

Dorota Pekasiewicz

Uniwersytet Łódzki

TESTY SEKWENCYJNE DLA PARAMETRÓW ROZKŁADU NORMALNEGO LEWOSTRONNIE UCIĘTEGO

1. Wstęp

Rozpatrywane w badaniach statystycznych cechy często są utożsamiane ze zmiennymi losowymi o rozkładzie normalnym. W praktyce zdarza się, że zmienne te przybierają tylko wartości nieujemne lub dodatnie większe od pewnej znanej liczby. Zatem w tych przypadkach analizowane zmienne mają rozkład normalny lewostronnie ucięty.

Do weryfikacji hipotez o wartościach parametrów rozkładu normalnego lewostronnie uciętego można stosować parametryczne testy sekwencyjne, których statystyka jest ilorazem funkcji wiarygodności oraz testy sekwencyjne oparte na ilorazie funkcji pseudowiarygodności.

Rozważmy zatem zmienną losową X o rozkładzie normalnym $N(\mu, \sigma)$. Niech $Y = X|X \geq c$. Zmienna losowa Y ma rozkład normalny lewostronnie ucięty o funkcji gęstości postaci:

$$f_Y(y) = \frac{\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi}\sigma\left(1-\Phi\left(\frac{c-\mu}{\sigma}\right)\right)} \quad \text{dla } y \geq c. \quad (1)$$

Stosując testy sekwencyjne, można weryfikować hipotezy o wartościach parametrów μ oraz σ tego rozkładu, a znając wielkości tych parametrów, można wyznaczyć zarówno wartość oczekiwaną, jak i odchylenie standardowe analizowanej zmiennej.

Znajomość klasy rozkładu zmiennej losowej Y wraz z wielkością c ograniczającą zbiór jej wartości stwarza możliwość wykorzystania do weryfikacji hipotez o parametrach rozkładu μ oraz σ ilorazowego testu sekwencyjnego.

W przypadku braku informacji o klasie rozkładu rozważanej zmiennej można stosować testy sekwencyjne oparte na ilorazie funkcji pseudowiarygodności. Konieczny jest tylko wybór, do statystyki testu, odpowiedniej funkcji z rodziny linowo-wykładniczej. W przypadku uciętego rozkładu normalnego wydaje się wskazane wykorzystanie funkcji gęstości rozkładu normalnego. Zatem zastosowanie tego typu testu będzie związane z pominięciem informacji o ucięciu rozkładu.

Zastosowanie testów sekwencyjnych opartych na ilorazie funkcji pseudowiarygodności może sprawić, że wielkości prawdopodobieństw błędów I i II rodzaju będą większe niż założone na początku badania. Przeprowadzone badania symulacyjne pozwalają oszacować w analizowanych przypadkach wielkości tych prawdopodobieństw i porównać je z wartościami przyjętymi.

Ponadto, stosując metody symulacyjne, możemy porównać obydwa rodzaje testów pod względem liczebności prób niezbędnych do podjęcia decyzji o akceptacji jednej z weryfikowanych hipotez.

2. Testy sekwencyjne dla parametrów rozkładu oparte na ilorazie wiarygodności

Rozważmy zmienną losową $Y = X | X \geq c$ o rozkładzie normalnym lewostronnie uciętym i następujące hipotezy dotyczące parametru μ :

$$H_0 : \mu = \mu_0, \quad (2)$$

$$H_1 : \mu = \mu_1, \text{ gdzie } \mu_1 > \mu_0. \quad (3)$$

Weryfikację hipotezy (2) wobec hipotezy (3) można przeprowadzić testem sekwencyjnym wykorzystującym informację o postaci rozkładu uciętego.

Statystyka ilorazowego testu sekwencyjnego obliczana na każdym etapie postępowania sekwencyjnego ma następującą postać [2]:

$$I_n = \frac{\mu_1 - \mu_0}{\sigma^2} \sum_{i=1}^n Y_i + \frac{(\mu_0^2 - \mu_1^2)n}{2\sigma^2} + n \ln \frac{1 - \Phi((c - \mu_0)/\sigma)}{1 - \Phi((c - \mu_1)/\sigma)}, \quad (4)$$

gdzie Φ jest dystrybuantą rozkładu normalnego standaryzowanego.

Wartość statystyki obliczaną na każdym kroku procedury sekwencyjnej porównujemy z wartościami $A = \ln \frac{1 - \beta}{\alpha}$ i $B = \ln \frac{\beta}{1 - \alpha}$. Gdy $I_n < B$, podejmujemy decyzję o akceptacji hipotezy H_0 , jeśli $I_n > A$, decyzję o akceptacji hipotezy H_1 , natomiast gdy $B \leq I_n \leq A$ – decyzję o zwiększeniu próby przez dołosowanie kolejnego elementu.

Dla tego testu możemy wyznaczyć wartość oczekiwaną liczebności próby koniecznej do podjęcia decyzji o akceptacji jednej ze sformułowanych hipotez ze wzoru [1]:

$$E_{\mu}(N) = \begin{cases} \frac{(1-\alpha)\ln\frac{\beta}{1-\alpha} + \alpha\ln\frac{1-\beta}{\alpha}}{E_{\mu}(Z)} & \text{dla } \mu = \mu_0, \\ \frac{\beta\ln\frac{\beta}{1-\alpha} + (1-\beta)\ln\frac{1-\beta}{\alpha}}{E_{\mu}(Z)} & \text{dla } \mu = \mu_1, \end{cases} \quad (5)$$

gdzie $E_{\mu}(Z)$ jest wartością oczekiwaną zmiennej losowej $Z = \ln \frac{f(y, \mu_1)}{f(y, \mu_0)}$.

