

Wyznaczanie standardowej niepewności pomiarów o rozkładzie trapezowym metodą maksymalizacji wielomianu

Zygmunt L. Warsza

Przemysłowy Instytut Automatyki i Pomiarów PIAP, Al. Jerozolimskie 202, 02-486 Warszawa

Serhii W. Zabolotnii

Cherkaskij Nacjonalnyj Tehnichnij Universitet, Ukraina

Streszczenie: Omówiono w skrócie rodzaje estymatorów parametrów mierzandów wyznaczanych z próbek danych pomiarowych pobranych z populacji o rozkładzie trapezowym. Zaproponowano użycie metody maksymalizacji wielomianu stochastycznego o symbolu PMM jako niekonwencjonalnego sposobu wyznaczania estymatorów wartości i odchylenia standardowego mierzandów dla próbek o rozkładach niegaussowskich. Na przykładach próbek z symetrycznego rozkładu trapezowego Trap oszacowano niepewność standardową dla wartości średniej, środka rozpięcia i estymatora mierzandów obliczonego metodą wielomianową PMM z użyciem kumulantów, które wyznaczono z danych próbki za pośrednictwem momentów centralnych. Metodą symulacji Monte Carlo (MC) dokonano analizy porównawczej ocen wariancji obliczanej klasycznie ze wzorów rozkładu, wg Przewodnika GUM [1], dla środka rozpięcia i metodą PPM. W funkcji liczby danych próbki i stosunku podstaw trapezu określono granice najefektywniejszego obszaru dla każdej z metod.

Słowa kluczowe: mierzand, estymata, odchylenie standardowe, rozkład trapezowy, kumulanty, metoda maksymalizacji wielomianu

1. Wprowadzenie

Podczas prowadzenia pomiarów, zarówno w obiekcie badanym, jak i w przyrządzie lub systemie pomiarowym i otoczeniu występują zjawiska losowe. Oddziałują one na wielkość mierzoną i na sygnały pomiarowe. Wskutek tego pojawia się składowa losowa w sygnale wyjściowym wielkości mierzonej, a ogólniej – w sygnałach mierzonych parametrów multimierzandów. Sygnał pomiarowy jest bądź ciągły w czasie i następnie podlega próbkowaniu, zwykle regularnemu, bądź też pomiary są wielokrotnie powtarzane, w tym automatycznie. W obu przypadkach otrzymuje się próbkę danych pomiarowych o losowym rozrzucie wartości. Wynik pomiaru podlega szacowaniu, zwykle według zaleceń międzynarodowego Przewodnika GUM (*Guide for the uncertainty in measurement*) [1]. Dane próbki pomiarowej oczyszcza się z błędów systematycznych i oblicza

się estymator wartości mierzandów wraz z oceną jego dokładności w postaci przedziału niepewności o wymaganym prawdopodobieństwie. Według Przewodnika GUM wynik pomiarów wyznacza się w postaci

$$X = \bar{X} \pm U \quad (1)$$

gdzie: \bar{X} – wartość średnia, $U = k_p u = k_p \sqrt{u_A^2 + u_B^2}$ – niepewność rozszerzona, k_p – współczynnik rozszerzenia dla przedziału o prawdopodobieństwie P , u – niepewność standardowa i jej składowe: u_A – obliczana z pomiarowych danych metodą statystyczną i u_B – szacowana na podstawie wiedzy o niezidentyfikowanych błędach, które mogą wystąpić w danym eksperymencie pomiarowym.

W praktyce pomiarowej niepewność typu A (u_A) liczy się tak, jak dla próbek z populacji o rozkładzie normalnym, tj. opisanym funkcją Gaussa, Przewodnik GUM [1] nie wskazuje, jak ją liczyć dla próbek danych z rozkładów innych niż normalny. Do wyznaczenia niepewności rozszerzonej U próbki o wymaganym prawdopodobieństwie określa się wartości współczynników k_p . Dla danych z populacji o rozkładzie normalnym i dużej liczbie n stosuje się współczynniki $k_{0,95} = 1,96$ i $k_{0,975} = 3$.

Ostatnio zaproponowano udoskonaloną wersję Przewodnika GUM 2 [2], opartą na podejściu Bayesa, w której zaleca się, aby po eliminacji znanych błędów systematycznych, niepewność rozszerzoną U wyznaczać jako

Autor korespondujący:

Zygmunt L. Warsza, zlw@op.pl

Artykuł recenzowany

nadesłany 26.09.2017 r., przyjęty do druku 14.11.2017 r.

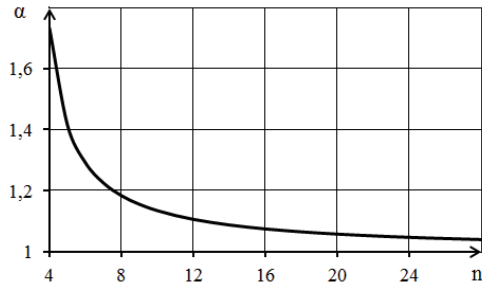


Zezwala się na korzystanie z artykułu na warunkach licencji Creative Commons Uznanie autorstwa 3.0

$$U = k_p u = k_p \sqrt{\frac{n-1}{n-3} u_A^2 + u_B^2} \quad (2)$$

gdzie: $n \geq 4$ – liczebność próbkki.

Współczynnik $\alpha \equiv (n-1)/(n-3)$ w (2) przy wariancji u_A^2 wzrasta ze zmniejszaniem się liczby n danych próbkki (rys. 1). Wynika to z rozkładu Studenta.



