УДК 621.019

Э.М. Фархадзаде, А.З. Мурадалиев, доктора техн. наук,
Ю.З. Фарзалиев канд. техн. наук
Азербайджанский научно-исследовательский
и проектно-изыскательский ин-т энергетики
(Азербайджанская Республика, Az1012 Баку, пр. Г. Зардаби, 94,
тел (+99412) 4316407, e-mail: fem1939@rambler.ru)

Распределение выборки непрерывной случайной величины

Приведены результаты исследования распределения статистик, характеризующих степень рассеяния выборки случайных величин, и эффективности разработанных на их основе критериев. При имитационном моделировании в качестве случайных величин введены численные значения вертикального отклонения статистической функции распределения совокупности многомерных случайных величин и статистической функции распределения выборки из этой совокупности.

Наведено результати дослідження розподілу статистик, які визначають ступінь розсіяння вибірки випадкових величин та ефективності розроблених на їх основі критеріїв. При імітаційному моделюванні в якості випадкових величин введено числові значення вертикального відхилення статистичної функції розподілу сукупності багатовимірних випадкових величин і статистичної функції розподілу вибірки із цієї сукупності.

К л ю ч е в ы е с л о в а: выборка, классификация степени рассеяния, критерий проверки гипотез.

Исходные предпосылки. К фундаментальным задачам обеспечения надежности электроустановок относится разработка системы автоматизированного мониторинга технического состояния и прогнозирования остаточного ресурса основного энергетического и электротехнического оборудования и устройств [1]. Решение этой задачи предусматривает возможность достоверной количественной оценки надежности работы, которая по статистическим данным эксплуатации традиционно проводится посредством классификации этих данных по некоторым разновидностям признаков. Такая классификация необходима для перехода при вычислениях от усредненных моделей надежности и оценок их показателей к неслучайно различающимся моделям и показателям надежности отдельных групп оборудования (кластеров). К моделям надежности относятся статистические функции распределения (СФР) таких случайных величин, как длительность безотказной работы, продолжительность простоя в аварийном ремонте, в резерве и др. Если показатель надежности оборудования используется для количественной оценки надежности электроустановок, например схем распределительных устройств, оценки резерва мощности и других, то модели надежности необходимы при имитационном моделировании процесса функционирования электроустановок и расчете гарантированных значений показателей надежности. Например, СФР являются исходными данными при имитационном моделировании совместных нерабочих состояний энергоблоков.

В работах [2—6] показано, что достоверность расчетов достигается при учете следующих факторов.

1. Предположение о неизменном соответствии статистических данных эксплуатации выборке из некоторой генеральной совокупности — ошибочно. Закон распределения данных случайным образом изменяется по мере их накопления. Само изменение обусловлено изменением воздействия соответствующих разновидностей признаков [2].

2. Именно поэтому достоверность решения при применении к статистическим данным эксплуатации классических методов теории проверки статистических гипотез, ориентированных, как правило, на большие выборки и нормальный закон распределения, не всегда удовлетворяет предъявляемым требованиям [3].

3. Такие исходные данные в прикладной математике называются конечной совокупностью многомерных данных. Выборки из этих данных могут быть как представительными, так и непредставительными [4].

4. Неконтролируемая классификация полученных статистических данных, т.е. без контроля на представительность, нецелесообразна, так как в большинстве случаев приводит к снижению достоверности оценок показателей надежности и увеличению риска ошибочного решения. Контроль целесообразности классификации позволяет исключить классификацию данных по незначимым признакам [5].

5. Различие возможных методов контроля целесообразности классификации статистических данных сводится к различию соответствующих критериев. Принято использовать критерий, мощность которого при заданной ошибке первого рода — наибольшая. Полное предпочтение одного критерия другим является ошибочным [5].

В предлагаемых методах оценки целесообразности классификации статистических данных об отказах и восстановлениях оборудования и устройств электроустановок при сопоставлении СФР конечной совокупности многомерных данных $F_{\Sigma}^{*}(X)$ и СФР выборки v из этой совокупности $F_{V}^{*}(X)$ в качестве статистик критериев оценки целесообразности классификации приняты статистики, характеризующие одно из основных свойств случайных величин — степень их рассеяния.

Степень рассеяния чаще всего характеризуется средним квадратическим отклонением $\sigma_{V}^{*}(X)$, реже используется коэффициент вариации $V_{V}^{*}(X)=\sigma_{V}^{*}(X)/M_{V}^{*}(X)$, где $M_{V}^{*}(X)$ — среднее статистическое случайных величин выборки, и еще реже — размах рассеяния $L_{V}^{*}(X)=(|X_{\max}|-|X_{\min}|)$.

Результаты экспериментальных исследований. Выполним оценку критических значений рассеяния вертикального отклонения СФР $F_{\Sigma}^{*}(X)$ и СФР выборки из этой совокупности $F_{V}^{*}(X)$. Величину вертикального отклонения определим по формуле

$$\Delta_i=\left|\frac{i}{n_V}-F_{\Sigma}^{*}(X_i)\right|,$$

где n_V — число случайных величин выборки; $i=1, n_V$. Поскольку величина Δ не зависит ни от типа закона распределения $F_{\Sigma}(X)$, ни от вида ($F_{\Sigma}(X)$ или $F_{\Sigma}^{*}(X)$), для упрощения расчетов примем СФР $F_{\Sigma}^{*}(X)$ соответствующей равномерному закону распределения в интервале [0,1]. При этом величину Δ_i вычислим по формуле

$$\Delta_i=\left|\frac{i}{n_V}-\xi_i\right|,\ i=1, n_V, \tag{1}$$

где ξ — случайная величина с равномерным распределением.