Wartość oczekiwana zmiennej losowej Z wyraża się wzorem:

$$E_{\mu}(Z) = \frac{\mu_1 - \mu_0}{\sigma^2} \cdot E_{\mu}(Y) + \frac{\mu_0^2 - \mu_1^2}{2\sigma^2} + \ln \frac{1 - \Phi\left(\frac{c - \mu_0}{\sigma}\right)}{1 - \Phi\left(\frac{c - \mu_1}{\sigma}\right)}, \quad (6)$$

natomiast

$$E_{\mu}(Y) = \mu + \frac{\sigma \exp\left(-\frac{(c - \mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi} \left(1 - \Phi\left(\frac{c - \mu}{\sigma}\right)\right)}. \quad (7)$$

Jeśli sformułujemy hipotezy dotyczące parametru σ postaci:

$$H_0 : \sigma = \sigma_0 \quad (8)$$

$$H_1 : \sigma = \sigma_1, \text{ gdzie } \sigma_1 > \sigma_0, \quad (9)$$

to, dla znanej wartości parametru μ możemy je weryfikować również parametrycznym, ilorazowym testem sekwencyjnym.

W tym przypadku statystyka testowa jest następująca:

$$I_n = \left(\frac{1}{2\sigma_0^2} - \frac{1}{2\sigma_1^2} \right) \sum_{i=1}^n (Y_i - \mu)^2 + n \ln \frac{\sigma_0}{\sigma_1} + n \ln \frac{1 - \Phi((c - \mu)/\sigma_0)}{1 - \Phi((c - \mu)/\sigma_1)}. \quad (10)$$

Analogicznie jak w przypadku testu dla parametru μ , na każdym etapie procedury sekwencyjnej obliczamy wartość tej statystyki i porównujemy ze stałymi A oraz B .

Wartość oczekiwana liczebności próby wyraża się wzorem:

$$E_{\sigma}(N) = \begin{cases} \frac{(1-\alpha)\ln\frac{\beta}{1-\alpha} + \alpha\ln\frac{1-\beta}{\alpha}}{E_{\sigma}(Z)} & \text{dla } \sigma = \sigma_0, \\ \frac{\beta\ln\frac{\beta}{1-\alpha} + (1-\beta)\ln\frac{1-\beta}{\alpha}}{E_{\sigma}(Z)} & \text{dla } \sigma = \sigma_1, \end{cases} \quad (11)$$

gdzie $E_{\sigma}(Z)$ jest wartością oczekiwaną zmiennej losowej $Z = \ln \frac{f(y, \sigma_1)}{f(y, \sigma_0)}$.

Wyznaczając, przy założeniu prawdziwości hipotez (8) i (9), warunkowe wartości oczekiwane zmiennej losowej Z ze wzoru

$$E_{\sigma}(Z) = \left(\frac{1}{2\sigma_0^2} - \frac{1}{2\sigma_1^2} \right) (E_{\sigma}(Y^2) - 2\mu E_{\sigma}(Y) + \mu^2) + \ln \frac{\sigma_0}{\sigma_1} + \ln \frac{1 - \Phi\left(\frac{c - \mu}{\sigma_0}\right)}{1 - \Phi\left(\frac{c - \mu}{\sigma_1}\right)}, \quad (12)$$

gdzie

$$E_{\sigma}(Y^2) = \sigma^2 + \mu^2 + \frac{(c + \mu)\sigma \exp\left(-\frac{(c - \mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi} \left(1 - \Phi\left(\frac{c - \mu}{\sigma}\right)\right)}, \quad (13)$$

$E_{\sigma}(Y)$ zaś wyraża się wzorem (7), możemy, analogicznie jak w przypadku testu dla parametru μ , obliczyć wartości oczekiwane liczebności próby koniecznej do podjęcia decyzji o akceptacji jednej z rozważanych hipotez.

3. Testy sekwencyjne dla parametrów rozkładu oparte na ilorazie funkcji pseudowiarygodności

Testy sekwencyjne oparte na ilorazie funkcji pseudowiarygodności mogą być stosowane do weryfikacji hipotez o wartościach parametrów μ oraz σ rozkładu normalnego uciętego lub też bezpośrednio do weryfikacji hipotez o wartości średniej i odchyleniu standardowym zmiennej losowej Y .

Związek między parametrem μ a wartością oczekiwaną określa wzór (7), natomiast między parametrem σ a wariancją następujący wzór:

$$D^2(Y) = \sigma^2 + \frac{\sigma(c-\mu)\exp\left(-\frac{(c-\mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi}\left(1-\Phi\left(\frac{c-\mu}{\sigma}\right)\right)} - \frac{\sigma^2\exp\left(-\frac{(c-\mu)^2}{\sigma^2}\right)}{2\pi\left(1-\Phi\left(\frac{c-\mu}{\sigma}\right)\right)^2}. \quad (14)$$

Jeśli weryfikujemy hipotezy o wartości parametru μ tym testem, to musimy przekształcić hipotezy (2) i (3) w hipotezy równoważne postaci:

$$H_0 : \mu_Y = \mu_{(0)} \quad (15)$$

$$H_1 : \mu_Y = \mu_{(1)}, \text{ gdzie } \mu_{(1)} > \mu_{(0)}, \quad (16)$$

gdzie μ_Y oznacza wartość oczekiwaną zmiennej losowej Y o rozkładzie lewostronnie uciętym.

Wartości $\mu_{(0)}$, $\mu_{(1)}$ wyznaczamy ze wzoru (7), podstawiając w miejsce μ odpowiednio μ_0 oraz μ_1 .