Rys. 1. Zależność współczynnika α od liczebności próbkki n
Fig. 1. Coefficient α as function of the number n of sample elements

Taki przebieg współczynnika α jak na rys. 1 dotyczy próbek z populacji o rozkładzie normalnym. Wzór (2) obowiązuje dla próbek o liczbie elementów $n \geq 4$ [3]. W komitecie JCGM d/s przewodników metrologicznych zdania o stosowaniu tego wzoru są podzielone i jeszcze nie rekomenduje się go w przewodnikach dla praktyki metrologicznej.

Wyznaczana wg Przewodnika GUM ocena niepewności mierzand, nie jest wystarczająco efektywna dla próbek pomiarowych modelowanych rozkładami niegaussowskimi. Do modelowania rozrzutu danych pomiarowych, w tym również, gdy wynika on ze zmian wartości mierzand w obiekcie mierzonym, poza rozkładem normalnym, stosuje się też inne rozkłady. W pracy [2], dla współczynnika k_p rozszerzenia niepewności standardowej jako wartości średniej populacji o rozkładzie symetrycznym jednomodalnym oraz o rozkładzie dowolnym niesymetrycznym proponuje się następujące wzory przybliżone

$$k_p \leq \frac{2}{3 \cdot \sqrt{1-p}} \quad k_p \leq \frac{1}{\sqrt{1-p}} \quad (3)$$

Z wzorów (3) dla $p = 0,95$ otrzymuje się bardzo duże wartości $k_p = 2,98$ i $k_p = 4,47$, gdy dla rozkładu normalnego jest $k_p = 1,96$.

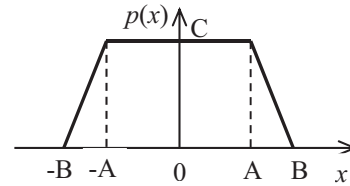
Ponadto w praktyce, obok przypadków, gdy znany jest rodzaj rozkładu populacji, z której pochodzą dane pomiarowe, występują sytuacje, gdy jest on *a priori* nieznan. Jednoznaczna jego identyfikacja jest możliwa tylko dla próbek o bardzo dużej liczbie danych n . Trzeba więc dla rozkładów niegaussowskich stosować metody alternatywne, które wykorzystują więcej informacji zawartych w próbce niż dwa pierwsze jej momenty.

Poniżej, na przykładzie próbek z symetrycznego rozkładu trapezowego Trap analizowane będą wyniki, jakie można uzyskać stosując podaną przez Y. Kunczenkę [13] metodę maksymalizacji wielomianu o symbolu PMM (ang. *Polynomial Maximization Method*). W metodzie tej stosuje się, jako prostszy, opis wzorów za pomocą kumulantów. Wyznacza się je pośrednio z momentów centralnych obliczonych z danych próbkki. Zastosowanie metody PMM do szacowania niepewności mierzand dla próbek z niegaussowskich rozkładów symetrycznych opisali autorzy w [14].

2. Estymatory mierzand o rozkładzie trapezowym

Symetryczny rozkład trapezowy o bokach liniowych (rys. 2), oznaczany jest w Suplemencie 1 do Przewodnika GUM sym-

bolem Trap. Rozkład ten występuje dość często w cyfrowych systemach pomiarowych i jest splotem dwu rozkładów równomiernych. Długości podstaw trapezu są sumą i różnicą rozstępów V_1, V_2 , czyli szerokości tych rozkładów.



Rys. 2. Rozkład trapezowy Trap o podstawach $2a = (-A, A)$, $2b = (-B, B)$
Fig. 2. PDF of trapeze Trap distribution: $2a = (-A, A)$, $2b = (-B, B)$

Charakterystycznym parametrem określającym kształt takiego trapezu jest stosunek długości jego podstaw $\beta \equiv a/b$. Przypadkami krańcowymi rodziny rozkładów Trap są:
– rozkład trójkątny ($\beta = 0$) – dla splecionych jednakowych rozkładów równomiernych,
– rozkład równomierny ($\beta = 1$), gdy szerokość jednego z tych rozkładów jest pomijalnie mała.

Odchylenia standardowe splecionych rozkładów równomiernych wynoszą $u_1 = V_1/2\sqrt{3}$ i $u_2 = \lambda u_1$, (gdzie $0 \leq \lambda \leq 1$). Standardowa niepewność typu A tworzonego przez nie rozkładu trapezowego Trap wynika z sumy ich wariancji i wynosi

$$u = \sqrt{u_1^2 + u_2^2} = u_1 \sqrt{1 + \lambda^2} \quad (4)$$

Współczynniki λ i β są powiązane zależnością

$$\lambda = \frac{1 - \beta}{1 + \beta} \quad (5)$$

Według p. 4.3.9 w GUM [1] środek $q_{V/2} = 0,5(b_+ - b_-)$ dla populacji o rozkładzie Trap ma wariancję

$$u^2(x_i) = b^2(1 + \beta^2)/6 \quad (6)$$

Do obliczania niepewności rozszerzonej dla wartości średniej \bar{X} rozkładu Trap Botsiura i Zakharov [12], podają następujący współczynnik rozszerzenia

$$k_{0,95} = \sqrt{3} \frac{1 + \lambda - \sqrt{0,2\lambda}}{\sqrt{1 + \lambda^2}} \quad (7)$$

Obliczona z danych pomiarowych wartość średnia jest najbardziej efektywnym estymatorem mierzand, tj. o najmniejszym odchyleniu standardowym (SD) tylko dla próbek danych pobranych z rozkładu normalnego (funkcja Gaussa). Dla innych rozkładów są inne lepsze estymatory.

