Дипломна робота на тему

Оцінювання характеристик частково спостережуваного ланцюга Маркова на двійкових послідовностях

Виконав
студент 4 курсу, групи ФІ-91
113 «Прикладна математика»
Цибульник А. В.

Науковий керівник
ст. викл. Наказной П. О.

Консультант
к.ф.-м.н., доцент Ніщенко І. І.

Актуальність та мета дослідження

Актуальність: вивчення еволюції систем за частковою інформацією про динаміку їхніх станів, які, своєю чергою, є наборами символів певної довжини (ДНК, мова жестів).

Мета дослідження: за зміною в часі набору функціоналів від двійкових послідовностей побудувати оцінки невідомих параметрів моделі.

План доповіді

1. Моделювання об'єкта дослідження
2. Побудова оцінок параметрів моделі
3. Результати чисельного експерименту

Моделювання об'єкта дослідження

Стан системи — двійкова послідовність довжини N.

Еволюція станів системи за узагальненою моделлю Еренфестів: навмання обраний символ стану X^t з імовірністю p не змінюється, а з імовірністю $(1-p)$ — змінюється.

Наприклад, при $N = 12$

$$
\begin{array}{ll}
 & x^t \\
t = 1 & 01\mathbf{0}011011101 \\
t = 2 & 01\mathbf{1}01101\mathbf{1}1101 \\
t = 3 & 0110110\mathbf{0}11\mathbf{0}1 \\
t = 4 & 01101100 11\mathbf{0}1
\end{array}
$$

Моделювання об'єкта дослідження

Спостерігаємо набір функціоналів

$$Y^t = \left(Y_1^t, \ldots, Y_L^t\right) = \left(\sum_{i \in I_1} X_i^t, \ldots, \sum_{i \in I_L} X_i^t\right),$$

де $I_1, \ldots, I_L$ є заданими підмножинами $\{1, 2, \ldots, N\}$.

Наприклад, при $N = 12$, $I_1 = (1, 2, 3)$, $I_2 = (6, 7, 10, 11, 12)$

	x^t	y^t
$t = 1$	010011011101	$(1, 3)$
$t = 2$	011011011101	$(2, 3)$
$t = 3$	011011001101	$(2, 3)$
$t = 4$	011011001101	$(2, 3)$

Побудова оцінок параметрів моделі

1. За наявними частковими спостереженнями $\{Y^t\}_{t=\overline{1,T}}$ про динаміку бінарних послідовностей оцінити керуючий параметр p заданої марковської моделі.

Метод максимальної правдоподібності:

$$\widehat{p} = \underset{p}{\operatorname{argmax}} \sum_{x \in E^T} P\left(X = x,\, Y = y \,|\, p\right) \equiv \underset{p}{\operatorname{argmax}} \sum_{x \in E^T} L_{p,x,y}$$

Ітераційний алгоритм Баума-Велша:

$$p^{(n+1)} = \underset{p}{\operatorname{argmax}}\, Q\left(p^{(n)}, p\right) = \underset{p}{\operatorname{argmax}} \sum_{x \in E^T} L_{p^{(n)},x,y} \ln L_{p,x,y}$$

Побудова оцінок параметрів моделі

Формула переоцінки параметра p, починаючи з деякого $p^{(0)}$:

$$p^{(n+1)} = p^{(n)} \cdot \frac{\sum\limits_{t=1}^{T-1} \sum\limits_{x \in E} \alpha_t(x)\, B_{xy^{t+1}}\, \beta_{t+1}(x)}{\sum\limits_{t=1}^{T-1} \sum\limits_{x \in E} \alpha_t(x)\, \beta_t(x)},$$

де

$$\alpha_t(x) = P\left(Y^1 = y^1, \ldots, Y^t = y^t,\, X^t = x \,|\, p^{(n)}\right)$$
$$\beta_t(x) = P\left(Y^{t+1} = y^{t+1}, \ldots, Y^T = y^T \,|\, X^t = x,\, p^{(n)}\right)$$

Побудова оцінок параметрів моделі

Крім того, побудовано змістовну та незміщену точкову оцінку параметра p за допомогою методу моментів:

$$\widehat{p} = 1 - \frac{N}{\left| \bigcup_{k=1}^{L} I_k \right|} \left(1 - \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{1}\left(Y^t = Y^{t+1} \right) \right)$$

Побудова оцінок параметрів моделі

2. За наявними спостереженнями $\{ Y^t \}_{t=\overline{1,T}}$ та оцінкою керуючого параметра $p^{(n)}$ відновити послідовність двійкових наборів.

Алгоритм Вітербі: пошук такої послідовності станів $\widehat{X}^1, \widehat{X}^2, \ldots, \widehat{X}^T$, яка найкращим чином описує наявні спостереження:

$$\widehat{X} = \underset{x \in E^T}{\operatorname{argmax}} P\left(X = x \,|\, Y = y, p^{(n)}\right)$$

Побудова оцінок параметрів моделі

3. За відомими значеннями набору функціоналів від деякої невідомої підмножини I_* стану прихованого ланцюга, оцінити потужність та набір елементів цієї підмножини.

