

С. В. Стафеев (Петрозаводск, ИПМИ КарНЦ РАН). **О размерности смешанных гауссовских деревьях с латентными переменными.**

УДК 519.237.7

DOI https://doi.org/10.52513/08698325_2022_33_1/00

Резюме: Рассматриваются смешанные гауссовские деревья с латентными переменными. Найдена размерность рассматриваемых моделей.

Ключевые слова: гауссовские деревья, латентные переменные, идентифицируемость, размерность модели.

Рассмотрим смешанное гауссовское дерево [1] с графом структуры $G(\mathbf{X}, \mathbf{E}_\mathbf{X})$. Множество вершин $\mathbf{X}$ представляется в виде объединения $\mathbf{Y} \cup \mathbf{H}$, где $\mathbf{Y}$ и $\mathbf{H}$ — множества листьев и внутренних вершин дерева G соответственно. Мы будем предполагать, что $\mathbf{Y}$ — наблюдаемые, а $\mathbf{H}$ — латентные (ненаблюдаемые) случайные величины.

Представим зависимость между элементами множества $\mathbf{Y}$ в виде системы регрессионных уравнений

$$Y_i = a_i H(Y_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad (1)$$

где $H(Y_i) \in \mathbf{H}$; $\varepsilon = \{\varepsilon_1, \dots, \varepsilon_n\}$ — вектор остатков, с нулевым вектором математических ожиданий и матрицей ковариаций $\Omega = (\omega_{ij})$; $\mathbf{H}$ и ε независимы. Мы будем предполагать, что все латентные переменные имеют нулевые математические ожидания и единичные дисперсии.

Обозначим $a_{ij} = \text{Cov}(H_i, H_j)$, $(H_i, H_j) \in \mathbf{E}_\mathbf{X}$ и пусть $\theta = \{a_i, i = 1, \dots, n, a_{ij}, (H_i, H_j) \in \mathbf{E}_\mathbf{X}, \omega_{ij}, (Y_i, Y_j) \in \mathbf{E}\}$. Определим параметрическое множество $\Theta = \{\theta | \Omega - \text{положительно определена}\}$.

Мы будем говорить, что простая цепь $P(X', X'')$ содержит тупиковую вершину X_i , если она содержит подцепь одного из четырех видов:

$$X_{i-1} \rightarrow X_i \leftarrow X_{i+1}, \quad X_{i-1} \rightarrow X_i \leftrightarrow X_{i+1}, \quad X_{i-1} \leftrightarrow X_i \leftarrow X_{i+1}, \quad X_{i-1} \leftrightarrow X_i \leftrightarrow X_{i+1}.$$

Пусть $\Sigma = (\sigma_{ij})$ — матрица ковариаций наблюдаемых случайных величин. Для смешанного гауссовского дерева мы имеем:

$$\sigma_{ij} = \begin{cases} \prod_{(H_k, H_l) \in P(Y_i, Y_j)} a_i a_j a_{kl}, & \text{если простая цепь } P(Y_i, Y_j) \\ & \text{не содержит тупиковых вершин;} \\ a_i a_j + \omega_{ij} & \text{если } i = j; \\ 0, & \text{в противном случае;} \end{cases} \quad (2)$$

Соотношение (2) задает отображение

$$\varphi : \Theta \rightarrow \mathbb{R}^{\frac{n(n+1)}{2}}. \quad (3)$$

Ранг R матрицы Якоби $J(\theta)$ отображения φ называется размерностью модели (1).

Пусть $\mathbf{Ne}(H_i)$ — множество вершин, соединенных с H_i неориентированными ребрами и пусть $\mathbf{Ch}(H_i)$ — множество вершин, в которые идут дуги из H_i .

Теорема. Пусть все латентные вершины модели (1) имеют не менее трех смежных вершин и $|\mathbf{Ne}(H_i) \cup \mathbf{Ch}(H_i)| \geq 1$. Тогда

$$R = |\mathbf{Y}| + |\mathbf{H}| - m - 1,$$

где m — число латентных вершин, для которых $|\mathbf{Ne}(H_i) \cup \mathbf{Ch}(H_i)| = 1$.

Финансовое обеспечение исследований осуществлялось из средств федерального бюджета на выполнение государственного задания КарНЦ РАН (Институт прикладных математических исследований КарНЦ РАН).

СПИСОК ЛИТЕРАТУРЫ

1. *Maathuis M., Drton M., Lauritzen S., Wainwright M. (Eds) Handbook of Graphical Models.* Chapman & Hall/CRC., 2018, 536 p.
2. *Shiers N., Zviernik P., Aston J. D., Smith J. Q.* The correlation space of Gaussian latent tree models and model selection without fitting. — *Biometrika*, 2016, v. 103, № 3, p. 531–545.

УДК 519.237.7

DOI https://doi.org/10.52513/08698325_2022_33_1/00

Stafeev S. V. (Petrozavodsk, Institute of Applied Mathematical Research of Karelian Research Centre Russian Academy of Sciences). **On the dimension of mixed gaussian latent tree models.**

Abstract: The problem of dimension of gaussian latent tree models is considered. The dimension for such models is obtained.

Keywords: gaussian tree models, latent variables, identifiability, model dimension.