Секция «Теория вероятностей и математическая статистика»

Статистические методы определения неоднородностей волкнистых материалов

Научный руководитель – Булинский Александр Вадимович

Димитров Денис Валерьевич
Аспирант
Московский государственный университет имени М.В.Ломоносова,
Механико-математический факультет, Кафедра теории вероятностей, Москва, Россия
E-mail: denis.dimitrov@math.msu.ru

В работе рассматривается применение оценок k-ближайших соседей дивергенции Кульбака–Лейблера [5] к поиску областей неоднородности в материалах. Асимптотические свойства таких оценок (а именно: асимптотическая несмещенность и L^2-состоятельность) были доказаны в недавней работе [2].

В статьях [1, 3] оценки дифференциальной энтропии Шеннона применяются для обнаружения неоднородностей в волокнистом материале, заполняющем параллелепипед $\Pi \subset \mathbb{R}^3$. Используется естественный подход, при котором выбирается скользящее окно достаточно малого размера, в пределах которого оценивается некоторая характеристика материала (фактически дифференциальная энтропия Шеннона распределения направлений волокон). Затем на основе собранных характеристик делается статистический вывод о свойствах этого материала.

В статье [4] оценки дивергенции Кульбака-Лейблера используются для обнаружения многомерных пространственных кластеров в модели Бернулли. Нами развиваются и обобщаются идеи этой работы для обнаружения произвольных неоднородностей в волокнистых материалах.

Вводится понятие *волокна* — пары (X, T), где X — случайный вектор со значениями в $\mathbb{R}^{d_1} \cap \Pi$, который по смыслу является центром волокна, а T — случайный вектор со значениями в $\mathbb{R}^{d_2}$, который по смыслу является меткой или некоторым свойством волокна (например, направлением в пространстве). Именно среди таких меток и будет искаться неоднородность. Здесь предполагается, что Π — некоторое объемлющее множество или материал, $\Pi \in \mathcal{B}(\mathbb{R}^{d_1})$, $\mu_{\mathbb{R}^{d_1}}(\Pi) < \infty$, $\mu_{\mathbb{R}^{d_1}}$ — мера Лебега в пространстве $\big(\mathbb{R}^{d_1}, \mathcal{B}(\mathbb{R}^{d_1})\big)$; $\mathsf{P}(X \in B) = \frac{\mu_{\mathbb{R}^{d_1}}(B)}{\mu_{\mathbb{R}^{d_1}}(\Pi)}$ для любого борелевского $B \subset \mathbb{R}^{d_1} \cap \Pi$, а T имеет следующую природу: $T = \xi \cdot \mathbf{1}\{X \in R_0\} + \eta \cdot \mathbf{1}\{X \in \Pi \setminus R_0\}$, где ξ, η – некоторые случайные векторы со значениями в $\mathbb{R}^{d_2}$, $\mathsf{law}(\xi) \neq \mathsf{law}(\eta)$, $p_\xi := \frac{d\mathsf{P}_\xi}{d\mu_{\mathbb{R}^{d_2}}}$, $p_\eta := \frac{d\mathsf{P}_\eta}{d\mu_{\mathbb{R}^{d_2}}}$. В этом случае назовём R_0 *областью неоднородности*. Все случайные элементы определены на пополненном вероятностном пространстве $(\Omega, \mathcal{F}, \mathsf{P})$.

Рассматривается модель (см. также Рис. 1), в которой наблюдателю доступны $\big\{X_i, T_i, i \in \{1, \ldots, \zeta_n\}\big\}$, $\zeta_n \sim \mathrm{Pois}(\lambda_n \cdot \mu_{\mathbb{R}^{d_1}}(\Pi))$, где для любого $n \in \mathbb{N}$, $\lambda_n > 0$, $\lambda_n \to \infty$, $n \to \infty$. Отметим также, что $\{X_i, T_i, \zeta_i, i \in \mathbb{N}\}$ предполагаются независимыми, $\mathsf{law}(X_i) = \mathsf{law}(X)$, $\mathsf{law}(T_i) = \mathsf{law}(T)$ для любого $i \in \mathbb{N}$. Для каждого множества R из некоторого набора множеств $\mathcal{R}$ строится оценка $\widehat{T}_n(R)$, основанная на оценках k-ближайших соседей дивергенции Кульбака–Лейблера между распределением меток T_j внутри окна R (то есть при $X_j \in R$) и вне его. Оценка области неоднородности $\widehat{R}_n$ полагается равной аргументу максимума такой специальной статистики $\widehat{T}_n(R)$, взятого по всем окнам $R \in \mathcal{R}$.

