

№ п/п	$I_0, \%$	$I_1, \%$	$I_2, \%$	$I_3, \%$
1	70	56	81	81
2	70	71	71	76
3	59	55	85	88
4	58	49	80	85
5	64	45	68	80
6	78	49	72	86
7	72	63	75	82
8	63	52	68	81
9	74	68	83	82
10	63	54	82	85
Среднее	67	56	76	82

Для оценки качества сокрытия данных в контейнере использован также программный инструмент StegSolve, позволяющий обнаружить артефакты и скрытую информацию в медиаконтейнерах. В результате анализа изображения-контейнера с помощью StegSolve восстановить скрытые изображения или обнаружить артефакты без раскрывающей сети не удалось, что свидетельствует о том, что данный метод имеет высокую надёжность передачи сообщений без возможности восстановления скрытых изображений участником, не имеющим доступа к раскрывающей сети.

ЛИТЕРАТУРА

1. *Baluja S.* Hiding Images in Plain Sight: Deep Steganography // Adv. Neural Inform. Processing Systems. 2017. No. 30. P. 1–11.
2. *Zhu J., Kaplan R., Johnson J., and Fei-Fei L.* HiDDeN: Hiding Data With Deep Networks. 2018. <https://arxiv.org/abs/1807.09937>.

УДК 519.23:519.6: 004.052

DOI 10.17223/2226308X/14/33

АДАПТАЦИЯ МЕТОДА РОЗЕНБЛАТТА — ПАРЗЕНА ДЛЯ ЭКСПЕРИМЕНТАЛЬНОЙ ОЦЕНКИ НАДЁЖНОСТИ ВЫЧИСЛИТЕЛЬНОЙ СИСТЕМЫ

В. С. Никулин

Отсутствие исходной информации о законе распределения случайных величин и их реализация в момент, близкий к началу наблюдения, а также наличие цензурированных данных вызывает необходимость адаптации непараметрического метода Розенблатта — Парзена. Для компенсации смещения и устранения нарушения условия нормировки рассмотрен метод зеркального отображения исходных данных. При построении плотности распределения случайных величин предлагается учитывать цензурированные данные. Проведённая оценка точности показывает уменьшение ошибки при аппроксимации эмпирической плотности распределения адаптированным методом Розенблатта — Парзена. На примере экспериментальной оценки показателей надёжности вычислительной системы продемонстрирована практическая реализация адаптированного метода Розенблатта — Парзена. Построение плотности и функции распределения времени до отказа позволяет рассчитать основные показатели безотказности объекта: интенсивность отказов, вероятность безотказной работы, среднее время наработки на отказ.

Ключевые слова: экспериментальный анализ надёжности, малые выборки, вычислительные системы, метод Розенблатта — Парзена.

Экспериментальная оценка показателей надёжности сводится к построению плотности распределения времени отказа. Применение параметрических методов анализа статистической информации ограничено требованием необходимости знания закона распределения наблюдаемых случайных величин. Например, в теории надёжности показано, что наработки до отказа однотипных радиоэлектронных элементов подчиняются экспоненциальному закону, при этом характер отказов предполагается внезапным, а сумма случайных величин, описываемых экспоненциальным распределением, подчиняется гамма-закону [1]. Другие факторы, обосновывающие применение того или иного семейства распределений, отсутствуют. Решением проблемы отсутствия информации о виде закона распределения наблюдаемых случайных величин является использование непараметрических методов.

Среди непараметрических методов построения плотности распределения широкое распространение находят гистограммные, проекционные, ядерные и ортогональные оценки. Методы построения этих оценок известны и хорошо изучены для случаев, когда в отношении результатов наблюдений выполняется предположение, что в каждом испытании реализуется наблюдаемый признак. Например, если речь идёт об обработке статистических данных с целью определения характеристик надёжности, то в качестве обрабатываемой информации используются полные наработки объектов до отказа.

