

Ідентифікації типу нелінійного перетворення в просторах однорідних спостережень за допомогою перетворення Джонсона.

В багатовимірному аналізі даних вимога нормального закону розподілу спостережень є ключовою, адже це значно спрощує вирішення задачі оцінювання ймовірнісних функцій. Проте, на практиці маргінальні розподіли можуть бути суттєво асиметричними, що вимагає пошуку методів для їх нормалізації. В даній роботі представлено результати дослідження ефективності перетворення Джонсона до даних отриманих за допомогою нелінійного перетворення x^α в просторах однорідних спостережень.

Перетворення Джонсона — це сімейство статистичних методів S_U, S_B, S_L [1, 2], які використовуються для перетворення даних, таким чином, що вони можуть бути апроксимовані нормальним розподілом. Вибір розподілу із сімейства Джонсона для нормалізації даних визначається за допомогою коефіцієнти асиметрії та ексцесу вхідних даних [3,4].

В ході дослідження реалізовано процес генерації одновимірних даних, що відповідають нормальному розподілу. Надалі ці дані піддавалися нелінійному перетворенню шляхом піднесення до степені з параметром α , значення якого змінювалися від 0.1 до 5 з кроком 0.1. Для відновлення початкового розподілу використовувалися методи Джонсона, які дозволяли зблизити спотворені дані до нормального розподілу.

Для відображення результатів для кожного α створено 1000 вибірок обсягом 1000 елементів та обсягом 100 елементів, побудовано графіки асиметрії та ексцесу до та після перетворення Джонсона, досліджена відносна частота кількості перетворень S_U, S_B, S_L , застосовано критерій згоди Колмогорова – Смирнова.

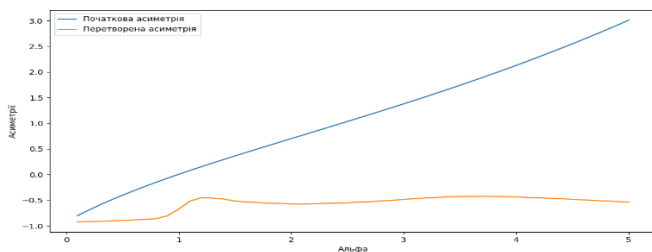


Рис.1. Графік усередненої асиметрії для кожного α до та після перетворення Джонсона, $N=1000$.

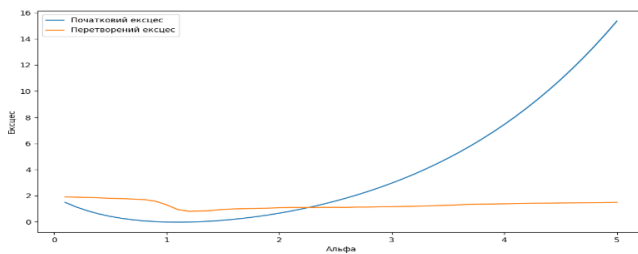


Рис.2. Графік усередненого ексцесу для кожного α до та після перетворення Джонсона, $N=1000$.

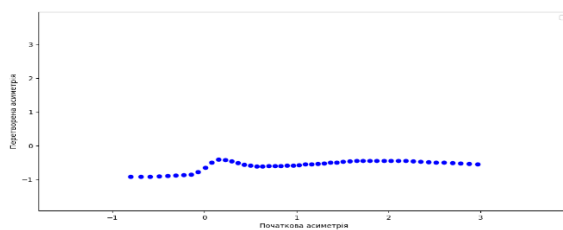


Рис.3. Графік залежності асиметрії перетворених даних від асиметрії початкових даних, $N=1000$.

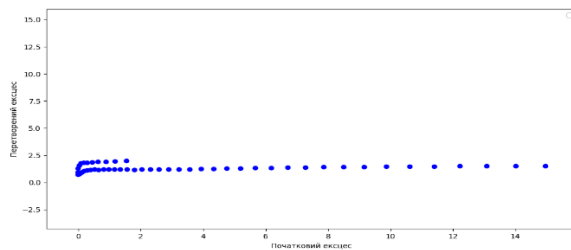


Рис.4. Графік залежності ексцесу перетворених даних від ексцесу початкових даних, $N=1000$.

На рис.3. та рис.4. видно, що початкова асиметрія та ексцес поступово зростає, проте їх значення після перетворення зберігаються на приблизно одному, низькому рівні.

Результати тих самих графіків, тільки з меншим обсягом вибірки $N=100$ суттєвої різниці не дали, відбувається незначне погіршення по усім показникам.



Рис.5. Графік відносних частот кількості нормально розподілених даних після перетворення, $N=1000$.



Рис.6. Графік відносних частот кількості нормально розподілених даних після перетворення, для вибірок менших обсягом, $N=100$.

У випадку маленької кількості елементів у вибірці, результат показує більше 95% нормально розподілених даних після перетворення Джонсона. Тому можна зазначити, що перетворення Джонсона більше підходить для перетворення вибірок меншим обсягом. Причиною такого результату може бути, наприклад те, що більша кількість даних краще відображає реальний розподіл даних, що може бути відміним від нормального.

Також, дослідження показали, що незалежно від розміру вибірок, найчастіше зустрічається перетворення S_B .

Висновок. Перетворення Джонсона демонструє хороші результати у перетворенні даних до розподілу, що є близьким до нормального, як на великих обсягах даних так і на невеликих. Проте критерій згоди на вибірках меншого

обсягу показав значно кращі результати в порівнянні з більшими обсягами вибірок. Проведені дослідження показали низьку обчислювальну складність, відповідно запропонований маргінальний метод можна рекомендувати до застосування в обробці багатовимірних даних.

Подальші дослідження можуть полягати у врахуванні появи аномалій та типів нелінійних спотворень, що відмінні від розглянутого степеневого.

Список літератури

1. Єременко В. С. Дослідження перетворення Джонсона для задач підвищення точності метрологічних характеристик стандартних зразків / В. С. Єременко, В. М. Мокійчук, О. В. Самойліченко // Системи обробки інформації. - 2010. - Вип. 4. - С. 36-42.
2. Жан Г. “Статистичні моделі в інженерних задачах ” / Жан Г, Шапиро С.. Пер. з англ. — М.: Мир, 1969. — 400 с.
3. Johnson, N.L. Tables to facilitate fitting S_U frequency curves / Johnson, N.L. // Biometrika. 1965. V. 52. P. 547-558.
4. Johnson N. L. Systems of frequency curves generated by methods of translation / Johnson N. L. // Biometrika. 1946. V. 36. P. 146-148.