

УДОСКОНАЛЕННЯ ЙМОВІРНІСНОЇ МОДЕЛІ ОЦІНЮВАННЯ ЧАСУ НАПРАЦЮВАННЯ НА ВІДМОВУ ЖОРСТКИХ ДИСКІВ

Копієвський Д.В.¹, Макарова Л.М., к.т.н., доцент²

^{1,2} Національний університет кораблебудування імені адмірала Макарова

^{1,2} Україна, м. Миколаїв

¹ kopievskydv@gmail.com, ² lidia.makarova@nuos.edu.ua, ORCID 0000-0003-2903-3001

Анотація. Розглянуто проблему оцінювання часу напрацювання на відмову жорстких дисків. Удосконалено ймовірнісну модель розподілу часу напрацювання на відмову жорстких дисків із застосуванням нормалізуючого перетворення Джонсона, що дало змогу підвищити достовірність оцінювання часу напрацювання на відмову відповідних пристроїв.

Ключові слова: ймовірнісна модель; нормалізуюче перетворення; перетворення Джонсона; час напрацювання на відмову; жорсткий диск.

Вступна частина. На сучасному етапі розвитку суспільства більша частина важливої інформації зберігається на цифрових носіях, починаючи від особистої інформації і закінчуючи важливими файлами та даними малого і великого бізнесу. Для збереження важливих даних існують сервіси, що надають послуги по створенню резервних копій у хмарі [1].

У зв'язку із швидким збільшенням об'єму даних та розширенням систем зберігання даних у хмарі, стабільна робота жорстких дисків та прогнозування їх відмов стає все більш важливою проблемою, яка має великий вплив на галузь [2]. Відмова диска може призвести до втрати важливих даних. Якщо є можливість точно передбачити відмову диска, то можна вчасно створити резервну копію та відновити дані. А це дозволить підвищити надійність системи зберігання даних [3]. Компаніям, що надають послуги резервного копіювання в хмару, важливо своєчасно замінювати несправні жорсткі диски на нові. Отже, проблема оцінювання й прогнозування часу напрацювання на відмову жорстких дисків є досить важливою, а вирішення даної проблеми, використовуючи традиційний підхід на основі нормального або експоненціального закону розподілу не дає задовільних результатів. Тому виникає необхідність в удосконаленні ймовірнісної моделі для оцінювання часу напрацювання на відмову жорстких дисків.

Ціль роботи: удосконалення ймовірнісної моделі оцінювання часу напрацювання на відмову жорстких дисків із використанням нормалізуючого перетворення Джонсона.

Основна частина. В теорії надійності, як аналітична модель закону розподілу часу напрацювання між відмовами різних пристроїв, прийнятий експоненціальний закон, який в свою чергу є наближеним рішенням [4]. Проте отриманні дані про напрацювання на відмову жорстких дисків не задовольняють цьому закону, що робить неможливим коректне оцінювання часу напрацювання на відмову жорстких дисків.

Для оцінки статистичних характеристик випадкової величини є можливість використовувати інтервальні або точкові оцінки. Однак більш надійним є оцінювання за допомогою довірчого інтервалу. Вибіркове середнє є точковою оцінкою математичного сподівання. При нормальному законі розподілу довірчий інтервал математичного сподівання визначають на основі t -розподілу Стюдента, що наведений в [5]. Оцінювання довірчих

інтервалів статистичних моментів для випадкової величини, розподіл якої відрізняється від нормального, відомо лише для небагатьох законів розподілу, наприклад, для експоненціального.

Для оцінювання довірчих інтервалів статистичних моментів тоді, коли випадкова величина не підпадає під нормальний чи експоненціальний закони розподілу, існує два підходи – підхід на основі нормалізуючих перетворень [6] або непараметричний підхід [5].

Саме тому діємо за наступним алгоритмом:

1. Виконуємо нормалізацію емпіричних даних про час напрацювання на відмову жорстких дисків.

2. Визначаємо довірчі інтервали точкових оцінок статистичних моментів для нормалізованої випадкової величини за допомогою традиційного способу.

3. Виконуємо зворотне перетворення.

4. Основуючись на зворотному перетворенні, отримуємо довірчі інтервали статистичних моментів початкової випадкової величини.

В результаті отримуємо удосконалену ймовірнісну модель розподілу часу напрацювання на відмову жорстких дисків із застосуванням нормалізуючого перетворення.