А л г о р и т м расчета случайных величин выборки $\{\Delta\}_{n_V}$ для ряда n_V следующий:

1. Моделируем n_V случайных величин ξ с равномерным распределением в интервале [0, 1].

2. Случайные числа $\{\xi\}_{n_V}$ размещаем в порядке возрастания.

3. Строим СФР выборки $F_{V}^{*}(X)$ и вычисляем n_V значений Δ по формуле (1).

Далее на основе выборки $\{\Delta\}_{n_V}$ проводим следующие расчеты:

А. Р а с ч е т СФР $F^{*}[\sigma_{V,i}^{*}(\Delta)]$:

4**А**. Вычисляется реализация $\sigma_{V,i}^{*}(\Delta)$ по формуле

$$\sigma_{V,i}^{*}(\Delta)=\sqrt{\frac{\sum_{j=1}^{n_V}[\Delta_{j,i}-M_{V,i}^{*}(\Delta)]^2}{(n_V-1)}}, \tag{2}$$

где $M^*_{V,i}(\Delta)$ — оценка среднего значения выборки из n_V случайных величин Δ, $M^*_{V,j} = n_V^{-1} \sum_{i=1}^{n_V} \Delta_{i,j}$.

5. Цикл вычислений 1—4 **А** выполняется N раз, т.е. N — число итераций.

6. N реализаций $\sigma^*_V(\Delta)$ размещается в порядке возрастания и j-му значению $\sigma^*_{V,j}(\Delta)$, $j = 1, N$, сопоставляется вероятность $F^*[\sigma^*_{V,i}(\Delta)] = j / N$.

7. Цикл расчетов 1—6 повторяется для очередного дискретного значения n_V.

В табл. 1 приведены результаты расчета квантилей распределения статистики $\sigma^*_V(\Delta)$ для ряда n_V и вероятностей $F^*[\sigma^*_V(\Delta)]$ с шагом 0,05 при

Таблица 1

N	$F^*[\sigma^*_V(\Delta)]$	Квантили распределения статистики $\sigma^*_V(\Delta)$ для ряда n_V								
		2	4	7	11	22	29	40	90	150
1	0,05	0,012	0,038	0,037	0,031	0,023	0,020	0,017	0,012	0,009
2	0,1	0,025	0,050	0,044	0,036	0,026	0,023	0,019	0,013	0,010
3	0,15	0,037	0,059	0,048	0,039	0,028	0,025	0,021	0,014	0,011
4	0,2	0,049	0,066	0,053	0,042	0,030	0,026	0,023	0,015	0,012
5	0,25	0,062	0,073	0,056	0,045	0,032	0,028	0,024	0,016	0,012
6	0,3	0,075	0,079	0,060	0,048	0,034	0,030	0,025	0,017	0,013
7	0,35	0,089	0,085	0,064	0,051	0,036	0,031	0,027	0,018	0,014
8	0,4	0,103	0,091	0,067	0,053	0,038	0,033	0,028	0,019	0,014
9	0,45	0,118	0,096	0,071	0,056	0,040	0,034	0,029	0,019	0,015
10	0,5	0,130	0,102	0,075	0,059	0,041	0,036	0,031	0,020	0,016
11	0,55	0,150	0,107	0,079	0,062	0,044	0,038	0,032	0,021	0,017
12	0,6	0,168	0,114	0,084	0,065	0,046	0,040	0,034	0,022	0,017
13	0,65	0,186	0,120	0,088	0,069	0,048	0,042	0,036	0,024	0,018
14	0,7	0,204	0,128	0,094	0,073	0,051	0,044	0,038	0,025	0,019
15	0,75	0,224	0,137	0,100	0,078	0,054	0,047	0,040	0,026	0,021
16	0,8	0,244	0,147	0,107	0,083	0,058	0,050	0,043	0,028	0,022
17	0,85	0,267	0,159	0,116	0,090	0,062	0,054	0,046	0,030	0,024
18	0,9	0,292	0,175	0,127	0,098	0,068	0,059	0,051	0,033	0,026
19	0,95	0,320	0,199	0,145	0,112	0,078	0,068	0,058	0,038	0,030
20	0,99	0,347	0,244	0,178	0,140	0,097	0,084	0,072	0,048	0,037

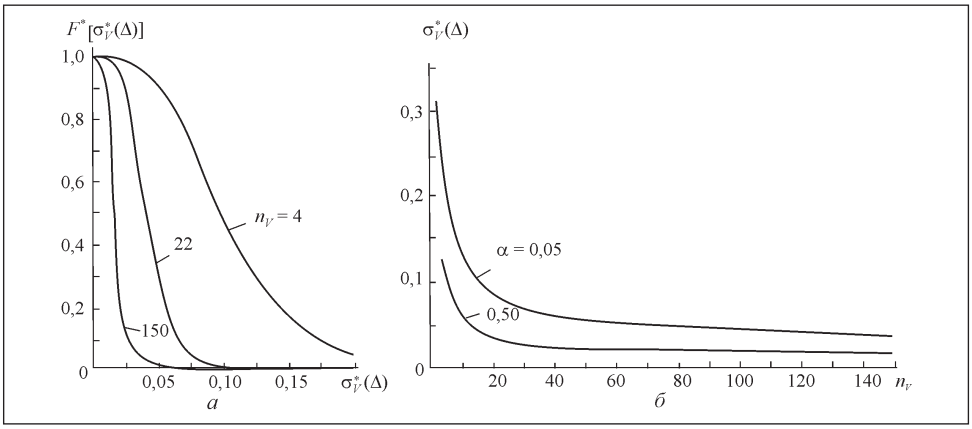

Рис. 1. Кривые распределения вероятностной $F^*[\sigma_V^*(\Delta)]$ (*а*) и изменения критических значений статистики $\sigma_V^*(\Delta)$ (*б*) в зависимости от n_V