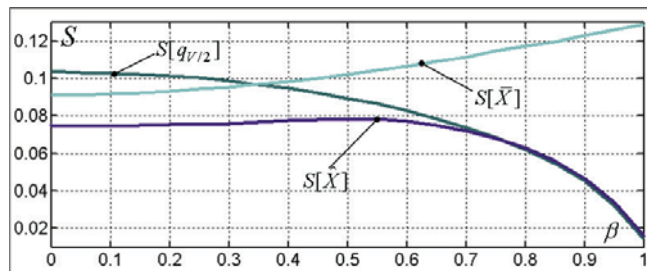
Standardowe odchylenie $S_{V/2}$ środka rozstępu $q_{V/2} = 0,5(x_{max} - x_{min})$ próbkki pobranej z równomiernego rozkładu o n elementach wg [4, 5] wynosi

$$S_{V/2} = \frac{V}{\sqrt{2(n-1)}} \cdot \sqrt{\frac{n+1}{n+2}} \quad (8)$$

Odchylenie standardowe $S_{V/2}$ jest mniejsze niż $S[\bar{X}]$ średniej. Środek rozpięcia jest więc estymatorem lepszym dla próbek z rozkładu równomiernego. Oba te odchylenia i ich względne niepewności standardowe maleją wraz ze wzrostem liczebności n próbkki. Natomiast $S[\bar{X}]$ jest bardziej dokładne od $S_{V/2}$. Sprawdzono to metodą Monte Carlo [6].

Metodą symulacyjną Monte Carlo (MC) zbadano też odchylenia standardowe wartości średniej \bar{X} , środka rozstępu $q_{V/2}$ oraz

mediany X_{med} jako jednoelementowych estymatorów wartości mierzalnego dla próbek danych pomiarowych z populacji o symetrycznych rozkładach trapezowych: Trap – o bokach liniowych i CTrap – o bokach krzywoliniowych wklęsłych [7–11]. Wyznaczono je jako funkcje stosunku długości podstaw trapezu $\beta \in [0; 1]$ i liczby danych n próbek. Dla zmiennej losowej o rozkładzie Trap podano je na rys 3. Odchylenie standardowe mediany w całym zakresie β jest większe od pozostałych estymatorów i nie podano go na wykresie. Odchylenie standardowe średniej $S[\bar{X}]$ najmniejsze jest tylko w przedziale $0 < \beta < 0,35$. Dla większych β , tj. w przedziale $0,35 < \beta \leq 1$, mniejsze od niego jest odchylenie $S[q_{V/2}]$ środka rozstępu.



Rys. 3. Odchylenia standardowe: $S[\bar{X}]$ – średniej i $S[\hat{X}]$ – środka rozstępu oraz $S[\hat{X}]$ – estymatora dwuelementowego $\hat{X}=0,5(\bar{X}+q_{V/2})$ próbek o $n = 400$ danych z rozkładu Trap(a, b) w funkcji stosunku podstaw trapezu β

Fig. 3. Standard deviations: $S[\bar{X}]$ of mean, $S[q_{V/2}]$ of mid range and $S[\hat{X}]$ of two-element estimator $\hat{X}=0,5(\bar{X}+q_{V/2})$ for the $n = 400$ data samples from linear trapeze PDF Trap(a, b) as function of a trapeze bases ratio β

Dla próbek z rozkładu Trap [6, 8] wykazano też, że odchylenie SD estymatora dwuelementowego $\hat{X} = 0,5(\bar{X} + q_{V/2})$ jest jeszcze mniejsze, co najmniej o 20% w całym zakresie β (rys. 3). Syntezę tych badań i porównanie estymatorów kilku innych rozkładów podano też w monografii [15].

3. Matematyczne sformułowanie problemu

Poniżej symbolem $\hat{\theta}$ oznacza się ocenę wartości mierzalnego wyznaczaną z próbki rozproszonych danych pomiarowych metodą PMM. Wartość tę estymuje się przez statystyczną analizę wektora $x = \{x_1, x_2, \dots, x_n\}$. Zakłada się, że dane losowe x_i są niezależne, o wartościach opisanych modelem $\xi = \theta + \xi_0$. W modelu tym ξ_0 jest wycelowaną zmienną losową. Jako przykład rozpatrzy się próbki o różnej liczbie danych n z populacji o rozkładzie Trap, tj. w postaci symetrycznego trapezu o liniowych bokach, szerokości dolnej podstawy $2b$ i stosunku długości podstaw β . Przez próby statystyczne dokonywane metodą MC co najmniej $M=10^4$ razy, należy zbadać estymatory parametrów próbki wyznaczone metodą maksymalizacji wielomianu PMM, ich dokładność i zbieżność rozkładu empirycznego do funkcji Gaussa, w zależności od liczby danych n próbek. Ponadto należy porównać te oceny z ocenami niepewności wartości średniej i środka rozpięcia jako estymatorów mierzalnego.

4. Wyznaczanie estymatorów metodą PMM

4.1. Podstawy teoretyczne metody PMM

W pracy, do wyznaczenia niepewności próbki z rozkładu Trap zastosuje się podaną przez Y. Kunchenkę metodę maksymalizacji wielomianu PMM [13]. Oszacowaniem poszukiwanej

wartości parametru θ jest rozwiązanie układu równań stochastycznych

$$\sum_{i=1}^r h_i(\theta) [\hat{\alpha}_i - \alpha_i(\theta)] \Big|_{\theta=\hat{\theta}} = 0 \quad (9)$$

gdzie: r – rząd wielomianu użytego do szacowania parametrów,

$\alpha_i(\theta)$ i $\hat{\alpha}_i = \frac{1}{n} \sum_{v=1}^n x_v^i$ – teoretyczne i początkowe momenty

i -tego rzędu próbki.

