

JAKUBIEC Beata¹
BĄBKA Robert²

ZASTOSOWANIE GENERATORÓW PSEUDOLOSOWYCH W WYZNACZANIU NIEPEWNOŚCI WYNIKÓW NUMERYCZNYCH

Wyznaczenie struktury i parametrów modelu matematycznego procesu zapisanego w postaci równań prowadzi do uzyskania wyników, których interpretacja pozwala na wysunięcie wniosków jakościowych i ilościowych w zakresie modelowanego zjawiska. W artykule przedstawiono zastosowanie generatorów pseudolosowych do oszacowania niepewności wyników numerycznych w celu oceny jakości modelu matematycznego pod względem dokładności otrzymanych wyników.

NUMERICAL METHOD UNCERTAINTY ESTIMATION BY MEANS OF PSEUDORANDOM NUMBER GENERATOR

Using mathematical equations as a model of the considered process, obtained result depends on its quality. The one of main problem includes the sensitivity of model parameters for accuracy of the numerical results. The paper deals with application of pseudorandom generators in testing dependencies between variation of model coefficients and uncertainty of obtained values estimated via computer simulation.

1. WSTĘP

Jakościowy i ilościowy opis rzeczywistości dokonywany jest za pomocą modeli matematycznych [1,2]. Podejście takie pozwala na sformalizowanie opisu zjawiska oraz wykorzystanie matematycznych narzędzi komputerowych do uzyskania wyników niezbędnych w konkretnej aplikacji. W zastosowaniach inżynierskich najistotniejszą grupą narzędzi wspomaganego jest oprogramowanie typu CAD (ang. *Computer Aided Design*) oraz CAS (ang. *Computer Algebra System*). Nazwa ta obejmuje programy komputerowe zaprojektowane tak, aby skutecznie wspierać rozwiązywanie problemów technicznych danej klasy. Niezależnie, czy wspomaganie komputerowe ma polegać na przetwarzaniu dużej ilości danych, czy realizować złożone obliczenia optymalizacyjne, działanie to sprowadza się do uzyskania wyników numerycznych wyznaczanych na podstawie zadanego algorytmu obliczeniowego [3]. Prawidłowe posługiwanie się narzędziami wspomaganego projektowania wymaga zatem świadomej analizy pojawiających się problemów numerycznych. Do analizy

¹Politechnika Częstochowska, Wydział Elektryczny; 42-200 Częstochowa; Al. Armii Krajowej 17.

²Techniczne Zakłady Naukowe; 42-200 Częstochowa; ul. Jasnogórska 84/90.

tego typu służy oprogramowanie matematyczno-inżynierskie, w którym istnieje możliwość badania właściwości algorytmów rozwiązywania zagadnień technicznych oraz testowania ich w celu późniejszej implementacji w programach CAD. Przykładem oprogramowania tego typu może być Matlab [6].

Opracowanie modelu matematycznego zjawiska lub procesu ma na celu przeniesienie zależności występujących w rzeczywistości na poziom opisu równaniami. Eksperymentalna weryfikacja takiego modelu prowadzi wówczas do sprawdzenia prawidłowego odwzorowania przez równania matematyczne [4,6]. Istotnym problemem, a często pomijanym w praktyce inżynierskiej jest ocena jakości modelu z punktu widzenia wrażliwości wyniku na zmianę wartości jego parametrów. Oznacza to w praktyce, że otrzymanie oczekiwanych wyników może nie być tak oczywiste jak przy założeniu stałych wartości parametrów układu. Przykładem może być na przykład nieuwzględnienie tolerancji wartości rezystancji rezystorów w procesie produkcyjnym podczas symulacji całego układu. Pominięcie takiej analizy może zniweczyć założenia projektu. Innym przykładem jest złe uwarunkowanie numeryczne, gdy macierze równań opisujących model nie są osobliwe, ale ich wyznaczniki są bliskie zeru. Otrzymany wynik przy zadanych parametrach modelu będzie praktycznie bezużyteczny.