Postać statystyki ilorazowego testu sekwencyjnego wyraża wzór:

$$I_n = \frac{\mu_{(1)} - \mu_{(0)}}{\sigma'^2} \sum_{i=1}^n Y_i + \frac{(\mu_{(0)}^2 - \mu_{(1)}^2)^n}{2\sigma'^2}, \quad (17)$$

gdzie jako σ'^2 przyjmujemy średnią arytmetyczną wartości wyznaczonych ze wzoru (14), przy założeniu prawdziwości hipotezy H_0 oraz H_1 lub wartość wariancji wyznaczoną z próby wstępnej.

W przypadku weryfikacji hipotez o wartości parametru σ , przy znanej wartości parametru μ , za pomocą testu sekwencyjnego z funkcją pseudowiarygodności przekształcamy hipotezy (8) i (9) na hipotezy równoważne postaci:

$$H_0 : \sigma_Y = \sigma_{(0)} \quad (18)$$

$$H_1 : \sigma_Y = \sigma_{(1)}, \text{ gdzie } \sigma_{(1)} > \sigma_{(0)}, \quad (19)$$

gdzie σ_Y oznacza odchylenie standardowe zmiennej losowej Y o rozkładzie lewostronnie uciętym.

Wartości $\sigma_{(0)}$, $\sigma_{(1)}$ wyznaczamy ze wzoru (14), podstawiając w miejsce σ odpowiednio σ_0 oraz σ_1 .

Statystyka testu sekwencyjnego weryfikującego hipotezy o wartości parametru σ rozkładu zmiennej Y wyraża się wzorem:

$$I_n = \left(\frac{1}{2\sigma_{(0)}^2} - \frac{1}{2\sigma_{(1)}^2} \right) \sum_{i=1}^n (Y_i - \mu)^2 + n \ln \frac{\sigma_{(0)}}{\sigma_{(1)}}. \quad (20)$$

W sytuacji, gdy mamy zweryfikować hipotezy o wartości oczekiwanej lub odchyleniu standardowym zmiennej losowej o rozkładzie normalnym uciętym, możemy ten rodzaj testu stosować, ustalając wartości parametrów w hipotezach zerowej i alternatywnej bez dodatkowych przekształceń.

4. Zastosowanie testów sekwencyjnych do weryfikacji hipotez o wartości parametru rozkładu

W celu zastosowania przedstawionych testów sekwencyjnych do weryfikacji hipotez o parametrach μ oraz σ rozkładu normalnego lewostronnie uciętego i porównania tych testów ze względu na liczebność próby wygenerowano populację o rozkładzie normalnym uciętym, przy czym wartością ucięcia było 0.

Rozważano 3 następujące grupy eksperymentów, które różnicowały rozkłady zmiennej losowej Y :

- I. rozkłady normalne ucięte o parametrach: $\mu = 10$ i $\sigma = 1; 2; \dots; 10$,
- II. rozkłady normalne ucięte o parametrach: $\mu = 5$ i $\sigma = 1; 2; \dots; 5$,
- III. rozkłady normalne ucięte o parametrach $\mu = 0$ i $\sigma = 0,5; 1; \dots; 3$.

W grupach I i II przy małych wartościach parametru σ rozkład zmiennej Y zbliżony był do rozkładu normalnego i ze względu na wielkość ucięcia rozkładu można spodziewać się, że zastosowanie obydwu testów przyniesie podobne rezultaty. Większe wartości odchyżeń standardowych sprawiały, że rozkład Y bardziej różnił się od rozkładu normalnego, dlatego analizowane były takie przypadki, w których współczynnik zmienności wynosił nawet 1.

W grupie III rozkład zmiennej losowej Y jeszcze mniej przypominał rozkład normalny, gdyż wartością ucięcia była mediana zmiennej losowej o rozkładzie normalnym.

W każdej grupie eksperymentów sformułowano hipotezy dotyczące parametrów μ oraz σ i weryfikowano je, stosując przedstawione testy sekwencyjne. Przeprowadzona analiza symulacyjna pozwoliła sprawdzić, czy przy zastosowaniu testów sekwencyjnych opartych na funkcji pseudowiarygodności założone wielkości prawdopodobieństw błędów I i II rodzaju nie zostały przekroczone.

Test sekwencyjny dla parametru μ oparty na funkcji pseudowiarygodności był stosowany przy założeniu, że wartość parametru σ , a tym samym odchylenia standardowego jest znana, oraz w przypadku, gdy wielkość tego parametru jest nieznana i była szacowana na podstawie próby wstępnej. W badaniu próba ta liczyła $n = 50$ elementów.

Aby możliwe było zastosowanie testu sekwencyjnego dla parametru μ opartego na ilorazie funkcji pseudowiarygodności i porównywanie z testem, opartym na ilorazie funkcji wiarygodności, sformułowane zostały hipotezy (15) i (16) równoważne hipotezom (2) i (3). Otrzymane na podstawie wzoru (7) wartości $\mu_{(0)}$ oraz $\mu_{(1)}$ i wyznaczoną z wzoru (14) wartość σ' i oszacowaną z próby wartość $\hat{\sigma}$ dla wybranych przypadków (grupa I dla $\sigma = 5; 6; 7; 8$; grupa II dla $\sigma = 2; 3; 4; 5$ oraz grupa III dla $\sigma = 0,5; 1; 2; 3$) przedstawiono w tab. 1.