Однако на практике информация от функционирующих объектов имеет ряд ограничений. Например, в момент ввода объекта в эксплуатацию возникают отказы, связанные с отладкой работы оборудования. Данный процесс вызывает проблему смещения плотности распределения. Непосредственно при эксплуатации объектов стараются не допускать отказов за счёт предупредительных профилактических мероприятий, в связи с этим возникает проблема малых выборок по полным наработкам на отказ. В процессе анализа надёжности приходится сталкиваться с ситуациями, когда определённая часть объектов или систем не отказывает за период наблюдения, а другая часть отказывает, но моменты отказов неизвестны. Такой процесс называется цензурированием данных. Учитывая описанные проблемы, применение непараметрических методов в явном виде не представляется возможным для задачи экспериментальной оценки показателей надёжности.

В данной работе предлагается адаптация непараметрического метода Розенблатта — Парзена. Этот метод обладает рядом преимуществ относительно метода построения гистограмм, проекционных и экспоненциальных оценок [2]. Его достоинством является положительная определённость (при выборе неотрицательного ядра), что не выполняется для оценок на основе ортогональных разложений. Важным моментом является наличие теоретических наработок по учёту малых выборок и цензурированных данных [3, 4]. В общем виде метод Розенблатта — Парзена применяется для построения плотности распределения $f(t)$ по полным наработкам на отказ [5]:

$$f(t) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{t - \varepsilon_i}{h}\right), \quad (1)$$

где $\varepsilon_1, \dots, \varepsilon_i$ — полные наработки на отказ элемента; $K(x)$ — ядро (чётная нормированная функция); h — параметр сглаживания; n — объём выборки наработок.

Анализ методов определения h показал, что значение данного параметра оказывает существенное влияние на вид оценки плотности распределения и её точность. При наличии малых выборок использование метода Сильвермана [6] для первоначального определения значения параметра h сохраняет универсальность метода построения плотности и не требует значительных вычислительных ресурсов. Формула вычисления

значения параметра h имеет вид

$$h = 0,9 \min(\hat{\sigma}, \mu/1,34) n^{-1/5},$$

где $\hat{\sigma}$ — стандартное отклонение; n — объём; μ — медиана выборки.

Выбор ядра обусловлен предпосылками, касающимися класса функции, которому принадлежит оцениваемая плотность, а также областью определения наработок на отказ. Рассмотрим функцию Гаусса в качестве первоначального ядра $f(t)$ [1]:

$$K(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2}.$$

Адаптация метода Розенблатта — Парзена

Известно, что выбранная в качестве $K(x)$ функция Гаусса определена на области $(-\infty, \infty)$, а при проведении анализа надёжности наблюдаемой величиной является время с областью определения $[0, \infty)$. В этом случае при наличии наработок на отказ, близких к нулю, присутствует смещение плотности с нарушением условия нормировки $F(0) = 0$. Для компенсации смещения плотности предлагается применить метод зеркального отображения исходных данных [4]. При этом выражение (1) примет следующий вид:

$$f(t) = \frac{1}{nh} \sum_{i=1}^n \left[K\left(\frac{t - \varepsilon_i}{h}\right) + K\left(\frac{t + \varepsilon_i}{h}\right) \right].$$

В случае цензурирования справа в качестве дополнительной исходной информации фигурирует массив из числа работоспособных элементов $\mathbf{V} = (v_1, \dots, v_s)$ на интервале $[l, \infty]$, где l — правая граница интервала эксплуатации, на котором не зафиксированно отказов. Выражение для учёта цензурированных справа данных получено в [6]. Итоговое выражение для восстановления плотности распределения наработок на отказ можно записать так:

$$\widehat{f}(t) = \frac{1}{hn} \left[\sum_{i=1}^n \left[K\left(\frac{t - \varepsilon_i}{h}\right) + K\left(\frac{t + \varepsilon_i}{h}\right) \right] + \sum_{j=1}^s \frac{v_j}{t^2} \int_0^1 K\left(\frac{1/t - u/l}{h}\right) du \right]. \quad (2)$$

Здесь первое слагаемое отвечает за полные наработки на отказ с компенсацией смещения плотности, а второе — за цензурированные данные.