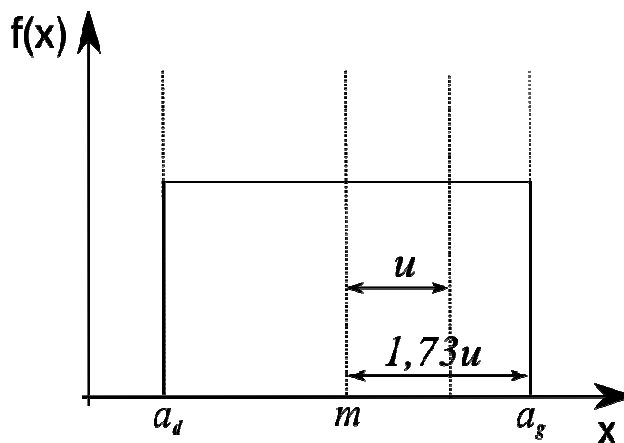
Z uwagi na zastosowanie praktyczne projektowania komputerowego modele wykorzystywane w warunkach przemysłowych mają zazwyczaj bardzo złożoną strukturę rozwiązywaną wyłącznie metodami numerycznymi. W celu sprawdzenia wrażliwości modelu tego typu stosowane są metody Monte Carlo polegające na statystycznej zmianie wartości parametrów oraz oszacowania na tej podstawie zmian wartości zmiennych wyjściowych. Skuteczność metod uwarunkowana jest głównie dokładnością odwzorowania zadanych empirycznych rozkładów prawdopodobieństwa za pomocą generatorów pseudolosowych. Dokładność ta zależy od metody generowania próbek, jak i ich liczby [3,4]. Poniżej przedstawiono praktyczny przykład zastosowania generatora pseudolosowego do wyznaczenia wrażliwości parametrów modelu na otrzymane wyniki symulacji komputerowej.

2. PSEUDOLOSOWA APROKSYMACJA ROZKŁADÓW PRAWDOPODOBIENSTWA

Generowanie przebiegu pseudolosowego związane jest z zachowaniem parametrów statystycznych zbioru próbek zgodnie z założonym rozkładem gęstości prawdopodobieństwa. Problem ten jest istotny nie tylko przy modelowaniu statystycznym procesów, lecz również w kryptografii. Jednym ze sprawdzonych algorytmów generacji liczb pseudolosowych jest oparty na liniowej rekurencji macierzowej i opracowany w 1997 roku przez Nishimurę i Matsumoto, algorytm oznaczony jako MT19937 [4]. Główną cechą tego generatora jest bardzo duży okres powtarzania, wynoszący 2219937 iteracji, przy dobrej szybkości działania. Ponieważ algorytm został zaprojektowany do modelowania metodami Monte Carlo, posiada również większą równomierność generowania liczb pseudolosowych niż inne algorytmy. W praktyce oznacza to uzyskanie dobrej niezależności zmiennych losowych. Cechy te zostały zweryfikowane, zarówno za pomocą dowodów twierzeń, jak i zestawu testów dla generatorów pseudolosowych określanych nazwą *diehard*.

Podstawowymi rozkładami losowymi mającymi zastosowanie w teorii propagacji niepewności jest rozkład normalny oraz prostokątny. Rozkład prostokątny służy jako model występowania wielkości z jednakowym prawdopodobieństwem w zadanych granicach od a_d do a_g (rys. 1). Jego gęstość prawdopodobieństwa przedstawia zależność [7]:

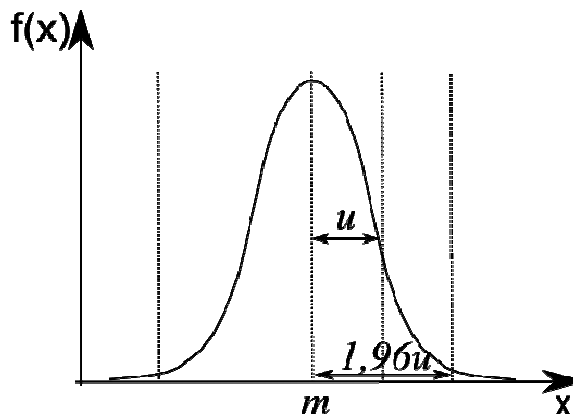
$$f(x) = \begin{cases} \frac{1}{a_g - a_d}, & x \in \langle a_d, a_g \rangle \\ 0 & , x \notin \langle a_d, a_g \rangle \end{cases} \quad (1)$$