Tabela 1. Wartości parametrów w hipotezach równoważnych dla testu sekwencyjnego opartego na funkcji pseudowiarygodności

Grupa	Hipotezy	$\mu_{(0)}$	$\mu_{(1)}$	σ'	$\hat{\sigma}$
I	$H_0 : \mu = 10$ $H_1 : \mu = 13$	10,3	13,1	4,81	4,84
		10,6	13,2	5,58	4,97
		11,1	13,5	6,28	5,50
		11,6	13,9	6,94	6,62
	$H_0 : \mu = 10$ $H_1 : \mu = 12$	10,3	12,1	4,78	4,84
		10,6	12,3	5,53	4,97
		11,1	12,7	6,21	5,50
		11,6	13,1	6,87	6,62
II	$H_0 : \mu = 5$ $H_1 : \mu = 7$	5,0	7,0	1,98	1,76
		5,3	7,1	2,81	2,49
		5,8	7,3	3,50	3,31
		6,4	7,8	4,14	4,37
	$H_0 : \mu = 5$ $H_1 : \mu = 6,5$	5,0	6,5	1,97	1,76
		5,3	6,6	2,79	2,49
		5,8	6,9	3,47	3,31
		6,4	7,4	4,10	4,37
III	$H_0 : \mu = 0$ $H_1 : \mu = 1$	0,4	1,0	0,39	0,24
		0,8	1,3	0,70	0,47
		1,6	2,0	1,30	0,95
		2,4	2,8	1,90	1,42
	$H_0 : \mu = 0$ $H_1 : \mu = 0,5$	0,4	0,6	0,35	0,24
		0,8	1	0,65	0,47
		1,6	1,8	1,25	0,95
		2,4	2,6	1,86	1,42

Źródło: obliczenia własne.

W celu zbadania własności testów i porównania ich wygenerowano populacje według opisanych rozkładów. Przedstawionymi testami sekwencyjnymi weryfikowano sformułowane hipotezy 10 000 razy, przy prawdopodobieństwach błędów I i II rodzaju równych 0,05.

W każdym eksperymencie wyznaczona została liczba prawidłowych i nieprawidłowych decyzji spośród 10 000 i tym samym oszacowane zostały prawdopodobieństwa błędów związanych z podjęciem niesłusznych decyzji. Eksperymenty

zostały tak zaplanowane, aby uzyskać oszacowania prawdopodobieństwa błędu I rodzaju – α . Liczby poprawnych i błędnych decyzji oznaczone odpowiednio m_0 oraz m_1 i oszacowania $\hat{\alpha}$ dla wybranych wartości parametrów σ uzyskane przy zastosowaniu rozważanych testów przedstawione zostały w tab. 2-4.

Tabela 2. Oszacowania prawdopodobieństw błędu I rodzaju oraz charakterystyk liczbowych liczebności próby parametrycznego testu sekwencyjnego dla parametru μ

Grupa	Hipotezy	σ	m_0	m_1	$\hat{\alpha}$	$E(N_1)$	$\bar{N}_1$	s_{N_1}	$N_{1,\min}$	$N_{1,\max}$
I	$H_0 : \mu = 10$ $H_1 : \mu = 13$	5	9646	354	0,035	15,83	18,64	12,75	2	118
		6	9640	360	0,036	24,47	27,89	19,47	4	195
		7	9620	380	0,038	35,79	39,48	26,46	5	236
		8	9599	401	0,040	49,98	54,62	37,43	7	415
	$H_0 : \mu = 10$ $H_1 : \mu = 12$	5	9607	393	0,039	36,81	40,24	27,93	5	240
		6	9595	405	0,041	56,15	61,18	39,78	8	436
		7	9598	402	0,040	82,28	88,08	61,07	11	656
		8	9591	409	0,041	115,76	122,42	86,23	12	893
II	$H_0 : \mu = 5$ $H_1 : \mu = 7$	2	9717	283	0,028	5,30	7,06	4,6	1	41
		3	9649	351	0,035	13,48	16,08	10,83	2	97
		4	9615	385	0,039	27,56	31,22	21,26	4	194
		5	9627	373	0,037	48,39	52,23	35,89	7	339
	$H_0 : \mu = 5$ $H_1 : \mu = 6,5$	2	9670	330	0,033	9,42	11,67	7,96	2	84
		3	9640	360	0,036	24,28	27,89	19,47	4	195
		4	9599	401	0,040	50,06	54,62	37,44	7	415
		5	9640	360	0,036	87,51	92,84	64,12	11	683
III	$H_0 : \mu = 0$ $H_1 : \mu = 1$	0,5	9824	176	0,018	2,23	3,49	1,82	1	20
		1	9723	277	0,028	10,93	13,63	8,44	1	79
		2	9691	309	0,031	53,17	56,04	37,92	7	386
		3	9639	361	0,036	118,10	124,81	84,46	8	760
	$H_0 : \mu = 0$ $H_1 : \mu = 0,5$	0,5	9723	277	0,028	13,39	13,64	8,44	1	79
		1	9691	309	0,031	3,01	56,04	37,92	7	386
		2	9617	383	0,038	213,62	223,03	156,55	21	1887
		3	9620	380	0,038	521,56	497,24	344,29	55	4157

Źródło: obliczenia własne.

Tabela 2 zawiera wielkości poprawnych i błędnych decyzji dla testu sekwencyjnego opartego na funkcji wiarygodności i zgodnie z założeniami liczba niepoprawnych decyzji nie przekroczyła ustalonej wartości 5%. Ponadto w tablicy tej przedstawione zostały obliczone wartości oczekiwane wielkości prób koniecznych do zakończenia procedury weryfikacji sformułowanych hipotez, średnie wielkości prób wyznaczone na podstawie 10 000 powtórzeń oraz odchylenia standardowe, minimalne i maksymalne liczebności prób.