Współczynniki $h_i(\theta)$ (dla $i=1$ s) wyznacza się przez rozwiązanie układu liniowych równań algebraicznych rzędu s podanych dla warunków minimalizacji wariancji poszukiwanej estymaty wartości parametru v [13], a mianowicie:

$$\sum_{i=1}^s h_i(\theta) F_{i,j}(\theta) = \frac{d}{d\theta} \alpha_j(\theta), \quad j = \overline{1, s}, \quad (10)$$

gdzie $F_{i,j}(\theta) = \alpha_{i+j}(\theta) - \alpha_i(\theta) \alpha_j(\theta)$.

W [13] wykazano, że estymatory $\hat{\theta}$ będące rozwiązaniami układu równań (9) są zgodne i asymptotycznie nieobciążone.

4.2. Estymatory dla zmiennych losowych o rozkładach symetrycznych wg metody PMM

Oszacowanie wartości składowej stałej metodą PMM w przypadku szczególnym, tj. przy stopniu wielomianu $s = 1$, jest równoważne oszacowaniu średniej arytmetycznej dla dowolnego rozkładu zmiennej losowej [10]. W pracy tej wykazano również, że przy symetrii rozkładu otrzymuje się dla nieparzystych kumulantów κ wartości współczynników równe zeru, zaś dla wielomianów stopnia $s = 2$ estymaty sprowadzają się do estymat liniowych. W pracy [14] wykazano, że przy użyciu wielomianów stopnia $s = 3$ znalezienie estymat parametrów wg algorytmu PMM dla rozkładów symetrycznych wymaga rozwiązania równania stochastycznego trzeciego stopnia o postaci:

$$a\theta^3 + b\theta^2 + c\theta + d \Big|_{\theta=\hat{\theta}} = 0, \quad (11)$$

gdzie: $a = \gamma_4$, $b = -3\gamma_4\hat{\alpha}_1$,

$c = 3\gamma_4\kappa_2 - (6 + 12\gamma_4 + \gamma_6)\kappa_2 + 3\gamma_4[\hat{\alpha}_2 - \kappa_2]$,

$d = (6 + 12\gamma_4 + \gamma_6)\kappa_2\hat{\alpha}_1 - \gamma_4\hat{\alpha}_3$;

kumulanty: $\kappa_1 = m_1$, $\kappa_2 = m_2$, $\kappa_3 = m_3$, $\kappa_4 = \mu_4 - 3\mu_2^2$,

i ich współczynniki: $\kappa_6 = \mu_6 - 15\mu_2\mu_4 - 10\mu_3^2 + 30\mu_2^3$.

We wzorze (11) statystyki $\hat{\alpha}_i$, $i = \overline{1, 3}$, podobnie jak i we wzorze (8), są momentami początkowymi próbki, zaś κ_2 , γ_4 i γ_6 – to kolejno: kumulant rzędu 2. i współczynniki kumulantów rzędów 4. i 6. zmiennej losowej ξ_0 . Wartości kumulantów rozkładu normalnego powyżej drugiego rzędu są równe zeru. Równanie stochastyczne (11) rozwiązuje się analitycznie stosując wzory Cardana [14].

Prostsze obliczeniowo jest rozwiązanie oparte na wykorzystaniu numerycznych iteracyjnych metod rozwiązywania równań nieliniowych. Jako pierwsze przybliżenie można posłużyć się estymacją średniej arytmetycznej (oszacowanie pierwszego momentu początkowego). Wraz ze wzrostem liczby elementów próbki wartość statystyki θ zbliża do rzeczywistej wartości esty-

mowanego parametru. Zwykle wystarcza tylko jedna iteracja. Stosując metodę Newtona-Raphsona, otrzyma się przybliżone rozwiązanie równania trzeciego stopnia (11) w postaci

$$\hat{\theta} \approx \hat{\alpha}_1 - \frac{a\hat{\alpha}_1^3 + b\hat{\alpha}_1^2 + c\hat{\alpha}_1 + d}{3a\hat{\alpha}_1^2 + 2b\hat{\alpha}_1 + c}. \quad (12)$$

Otrzymywany metodą PMM estymator $\hat{\theta}$ wartości menzurandu, podobnie jak inne estymatory dla rozkładów trapezowych [7–11], będzie niewiele różnić się od wartości średniej próbki. Zwykle mieści się on przedziale $\pm S[\bar{X}]$ jej niepewności standardowej.

4.3. Dokładność estymat odchylenia standardowego według metody PMM dla rozkładów symetrycznych

Estymator wariancji σ_θ^2 średniej arytmetycznej (identyczny jak dla metody PMM przy $s = 1$) nie zależy od wartości oszacowania parametru θ . Określa go wariancja drugiego rzędu próbki m_2 (równa jej kumulantowi κ_2) podzielona przez liczbą jej danych n , tj.

$$\sigma_{(\theta)1}^2 = \frac{\kappa_2}{n} \quad (13)$$

Ze wzoru opisującego wielkość uzyskanej informacji o badanym parametrze, w pracach [13, 14] otrzymano wyrażenie analityczne dla asymptoty wariancji $\sigma_{(\alpha)3}^2$ przy $n \rightarrow \infty$, oszacowanej metodą PMM jako

$$\sigma_{(\theta)3}^2 = \frac{\kappa_2}{n} \left[1 - \frac{\gamma_4^2}{6 + 9\gamma_4 + \gamma_6} \right] \quad (14)$$

Z (11) i (12) otrzymuje się współczynnik stosunku wariancji:

$$g_{(\theta)3} = \frac{\sigma_{(\theta)3}^2}{\sigma_{(\theta)1}^2} = 1 - \frac{\gamma_4^2}{6 + 9\gamma_4 + \gamma_6} \quad (15)$$

Wartości $g_{(\theta)3}$ należą do przedziału (0; 1]. Zależą one tylko od właściwości rozkładu prawdopodobieństwa określonych przez współczynniki kumulantów γ_4 i γ_6 .