N = 25000. Полученные результаты позволяют установить критические значения статистики $\sigma^*_{V,\alpha}(\Delta)$ для заданного уровня значимости $\alpha =$ $= R^*[\sigma^*_{V,\alpha}(\Delta)] = [1 - F^*[\sigma^*_{V,\alpha}(\Delta)]]$.

На рис. 1, *а*, приведены графики распределения вероятностей $F^*[\sigma_V^*(\Delta)]$ для ряда n_V, а на рис. 1, *б*, — графики закономерности изменения критических значений статистики $\sigma^*_{V,\alpha}(\Delta)$ при α = 0,05 и α = 0,50 в функции числа случайных величин выборки n_V, рассчитанные по данным табл. 1.

Интерес представляет возможность аналитической оценки критических значений $\sigma^*_{V,\alpha}(\Delta)$ в зависимости от n_V. Установлено, что наименьшая погрешность (наибольший коэффициент детерминации R^2) наблюдается для степенной функции $\sigma^*_{V,\alpha}(\Delta) = An_V^{-b}$.

Результаты расчетов постоянных коэффициентов регрессии A и b и коэффициента детерминации R^2 приведены в табл. 2.

Пример 1. Пусть α = 0,5, n_V = 11; A = 0,197, b = 0,5. При этом $\sigma^*_{V;0,5}(\Delta) =$ $= 0{,}197/\sqrt{11} = 0{,}059$. Аналогичный результат представлен в табл. 1.

Пример 2. Пусть α = 0,05, n_V = 4; A = 0,428, b =0,54. При этом $\sigma^*_{V;0,05}(\Delta) = 0{,}428/n^{0,54} = 0{,}428/2{,}11 = 0{,}202$. По данным табл. 1 $\sigma^*_{V;0,05}(\Delta) =$ $= 0{,}199$. Расхождение составляет 1,5%.

В. Расчет СФР $F^*[V_V^*(\Delta)]$.

4**В**. Вычисляется реализация $V_V^*(\Delta)$ по формуле $V_V^*(\Delta) = \sigma_V^*(\Delta)/M_V^*(\Delta)$.

5. Цикл 1—4**В** выполняется N раз.

6. Вычисляется $F^*[V_V^*(\Delta)]$.

7. Цикл 1—6 повторяется для очередного дискретного значения n_V.

Таблица 2

i	$\alpha_i = R_i^*[\sigma_V^*(\Delta)]$	A_i	b_i	R_i^2
1	0,05	0,428	0,54	0,997
2	0,1	0,385	0,55	0,996
3	0,2	0,322	0,545	0,996
4	0,3	0,276	0,54	0,998
5	0,4	0,237	0,53	0,999
6	0,5	0,197	0,50	0,999
7	0,6	0,163	0,48	0,995
8	0,7	0,129	0,46	0,991
9	0,8	0,094	0,43	0,996
10	0,9	0,061	0,40	0,992

Таблица 3

N	$F^*[V_V^*(\Delta)]$	Квантили распределения статистики $V_V^*(\Delta)$ для ряда n_V										
		2	3	4	5	7	11	22	29	40	90	150
1	0,05	0,055	0,203	0,289	0,336	0,380	0,426	0,463	0,473	0,479	0,488	0,490
2	0,1	0,109	0,281	0,353	0,392	0,435	0,476	0,510	0,518	0,524	0,533	0,536
3	0,15	0,167	0,334	0,396	0,435	0,475	0,510	0,542	0,548	0,556	0,562	0,565
4	0,2	0,221	0,378	0,435	0,473	0,507	0,539	0,567	0,571	0,580	0,586	0,588
5	0,25	0,276	0,417	0,471	0,507	0,535	0,564	0,588	0,593	0,600	0,606	0,609
6	0,3	0,326	0,452	0,506	0,538	0,560	0,587	0,609	0,611	0,619	0,625	0,625
7	0,35	0,374	0,488	0,539	0,565	0,583	0,608	0,625	0,629	0,636	0,640	0,643
8	0,4	0,420	0,524	0,569	0,590	0,605	0,627	0,643	0,645	0,652	0,655	0,657
9	0,45	0,464	0,562	0,601	0,615	0,638	0,646	0,659	0,660	0,667	0,667	0,667
10	0,5	0,510	0,601	0,632	0,640	0,649	0,667	0,674	0,676	0,682	0,684	0,688
11	0,55	0,563	0,641	0,661	0,667	0,672	0,683	0,691	0,692	0,698	0,700	0,700
12	0,6	0,622	0,681	0,689	0,692	0,694	0,703	0,707	0,708	0,714	0,714	0,714
13	0,65	0,686	0,723	0,719	0,720	0,718	0,723	0,725	0,725	0,730	0,731	0,733
14	0,7	0,758	0,765	0,750	0,749	0,741	0,744	0,744	0,743	0,750	0,750	0,750
15	0,75	0,842	0,809	0,786	0,780	0,769	0,767	0,765	0,762	0,767	0,765	0,767
16	0,8	0,939	0,852	0,826	0,816	0,800	0,792	0,787	0,784	0,787	0,786	0,786
17	0,85	1,047	0,902	0,876	0,860	0,837	0,824	0,812	0,810	0,812	0,810	0,812
18	0,9	1,167	0,972	0,943	0,915	0,885	0,864	0,849	0,843	0,844	0,842	0,846
19	0,95	1,284	1,114	1,048	1,000	0,959	0,925	0,902	0,895	0,895	0,889	0,895
20	0,99	1,388	1,410	1,267	1,196	1,101	1,043	1,000	1,000	1,000	1,000	1,000