В целях экспериментальной проверки метода Розенблатта — Парзена и его адаптированного варианта проведена серия экспериментов с моделированием наработок по закону распределения Вейбулла с параметрами $m = 1,1$, $\theta = 10^3$ с областью определения $[0, \infty)$. На смоделированных данных проведено исследование точности оценок. Выводы о точности основывались на вычислении ошибки оценивания e_n в метрике L_1 -пространства по выражению

$$e_n = \int_0^{\infty} |f(t) - \widehat{f}(t)| dt,$$

где $f(t)$ — смоделированная плотность распределения времени до отказа; $\widehat{f}(t)$ — экспериментальная плотность распределения времени до отказа.

По результатам эксперимента максимальная ошибка оценивания e_n составляет:

- 0,31 при расчёте $\widehat{f}(t)$ по выражению (1);
- 0,188 при расчёте $\widehat{f}(t)$ по выражению (2).

Таким образом, адаптация метода Розенблатта — Парзена позволяет повысить точность восстанавливаемой плотности распределения за счёт компенсации смещения и учёта цензурированных данных. Применение предложенной адаптации целесообразно при отсутствии априорной информации о законе распределения случайных величин, наличии малых выборок, цензурированных данных и необходимости компенсации смещения.

Рассмотрим адаптированный метод Розенблатта — Парзена на примере оценки показателей надёжности вычислительной системы. Выражение (2) позволяет рассчитать основные показатели надёжности сложных систем. Как правило, надёжность объекта определяется надёжностью комплектующих элементов, поэтому для каждого элемента из состава объекта рассчитываются следующие показатели:

- функция распределения времени на отказ:

$$F(t) = \int_0^T f(t)dt; \quad (3)$$

- вероятность безотказной работы:

$$P(t) = 1 - F(t); \quad (4)$$

- интенсивность отказов:

$$\lambda(t) = \frac{f(t)}{P(t)}; \quad (5)$$

- средняя наработка времени между отказами:

$$T_{\text{ср}} = \frac{1}{n} \sum_{j=1}^n t_j. \quad (6)$$

Здесь n — объём выборки наработок; t_j — время штатного функционирования между $(j - 1)$ -м и j -м отказами.

В качестве объекта рассматривается вычислительная система из 20 параллельно функционирующих вычислительных узлов (элементов). Период полной подконтрольной эксплуатации объекта составляет $T = 35040$ ч (4 года). В период эксплуатации система мониторинга фиксирует информацию об отказах по каждому элементу. При заданном режиме функционирования отказ объекта наступает в результате отказа всех составных элементов. Полученная информация предварительно обрабатывается в соответствии с методикой подготовки данных [7] для последующей оценки плотности распределения времени до отказа и представляет:

- $\mathbf{E}_T = (\varepsilon_1, \dots, \varepsilon_i)$ — выборки полных наработок на отказ элементов за период T , где ε_i — время отказа оборудования;
- $\mathbf{V} = (v_1, \dots, v_s)$ — массив из числа работоспособных элементов на интервале $[l, \infty]$, где l — правая граница периода T , на котором не зафиксировано отказов.

Для каждого элемента вычислительной системы по исходным данным и выражениям (3)–(6) рассчитаны основные показатели надёжности (с нижним индексом (el)) (таблица).

№ эл-та	$P_{(el)}(t)$	$\lambda_{(el)}(t)$	$T_{cp(el)}(t)$	№ эл-та	$P_{(el)}(t)$	$\lambda_{(el)}(t)$	$T_{cp(el)}(t)$
1	0,91	2,28e-04	4379	11	0,72	2,51e-04	3980
2	0,69	1,83e-04	5474	12	0,83	1,83e-04	5474
3	0,65	2,06e-04	4865	13	0,65	2,28e-04	4379
4	0,79	2,97e-04	3368	14	0,60	2,28e-04	4379
5	0,72	2,74e-04	3649	15	0,65	1,60e-04	6256
6	0,77	2,06e-04	4865	16	0,72	1,83e-04	5474
7	0,72	2,74e-04	3649	17	0,57	1,37e-04	7299
8	0,67	1,37e-04	7299	18	0,65	1,83e-04	5474
9	0,66	1,60e-04	6256	19	0,60	9,13e-05	10949
10	0,69	2,28e-04	4379	20	0,83	1,37e-04	7299

