Порівняння параметричних та напівпараметричних моделей виживання з моделями, в яких застосовуються методи jackknife і bootstrap

Виконала:

студентка 4-го курсу групи КА-31

Воронцева А. І.

Керівник: к. ф.-м. н., доцент Каніовська І. Ю.

Об’єкт дослідження

- Оцінки параметрів моделей виживання

Предмет дослідження

- Застосування методів повторних вибірок для покращення оцінок параметрів моделей виживання

Мета дослідження

- Застосувати методи jackknife і bootstrap до моделей аналізу виживання та отримати оцінки параметрів.
- Порівняти отримані оцінки з оцінками, отриманими класичними методами та зробити висновок щодо можливості, доцільності та умов застосування.

Актуальність

- Класичні моделі аналізу виживання часто дають неточні та зміщені оцінки через малий розмір вибірки, який часто виникає в клінічних дослідженнях.
- Використовувати регресійні моделі неможливо через наявність цензурування.
- За допомогою методів повторної вибірки можна отримати точні незміщені оцінки і таким чином підвищити якість аналізу результатів багатьох досліджень, у тому числі в медицині.

Моделі аналізу виживання

Модель пропорційних ризиків

$$\lambda_i(t) = \lambda(t)\exp(x_i^T\beta),$$

$\lambda_i(t)$ – функція ризику

$\lambda(t)$ - базова функція ризику

x_i^T - вектор коваріат

Модель прискорення часу відмови

$$\log T_i = x_i^T\beta + W_i,$$

T_i - час виживання

x_i^T - вектор коваріат

W_i - похибка

Методи повторних вибірок

Bootstrap

- Згенерувати велику кількість (~10000) bootstrap-вибірок з початкових даних того ж розміру шляхом вибору з повторенням
- Обчислити бажану статистику
- Отримані оцінки формують її вибірковий розподіл