Для того, щоб апроксимувати емпіричні дані про час напрацювання на відмову жорстких дисків було запропоновано використати нормалізуюче перетворення Джонсона [6], яке у загальному випадку, має наступний вигляд:

$$z = \gamma + \eta h(x, \varphi, \lambda); \eta > 0; -\infty < \gamma < \infty; \lambda > 0; -\infty < \varphi < \infty, \quad (1)$$

де z – нормована нормально розподілена випадкова величина; $\gamma, \eta, \varphi, \lambda$ – параметри перетворення; x – випадкова величина, що нормалізується; h – функція певної сім'ї розподілу Джонсона.

Перетворення (1) має зворотне перетворення:

$$x = \varphi + \lambda h^{-1}(z, \gamma, \eta); \eta > 0; -\infty < \gamma < \infty; \lambda > 0; -\infty < \varphi < \infty. \quad (2)$$

Сім'я розподілів Джонсона обирається виходячи із значень квадрата асиметрії A^2 і ексцесу ϵ вибірки емпіричних даних про час напрацювання на відмову жорстких дисків за формулою, що наведена в [7]. Використовуючи метод максимальної правдоподібності, знаходимо значення невідомих параметрів розподілу Джонсона.

Далі будуємо ймовірнісну модель розподілу часу напрацювання на відмову жорстких дисків, час роботи вимірювався у днях.

Використовуючи критерій згоди χ^2 Пірсона, з довірчою ймовірністю 0,95 було встановлено, що гіпотеза про відповідність емпіричних даних експоненціальному закону розподілу не приймається: $\chi^2 = 169,41$ при критичному значенні $\chi^2_{кр} = 12,59$.

Ймовірнісні характеристики емпіричного розподілу: $n=147$; $\bar{x}=684,01$; $D=183503$; $\sigma=428,37$ $A=0,5243$; $\epsilon=2,0809$. Ґрунтуючись на оцінках A^2 і ϵ , була обрана сім'я розподілів S_B Джонсона.

Нормалізація випадкової величини x була виконана у відповідності з перетворенням (1), параметри перетворення: $\gamma = 0,3888$; $\eta = 0,7580$; $\varphi = -7,2088$; $\lambda = 1717,45$.

Використовуючи критерій згоди χ^2 Пірсона, з довірчою ймовірністю 0,95 було встановлено, що гіпотеза, про відповідність перетворених даних нормальному закону розподілу, приймається: $\chi^2 = 8,76$ при критичному значенні $\chi^2_{кр} = 9,49$.

Ймовірнісні характеристики нормалізованої вибірки: $n=147$; $\bar{z}_x=0,0002$; $D_z=1,0001$; $\sigma_z=1,0001$; $A_z=0,0000$; $\varepsilon_z=3,0000$.

Функція щільності ймовірності для сім'ї розподілів S_B Джонсона має вигляд [6]:

$$f_B(x) = \frac{\eta\lambda}{\sqrt{2\pi}(x-\varphi)(\lambda+\varphi-x)} \exp\left\{-\frac{1}{2}\left[\gamma+\eta\ln\left(\frac{x-\varphi}{\lambda+\varphi-x}\right)\right]^2\right\}, \quad (3)$$

де $\varphi < x < \varphi + \lambda$; $\eta > 0$; $-\infty < \gamma < \infty$; $-\infty < \varphi < \infty$.

В результаті підстановки параметрів перетворення Джонсона у формулу (3), отримуємо функцію щільності ймовірності розподілу часу напрацювання на відмову жорстких дисків:

$$f_B(x) = \frac{1301,8271}{\sqrt{2\pi}(x+7,2088)(1710,2412-x)} \exp\left\{-\frac{1}{2}\left[0,3888+0,758 \cdot \ln\left(\frac{x+7,2088}{1710,2412-x}\right)\right]^2\right\}.$$

Проведемо розрахунок 95% довірчих інтервалів точкових оцінок вибіркового середнього і параметру експоненціального розподілу λ випадкової величини x , використовуючи непараметричне оцінювання, експоненціальний закон розподілу та нормалізуюче перетворення Джонсона сім'ї S_B . Зазначені довірчі інтервали наведено в таблиці 1.