5. Statystyczne modelowanie estymatorów

Implementacja modelowania statystycznego została wykonana za pomocą pakietu oprogramowania w środowisku MATLAB. Wykorzystuje się tu metodę Monte-Carlo (MC) opartą na wielokrotnie powtarzanych testach o losowo zmienianych danych. Umożliwiła ona przeprowadzenie analizy dokładności proponowanych algorytmów statystycznej estymacji wielomianowej i zbadanie właściwości probabilistycznych uzyskanych oszacowań. Jako kryteria do porównywania skuteczności zastosowano wyznaczane eksperymentalne stosunki wariancji

$$\hat{g}_{(\theta)3} = \frac{\hat{\sigma}_{(\theta)3}^2}{\hat{\sigma}_{(\theta)1}^2}, \quad \hat{q}_{(\theta)3} = \frac{\hat{\sigma}_{(\theta)3}^2}{\hat{\sigma}_{(\theta)V/2}^2}, \quad (16)$$

gdzie $\hat{\sigma}_{(\theta)V/2}^2$, $\hat{\sigma}_{(\theta)1}^2$, $\hat{\sigma}_{(\theta)3}^2$ – uśrednione dla M prób MC wartości estymatorów wariancji parametrów θ . Są one obliczane dla statystyk środka rozpięcia $q_{V/2}$, średniej arytmetycznej \bar{X} i dla estymatora metody PMM o stopniu wielomianu $r = 3$.

Wiarygodność wyników symulacji algorytmami estymacji statystycznej zależy od dwu czynników: wielkości próbki, czyli liczby n elementów wektora x i od liczby M statystycznych eksperymentów wykonanych w jednakowych warunkach początkowych (stała wartość β trapezu).

Obliczenia wartości estymat metodą wielomianową PMM nie wymagają informacji o rodzaju rozkładu. Korzysta się z wartości trzech parametrów modelu: κ_2 , γ_4 i γ_6 . W badaniach opisanych w pracy [14] wartości tych parametrów modelu obliczono z analitycznych wyrażeń wiążących parametry rozkładu gęstości prawdopodobieństwa (pdf), tj. momenty początkowe, odpowiednie kumulanty i współczynniki kumulantów. Jednak w praktyce, gdy informacje na temat rozkładu pdf i/lub wartości ich parametrów nie jest znana *a priori*, można zastosować podejście adaptacyjne wykorzystując w danym badaniu oszacowania *a posteriori*. Użykuje się je z zależności asymptotycznych dla populacji

$$\hat{\kappa}_2 = \hat{\mu}_2, \quad \hat{\gamma}_4 = \frac{\hat{\mu}_4}{\hat{\mu}_2^2} - 3, \quad \hat{\gamma}_6 = \frac{\hat{\mu}_6}{\hat{\mu}_2^3} - 15 \frac{\hat{\mu}_4}{\hat{\mu}_2^2} + 30 \quad (17)$$

gdzie $\hat{\mu}_i$ – moment centralny i -tego rzędu próbki

$$\hat{\mu}_i = \frac{1}{n} \sum_{v=1}^n (x_v - \bar{x})^i \quad (18)$$

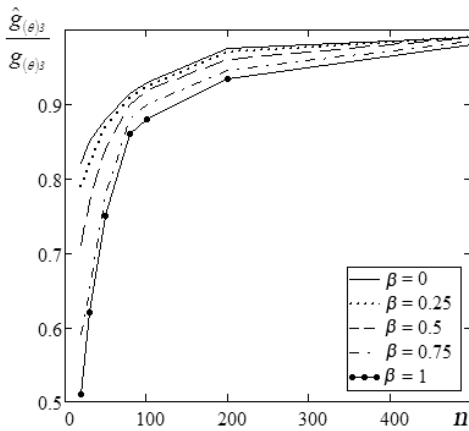
Tabela 1 przedstawia wyniki badań otrzymanych z symulacji Monte Carlo.

Tabela 1. Współczynniki stosunku wariancji estymat
Table 1. The coefficients of the variance ratio of estimates

β	Teoret.	Rezultaty modelowania					
		$\hat{g}_{(\theta)3} = \hat{\sigma}_{(\theta)3}^2 / \hat{\sigma}_{(\theta)1}^2$			$\hat{q}_{(\theta)3} = \hat{\sigma}_{(\theta)3}^2 / \hat{\sigma}_{(\theta)V/2}^2$		
		n					
		20	50	200	20	50	200
$\beta = 1$	0,3	0,56	0,36	0,32	2,15	3,51	10,4
$\beta = 0,75$	0,36	0,61	0,45	0,38	1,53	1,29	1,04
$\beta = 0,5$	0,55	0,78	0,63	0,57	1,02	0,85	0,74
$\beta = 0,25$	0,76	0,97	0,86	0,79	0,9	0,77	0,71
$\beta = 0$	0,84	1,03	0,95	0,87	0,84	0,76	0,69

Współczynniki ilorazu dwu wariancji $\hat{g}_{(\theta)3}$, $\hat{q}_{(\theta)3}$ uzyskano doświadczalnie z $M = 10^4$ prób wykonanych dla różnych wartości parametru β . Analiza danych otrzymanych metodą Monte Carlo wykazuje dużą korelację między obliczeniami analitycznymi i wynikami modelowania statystycznego. Wraz ze wzrostem liczby danych n w próbce x , zmniejsza się różnica między teoretycznymi wartościami (dla populacji) współczynników stosunku wariancji $g_{(\theta)3}$ i wartościami eksperymentalnymi $\hat{g}_{(\theta)3}$ wyznaczanymi z próbki. Na przykład dla $n = 50$ różnica ta nie przekracza 20%, a przy $n = 200$ spada już poniżej 10%. Zmiany stosunku $\hat{g}_{(\theta)3} / g_{(\theta)3}$ w funkcji liczby elementów próbki n dla kilku wartości współczynnika β podaje rys. 4.