Отже, спостерігаємо значення $Y^t_{I_*} = \sum\limits_{i \in I_*} X^t_i$.

Наприклад, $N = 12$, $I_1 = (1, 2, 3)$, $I_2 = (6, 7, 10, 11, 12)$, $I_* = ?$

	x^t	y^t	$y^t_{I_*}$
$t = 1$	010011011101	$(1, 3)$	3
$t = 2$	011011011101	$(2, 3)$	3
$t = 3$	011011001101	$(2, 3)$	2
$t = 4$	011011001101	$(2, 3)$	2

Побудова оцінок параметрів моделі

За набором спостережуваних «сигналів» $Y^1_{I_*}, \ldots, Y^T_{I_*}$, оцінкою параметра $p^{(n)}$ та декодованим ланцюгом станів $\left\{ \widehat{X}^{\,t} \right\}_{t=\overline{1,T}}$:

- побудовано змістовну та незміщену точкову оцінку потужності множини I_* :

$$\widehat{|I_*|} = \frac{N}{1-p} \left(1 - \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbb{1}\Big(Y^t_{I_*} = Y^{t+1}_{I_*} \Big) \right)$$

- розроблено алгоритм визначення компонент множини I_*.

Побудова оцінок параметрів моделі

4. Спостереження на множинах $I_1, \ldots, I_L$ спотворюються з імовірностями $q_1, \ldots, q_L$:

$$Y^t = \left(Y_k^t\right)_{k=\overline{1,L}} = \left(\sum_{i \in I_k} \widetilde{X}_i^{\,t}\right)_{k=\overline{1,L}}$$

де для $i \in I_k$

$$\widetilde{X}_i^{\,t} = \begin{cases} 1 - X_i^t, & \text{з імовірністю } q_k \\ X_i^t, & \text{з імовірністю } 1 - q_k \end{cases}$$

Побудова оцінок параметрів моделі

Наприклад, при $N = 12$, $I_1 = (1,2,3)$, $I_2 = (6,7,10,11,12)$

	x^t	$\widetilde{x}^t$	y^t	q
$t = 1$	010011011101	000011011101	$(0,3)$	(q_1, q_2)
$t = 2$	011011011101	010011011101	$(1,3)$	(q_1, q_2)
$t = 3$	011011001101	111011001111	$(3,4)$	(q_1, q_2)
$t = 4$	011011001101	011011001000	$(2,1)$	(q_1, q_2)

Задача: за спотвореними спостереженнями оцінити керуючий параметр моделі p та вектор ймовірностей спотворення q, використовуючи ітераційний алгоритм Баума-Велша.

Побудова оцінок параметрів моделі

Починаючи з деякого наближення $p^{(0)}$ та $q^{(0)}$, формула переоцінки параметра p :

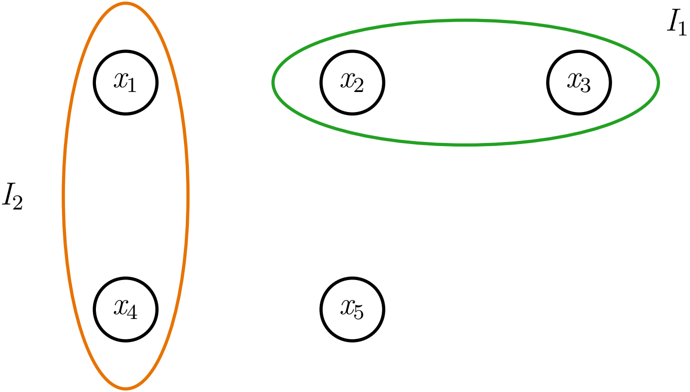
$$p^{(n+1)} = p^{(n)} \cdot \frac{\sum\limits_{t=1}^{T-1} \sum\limits_{x \in E} \alpha_t(x)\, B_{xy^{t+1}}^{q^{(n)}}\, \beta_{t+1}(x)}{\sum\limits_{t=1}^{T-1} \sum\limits_{x \in E} \alpha_t(x)\, \beta_t(x)}$$

Формула переоцінки компонент вектора q :

$$q_k^{(n+1)} = q_k^{(n)} \cdot \frac{\sum\limits_{t=1}^{T} \sum\limits_{x \in E} \sum\limits_{x' \in E} \alpha_{t-1}(x')\, A_{x'x}^{(n)}\, \beta_t(x) \sum\limits_{i \in I_k} P_{x,i,1}^{q^{(n)}}}{|I_k| \sum\limits_{t=1}^{T} \sum\limits_{x \in E} \alpha_t(x)\, \beta_t(x)}$$