Analogiczne charakterystyki dla testów sekwencyjnych opartych na funkcji pseudowiarygodności odpowiednio ze znanym oraz szacowanym, na podstawie próby wstępnej, odchyleniem standardowym przedstawiono w tab. 3 i 4. W tabe-

lach tych nie ma wartości oczekiwanych liczebności prób, gdyż wielkości te są wyznaczane w oparciu o rozkład analizowanej zmiennej, a przy wykorzystaniu funkcji pseudowiarygodności zwykle takich informacji nie mamy.

Tabela 3. Oszacowania prawdopodobieństw błędu I rodzaju oraz charakterystyk liczbowych liczebności próby dla testu sekwencyjnego dla μ , z funkcją pseudowiarygodności, ze znanym σ

Grupa	Hipotezy	σ	m_0	m_1	$\hat{\alpha}$	$\bar{N}_2$	s_{N_2}	$N_{2,\min}$	$N_{2,\max}$
I	$H_0 : \mu = 10$ $H_1 : \mu = 13$	5	9665	335	0,034	17,95	12,24	2	123
		6	9635	365	0,037	27,92	19,56	4	195
		7	9687	313	0,032	39,46	26,70	5	233
		8	9594	406	0,041	53,27	57,08	5	432
	$H_0 : \mu = 10$ $H_1 : \mu = 12$	5	9635	365	0,037	40,37	27,75	5	221
		6	9584	416	0,042	62,26	43,81	7	449
		7	9629	371	0,037	84,01	57,33	9	490
		8	9566	434	0,043	121,63	86,50	12	809
II	$H_0 : \mu = 5$ $H_1 : \mu = 7$	2	9685	315	0,032	6,98	4,59	1	41
		3	9659	341	0,034	15,33	10,32	2	72
		4	9627	373	0,037	32,45	22,18	5	203
		5	9607	393	0,039	51,94	36,06	7	318
	$H_0 : \mu = 5$ $H_1 : \mu = 6,5$	2	9617	383	0,038	11,47	7,87	2	85
		3	9635	365	0,037	27,92	19,56	4	195
		4	9588	412	0,041	58,12	40,18	7	376
		5	9563	437	0,044	98,51	70,33	10	771
III	$H_0 : \mu = 0$ $H_1 : \mu = 1$	0,5	9894	106	0,011	3,24	1,59	1	16
		1	9818	182	0,018	12,14	7,38	1	81
		2	9734	266	0,027	58,63	38,31	9	384
		3	9694	306	0,031	119,84	79,90	15	749
	$H_0 : \mu = 0$ $H_1 : \mu = 0,5$	0,5	9818	182	0,018	18,45	1,52	1	120
		1	9734	266	0,027	58,63	38,31	9	384
		2	9672	328	0,033	202,39	136,93	23	1165
		3	9709	291	0,029	430,78	287,34	39	2730

Źródło: obliczenia własne.

Tabela 4. Oszacowania prawdopodobieństw błędu I rodzaju oraz charakterystyk liczbowych liczebności próby dla testu sekwencyjnego dla μ , opartego na funkcji pseudowiarygodności, z oszacowanym odchyleniem standardowym

Grupa	Hipotezy	σ	m_0	m_1	$\hat{\alpha}$	$\bar{N}_3$	s_{N_1}	$N_{3,\min}$	$N_{3,\max}$
I	$H_0 : \mu = 10$ $H_1 : \mu = 13$	5	9691	309	0,031	18,24	12,53	2	123
		6	9363	637	0,064	21,25	15,20	2	159
		7	9327	673	0,067	28,40	20,09	3	202
		8	9477	523	0,052	47,41	33,29	5	375
	$H_0 : \mu = 10$ $H_1 : \mu = 12$	5	9672	328	0,033	41,65	28,86	5	282
		6	9269	731	0,073	47,26	34,87	5	375
		7	9310	690	0,069	61,82	44,69	7	478
		8	9462	538	0,054	110,47	79,40	11	810
II	$H_0 : \mu = 5$ $H_1 : \mu = 7$	2	9434	566	0,057	5,53	3,80	1	41
		3	9348	652	0,065	11,64	8,22	1	93
		4	9483	517	0,052	28,57	20,19	3	201
		5	9721	279	0,028	58,39	40,6	7	349
	$H_0 : \mu = 5$ $H_1 : \mu = 6,5$	2	9379	621	0,062	9,11	6,35	1	58
		3	9363	637	0,064	21,25	15,20	2	159
		4	9474	526	0,053	51,69	36,51	6	361
		5	9708	292	0,029	115,39	79,33	15	778
III	$H_0 : \mu = 0$ $H_1 : \mu = 1$	0,5	9303	697	0,070	1,48	0,8	1	9
		1	8939	1061	0,106	5,12	3,42	1	36
		2	8869	1131	0,113	26,21	19,16	1	180
		3	8877	1123	0,112	56,35	41,77	4	390
	$H_0 : \mu = 0$ $H_1 : \mu = 0,5$	0,5	8919	1081	0,108	7,48	5,15	1	55
		1	8869	1131	0,113	26,21	19,16	1	180
		2	8878	1122	0,112	97,45	73,17	7	744
		3	8887	1113	0,111	209,55	158,12	20	1491

Źródło: obliczenia własne.