Rys. 1. Funkcja gęstości prawdopodobieństwa rozkładu prostokątnego

Kolejnym rozkładem odzwierciedlającym z kolei zmienność parametrów zjawisk fizycznych jest rozkład normalny (rys. 2). Rozkład ten jest jednocześnie rozkładem granicznym dla splotu rozkładów od wielu zmiennych, dlatego też służy do modelowania parametrów fizycznych dla których występuje jednoczesny wpływ kilku czynników na raz. Gęstość prawdopodobieństwa rozkładu normalnego opisuje równanie [7]:

$$f(x) = \frac{1}{\sqrt{2\pi}u} \exp\left(\frac{-(x-m)^2}{2u^2}\right) \quad (2)$$



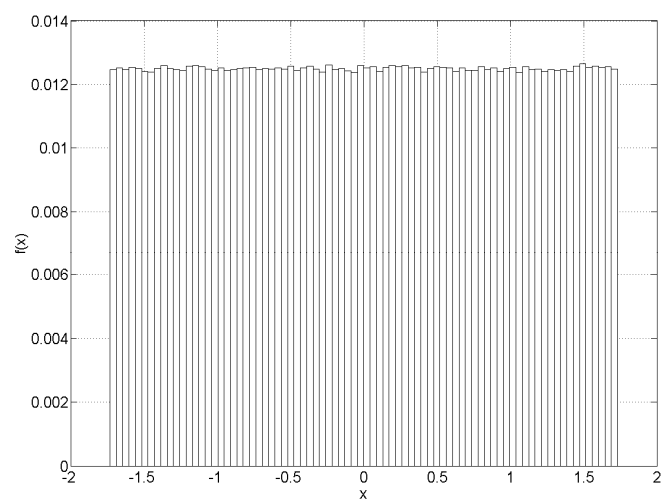
Rys. 2. Funkcja gęstości prawdopodobieństwa rozkładu normalnego

Do opisu jednego jak i drugiego rozkładu prawdopodobieństwa służą momenty pierwszego oraz drugiego rzędu. Pierwszą wielkością jest wartość oczekiwana – m będąca jednocześnie medianą dla każdego z nich. Wielkość u określa zakres zmienności zmiennej losowej względem wartości wielkości m . Istotne znaczenie wymienionych parametrów ma możliwość ich empirycznej estymacji na podstawie losowanych próbek, co wyrażają zależności [7]:

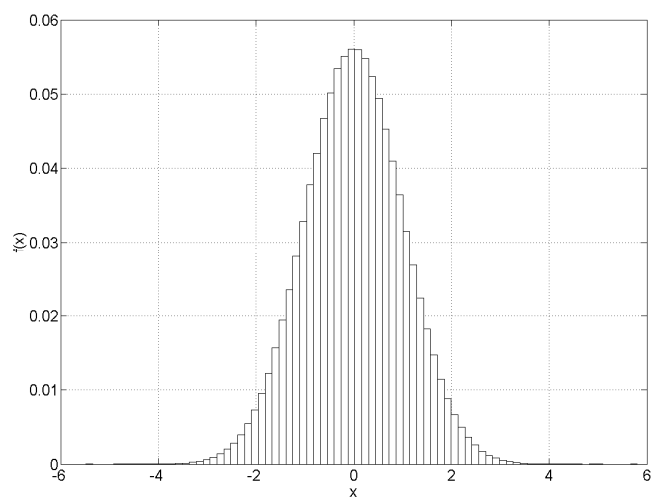
$$\hat{m} = \frac{1}{N} \sum_{i=1}^N x_i \quad (3)$$

$$\hat{u} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \hat{m})^2} \quad (4)$$