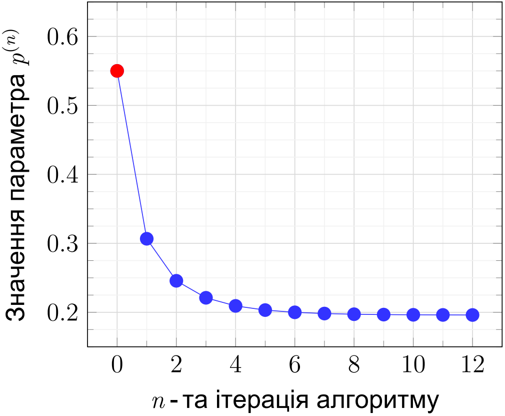
Результати чисельного експерименту

Було згенеровано прихований ланцюг Маркова протягом $T = 200$ моментів часу, $N = 5$ та $p = 0.2$. Множина спостережуваних індексів $I = \{I_1, I_2\} = \{(2,3),(1,4)\}$:



Результати чисельного експерименту

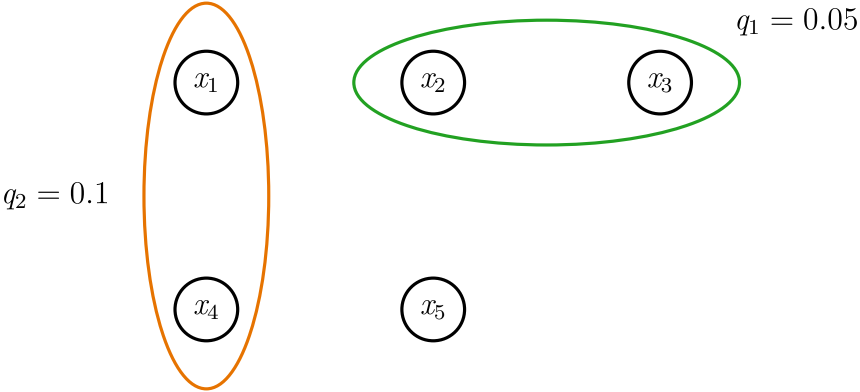
- За наявними частковими спостереженнями про динаміку бінарних послідовностей оцінено керуючий параметр p :



p	$p^{(12)}$	$\lvert p - p^{(12)} \rvert$
0.2	0.1959	0.0041

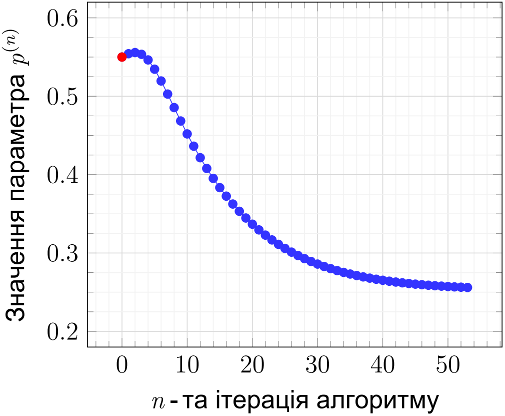
Результати чисельного експерименту

Для спостережуваних множин I_1, I_2 згенерованого ланцюга було задано такі коефіцієнти спотворення:



Результати чисельного експерименту

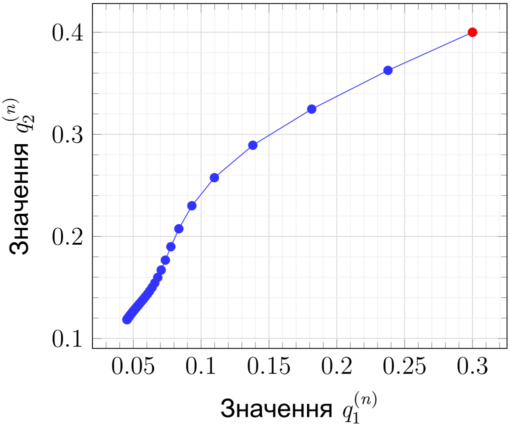
- За спотвореними спостереженнями оцінено керуючий параметр моделі p та вектор ймовірностей q :



p	$p^{(53)}$	$\lvert p - p^{(53)} \rvert$
0.2	0.2559	0.0559

Результати чисельного експерименту

- За спотвореними спостереженнями оцінено керуючий параметр моделі p та вектор ймовірностей q :



q_1	$q_1^{(53)}$	$\lvert q_1 - q_1^{(53)}\rvert$
0.05	0.0454	0.0046

q_2	$q_2^{(53)}$	$\lvert q_2 - q_2^{(53)}\rvert$
0.1	0.1184	0.0184

Висновки

Невідомі параметри заданої моделі були оцінені

- або шляхом побудови змістовних та незміщених оцінок за допомогою методу моментів;
- або за допомогою ітераційного алгоритму Баума-Велша.

Результати чисельного експерименту продемонстрували ефективність використаних методів, зокрема збіжність побудованих оцінок до істинних значень параметрів при збільшенні кількості спостережень.

Апробація результатів та публікації

- *Цибульник А. В., Ніщенко І. І.,* XXI Всеукраїнська науково-практична конференція студентів, аспірантів та молодих вчених «Теоретичні і прикладні проблеми фізики, математики та інформатики».

  Секція «Математичне моделювання та аналіз даних» (стр. 419 – 432).

  11-12 травня 2023 р., м. Київ.