W przypadku zastosowania testu sekwencyjnego opartego na funkcji pseudowiarygodności ze znaną wartością σ oszacowane wielkości prawdopodobieństw błędu I rodzaju w każdej grupie eksperymentów nie przekroczyły założonej wartości 0,05. Zastosowanie testu sekwencyjnego opartego na funkcji pseudowiarygodności z oszacowaną wartością odchylenia standardowego dało gorsze rezultaty. Liczba błędnych decyzji była większa niż założona, szczególnie w III grupie eksperymentów była nawet dwukrotnie wyższa. W I i II grupie eksperymentów oszacowane wielkości prawdopodobieństw błędu I rodzaju w 12 przypadkach na 16 przekroczyły 0,05, ale wzrost ten nie był tak znaczący.

W tabeli 5 zestawiono obliczone średnie liczebności prób $\bar{N}_1$, $\bar{N}_2$ i $\bar{N}_3$, przy których podejmowane były decyzje o przyjęciu jednej z weryfikowanych hipotez dla kolejno analizowanych testów sekwencyjnych. Przez obliczenie miar $e_2 = \bar{N}_2 / \bar{N}_1$ oraz $e_3 = \bar{N}_3 / \bar{N}_1$ porównane zostały średnie liczebności dla testów

opartych na funkcjach pseudowiarygodności ze średnimi liczebnościami prób dla testów sekwencyjnych z ilorazem funkcji wiarygodności. Efektywność testów mierzona ilorazem średnich liczebności prób dla pierwszych dwóch testów jest bliska 1, co oznacza, że testy te wymagały bardzo zbliżonej wielkości próby sekwencyjnej. Zastosowanie testu sekwencyjnego z ilorazem funkcji pseudowiarygodności, z oszacowanym na podstawie próby wstępnej odchyleniem standardowym, wymagało zwykle mniejszej liczby elementów próby (czasem o połowę mniejszej – grupa III), ale w tych przypadkach wiązało się to z podjęciem większej liczby błędnych decyzji (ponad połowę większej niż założona).

Tabela 5. Porównanie efektywności testów sekwencyjnych dla parametru μ

Grupa	Hipotezy	σ	$\bar{N}_1$	$\bar{N}_2$	$\bar{N}_3$	e_2	e_3
I	$H_0 : \mu = 10$ $H_1 : \mu = 13$	5	18,64	17,95	18,24	0,96	0,98
		6	27,89	27,92	21,25	1,00	0,76
		7	39,48	39,46	28,40	1,00	0,72
		8	54,62	53,27	47,41	0,98	0,87
	$H_0 : \mu = 10$ $H_1 : \mu = 12$	5	40,24	40,37	41,65	1,00	1,04
		6	61,18	62,26	47,26	1,02	0,77
		7	88,08	84,01	61,82	0,95	0,70
		8	122,42	121,63	110,47	0,99	0,90
II	$H_0 : \mu = 5$ $H_1 : \mu = 7$	2	7,06	6,98	5,53	0,99	0,78
		3	16,08	15,33	11,64	0,95	0,72
		4	31,22	32,45	28,57	1,04	0,92
		5	52,23	51,94	58,39	0,99	1,12
	$H_0 : \mu = 5$ $H_1 : \mu = 6,5$	2	11,67	11,47	9,11	0,98	0,78
		3	27,89	27,92	21,25	1,00	0,76
		4	54,62	58,12	51,69	1,06	0,95
		5	92,84	98,51	115,39	1,06	1,24
III	$H_0 : \mu = 0$ $H_1 : \mu = 1$	0,5	3,49	3,24	1,48	0,93	0,42
		1	13,63	12,14	5,12	0,89	0,38
		2	56,04	58,63	26,21	1,05	0,47
		3	124,81	119,84	56,35	0,96	0,45
	$H_0 : \mu = 0$ $H_1 : \mu = 0,5$	0,5	13,64	18,45	7,48	1,35	0,55
		1	56,04	58,63	26,21	1,05	0,47
		2	223,03	202,39	97,45	0,91	0,44
		3	497,24	430,78	209,55	0,87	0,42

Źródło: obliczenia własne.

Wygenerowane populacje zostały również wykorzystane do symulacyjnej analizy testów sekwencyjnych dla parametru σ rozkładu normalnego lewostronnie uciętego wartością 0. Sformułowane hipotezy były weryfikowane z użyciem obydwu rodzajów testów. Liczby prawidłowych i nieprawidłowych decyzji oraz charakterystyki liczbowe liczebności prób dla wybranych hipotez dla testu sekwencyjnego z funkcją wiarygodności przedstawione są w tab. 6, natomiast dla testu sekwencyjnego z funkcją pseudowiarygodności – w tab. 7. Prawdopodobieństwa

błędów ustalone zostały, podobnie jak w przypadku weryfikacji hipotez o parametrze μ , na poziomie 0,05.

Tabela 6. Oszacowania prawdopodobieństw błędu I rodzaju oraz charakterystyk liczbowych liczebności próby parametrycznego testu sekwencyjnego dla parametru σ