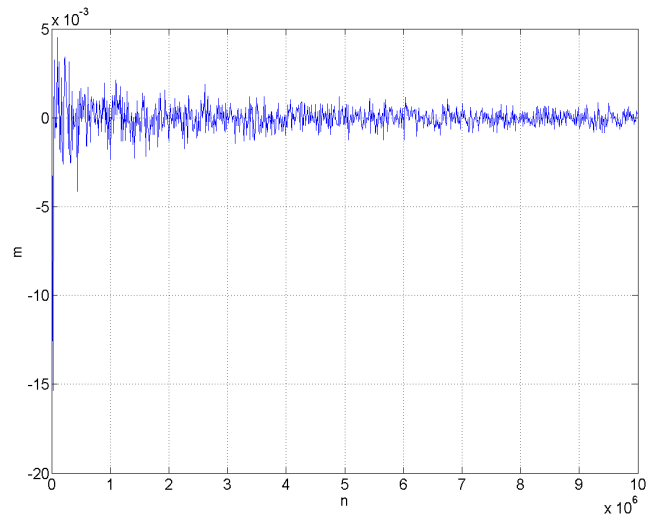
Wyznaczenie estymatorów poszczególnych rozkładów pozwala na ocenę przydatności generowanych próbek za pomocą generatora pseudolosowego. Na rysunkach 3 i 4 przedstawiono histogramy dla wymienionych rozkładów losowych wygenerowanych generatorem MT19937. Wylosowano 1e6 próbek dla każdego rozkładu. Widoczne na wykresie zmiany wysokości histogramu wynikają zarówno ze zliczania wartości próbek dla klas w histogramie, jak i z niedokładności samego algorytmu. Przy zadanych wartościach $m=0$ oraz $u=1$ otrzymane parametry rozkładów losowanych komputerowo nie odbiegają do zadanych o więcej niż 0,1%. W przypadku zmieniającej się liczby próbek dokładność ta się zmienia. Zależność tych zmian dla rozkładu prostokątnego i wymienionego generatora losowego przedstawiono na rysunkach 5 i 6. Wykresy pozwalają na oszacowanie liczby próbek wystarczającej do uzyskania zadanej dokładności szacowania parametrów wybranego generowanego rozkładu losowego.



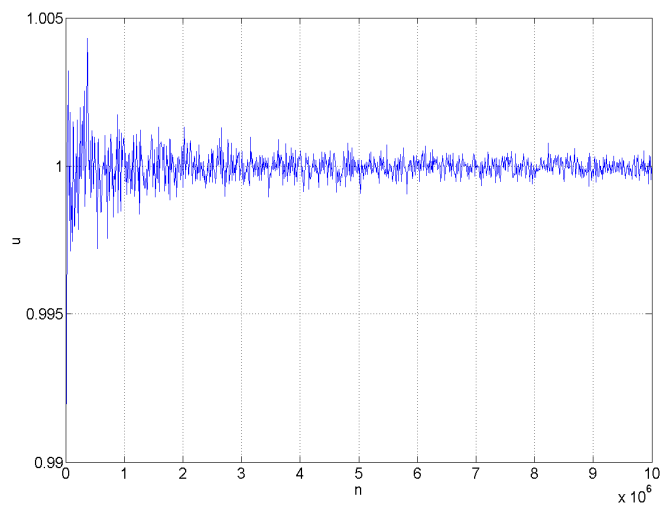
Rys. 3. Histogram numerycznego prostokątnego rozkładu gęstości prawdopodobieństwa dla $1e6$ próbek i założonych parametrów $m=0$, $u=1$. Wyznaczona z próbek wartość oczekiwana wynosi $m=0,0011$ i $u=1,001$



Rys. 4. Histogram numerycznego normalnego rozkładu gęstości prawdopodobieństwa dla $1e6$ próbek i założonych parametrów $m=0$, $u=1$. Wyznaczona z próbek wartość oczekiwana wynosi $m=2.7564e-4$ i $u=1,0002$



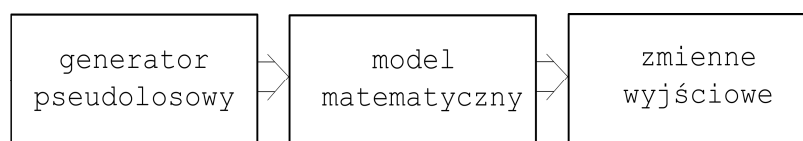
Rys. 5. Zmiany wartości estymatora parametru m dla rozkładu losowego w funkcji liczby generowanych próbek w zbiorze testowym