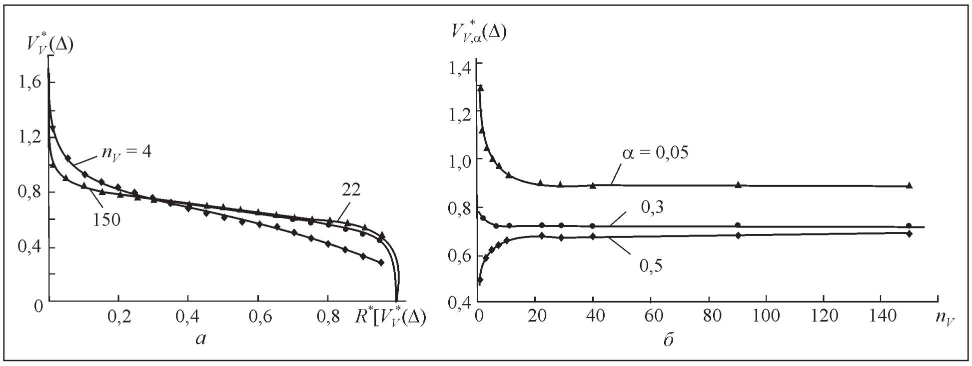

Рис. 2. Графики распределений $R^*[V_V^*(\Delta)=1-F^*[V_V^*(\Delta)]$ (*а*) и критических значений коэффициента вариации $V_{V,\alpha}^*(\Delta)$ в зависимости от n_V (*б*)

Результаты расчетов распределения статистики $V_V^*(\Delta)$ для ряда n_V и вероятностей $F^*[V_V^*(\Delta)]$ приведены в табл. 3 и на рис. 2, *а*, откуда следует, что при $n_V > 20$ расхождение между квантилями СФР $F^*[V_V^*(\Delta)]$ не превышает нескольких процентов. Кривые, представленные на рис. 2, *б*, свидетельствуют о том, что при $n_V > 20$ значение $V_{V,\alpha}^*(\Delta)$ практически не зависит от n_V. Как видим, при уменьшении значений n_V и α коэффициент вариации $V_{V,\alpha}^*(\Delta)$ возрастает, а при уменьшении n_V и увеличении α — уменьшается.

Пример 3. Пусть задана выборка ($n_V = 4$) случайных величин ξ с равномерным распределением ε_{ip} в интервале [0, 1]. Известно, что соответствие ξ равномерному закону не означает соответствие этому закону случайной выборки. Оценим вероятность соответствия СФР $F^*(\xi)$ равномерному закону. Для этого вычислим экспериментальные значения $V_{V\text{э}}^*(\Delta)$ и, используя табл. 3, оценим вероятность $F^*[V_{V\text{э}}^*(\Delta)]$.

Результаты расчетов приведены в табл. 4, из которой видно, что ошибка первого рода составляет $\alpha < 0{,}01$. Таким образом, соответствие распределения равномерному закону в интервале [0,1] маловероятно.

С. Расчет СФР $F^*[L_V^*(\Delta)]$.

4**С**. Вычисляется реализация $L_V^*(\Delta)$ по формуле $L_V^*(\Delta)=(\Delta_{\max}-\Delta_{\min})$, где $\Delta_{\max}=\max\{|\Delta_1|,|\Delta_2|,...,|\Delta_{n_V}|\}$, а $\Delta_{\min}=\min\{|\Delta_1|,|\Delta_2|,...,|\Delta_{n_V}|\}$.

5. Цикл 1—4**С** выполняется N раз.

6. Вычисляется $F^*[L_V^*(\Delta)]$.

7. Цикл 1—6 повторяется для заданных значений n_V.

Результаты моделирования $L_V^*(\Delta)$ для вероятностей $F^*[L_V^*(\Delta)]$ и случайных величин n_V приведены в табл. 5, а на рис. 3 представлены закономерности распределения $F^*[L_V^*(\Delta)]$ и критических значений $L_{V,\alpha}^*(\Delta)$. Как