Grupa	Hipotezy	Rozkład X	m_0	m_1	$\hat{\alpha}$	$E(N_4)$	$\bar{N}_4$	s_{N_4}	$N_{4,\min}$	$N_{4,\max}$
I	$H_0 : \sigma = 7$ $H_1 : \sigma = 8$	$N(10,7)$	9714	286	0,028	231,57	247,17	159,80	7	1715
	$H_0 : \sigma = 7$ $H_1 : \sigma = 9$	$N(10,7)$	9736	264	0,026	65,28	79,27	47,66	1	604
	$H_0 : \sigma = 8$ $H_1 : \sigma = 9$	$N(10,8)$	9699	301	0,030	297,94	309,87	199,96	22	2460
	$H_0 : \sigma = 8$ $H_1 : \sigma = 10$	$N(10,8)$	9732	268	0,027	81,91	97,31	59,67	4	575
II	$H_0 : \sigma = 2$ $H_1 : \sigma = 3$	$N(5,2)$	9762	238	0,024	22,23	30,22	18,10	2	172
	$H_0 : \sigma = 2$ $H_1 : \sigma = 4$	$N(5,2)$	9823	177	0,018	8,28	13,28	6,95	1	74
	$H_0 : \sigma = 3$ $H_1 : \sigma = 4$	$N(5,3)$	9741	259	0,026	50,53	61,62	36,97	1	397
	$H_0 : \sigma = 3$ $H_1 : \sigma = 5$	$N(5,3)$	9818	182	0,018	16,20	23,65	12,67	1	130
III	$H_0 : \sigma = 2$ $H_1 : \sigma = 3$	$N(0,2)$	9798	202	0,020	16,41	22,65	12,53	1	119
	$H_0 : \sigma = 2$ $H_1 : \sigma = 4$	$N(0,2)$	9871	129	0,013	5,81	9,70	4,55	1	49
	$H_0 : \sigma = 3$ $H_1 : \sigma = 4$	$N(0,3)$	9774	226	0,023	32,31	40,68	24,01	1	259
	$H_0 : \sigma = 3$ $H_1 : \sigma = 5$	$N(0,3)$	9831	169	0,017	10,45	15,56	8,11	1	83

Źródło: obliczenia własne.

Tabela 7. Oszacowania prawdopodobieństw błędu I rodzaju oraz charakterystyk liczbowych liczebności próby dla testu sekwencyjnego dla σ , z funkcją pseudowiarygodności, ze znanym μ

Grupa	Hipotezy	Rozkład X	m_0	m_1	$\hat{\alpha}$	$\bar{N}_5$	s_{N_5}	$N_{5,\min}$	$N_{5,\max}$
I	$H_0 : \sigma = 7$ $H_1 : \sigma = 8$	$N(10,7)$	9778	222	0,022	281,27	179,46	35	1669
	$H_0 : \sigma = 7$ $H_1 : \sigma = 9$	$N(10,7)$	9755	245	0,025	88,40	52,72	3	523
	$H_0 : \sigma = 8$ $H_1 : \sigma = 9$	$N(10,8)$	9723	277	0,028	353,48	224,69	22	2006
	$H_0 : \sigma = 8$ $H_1 : \sigma = 10$	$N(10,8)$	9740	260	0,026	106,60	65,96	7	617
II	$H_0 : \sigma = 2$ $H_1 : \sigma = 3$	$N(5,2)$	9799	201	0,020	33,97	20,17	1	235
	$H_0 : \sigma = 2$ $H_1 : \sigma = 4$	$N(5,2)$	9872	128	0,013	14,67	7,55	1	72
	$H_0 : \sigma = 3$ $H_1 : \sigma = 4$	$N(5,3)$	9811	189	0,019	69,79	41,88	2	502
	$H_0 : \sigma = 3$ $H_1 : \sigma = 5$	$N(5,3)$	9841	159	0,016	25,31	13,62	1	168
III	$H_0 : \sigma = 2$ $H_1 : \sigma = 3$	$N(0,2)$	9452	548	0,055	20,46	11,34	1	117
	$H_0 : \sigma = 2$ $H_1 : \sigma = 4$	$N(0,2)$	9619	381	0,038	9,02	4,19	1	44
	$H_0 : \sigma = 3$ $H_1 : \sigma = 4$	$N(0,3)$	9419	581	0,058	36,32	21,64	1	264
	$H_0 : \sigma = 3$ $H_1 : \sigma = 5$	$N(0,3)$	9529	471	0,047	14,28	7,43	1	95

Źródło: obliczenia własne.

W przedstawionych przykładach wykorzystanie ilorazowego testu sekwencyjnego z funkcją pseudowiarygodności dało zadowalające rezultaty. Liczba błędnych decyzji, poza dwoma przypadkami, nie przekroczyła ustalonej wielkości (tab. 6). Jednak przy zastosowaniu tego testu do podjęcia decyzji zwykle potrzebna była liczniejsza próba (tab. 8), zatem efektywność testu sekwencyjnego z funkcją pseudowiarygodności była mniejsza.

Tablica 8. Porównanie efektywności testów sekwencyjnych dla parametru σ

Grupa	Hipotezy	Rozkład X	$\bar{N}_4$	$\bar{N}_5$	$e = \frac{\bar{N}_5}{\bar{N}_4}$
I	$H_0 : \sigma = 7$ $H_1 : \sigma = 8$	$N(10,7)$	247,17	281,27	1,14
	$H_0 : \sigma = 7$ $H_1 : \sigma = 9$	$N(10,7)$	79,27	88,40	1,12
	$H_0 : \sigma = 8$ $H_1 : \sigma = 9$	$N(10,8)$	309,87	353,48	1,14
	$H_0 : \sigma = 8$ $H_1 : \sigma = 10$	$N(10,8)$	97,31	106,60	1,10
II	$H_0 : \sigma = 2$ $H_1 : \sigma = 3$	$N(5,2)$	30,22	33,97	1,24
	$H_0 : \sigma = 2$ $H_1 : \sigma = 4$	$N(5,2)$	13,28	14,67	1,10
	$H_0 : \sigma = 3$ $H_1 : \sigma = 4$	$N(5,3)$	61,62	69,79	1,13
	$H_0 : \sigma = 3$ $H_1 : \sigma = 5$	$N(5,3)$	23,65	25,31	1,07
III	$H_0 : \sigma = 2$ $H_1 : \sigma = 3$	$N(0,2)$	22,65	20,46	0,90
	$H_0 : \sigma = 2$ $H_1 : \sigma = 4$	$N(0,2)$	9,70	9,02	0,93
	$H_0 : \sigma = 3$ $H_1 : \sigma = 4$	$N(0,3)$	40,68	36,32	0,89
	$H_0 : \sigma = 3$ $H_1 : \sigma = 5$	$N(0,3)$	15,56	14,28	0,92

Źródło: obliczenia własne.

