ИНДИВИДУАЛЬНОЕ ДОМАШНЕЕ ЗАДАНИЕ

по дисциплине «Безопасность, надёжность технических систем, методы оценки и управления риском»

(методические указания)

Целью индивидуального домашнего задания является выбор вероятностной модели надежности оборудования по эксплуатационным данным о наработках до отказа.

В соответствии с вариантом для анализа предоставлена выборка наработок до отказа оборудования объема N, а также набор моделей-кандидатов, задаваемых функциями распределений.

<u>Основные этапы работы:</u>

1. Определение точечных оценок параметров вероятностных моделей надежности (ВМН).
2. Определение интервальных оценок ВМН.
3. Сокращение набора моделей-кандидатов.
4. Верификация (выбор наилучшей ВМН).

Для решения задач 1-го и 2-го этапов предлагается использовать метод максимального правдоподобия и метод информационной матрицы Фишера, соответственно. Процедура сокращения набора моделей-кандидатов должна проводиться с использованием информационных критериев Ака́йке (AIC) и Шварца (BIC). Результатом 3-го этапа работы должен быть набор из трех моделей-кандидатов, обладающих наилучшими значениями информационных критериев.

Верификация результатов заключается в проверке простой гипотезы о принадлежности выборки данных некоторому известному распределению. Для выполнения 4-го этапа необходимо рассчитать значения следующих критериев согласия:

- критерий χ^2 Пирсона;
- критерий Колмогорова-Смирнова;
- критерий Крамера- фон Мизеса-Смирнова;
- критерий Андерсона-Дарлинга;

и сравнить эти значения с соответствующими критическими для заданного по варианту уровня значимости $\alpha_{знач}$.

Пояснения к выполнению ИДЗ

$$***$$

Для того, чтобы проверяемая гипотеза была простой, необходимо, чтобы оценка параметров и верификация проводились по разным выборкам. В связи с этим требуется разделить имеющуюся выборку объема N на две части с объемами $\frac{1}{3}N$ и $\frac{2}{3}N$. Ме́ньшая выборка будет использоваться для оценки параметров ВМН; бо́льшая – для расчета значений критериев согласия. Исходная выборка должна быть разделена случайным образом, однако следует стремиться к тому, чтобы выборочные средние и выборочные дисперсии обоих частей выборки имели близкие значения.

$$***$$

В качестве ВМН может использоваться конечная смесь распределений, функция распределения которой задается выражением

$$F_{mix}(x) = \sum_{i=1}^{K} w_i F_i(x, \boldsymbol{\Theta_i}),$$

где $F_i(x, \boldsymbol{\Theta_i})$ – функция распределения i-го компонента смеси; $\boldsymbol{\Theta_i}$ – вектор параметров i-го компонента смеси; $0 < w_i < 1$ – весовой коэффициент i-го компонента смеси. Причем

$$\sum_{i=1}^{K} w_i = 1.$$

Распределения, участвующие в смеси, могут быть однотипными (например, распределения Вейбулла) или различными. Количество параметров смеси равно сумме числа параметров всех составляющих ее компонентов и всех весовых коэффициентов. При этом один весовой коэффициент (например, K-й) можно исключить, заменив его на величину

$$w_K = 1 - \sum_{i=1}^{K-1} w_i.$$

Смеси распределений могут использоваться для выборок, в составе которых присутствуют несколько различающихся групп (типов) изделий. В этом случае коэффициент w_i указывает на долю изделий i-го типа в общей выборке.

***

В случаях, когда наработки до отказа в выборке обусловлены различными причинами отказов, возможно использовать так называемые ***модели конкурирующих рисков (МКР)*** в качестве ВМН. Функция распределения модели конкурирующих рисков задается выражением

$$F_{CRM}(x) = 1 - \prod_{i=1}^{K}\left(1 - F_i(x, \boldsymbol{\Theta_i})\right),$$

где $F_i(x, \boldsymbol{\Theta_i})$ – функция распределения наработки до отказа по i-й причине смеси; $\boldsymbol{\Theta_i}$ – вектор параметров i-го распределения.