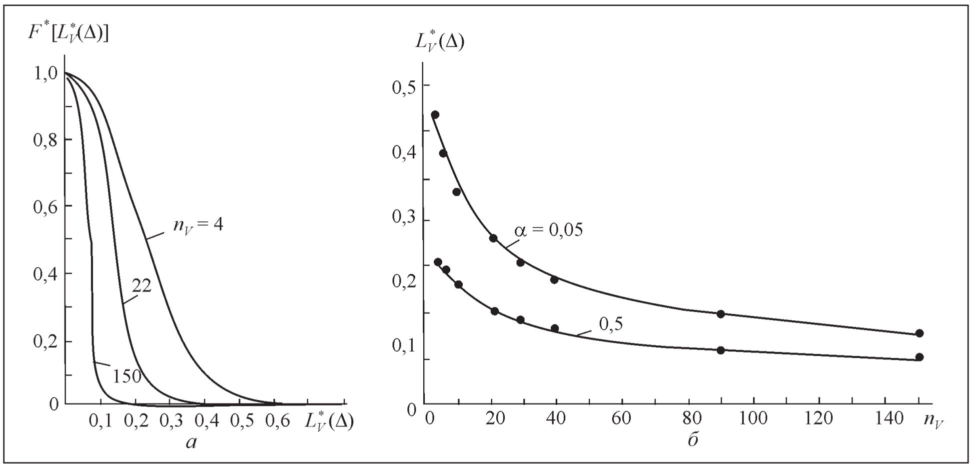

Рис. 3. Графики распределений $R^*[L^*_V(\Delta)]=1-F^*[L^*_V(\Delta)]$ (*а*) и критических значений $L^*_{V,\alpha}(\Delta)$ в зависимости от n_V (*б*)

следует из рис. 3, при уменьшении значения n_V асимметрия распределения случайных значений $L^*_V(\Delta)$ возрастает.

Пример 4. По данным табл. 4 определим реализацию размаха рассеяния $L^*_V(\Delta)$. Поскольку $\Delta_{\max}$ = 0,311, а $\Delta_{\min}$ = 0,013, $L^*_{V\text{э}}(\Delta)$ = 0,298, $R^*[L^*_{V\text{э}}(\Delta)]=\alpha=0{,}27>>0{,}05$. Следовательно, выборка представительна.

Заметим, что при проведении расчетов для повторной выборки случайных чисел с равномерным распределением в интервале [0,1] (0,166; 0,211; 0,168; 0,327) результат проверки предположения о репрезентативности выборки будет противоположным, а именно: $V^*_{V\text{э}}(\Delta)=0{,}632$ и $R^*[V^*_{V\text{э}}(\Delta)]=0{,}47$, $L^*_{V\text{э}}(\Delta)=0{,}589$ и $R^*[L^*_{V\text{э}}(\Delta)]=0{,}005$, $\sigma^*_{V\text{э}}(\Delta)=0{,}257$ и $R^*[\sigma^*_{V\text{э}}(\Delta)]=0{,}008$.

Таблица 4

i	ξ_i		$F^*(\xi_{i\text{р}})$	Δ_i	$[\Delta_i - M^*(\Delta)]$	$[\Delta_i - M^*(\Delta)]^2$	Примечание
	Исходное значение	$\xi_{i\text{р}}$					
1	0,1885	0,1748	0,25	0,075	0,03	0,0009	$M^*_V(\Delta)=0{,}105$
2	0,9765	0,1885	0,50	0,311	0,206	0,0424	$\sigma^*_V(\Delta)=0{,}14$
3	0,7368	0,7368	0,75	0,013	0,092	0,0084	$V^*_V(\Delta)=1{,}13$
4	0,1748	0,9765	1,00	0,023	0,082	0,0067	
Σ	—	—	—	0,422	—	0,0584	$R^*[V^*_{V\text{э}}(\Delta)]<0{,}01$

Сопоставление эффективности критериев проверки гипотезы о целесообразности классификации данных сводится к выбору одного из трех рассмотренных выше критериев для конкретной выборки. Предлагается выбрать критерий, статистике разброса которого соответствует минимальная ошибка первого рода.

А л г о р и т м статистических исследований достоверности этого предложения следующий:

1. Моделирование n_V случайных чисел ξ с равномерным распределением $F_\Sigma(\xi)$ в интервале [0,1].

2. Построение СФР $F_V^*(\xi)$.

3. Вычисление n_V абсолютных значений вертикальных расхождений Δ_i, $i = 1, n_V$, $F_\Sigma(\xi)$ и $F_V^*(\xi)$ по формуле (1).

4. Вычисление статистики разброса $\sigma_V^*(\Delta)$, $L_V^*(\Delta)$ и $V_V^*(\Delta)$ для n_V значений Δ, т.е. с использованием метода общих случайных чисел.