Rys. 6. Zmiany wartości estymatora parametru u dla rozkładu losowego w funkcji liczby generowanych próbek w zbiorze testowym

3. ANALIZA PROBLEMU NUMERYCZNEGO

Koncepcja wyznaczania niepewności wyniku numerycznego została przedstawiona na rysunku 7. Generator losowy służy do wyznaczania próbek zmiennych wejściowych wprowadzanych do modelu matematycznego. Uzyskane w ten sposób zmienne wyjściowe mają charakter losowy i ich parametry mogą być szacowane na podstawie zależności (3,4). W oprogramowaniu inżynierskim, zmiennymi wejściowymi są parametry elementów konstrukcji lub schematu elektrycznego. Przeprowadzając analizę tego typu można wyznaczyć jak zachowa się dane rozwiązanie techniczne przy założonej dokładności parametrów materiałowych.



Rys. 7. Wyznaczanie niepewności wyniku obliczeń z zastosowaniem generatora zmiennych pseudolosowych

Innym przypadkiem jest opracowanie struktury modelu matematycznego i dobór jego parametrów na podstawie pomiarów z obiektu rzeczywistego. Jest to główne zadanie identyfikacji, na przykład w procesach automatycznej regulacji. Otrzymane współczynniki równań przy nieortogonalnych seriach pomiarowych mogą prowadzić do uzyskania macierzy osobliwych, co w przypadku prostych modeli liniowych jest łatwo wykrywalne. Przy modelach złożonych i nieliniowych występujących w praktyce inżynierskiej zazwyczaj konieczna jest analiza wrażliwości parametrów modelu za pomocą metody symulacyjnej. Odpowiedni dobór generatora pseudolosowego pozwala na wyznaczenie wrażliwości tych parametrów na końcowy wynik obliczeń.

W celu zobrazowania zagadnienia zaprezentowany zostanie prosty problem numeryczny. W przykładzie wybrano model liniowy ze względu na istotne uproszczenie analizy matematycznej. Korzystając z modelu układu równań liniowych zapisanych wyrażeniem (5) należy wyznaczyć wartości niewiadomych, oznaczonych symbolem \mathbf{x} . Macierzowy zapis równań został dostosowany do narzędzia programowego jakim jest pakiet Matlab. Przyjmując wartości parametrów macierzy \mathbf{A} jak w równaniu (6) oraz macierzy \mathbf{B} jak w równaniu (7), rozwiązanie zależności (5) opisuje równanie (8), gdzie podano również obliczone analitycznie wartości \mathbf{x} .