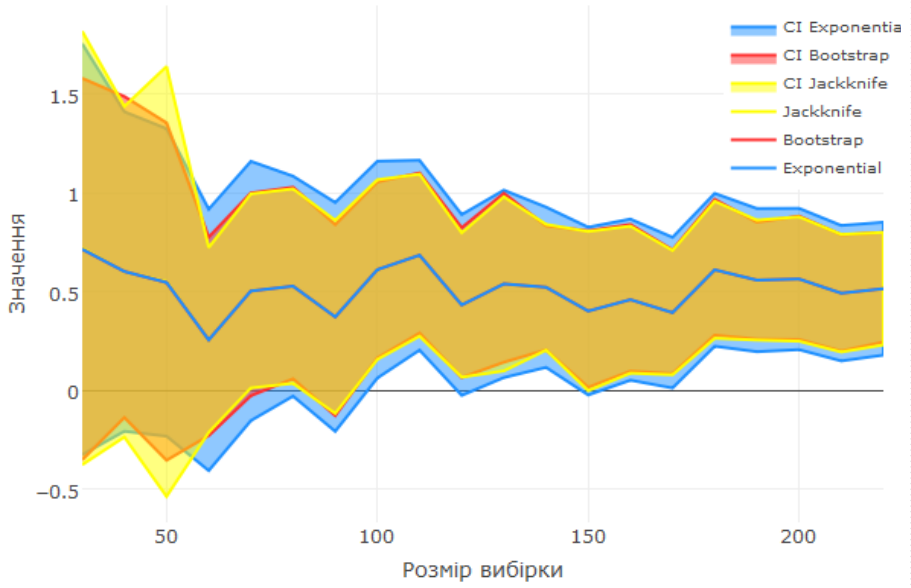
Jackknife

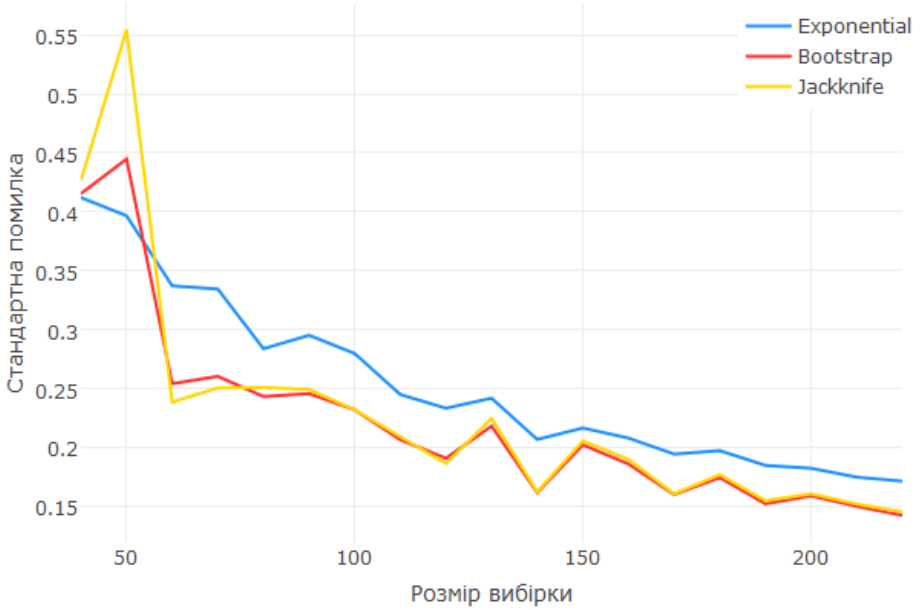
- Jackknife-вибірки отримуються шляхом взяття вихідного вектора даних і видалення одного спостереження з набору. Кількість отриманих вибірок рівна розміру вхідного вектора
- Обчислити бажану статистику
- Отримані оцінки формують її вибірковий розподіл

Суть запропонованого методу

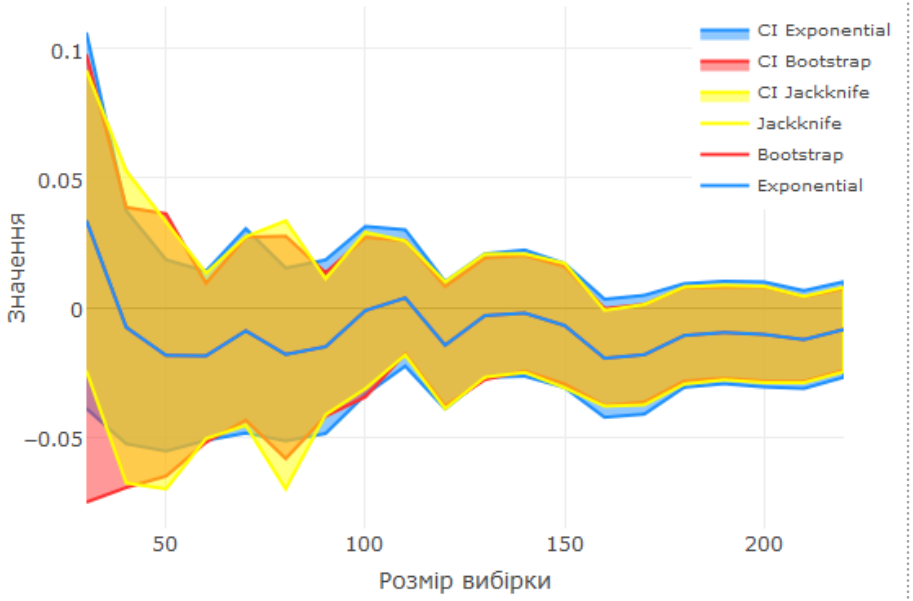
- Побудуємо модель прискорення часу відмови з експоненційним розподілом та модель пропорційних ризиків
- Проведемо моделювання bootstrap і jackknife для оцінок параметрів моделей на вибірках різного розміру
- Проведемо моделювання на двох наборах реальних даних клінічних досліджень та проаналізуємо отриманий результат