Таблица 5

N	$F^*[L_V^*(\Delta)]$	Квантили распределения статистики $L_V^*(\Delta)$ для ряда n_V								
		2	4	7	11	22	29	40	90	150
1	0,05	0,016	0,084	0,099	0,095	0,08	0,073	0,065	0,048	0,038
2	0,1	0,033	0,11	0,117	0,111	0,091	0,083	0,073	0,053	0,042
3	0,15	0,05	0,129	0,131	0,123	0,099	0,09	0,079	0,057	0,045
4	0,2	0,069	0,145	0,142	0,132	0,106	0,096	0,085	0,06	0,048
5	0,25	0,087	0,160	0,153	0,141	0,113	0,102	0,09	0,064	0,051
6	0,3	0,106	0,174	0,165	0,15	0,119	0,107	0,095	0,067	0,053
7	0,35	0,127	0,187	0,176	0,159	0,125	0,112	0,099	0,07	0,056
8	0,4	0,147	0,2	0,186	0,167	0,131	0,118	0,104	0,073	0,058
9	0,45	0,167	0,213	0,197	0,175	0,138	0,124	0,108	0,076	0,060
10	0,5	0,189	0,225	0,208	0,184	0,144	0,130	0,113	0,079	0,063
11	0,55	0,213	0,237	0,22	0,194	0,151	0,136	0,118	0,083	0,066
12	0,6	0,237	0,249	0,232	0,204	0,158	0,142	0,123	0,086	0,068
13	0,65	0,261	0,267	0,245	0,214	0,166	0,149	0,129	0,09	0,072
14	0,7	0,287	0,287	0,259	0,225	0,174	0,156	0,136	0,094	0,075
15	0,75	0,314	0,309	0,275	0,239	0,184	0,165	0,143	0,099	0,079
16	0,8	0,345	0,333	0,294	0,254	0,195	0,174	0,151	0,105	0,083
17	0,85	0,378	0,362	0,316	0,273	0,208	0,186	0,161	0,111	0,088
18	0,9	0,414	0,397	0,346	0,297	0,225	0,202	0,174	0,121	0,096
19	0,95	0,455	0,450	0,393	0,334	0,255	0,226	0,195	0,134	0,107
20	0,99	0,489	0,557	0,482	0,405	0,350	0,273	0,237	0,163	0,127

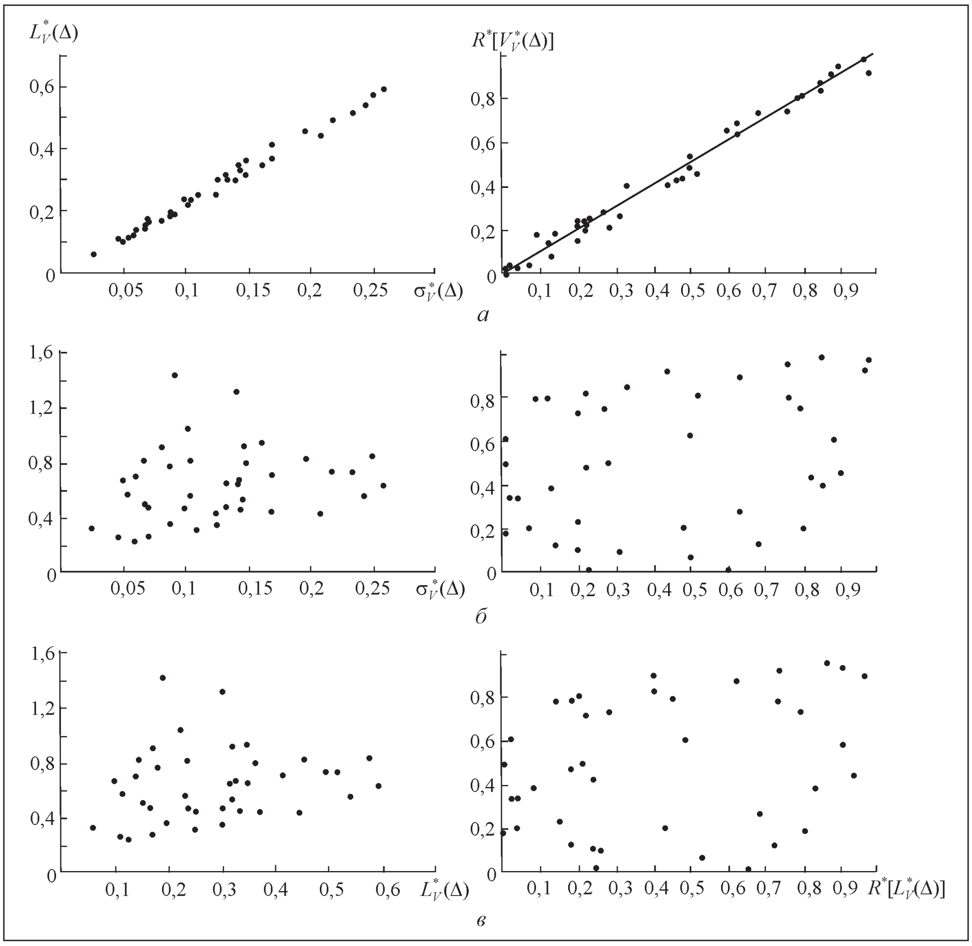

Рис. 4. Взаимосвязь оценок коэффициентов корреляции: *а* — $\sigma_V^*(\Delta)$ и $L_V^*(\Delta)$, $R^*[\sigma_V^*(\Delta)]$ и $R^*[L_V^*(\Delta)]$; *б* — $\sigma_V^*(\Delta)$ и $V_V^*(\Delta)$, $R^*[\sigma_V^*(\Delta)]$ и $R^*[V_V^*(\Delta)]$; *в* — $L_V^*(\Delta)$ и $V_V^*(\Delta)$, $R^*[L_V^*(\Delta)]$ и $R^*[V_V^*(\Delta)]$

5. Определение соответствующей вероятности $F^*[\sigma_V^*(\Delta)]$, $F^*[L_V^*(\Delta)]$ и $F^*[V_V^*(\Delta)]$ для каждой из этих статистик.

6. Пункты 1—5 повторяются N раз.

7. Определение числа выборок, для которых каждая статистика удовлетворяла условиям $\sigma_V^*(\Delta) \geq \sigma_{V\alpha}^*(\Delta)$, $L_V^*(\Delta) \geq L_{V\alpha}^*(\Delta)$ и $V_V^*(\Delta) \geq V_{V\alpha}^*(\Delta)$.

В результате вычислений установлено следующее.

