbb{E} d^2(Y_t', \widehat{\mathcal{M}}) - \frac{1}{T} \sum_{t=1}^{T} \mathbb{E} d^2(Y_t', \mathcal{M}^*).$$

Для случая эргодического марковского процесса $\{X_t\}$ была доказана следующая теорема.

Теорема 1.3 *Пусть выполнены предположения* (A1), (A2) *и* (A4). *Тогда минимизатор эмпирического риска* (2)*удовлетворяет неравенству*

$$\frac{1}{T} \sum_{t=1}^{T} \mathbb{E} d^2(Y_t', \widehat{\mathcal{M}}) - \inf_{\mathcal{M} \in \mathscr{M}_\varkappa^d} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E} d^2(Y_t', \mathcal{M}) = \begin{cases} O\left(\sqrt{\frac{D \log T}{T \log(1/(1-\rho))}}\right), & d < 4, \\ O\left(\frac{\sqrt{D} \log^{3/2} T}{\sqrt{T \log(1/(1-\rho))}}\right), & d = 4, \\ O\left(\left(\frac{\log T}{T \log(1/(1-\rho))}\right)^{2/d}\right), & d > 4. \end{cases}$$

Результат Теоремы 1.3 значительно улучшает верхние оценки $O((\log^4 T/T)^{1/(d+4)})$ и $O((\log^4 T/T)^{2/(d+4)})$, полученные в работах [5] и [2] соответственно, для случая, когда $Y_1, \ldots, Y_T$ образуют простую выборку. В случае, когда процесс $\{X_t\}$ не является эргодическим, было доказано следующее.

Теорема 1.4 *Пусть выполнены условия* (A1), (A3) *и* (A4). *Предположим, что ортогональная и касательная к* $\mathcal{M}^*$ *компоненты векторов* $\varepsilon_1, \ldots, \varepsilon_T$ *независимы. Пусть существует константа* $c \in (0, 1)$, *такая что* $\sigma_1, \ldots, \sigma_T$ *удовлетворяют неравенству*

$$\frac{8}{T} \sum_{t=1}^{T} \sigma_t^2 + 64 D \sigma_{\max}^2 \leqslant \frac{p_0}{8k} \left(\frac{c\varkappa}{4}\right)^{d+2}.$$

Тогда с вероятностью хотя бы $1 - 8/T$ *выполнено неравенство*

$$\frac{1}{T}\sum_{t=1}^{T} d^2(X_t, \widehat{\mathcal{M}}) = O\left(\psi_T + \frac{D\left(\sigma_{\max}\sqrt{\log T} \vee (\log T/T)^{1/d}\right)}{T}\sum_{t=1}^{T}\sigma_t^2\right) + O\left(\frac{\max\left\{\sigma_{\max}^4 \log^2 T, (\log T/T)^{4/d}\right\}}{\varkappa^2}\right),$$

где $\sigma_{\max} = \max\limits_{1\leqslant t\leqslant T} \sigma_t$ *и*

$$\psi_T = \begin{cases} \frac{1}{T}\sqrt{\sum\limits_{t=1}^{T}\sigma_t^2}, & d < 4, \\ \frac{\log T}{T}\sqrt{\sum\limits_{t=1}^{T}\sigma_t^2}, & d = 4, \\ T^{-2/d}\sqrt{\frac{\sum\limits_{t=1}^{T}\sigma_t^2}{T}}, & d > 4. \end{cases}$$

2 Опубликованные и поданные в печать работы

В отчетном году было опубликовано и подано в печать две работы.

- N. Puchkin and N. Zhivotovskiy. Exponential Savings in Agnostic Active Learning Through Abstention. In *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pages 3806–3832, 2021.
- N. Puchkin and V. Spokoiny. Structure-adaptive Manifold Estimation. *Journal of Machine Learning Research (в печати).*

3 Участие в конференциях и летних школах

В 2021 году выступил с докладом на трех конференциях:

- 34th Annual Conference on Learning Theory, Боулдер, Колорадо (+онлайн), 15–19 августа 2021г.;
- New Trends in Mathematical Stochastics, Санкт-Петербург, 30 августа – 3 сентября 2021г.;
- Statistics, Artificial Intelligence, Machine Learning, Probability, Learning Theory Event, Геленджик, 26–30 октября 2021г.

Принял участие в качестве преподавателя в летней школе "Современные методы теории информации, оптимизации и управления," проходившей в Образовательном центре "Сириус", г. Сочи, с 19 июля по 8 августа 2021г., а также в качестве слушателя в двух конференциях и одной летней школе:

- конференция "Новые вызовы в современной теории вероятностей в пространствах высокой размерности и ее применениях в машинном обучении", Сочи, 12–16 мая 2021г.;

- летняя школа “Информация, управление и оптимизация”, Вороново, 10 – 17 июня 2021г.;
- конференция “Оптимизация без границ”, Сочи, 11–18 июля 2021г.

4 Работа в научных центрах

Являюсь младшим научным сотрудником Международной лаборатории стохастических алгоритмов и анализа многомерных данных НИУ ВШЭ, а также и.о. младшего научного сотрудника в Институте проблем передачи информации им. А. А. Харкевича РАН.

5 Педагогическая деятельность

В 2021 году было принято участие в семи курсах в качестве лектора, семинариста или учебного ассистента.

- Онлайн-методы машинного обучения, МФТИ, весенний семестр, лектор.
- Статистическая теория машинного обучения, МФТИ, осенний семестр, лектор.
- Advanced statistical methods, НИУ ВШЭ и Сколтех, весенний семестр, семинарист.
- Random matrix theory, НИУ ВШЭ, осенний семестр, семинарист.
- Математическая статистика, МФТИ, осенний семестр, семинарист.
- Введение в теорию случайных процессов, НИУ ВШЭ, весенний семестр, учебный ассистент.
- Введение в теорию вероятностей, НИУ ВШЭ, осенний семестр, учебный ассистент.

Список литературы

[1] E. Aamari and C. Levrard. Nonasymptotic rates for manifold, tangent space and curvature estimation. *Ann. Statist.*, 47(1):177–204, 2019.

[2] C. Fefferman, S. Mitter, and H. Narayanan. Testing the manifold hypothesis. *J. Amer. Math. Soc.*, 29(4):983–1049, 2016.

[3] C. R. Genovese, M. Perone-Pacifico, I. Verdinelli, and L. Wasserman. Manifold estimation and singular deconvolution under Hausdorff loss. *Ann. Statist.*, 40(2):941–963, 2012.

[4] C. R. Genovese, M. Perone-Pacifico, I. Verdinelli, and L. Wasserman. Minimax manifold estimation. *J. Mach. Learn. Res.*, 13:1263–1291, 2012.

[5] H. Narayanan and S. Mitter. Sample complexity of testing the manifold hypothesis. In *Advances in Neural Information Processing Systems 23*, pages 1786–1794. 2010.

[6] J.-H. Park, T. Sriram, and X. Yin. Dimension reduction in time series. *Statistica Sinica*, 20, 04 2010.

[7] M. Simchowitz, H. Mania, S. Tu, M. Jordan, and B. Recht. Learning without mixing: Towards a sharp analysis of linear system identification. 02 2018.