Количество параметров МКР равно сумме числа параметров распределений всех причин отказов.

***

Информационные критерии позволяют оценить относительное качество моделей-кандидатов, то есть указать на модели наиболее (и наименее) подходящие для описания выборки. При этом, если модели-кандидаты с «худшими» значениями критериев можно смело исключить, «хорошие» значения критериев не дают основание считать модель адекватной: необходима дальнейшая верификация (критерии согласия).

В данной работе необходимо использовать информационные критерии для выбора трех наилучших моделей-кандидатов. В спорных случаях критерий BIC следует предпочесть критерию AIC. При равенстве значений критериев для разных моделей следует выбирать модели с ме́ньшим числом параметров.

***

При проверке соответствия выборки теоретическим распределениям с помощью критерия хи-квадрат Пирсона необходимо сравнивать гистограмму времен наблюдаемых отказов и «теоретическую» гистограмму, то есть такую, в которой высота столбцов пропорциональна вероятности появления события в соответствующих интервалах. При этом рекомендуется разделить носители распределений на $k \geq m + 2$ непересекающихся равновероятных интервалов, где m – максимальное число параметров в рассматриваемом списке моделей-кандидатов.

$$\operatorname{supp} F(x) = \bigcup_{i=1}^{k} [x_{i-1}; x_i) \,;$$

где $x_0 = 0$; $x_k = \infty$;

$$p_i = F(x_i) - F(x_{i-1}) = \frac{1}{k}.$$

В данном ИДЗ рекомендуется использовать значение $k = 8$ или $k = 10$.

<u>Ход работы:</u>

1. Разделить исходную выборку на две части: выборка V_1 (объем $N_1 = \frac{1}{3}N$) и выборка V_2 (объем $N_2 = \frac{2}{3}N$). Определить значения выборочных средних и несмещенных выборочных дисперсий (или среднеквадратичных отклонений) и построить гистограммы распределения для каждой части. Сохранить разделенные выборки в файлах Excel.
2. Определить точечные оценки параметров моделей-кандидатов методом максимального правдоподобия.
3. Определить интервальные оценки параметров моделей-кандидатов методом информационной матрицы Фишера для заданного по варианту уровня значимости $\alpha_{знач}$.
4. Определить значения информационных критериев AIC и BIC для всех моделей-кандидатов.
5. Представить результаты пп. 2-4 в табличном виде.
6. Выбрать три наилучшие модели с учетом значений информационных критериев.
7. Для каждой из оставшихся моделей-кандидатов определить границы равновероятных интервалов. Определить количество элементов выборки V_2, попавшее в каждый интервал. Записать в табличном виде гистограммы (матрицы) отдельно для каждой из трех моделей-кандидатов.
8. Рассчитать значение критерия χ^2 Пирсона для каждой модели-кандидата.
9. Рассчитать значение статистики Колмогорова S_K для каждой модели-кандидата.
10. Рассчитать значение статистики Крамера- фон Мизеса-Смирнова S_ω для каждой модели-кандидата.
11. Рассчитать значение статистики Андерсона-Дарлинга S_Ω для каждой модели-кандидата.
12. Определить критические значения всех рассчитанных критериев согласия для заданного по варианту уровня значимости $\alpha_{знач}$; сравнить рассчитанные значения с критическими.
13. Представить результаты пп. 8-12 в табличном виде.
14. Сделать вывод.