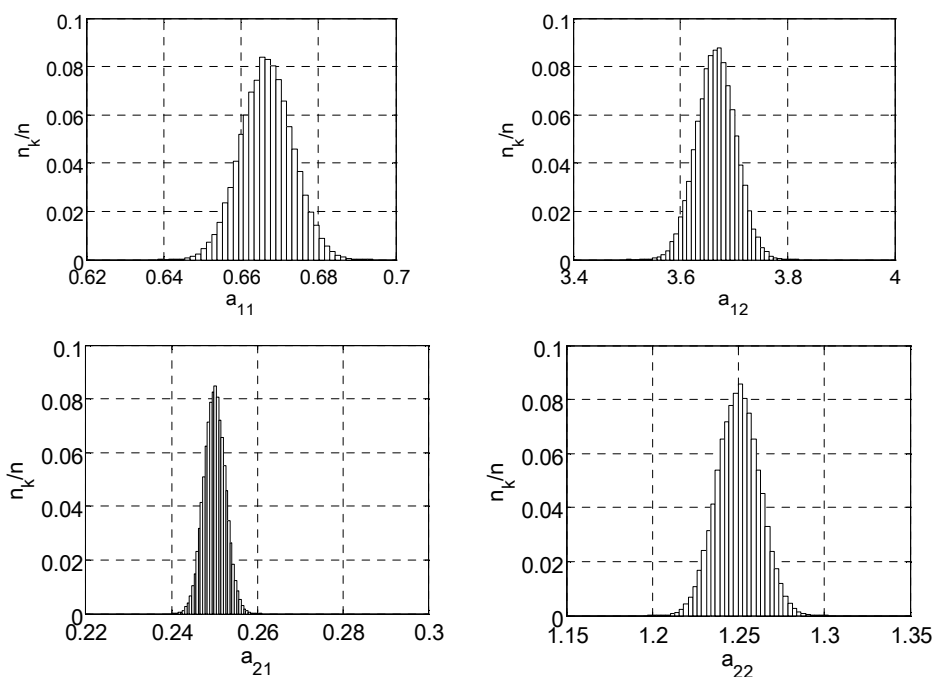
$$\mathbf{Ax} = \mathbf{B} \quad (5)$$

$$\mathbf{A} = \begin{bmatrix} \frac{2}{3} & \frac{11}{3} \\ \frac{1}{4} & \frac{5}{4} \end{bmatrix} \quad (6)$$

$$\mathbf{B} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad (7)$$

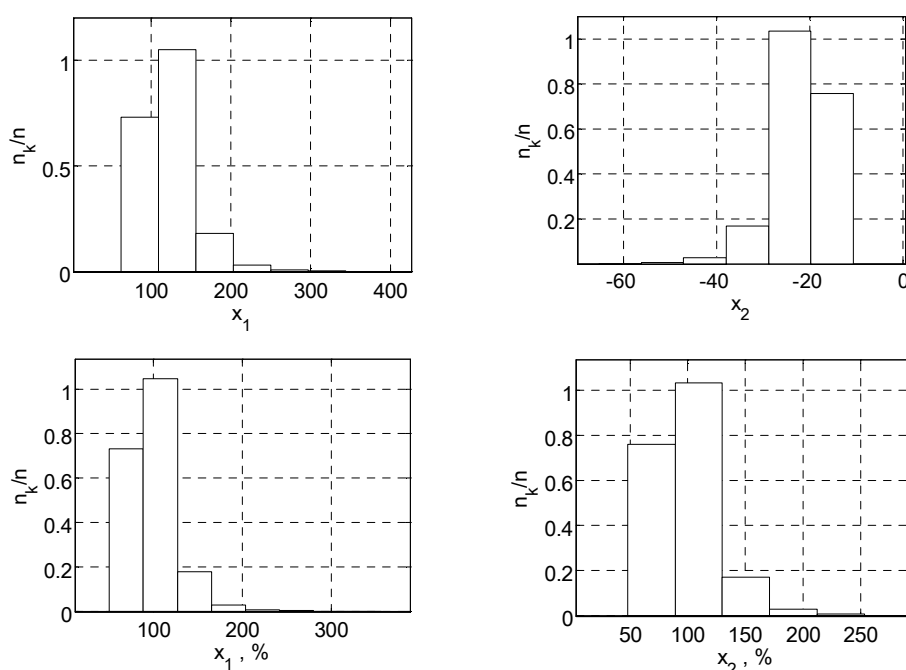
$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{B} = \begin{bmatrix} 117 \\ -21 \end{bmatrix} \quad (8)$$

Przedstawione zadanie, mimo że z matematycznego punktu widzenia jest banalne, w praktyce inżynierskiej może nie dać tak jednoznacznie pewnego wyniku. Elementy macierzy \mathbf{A} opisującej parametry równania liniowego zapisano w postaci ułamkowej dla zachowania dokładności numerycznej. W rzeczywistości, macierzy \mathbf{A} mogą odpowiadać zmierzone parametry pewnego procesu określone z pewną dokładnością wymierną, zdeterminowaną właściwościami przyrządów pomiarowych. Oznacza to, że z punktu widzenia wiarygodności otrzymanego rozwiązania istotnym jest sprawdzenie wpływu zmian wartości tych parametrów na wynik. W celu zbadania znaczenia opisanego zjawiska opracowano statystyczny model numeryczny pozwalający, nie tylko na oszacowanie zakresu zmian wyniku, ale również na określenie rozkładów losowych dla zmiennej \mathbf{x} . W celu uproszczenia eksperymentu założono niezależność losową parametrów macierzy \mathbf{A} , a ich rozkład jako normalny - rys. 8.