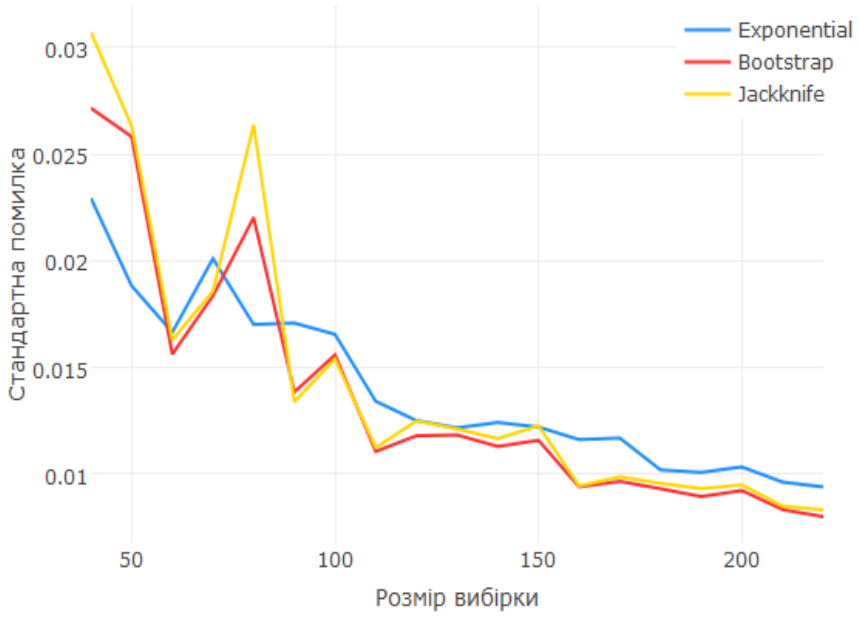
Отримані результати: приклад





Отримані результати: приклад





Отримані результати: опис

- На вибірках розміром більше 100 елементів bootstrap та jackknife мають стандартну помилку меншу за звичайну оцінку.
- Якщо параметр має досить високу значущість, стандартна помилка методів bootstrap та jackknife буде вищою в порівнянні зі звичайною оцінкою.
- Bootstrap дає дуже широкий довірчий інтервал та високу стандартну помилку на малих вибірках, а jackknife хоча і дає кращі оцінки на малих вибірках, дає більш широкий довірчий інтервал ніж bootstrap при тій же стандартній помилці.

Отримані результати: опис

- Bootstrap є точнішим за jackknife, але при цьому потребує значних обчислень.
- На малих вибірках оцінки bootstrap мають дуже широкий довірчий інтервал, на відміну від звичайної та jackknife оцінок, проте зі збільшенням розміру вибірки довірчий інтервал bootstrap стає вужчим за інші.
- За допомогою методів bootstrap та jackknife можна оцінювати зміщення статистики та використовувати дане значення для його усунення, однак це не можна робити на малих вибірках

У роботі було виконано порівняння оцінок параметричної і напівпараметричної моделей, та оцінками, отриманими з застосуванням методів bootstrap та jackknife.

Наукова новизна

Було запропоновано використовувати методи bootstrap та jackknife для підвищення точності оцінок, доцільність показана на двох клінічних дослідженнях

Практична значимість результатів

В аналізі виживання дуже часто доводиться працювати з малими вибірками (не більше 500 елементів), в цих умовах стандартні моделі можуть давати невірні або неточні оцінки, які є зміщеними. Запропонований підхід дає змогу отримати точні незміщені оцінки і таким чином підвищити якість аналізу результатів

Перспективи подальших досліджень

- У подальших дослідженнях планується порівняння оцінок методів bootstrap та jackknife з оцінками параметричних моделей з іншими розподілами, такими як Вейбулла, Гамма, Логнормальний, Логістичний та інші.
- Також планується застосування даних методів для аналізу нових даних клінічних досліджень для отримання висновків про ефективність лікування або фактори, що впливають на розвиток певної хвороби.

Результати даної роботи доповідались на інтернет-конференції The IV-th International Scientific and Practical Conference “Open Evolving Systems”, яка проходила 18 – 21 травня 2016 року

Дякую за увагу!