1. Относительное число выборок α_V^*, для которых вероятность $R^*[\ldots]$ оказалась меньше, чем принятый для каждого из трех критериев уровень значимости α.

2. Величина α^*_V распределена между статистиками $\sigma^*_V(\Delta)$, $L^*_V(\Delta)$ и $V^*_V(\Delta)$ неравномерно. Приблизительно половина общего числа непредставительных выборок установлена по критериям на основе статистик $\sigma^*_V(\Delta)$ и $L^*_V(\Delta)$, а вторая половина — на основе статистики $V^*_V(\Delta)$. При этом их относительные значения достаточно близки к значению α.

Для выяснения причин указанных особенностей были построены корреляционные зависимости (рис. 4), значения коэффициентов корреляции которых приведены в табл. 6.

Таким образом, в результате моделирования установлено следующее.

Неравномерный характер распределения числа непредставительных выборок между рассматриваемыми критериями обусловлен тесной взаимосвязью численных значений статистик $\sigma^*_V(\Delta)$ и $L^*_V(\Delta)$ ($L^*_V(\Delta)=2{,}2\sigma^*_V(\Delta)$) и независимостью статистик $\sigma^*_V(\Delta)$ и $L^*_V(\Delta)$ от статистики $V^*_V(\Delta)$. Следовательно, необходим выбор между статистиками $\sigma^*_V(\Delta)$ и $L^*_V(\Delta)$.

Проведенные экспериментальные исследования показали, что если выборки $\{\xi\}_{n_V}$ представительны, а значения α_{L_V} и α_{σ_V} меньше заданного критического значения $\alpha_{\text{к}}$, то ошибка второго рода β_{σ_V} всегда оказывается меньше ошибки второго рода β_{L_V}, т.е. мощность критерия, основанного на статистике σ_V, выше. При этом число независимых статистик, характеризующих степень рассеяния реализаций выборки $\{\Delta\}_{n_V}$, равно двум, а именно: $\sigma^*_V(\Delta)$ и $V^*_V(\Delta)$. Вероятность того, что реализации статистик $\sigma^*_V(\Delta)$ и $V^*_V(\Delta)$ не превысят критических значений с ошибкой первого рода, имеет вид

$$\alpha_\Sigma=\{1-R^*[\sigma^*_V(\Delta)]\,R^*[V^*_V(\Delta)]\}=1-(1-\alpha_{\sigma_V})(1-\alpha_{V_V})\cong\alpha_{\sigma_V}+\alpha_{V_V}).$$

Поскольку уровень значимости критериев принят одинаковым, $\alpha_{\sigma_V}=$ $=\alpha_{V_V}$, ошибка первого рода α_V, будет в два раза превышать значение α, т.е. $\alpha_{\sigma_V}=\alpha_{V_V}\cong 0{,}5\alpha_\Sigma$.

Таблица 6

N	Зависимость	Коэффициент корреляции
1	$L^*_V(\Delta)=f[\sigma^*_V(\Delta)]$	0,950
2	$R^*[L^*_V(\Delta)]=f\{R^*[\sigma^*_V(\Delta)]\}$	0,995
3	$V^*_V(\Delta)=f[\sigma^*_V(\Delta)]$	0,196
4	$R^*[V^*_V(\Delta)]=f\{R^*[\sigma^*_V(\Delta)]\}$	0,278
5	$V^*_V(\Delta)=f[L^*_V(\Delta)]$	0,176
6	$R^*[V^*_V(\Delta)]=f\{R^*[L^*_V(\Delta)]\}$	0,273

Пример 5. По данным N = 40 реализаций выборок при n_V = 4 непредставительность выборок для $\alpha_{L_V} = \alpha_{V_V} = 0{,}20$ проявилась в семи случаях для статистики $L_V^*(\Delta)$ и в восьми случаях для статистики $V_V^*(\Delta)$. При этом ошибка первого рода составила $\alpha_V = 1 - 0{,}8 \cdot 0{,}8 = 0{,}36$, а по данным эксперимента — 0,38=15/40. Следует заметить, что эти цифры хорошо согласуются с поправкой, полученной в [6].

Пример 5 не только подтверждает необходимость проверки гипотезы о целесообразности классификации статистических многомерных данных несколькими критериями [7], но и обосновывает ее. Основанием этого является характер и независимость отдельных статистик. Кроме того, проверка гипотезы о целесообразности классификации статистических данных лишь одним или неполным числом критериев приводит к тому, что непредставительная выборка по результатам проверки гипотезы считается представительной. Например, если проверить гипотезу критериями, в основе которых находятся статистики $V_V^*(\Delta)$ и $\sigma_V^*(\Delta)$, но не учесть статистику $M_V^*(\Delta)$, то при характеристиках разброса, не превышающих критических значений для представительных выборок, $\sigma_{V\text{к}}(\Delta)$ и $V_{V\text{к}}(\Delta)$, выборка $\{\Delta\}_{n_V}$ может иметь оценку среднего значения $M_V^*(\Delta) >> M_{V\text{к}}^*(\Delta)$, т.е. быть на самом деле непредставительной.

Выводы

1. Рассеяние выборки случайных величин X в наибольшей степени характеризуется коэффициентом вариации $V_V^*(\Delta)$ и среднеквадратическим отклонением $\sigma_V^*(\Delta)$.