Таблиця 1 - Довірчі інтервали точкових оцінок вибіркового середнього і параметру експоненціального розподілу λ випадкової величини x

Параметр	Непараметричне оцінювання		Експоненціальний закон розподілу		Нормалізуюче перетворення Джонсона	
	нижня границя	верхня границя	нижня границя	верхня границя	нижня границя	верхня границя
$[\hat{\alpha}_1(x)]$	588,31	814,12	597,35	774,38	614,19	753,84
$[\lambda]$	$1,2283 \cdot 10^{-3}$	$1,6998 \cdot 10^{-3}$	$1,2914 \cdot 10^{-3}$	$1,6741 \cdot 10^{-3}$	$1,3265 \cdot 10^{-3}$	$1,6282 \cdot 10^{-3}$

З наведених вище даних випливає, що непараметричне оцінювання призводить до збільшення у 1,62 та 1,56 рази довірчих інтервалів вибіркового середнього та параметру експоненціального розподілу відповідно в порівнянні з підходом на основі нормалізуючого перетворення Джонсона сім'ї S_B . Також, при застосуванні нормалізуючого перетворення Джонсона, дістаємо несиметричний довірчий інтервал вибіркового середнього, що демонструє асиметрію вихідної вибірки на відміну від непараметричного оцінювання. Також у 1,27 рази довірчий інтервал вибіркового середнього, що базується на експоненціальному законі, більший в порівнянні з підходом, що ґрунтується на нормалізуючому перетворенні Джонсона.

Можна зробити висновок, що коли емпіричний розподіл випадкової величини відрізняється від нормального, використання непараметричного оцінювання може призвести до помітно викривлених результатів. У такому випадку кращим варіантом є застосування нормалізуючих перетворень.

Висновки. Удосконалено ймовірнісну модель розподілу часу напрацювання на відмову жорстких дисків із застосуванням нормалізуючого перетворення Джонсона. У подальшому планується розробка відповідної програми для автоматизації процесу розрахунків.

REFERENCES

- [1] Amrit Singh (2020). Protecting Your Business: Cloud Backup Vs. Cloud Sync. Взято з <https://www.backblaze.com/blog/business-cloud-backup-vs-cloud-sync/>.
- [2] Liu, W., Xue, Y., Liu, P. (2020). *SMART HDD Anomaly Risk Prediction. Communications in Computer and Information Science*, vol 1261. Singapore: Springer, Singapore.
- [3] Qiang, Li, Hui, Li, Kai, Zhang. (2019). *Prediction of HDD Failures by Ensemble Learning. 2019 IEEE 10th International Conference on Software Engineering and Service Science*. Beijing, China, China: Institute of Electrical and Electronics Engineers.
- [4] Острейковский, В.А. (2003). *Теория надежности: Учебник для вузов*. Москва: Высшая школа.
- [5] Орлов, А.И. (2004). *Прикладная статистика*. Москва: Экзамен.
- [6] Приходько, С.Б. (2011). Інтервальне оцінювання статистичних моментів негаусівських випадкових величин на основі нормалізуючих перетворень. *Математичне моделювання: науковий журнал*. – Дніпродзержинськ, 1(24), 9-13.
- [7] Приходько, С.Б. (2016). Аналитическая зависимость для выбора семейства распределений Джонсона. *Проблемы інформаційних технологій*. – Херсон: ХНТУ, 2(20), 105-110.

Dmytro Kopievskiy, Lidiia Makarova

Improvement of the probabilistic model for estimation of the hard disk drive failure time

Abstract. *The problem of hard disk failure time estimation is considered. The probabilistic model of hard disk drive failure time allocation using Johnson's normalizing transformation has been improved, which allowed to increase reliability of the respective devices failure time estimation.*

Keywords: *probabilistic model; normalizing transformation; Johnson transformation; time to failures; hard disk drive.*

Копиевский Д.В., Макарова Л.Н.

Усовершенствование вероятностной модели оценки времени наработки на отказ жестких дисков

Аннотация. *Рассмотрена проблема оценки времени наработки на отказ жестких дисков. Усовершенствована вероятностная модель распределения времени наработки на отказ жестких дисков с использованием нормализующего преобразования Джонсона, что позволило повысить достоверность оценивания времени наработки на отказ соответствующих устройств.*

Ключевые слова: *вероятностная модель; нормализующее преобразование; преобразование Джонсона; время наработки на отказ; жесткий диск.*