Варианты заданий

№	ФИО студента	$\alpha_{знач}$	Модели-кандидаты					
			2 параметра		3 параметра		4	5
1	Балахнин Илья Александрович	0,1	W	GR	ECEG	2R	WR	2W
2	Боровской Артём Романович	0,015	W	CEG	Kw-R	2E	WR	2W
3	Долгих Владимир Алексеевич	0,085	W	EE	GCRG	2R	CRM	2W
4	Жэнь Юйфэй	0,095	W	GE	ECRG	2R	WE	2W
5	Ивлев Алексей Александрович	0,025	W	CRG	GW	RE	CRM	2W
6	Киргефнер Михаил Сергеевич	0,01	W	ER	EW	2E	WE	2W
7	Лаврентьев Виктор	0,07	W	GR	CWG	2E	CRM	2W
8	Ларина Анастасия Валерьевна	0,03	W	CEG	Kw-E	2E	WR	2W
9	Лесных Глеб Игоревич	0,035	W	EE	GCEG	2E	WE	2W
10	Парфенов Павел Васильевич	0,075	W	GE	ECEG	2R	WR	2W
11	Стрельникова Виктория Анатольевна	0,09	W	CRG	Kw-R	RE	WE	2W
12	Суворов Данил Владиславович	0,06	W	ER	GCRG	RE	WR	2W
13	Шадиянов Ильшат Рашитович	0,065	W	GR	ECRG	2E	WR	2W
14	Щербашин Никита Геннадьевич	0,04	W	CEG	GW	2R	CRM	2W

Обозначения распределений

W – распределение Вейбулла;

CEG – комплементарное экспоненциально-геометрическое распределение;

CRG - комплементарное Рэлей-геометрическое распределение;

EE – экспоненциированное экспоненциальное распределение;

ER – экспоненциированное распределение Рэлея;

GE – обобщенное экспоненциальное распределение;

GR – обобщенное распределение Рэлея;

Kw-E – Кумарасвами-экспоненциальное распределение;

Kw-R – распределение Кумарасвами-Рэлея;

GCEG – обобщенное комплементарное экспоненциально-геометрическое распределение;

GCRG – обобщенное комплементарное Рэлей-геометрическое распределение;

GW – обобщенное распределение Вейбулла;

ECEG - экспоненциированное комплементарное экспоненциально-геометрическое распределение;

ECRG - экспоненциированное комплементарное Рэлей-геометрическое распределение;

EW – экспоненциированное распределение Вейбулла;

CWG – комплементарное Вейбулл-геометрическое распределение;

CRM – модель конкурирующих рисков на основе распределений Вейбулла

$$F_{CRM}(x) = 1 - e^{-\left(\frac{x}{\eta_1}\right)^{\beta_1}} \cdot e^{-\left(\frac{x}{\eta_2}\right)^{\beta_2}};$$

2E – двухкомпонентная смесь экспоненциальных распределений

$$F_{2E}(x) = w \cdot \left(1 - e^{-\lambda_1 x}\right) + (1 - w) \cdot \left(1 - e^{-\lambda_2 x}\right);$$

2R - двухкомпонентная смесь распределений Рэлея

$$F_{2R}(x) = w \cdot \left(1 - e^{-\frac{x^2}{2\sigma_1^2}}\right) + (1 - w) \cdot \left(1 - e^{-\frac{x^2}{2\sigma_2^2}}\right);$$

2W - двухкомпонентная смесь распределений Вейбулла

$$F_{2R}(x) = w \cdot \left(1 - e^{-\left(\frac{x}{\eta_1}\right)^{\beta_1}}\right) + (1 - w) \cdot \left(1 - e^{-\left(\frac{x}{\eta_2}\right)^{\beta_2}}\right);$$

RE – смесь распределения Рэлея и экспоненциального

$$F_{RE}(x) = w \cdot \left(1 - e^{-\frac{x^2}{2\sigma^2}}\right) + (1 - w) \cdot \left(1 - e^{-\lambda x}\right);$$

WE – смесь распределения Вейбулла и экспоненциального

$$F_{WE}(x) = w \cdot \left(1 - e^{-\left(\frac{x}{\eta}\right)^{\beta}}\right) + (1 - w) \cdot \left(1 - e^{-\lambda x}\right);$$

WR – смесь распределений Вейбулла и Рэлея

$$F_{WR}(x) = w \cdot \left(1 - e^{-\left(\frac{x}{\eta}\right)^{\beta}}\right) + (1 - w) \cdot \left(1 - e^{-\frac{x^2}{2\sigma^2}}\right).$$