Analiza wyników symulacji statystycznej MC z tabeli 1 potwierdza też, że efektywność zastosowania danej statystyki do oszacowania niepewności parametrów rozkładu istotnie zależy od przyjętego modelu opisującego losowy rozrzut pomiarów i od liczebności n próbek.

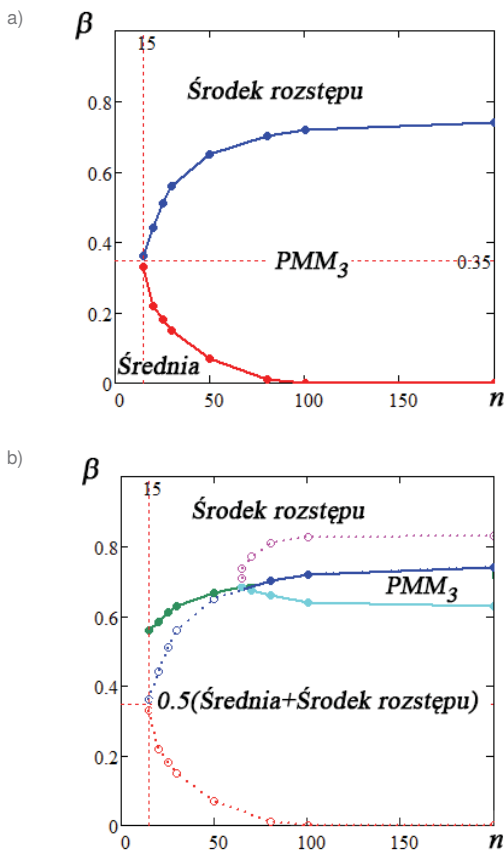


Rys. 4. Stosunek wartości współczynników wariancji próbek o n danych obliczonych w symulacji MC oraz dla populacji Trap

Fig. 4. Ratio of the variance coefficients calculated for n -element samples in MC simulation and theoretical of population Trap

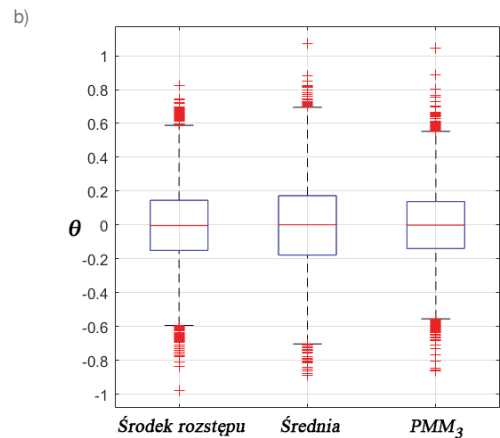
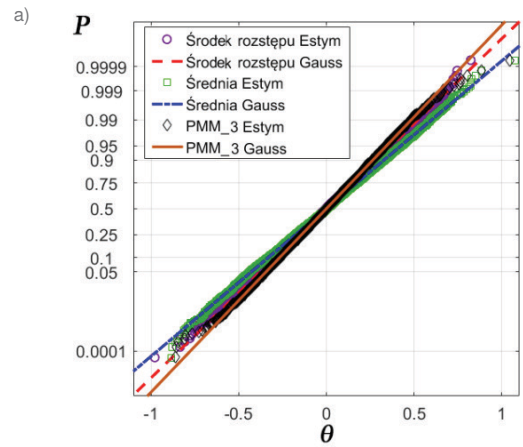
Z wyników wielu powtórzonych obliczeń metodą MC próbek o $n = 15-200$ danych otrzymano rysunki 5a, b. Rysunek 5a przedstawia dla wartości parametrów n i β granice obszarów największej efektywności trzech estymatorów menzurandu: parametru θ wg metody wielomianowej PMM, średniej arytmetycznej \bar{X} i środka rozstępu $q_{V/2}$. Otrzymano je przy porównywaniu według kryterium minimalnej wariancji. Dla próbek z rozkładu Trap występuje tu podana [7, 8, 15] granica obszaru $1 \geq \beta > 0,35$ o mniejszej wartości SD dla środka rozpięcia niż średniej arytmetycznej.

Na rysunku 5b podano też estymator dwuelementowy $\hat{X} = 0,5(\bar{X} + q_{V/2})$. W całym zakresie stosunku podstaw β jest on bardziej efektywny niż estymatory jednoelementowe (rys. 3).