W przypadku testów dla parametru μ otrzymane wyniki sugerowały większą lub porównywalną efektywność testów sekwencyjnych z funkcją pseudowiarygodności (tab. 5) w stosunku do testu sekwencyjnego z funkcją wiarygodności.

5. Uwagi końcowe

W pracy przedstawione zostały testy sekwencyjne, które można stosować do weryfikacji hipotez o wartościach parametrów μ lub σ zmiennej losowej o rozkładzie normalnym lewostronnie uciętym.

Pierwszy z analizowanych testów wymaga znajomości funkcji gęstości rozkładu normalnego uciętego wraz z wartością występującego parametru, odpowiednio σ w przypadku testu dla parametru μ oraz μ w przypadku testu dla parametru σ . Ponadto niezbędna jest znajomość wartości ucięcia rozkładu. Sprawdzianem tego testu jest statystyka określona jako iloraz łącznych funkcji gęstości odpowiednio, przy założeniu prawdziwości hipotezy alternatywnej i hipotezy zerowej.

Stosując drugi z testów sekwencyjnych, czyli test oparty na funkcji pseudowiarygodności, nie musimy znać postaci rozkładu zmiennej losowej, ale musimy wybrać odpowiednią funkcję z rodziny liniowo-wykładniczej wykorzystywaną do konstrukcji funkcji pseudowiarygodności. W przypadku rozkładów normalnych lewostronnie uciętych taką funkcją może być funkcja gęstości rozkładu normalnego. Przy stosowaniu tego typu testów sekwencyjnych hipotezy dotyczą wartości średniej zmiennej losowej, a nie parametru μ , lub odchylenia standardowego, a nie parametru σ . Konieczna jest zatem znajomość wartości odchylenia standardowego lub jego oszacowania w pierwszym przypadku oraz wartości średniej w drugim.

Przeprowadzone badania symulacyjne pozwalają stwierdzić, że gdy wartość odchylenia standardowego zmiennej losowej była znana, to zastosowanie testu sekwencyjnego dla parametru μ z funkcją pseudowiarygodności dało, w analizowanych przypadkach, zadowalające rezultaty, tzn. prawdopodobieństwa błędów związanych z podjęciem niesłusznych decyzji były nie większe od ustalonych. Wielkość ucięcia rozkładu normalnego nie miała tu szczególnego znaczenia. Zarówno w przypadku „małych” ucięć (np. ucięcie rozkładu normalnego $N(10,5)$), jak i „dużych” (np. ucięcie rozkładu normalnego $N(0,2)$), decyzje o przyjęciu prawdziwej hipotezy zależały głównie od wartości odchylenia standardowego. Dlatego, gdy wartość oszacowanego na podstawie próby wstępnej odchylenia standardowego była znacznie niższa niż rzeczywista, liczba błędnych decyzji była dużo większa niż zakładana na początku badania. Analizując efektywność testów (przy znanej wartości parametru σ) mierzoną ilorazem średnich liczebności prób, należy stwierdzić, że nie da się zauważyć wyraźnej przewagi żadnego z testów, gdyż test parametryczny wymaga nieznacznie mniejszej liczebności próby, a drugi rodzaj testu oparty jest na funkcji pseudowiarygodności.

Podobne wyniki uzyskane zostały dla testów sekwencyjnych dla parametru σ , chociaż w tym przypadku można zauważyć większą efektywność testu parametrycznego.

Jeśli badanie statystyczne dotyczy wartości średniej lub odchylenia standardowego zmiennej losowej o rozkładzie uciętym (nie parametru μ czy σ), to test sekwencyjny oparty na funkcji pseudowiarygodności wydaje się wygodniejszy, gdyż od razu weryfikujemy hipotezy o wartości odpowiedniego parametru, a poza tym statystyki będące sprawdzianami testów mają trochę prostszą postać.

Literatura

- [1] Marek T., Noworol Cz., *Analiza sekwencyjna w badaniach empirycznych*, PWN, Warszawa 1982.
- [2] Pekasiewicz D., *Sequential Tests for Truncated Distribution Parameters*, Folia Oeconomica, Łódź 2006, s. 47-56.

THE SEQUENTIAL TESTS FOR PARAMETERS OF LEFT-TRUNCATED NORMAL DISTRIBUTION

Summary

The sequential probability ratio tests, presented in this paper, are used to verify hypotheses about parameters μ and σ of density function of the left-truncated normal distribution $N(\mu, \sigma)$. The statistic of the first presented test is the likelihood functions ratio, calculated under the assumption that respectively: alternative and null hypotheses are true. In the case of the independent sample sampling scheme, likelihood function is a product of density functions of the normal truncated distribution. In the second test, the statistic is based on the pseudo-likelihood functions ratio with density functions of the normal distribution $N(\mu, \sigma)$.

Apart from theoretical considerations, the results of simulation analysis of the properties of both sequential tests are presented in the paper. Estimated numerical characteristics of the sample sizes, necessary to take a decision about the acceptance of one of the verified hypotheses, allow to compare regarded tests in respect to their efficiencies.