2. Коэффициент корреляции взаимосвязи вероятности оценки среднего квадратического отклонения $\sigma_V^*(\Delta)$ и оценки размаха $L_V^*(\Delta)$ для n_V = 4 равен 0,995. С увеличением значения n_V коэффициент корреляции не уменьшается.

3. Проверка предположения о представительности выборки при классификации многомерных данных с привлечением нескольких критериев требует уточнения величины уровня значимости критериев, что может быть выполнено с помощью поправки Банферрони [6].

СПИСОК ЛИТЕРАТУРЫ

1. *Воропай Н.И., Ковалев Г.Ф.* Об основных положениях концепции обеспечения надежности в электроэнергетике//Энергетическая политика.— 2010.— Вып. 3. — С. 7—10.
2. *Фархадзаде Э.М., Мурадалиев А.З., Фарзалиев Ю.З.* Повышение точности оценки показателей индивидуальной надежности энергоблоков// Электрон. моделирование. — 2007. — **29**, № 5. — С. 75—83.

3. *Рябинин И.А.* Основы теории и расчета надежности судовых электроэнергетических систем. 2-е изд. — Ленинград: Судостроение, 1971. — 456 с.
4. *Кендалл М., Стьарт А.* Статистические выводы и связи. — М. : Наука, 1973. — 900 с.
5. *Farhadzadeh E.M., Farzaliyev Y.Z., Muradaliyev A.Z.* Principles of classification statistical data about reliability of the electric equipment of power supply systems//Reliability: Theory&applications. — 2013. — Vol. 8, No. 3(30). — P. 56—74.
6. *Bonferroni C.E.* Il calcolo delle assicurazioni su gruppi di teste//In Studi in Onore del Professore Salvatore Ortu Carboni. — Italy, Rome. — 1935. — P. 13—60.
7. *Гнеденко Б.В., Беляев Ю.К., Соловьев А.Д.* Математические методы в теории надежности. — М. : Наука, 1965.

E.M. Farhadzadeh, A.Z. Muradaliyev, Y.Z. Farzaliyev

DISTRIBUTION OF SAMPLE OF A CONTINUOUS RANDOM VARIABLE

The research results have been presented for distribution of statistics, describing a degree of dispersion of the sample of random variables, and efficiency of the criteria developed on their basis. Numerical values of a vertical deviation of statistical function of distribution of a set of multivariate random variables and statistical function of distribution of the sample of this set are entered as random variables under imitating modeling.

K e y w o r d s: sample, classification of a degree of dispersion, criterion check of hypotheses.

REFERENCES

1. Voropay, N.I. and Kovalyov, G.F. (2010), “About substantive provisions of the concept of maintenance of reliability in electric power industry”, *Energeticheskaya politika*, Iss. 3, pp. 7-10.
2. Farhadzadeh, E.M., Muradaliyev, A.Z. and Farzaliyev, Yu.Z. (2007), “Increase of accuracy of an estimation of parameters of individual reliability of power units”, *Elektronnoe modelirovanie*, Vol. 29, no. 5, pp. 75-83.
3. Ryabinin, I.A. (1971), *Osnovy teorii rascheta nadyozhnosti sudovykh elektroenergeticheskikh* system [Bases of the theory of reliability calculation of the vessel electric power systems, 2nd ed.], Sudostroenie, Leningrad, Russia.
4. Kendall, M. and Stuart, A. (1973), *Statisticheskie vyvody i svyazi* [Statistical conclusions and communications], Nauka, Moscow, Russia.
5. Farhadzadeh, E.M., Farzaliyev, Yu.Z. and Muradaliyev, A.Z. (2013), “Principles of classification of statistical data about reliability of the electric equipment of power supply systems”, *Reliability: Theory&Applications*, Vol. 8, no. 3 (30), pp. 56-74.
6. Bonferroni, C.E. (1935), “Il calcolo delle assicurazioni su gruppi di teste”, *In Studi in Onore del Professore Salvatore Ortu Carboni*, Italy, Rome, pp. 13-60.
7. Gnedenko, B.V., Beljaev, Yu.K. and Solovyov, A.D. (1965), *Matematicheskie metody v teorii nadyozhnosti* [Mathematical methods in the theory of reliability], Nauka, Moscow, Russia.

Поступила 31.08.15

ФАРХАДЗАДЕ Эльмар Мехти оглу, д-р техн. наук, профессор, руководитель лаборатории «Надежность энергетического оборудования» Азербайджанского научно-исследовательского и проектно-изыскательского ин-та энергетики (г. Баку). В 1961 г. окончил Азербайджанский ин-т нефти и химии. Область научных исследований — надежность и эффективность электроэнергетических систем.

МУРАДАЛИЕВ Айдын Зураб оглу, д-р техн. наук, руководитель отдела «Надежность энергетического оборудования» Азербайджанского научно-исследовательского и проектно-изыскательского ин-та энергетики (г. Баку). В 1982 г. окончил Азербайджанский ин-т нефти и химии. Область научных исследований — количественная оценка индивидуальной надежности оборудования и устройств электроэнергетических систем.

ФАРЗАЛИЕВ Юсиф Зейни оглу, канд. техн. наук, ст. науч. сотр. лаборатории «Надежность энергетического оборудования» Азербайджанского научно-исследовательского и проектно-изыскательского ин-та энергетики (г. Баку). В 1985 г. окончил Азербайджанский госуниверситет. Область научных исследований — точность и достоверность оценок показателей индивидуальной надежности оборудования и устройств энергетических систем.