Rys. 5 a, b. Obszary efektywności metod znajdowania estymat standardowego odchylenia rozkładu trapezowego

Fig. 5 a, b. Areas of effectiveness of methods for finding estimates of the trapezoidal distribution standard deviation



Rys. 6. Empiryczne rozkłady estymat parametru SD: a) wykres probabilistyczny (Q-Q wykres) przybliżenia do funkcji Gaussa; b) wykres typu Box-plot (dla przedziału ufności 99%)

Fig. 6. The empirical distribution of estimates of the SD parameter: a) the probabilistic graph (Q-Q plot) of Gaussian approximation; b) a plot of the type Box-plot (99% confidence interval)

Przykłady wyników modelowania podano w postaci wykresów na rysunkach 6 a, b. Rysunki te zawierają oceny parametru $\theta = 0$ dla rozkładu Trap przy $\beta = 0,5$ i $\lambda = 6$. Przedstawiono je jako rozkłady otrzymanych eksperymentalnie ocen liczbowych dla wartości menzurandu wg metody PMM_3 ($r = 3$) oraz dla środka rozpięcia i średniej arytmetycznej. Otrzymano je dla $M = 10^4$ prób metodą MC dla próbek o $n = 50$ danych.

Hipotezę o gaussowskim rozkładzie estymat metody PMM sprawdzono także za pomocą wbudowanego w oprogramowanie MATLAB testu Lillieforsa. Opiera się na statystyce Kolmogorowa-Smirnowa. Wyniki tych badań przedstawiono w tabeli 2.

Tabela 2. Wyniki kontroli hipotez o gaussowości estymat rozkładu empirycznego wyznaczonych metodą PMM_3 ($s = 3$)

Table 2. Results of testing hypotheses of Gaussianity of the empirical distribution estimates, found by PMM_3 method (for $s = 3$)

β	CV	Wyniki testu Lillieforsa					
		LSTAT			P		
		n					
		20	50	200	20	50	200
$\beta = 1$	0,009	0,029	0,021	0,008	0,001	0,001	0,12
$\beta = 0.75$		0,028	0,016	0,007	0,001	0,002	0,18
$\beta = 0.5$		0,014	0,007	0,005	0,002	0,18	0,5
$\beta = 0.25$		0,01	0,008	0,005	0,006	0,19	0,5
$\beta = 0$		0,008	0,007	0,005	0,18	0,21	0,5

Jest to zestaw parametrów wyjściowych testu: CV – krytyczna wartość statystyki testu; $LSTAT$ – wybrana wartość statystyki testowej; P – poziom istotności. Jeśli $LSTAT < CV$ to przy zadanym poziomie istotności $\alpha_0 = 0,05$ hipotezy zerowej (Gaussa) nie odrzuca się.

Ważnym rezultatem modelowania statystycznego jest też potwierdzenie tezy, że rozkład estymat parametrów wielomianu optymalizowanego metodą PMM dąży asymptotycznie (przy $n \rightarrow \infty$) do rozkładu Gaussa. Można to zastosować do wyznaczania niepewności rozszerzonej.

6. Wnioski końcowe

We wprowadzeniu dokonano przeglądu estymatorów służących do oceny wyniku pomiaru mierzand dla próbek z rozkładu równomiernego i rozkładu trapezowego Trap jako splotu dwu rozkładów równomiernych. Następnie omówiono zastosowanie do tego celu niekonwencjonalnej Metody Maksymalizacji Wielomianu o akronimie PMM. Metoda ta, zwana w skrócie wielomianową, opiera się na pierwszych kilku kumulantach wyznaczanych pośrednio z momentów centralnych próbek.

Za pomocą symulacji metodą Monte Carlo zbadano wyniki uzyskiwane przy stosowaniu metody PMM do oceny estymatorów parametrów mierzand w sytuacji, gdy rozkład populacji, z której pochodzi próbka, nie jest znany a priori, a liczba danych pomiarowych jest za mała do jednoznacznej jego identyfikacji. Jako przykład wykonano badania symulacyjne dla wielomianu rzędu 2. i próbek danych pobranych z kilku rozkładów rodziny trapezów liniowych Trap.

Łączna analiza uzyskanych wyników wykazuje, że dla określonych wartości parametrów opisujących właściwości probabilistyczne rozkładu Trap (w całym zakresie $0 \leq \beta \leq 1$ stosunku długości podstaw trapezu), estymaty odchylenia standardowego SD obliczone metodą Monte Carlo dla metody wielomianowej PMM₃ są bardziej efektywne (mają mniejszą wariancję) w porównaniu do estymat SD obliczonych dla średniej arytmetycznej i środka rozstępu.

Względne zmniejszenie się wariancji estymat, jak i stopień normalizacji ich rozkładu w dużej mierze zależy od liczby n danych próbki. Na przykład dla próbek o liczbie elementów $n > 50$, z populacji TRAP o stosunku długości podstaw trapezu $\beta < 0,7$, metodą PMM otrzymuje się bardziej efektywną wartość mierzand, tj. o mniejszej wariancji niż średnia wg Przewodnika GUM i środek rozpięcia – patrz granice obszaru na rys 2.

Omówione badania pozwalają sformułować wniosek, iż metodą wielomianową PMM można wykorzystywać z powodzeniem do oszacowania przedziału niepewności standardowej u_A oraz niepewności rozszerzonej U mierzand dla próbek danych z rodziny rozkładów trapezowych Trap.