Rys. 8. Losowe rozkłady wartości elementów macierzy \mathbf{A} przy założeniu 1% niepewności ich wyznaczania

Przyjęto również, że w wyniku występujących błędów losowych parametrów macierzy \mathbf{A} , odchylenie standardowe od ich wartości rzeczywistych wynosi 1%. Wyniki obliczeń przedstawiono na rysunku 9. Dokładnym, czyli analitycznym rozwiązaniem równania (5) jest wektor $\mathbf{x}=[x_1, x_2]^T = [117, -21]^T$, podczas gdy po uwzględnieniu zamodelowanej niedokładności elementów macierzy \mathbf{A} , x_1 może z dużym prawdopodobieństwem przyjmować wartości 50 albo 200, natomiast x_2 -30 lub -10. Względne wartości graniczne błędu dla rozważanego przypadku mogą więc wynieść 50 %, albo nawet 200%. Przedstawione wykresy wskazują również na asymetrię rozkładu prawdopodobieństwa zaistniałych błędów.



Rys. 9. Zmiany wartości otrzymanego wyniku z powodu 1% niepewności wyznaczenia parametrów macierzy \mathbf{A}

4. WNIOSKI

Komputerowe wspomaganie projektowania złożonych procesów lub urządzeń nie zwalnia od konieczności zrozumienia algorytmów w rozwiązywaniu problemów technicznych. W wielu przypadkach praktyki inżynierskiej ich dobór opiera się jedynie na bezkrytycznym zaufaniu do oprogramowania, traktowanego jako uniwersalne narzędzie rozwiązania zagadnienia projektowego. Pominięcie weryfikacji otrzymanych wyników pod względem wrażliwości na zmianę wartości parametrów modelu matematycznego może prowadzić do uzyskania nieoczekiwanych wyników. Zastosowanie metody testowania wrażliwości z

zastosowaniem generatorów pseudolosowych pozwala na wykrycie osobliwości na wczesnym etapie projektowania nawet dla projektów złożonych. Przytoczone w artykule wartości niepewności dla prostego modelu liniowego wynoszące 1%, założone jedynie w celu demonstracyjnym, w rzeczywistości przy danej technologii lub złych warunkach technicznych mogą być wielokrotnie większe. Aby uzmysłowić skalę zjawiska wystarczy zauważyć, że niepewności wielu przyrządów pomiarowych, standardowo wykorzystywanych w przemyśle, wynoszą niejednokrotnie kilka procent. Doświadczenie przemysłowe pokazuje również, że często zaniedbywanie błędów metod pomiarowych, powoduje wzrost błędów wyznaczania wartości parametrów modeli matematycznych dla projektowanych konstrukcji. Zastosowanie w takich warunkach narzędzi numerycznych z użyciem metod statystycznych pomaga w prawidłowym doborze struktury modelu w celu odzwierciedlenia warunków rzeczywistych. Z uwagi na wzrastającą moc obliczeniową komputerów, metody takie mogą być akceptowalne również pod względem czasu obliczeń. Oznacza to, że mogą one być wykorzystane jako potężne narzędzie analizy problemów stanowiących niejednokrotnie pułapkę dla inżynierów konstruktorów.

5. BIBLIOGRAFIA

- [1] Brzózka J., Dorobczyński L.: *Matlab. Środowisko obliczeń naukowo-technicznych*, Wydawnictwo MIKOM, Warszawa 2005.
- [2] Gutenbaum J.: *Modelowanie matematyczne systemów*, Akademicka Oficyna Wydawnicza EXIT, 2003.
- [3] Kincaid D., Cheney W.: *Analiza numeryczna*, WNT, Warszawa 2006.
- [4] Kołodziński E.: *Symulacyjne metody badania systemów*, PWN, Warszawa 2002.
- [5] Matsumoto M., Nishimura T.: *Mersenne-Twister a 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generator*, ACM Transactions on Modeling and Computer Simulation, vol. 8, No. 1, January 1998, str. 3-30.
- [6] Osowski S.: *Modelowanie układów dynamicznych z zastosowaniem języka SIMULINK*, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 1999.
- [7] Plucińska A., Pluciński E.: *Probabilistyka*, Wydawnictwo Naukowo-Techniczne, Warszawa 2000.