Связь показателей надёжности параллельно функционирующих элементов с показателями надёжности объекта определяется следующими выражениями:

— вероятность безотказной работы:

$$P_{(ob)}(t) = 1 - \prod_{i=1}^N (1 - P_{(el)}(t)_i); \quad (7)$$

— интенсивность отказа:

$$\lambda_{(ob)}(t) = \sum_{i=1}^N \frac{\lambda_{(el)}(t)_i}{\sum_{j=1}^N \frac{1}{j}}; \quad (8)$$

— средняя наработка времени между отказами:

$$T_{cp(ob)}(t) = \frac{1}{\lambda_{(ob)}(t)}. \quad (9)$$

По данным из таблицы и выражениям (7)–(9) рассчитаны эксплуатационные показатели надёжности объекта:

- вероятность безотказной работы $P_{(ob)}(t) = 0,999$;
- интенсивность отказа объекта $\lambda_{(ob)}(t) = 5,52e-05$;
- средняя наработка времени между отказами $T_{cp(ob)}(t) = 18108$.

По полученным оценкам построены кривые, представленные на рис. 1.

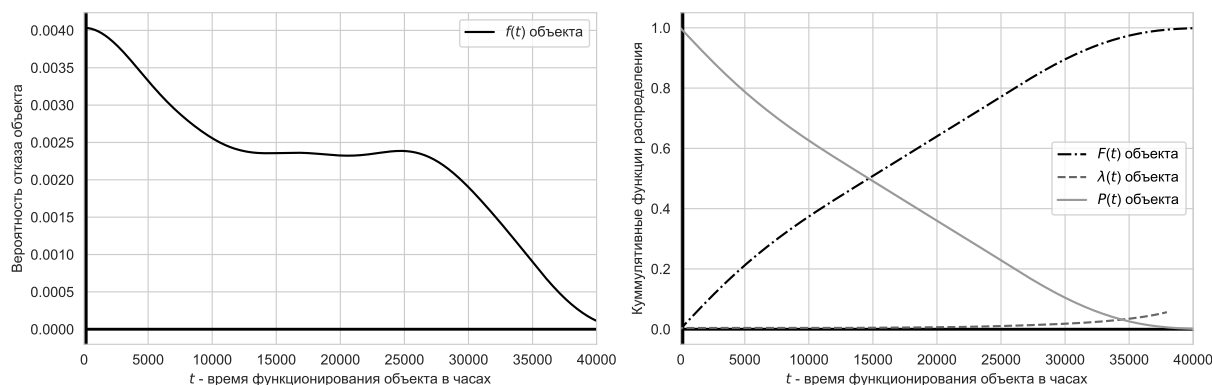


Рис. 1. Кривые, построенные по оценкам показателей надёжности объекта за период эксплуатации

Рассчитанные экспериментальные оценки показателей надёжности вычислительной системы необходимы для принятия управляющих решений при эксплуатации и сохранении работоспособности объекта.

ЛИТЕРАТУРА

1. Половко А. М., Гуров С. В., Основы теории надёжности. 2-е изд. СПб.: БХВ-Петербург, 2008.
2. Захаров Д. Н., Никулин В. С., Анализ методов статистической оценки эксплуатационной надёжности вычислительных комплексов // Научные технологии в космических исследованиях Земли. 2020. Т. 12. № 1. С. 64–69.
3. Чепурко В. А. Ядерная оценка параметра потока отказов. Диагностика и прогнозирование состояния сложных систем // Сб. научных трудов каф. АСУ НИЯУ МИФИ. 2004. № 15. С. 19–31.
4. Антонов А. В., Никулин М. С. Статистические модели в теории надёжности. М.: Абрис, 2012. 390 с.
5. Parzen E. On estimation of a probability density function and mode // Ann. Math. Statistics. 1962. V. 33. P. 1065–1076.
6. Silverman B. Density estimation for Statistics and Data Analysis. London, N.Y.: Chapman & Hall/CRC, 1986.
7. Никулин В. С. Методика подготовки данных для интеллектуального анализа надёжности вычислительных комплексов // Вестник СИБГУТИ. 2020. № 3(51). С. 26–37.