При широких условиях удается доказать состоятельность предложенной процедуры (то есть удается доказать, что в некотором смысле $\widehat{R}_n$ близко к R_0 для достаточно больших $n \in \mathbb{N}$).

Источники и литература

1) Alonso-Ruiz P.; Spodarev E. Entropy-based inhomogeneity detection in fiber materials // In Methodology and Computing in Applied Probability, 2018, 20, P. 1223–1239.
2) Bulinski A.; Dimitrov D. Statistical Estimation of the Kullback–Leibler Divergence // In Mathematics, 2021, 9, 544, P. 1–36.
3) Dresvyanskiy D.; Karaseva T.; Makogin V.; Mitrofanov S.; Redenbach C.; Spodarev E. Detecting anomalies in fibre systems using 3-dimensional image data // In Statistics and Computing, 2020, 30, 4, P. 817–837.
4) Walther G. Optimal and fast detection of spatial clusters with scan statistics // In The Annals of Statistics, 2010, 38, P. 1010–1033.
5) Wang Q.; Kulkarni S.R.; Verdú S. Divergence estimation for multidimensional densities via k-nearest-neighbor distances // In IEEE Transactions on Information Theory, 2009, 55, P. 2392–2405.

Иллюстрации

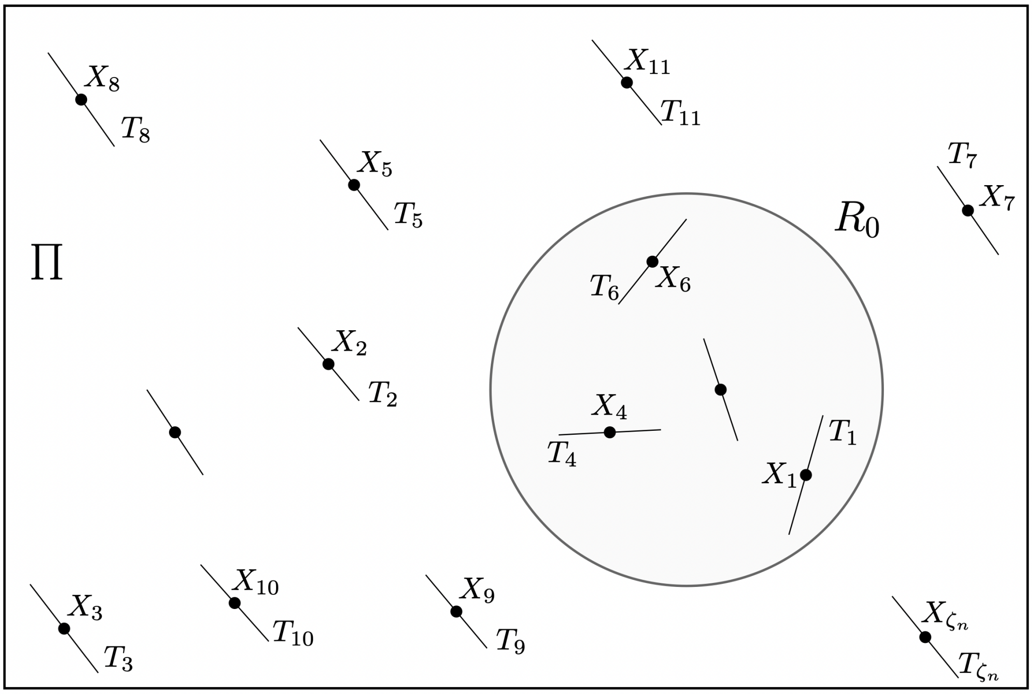

Рис. 1. Исследуемый волокнистый материал Π с областью неоднородности R_0.