Bibliografia

1. Evaluation of measurement data – Guide to the expression of uncertainty in measurement (GUM), BIPM, JCGM 100 (2008) + Supplement 1 – Propagation of distributions using a Monte Carlo method (2007).
2. Bich W., Cox M., Michotte C., Towards a new GUM-an update. *Metrologia* 53 (2016) S149–S159 IOP publ. BIPM.
3. Cox M., Shirono K., *Informative Bayesian Type A uncertainty evaluation, especially applicable to a small number of observations*. “Metrologia”, Vol. 54, No. 5, 642–652.
4. Cramer H., *Mathematical Methods of Statistics*, Stockholm University 1946 [*Metody matematyczne w statystyce*, z ang. tł. W. Oktaba, PWN Warszawa 1958].
5. Novitski P.V., Zograf I.A., *Ocenka pogreshnostiej rezultatov izmerenia*. Energoatomizdat, Leningrad 1985.
6. Kubisa S., Warsza Z.L., *Środek rozstępu jako estymator mierzand dla próbek z populacji o rozkładzie równomiernym i płasko-normalnym*. „Pomiary Automatyka Kontrola”, Vol. 60, No. 6, 2014, 398–401.
7. Warsza Z.L., Galovska M., *About the best measurand estimators of trapezoidal probability distributions*. “Przegląd Elektrotechniczny – Electrical Review”, No. 5, 2009, 86–91.
8. Warsza Z.L., Galovska M., *The best measurand estimators of trapezoidal PDF*. Proceedings of IMEKO World Congress “Fundamental and Applied Metrology”, September 2009, Lisbon Portugal, 2405–2410.
9. Warsza Z.L., *Dwuelementowe estymatory wartości mierzand próbek danych pomiarowych o trapezowych rozkładach prawdopodobieństwa – przegląd prac*, „Pomiary Automatyka Kontrola”, Vol. 57, No. 1, 2011, 105–108.
10. Warsza Z.L., Wyznaczanie niepewności próbek pomiarowych o kilku niegaussowskich rozkładach prawdopodobieństwa. *Elektronika (Technika Sensorowa)*, Nr 9, 2012, 135–140.
11. Warsza Z.L., *Effective Measurand Estimators for Samples of Trapezoidal PDFs*. “Journal of Automation, Mobile Robotics and Intelligent Systems”, Vol. 6, No. 1, 2012, 35–41.
12. Botsiura O.A., Zakharov I.P., *Peculiarities of evaluation of measurement uncertainty type A based on a Bayesian approach*. “Sistemy Obrotki Informacji (SOI)”, Kharkiv 2015, Vypusk 6 (131) 17–20 (in Russ.).
13. Kunchenko Y., *Polynomial Parameter Estimations of Close to Gaussian Random Variables*. Shaker Verlag Aachen Germany, 2002.
14. Zabolotni S.W., Warsza Z.L., *A polynomial estimation of measurand parameters for samples of non-Gaussian symmetrically distributed data*. R. Szewczyk et al (ed.) Proceedings of Automation 2017 – Innovations in Automation, Robotics and Measurement Techniques. Advances in Intelligent Systems and Computing 550. Springer International Publishing AG 2017, 470–482.
15. Warsza Z.L., *Metody rozszerzenia analizy niepewności pomiarów*. Monografie • Studia • Rozprawy. Oficyna Wydawnicza PIAP Warszawa 2016.

Evaluation of the Uncertainty of Trapeze Distributed Measurements by the Polynomial Maximization Method

Abstract: The types of measurand parameter estimators derived from samples of measured data taken from the symmetrical trapezoidal population Trap are briefly discussed. A non-standard approach to finding estimates of the non-Gaussian distributions parameters based on the unconventional method of maximizing the stochastic polynomial (PMM) and using a moment-cumulant description of random variables is proposed. By means of multiple statistical tests of Monte Carlo method, the properties of polynomial estimators are investigated and an analysis of their accuracy is made with compare to estimates of the distributions with arithmetic mean or the mid-range as their centers. As a function of the number of sample data and the basis of trapeze ratio, the boundaries of the areas where these methods are most effective are determined. The PPM method has been proposed to use for determining estimated values of the standard deviation and uncertainties of measurand when distribution of the random errors population is a priori unknown and first few cumulants have to be find from the sample data.

Keywords: measurand, estimate, standard deviation, trapezoidal distribution, cumulant, polynomial maximization method

doc. dr inż. Zygmunt Lech Warszawa

zlw1936@gmail.com

Ukończył Wydział Elektryczny Politechniki Warszawskiej 1959, doktorat 1967, docent od 1970. Praca: Instytut Elektrotechniki 1958–1963 i 1994–1995, Politechnika Warszawska 1960–70, Politechnika Świętokrzyska 1970–1978 (organizator i dziekan Wydziału Transportu w Radomiu), Organizator i kierownik: Ośrodka Aparatury Pomiarowej w Instytucie Meteorologii i Gospodarki Wodnej 1978–1982 oraz Zakładu Automatykacji i Pomiarów w Instytucie Chemii Przemysłowej 1983–1992. Doradca Ministra Edukacji Narodowej 1992–1995, Politechnika Radomska 1983–2002. Obecnie główny specjalista w Przemysłowym Instytucie Automatyki i Pomiarów PIAP. Autor ponad 250 publikacji, 6 monografii, kilkadziesiątu prac badawczych i konstrukcyjnych, 11 patentów oraz promotor 2 doktorów. Prezes Polskiego Towarzystwa Metrologicznego. Członek stowarzyszenia PolSPAR oraz Akademii Metrologii Ukrainy.



Prof. Dr Tech. Sc., El. Eng.

Serhii V. Zabolotnii

s.zabolotnii@chdtu.edu.ua

Urodzony w 1973 r. w Cherkasach, Ukraina. Ukończył studia na Wydziale Techniki i Systemów Informatycznych Narodowego Uniwersytetu Technologicznego w Cherkasach w 1995 r. Stopień kandydata nauki (dr) z systemów informacyjno-pomiarowych 2000 r., dr nauk techn. (równoważny habilitacji) 2015 r. Praca – Wydział Radio-techniki Narodowego Uniwersytetu Technologicznego w Cherkasach: asystent 2003 r., docent od 2003 r., stanowisko profesora NUT w Cherkasach 2015 r. Autor lub współautor ponad 100 publikacji, 1 monografii i 7 patentów.



