

Sergiusz SIENKOWSKI

UNIwersytet Zielonogórski, Instytut Metrologii Elektrycznej,
ul. Podgórna 50, 65-246 Zielona Góra

Sprawdzanie dokładności szacowania wartości oczekiwanej metodą Monte Carlo

Dr inż. Sergiusz SIENKOWSKI

Absolwent Wydziału Podstawowych Problemów Techniki Politechniki Zielonogórskiej (2001 r.) oraz Wydziału Elektrotechniki, Informatyki i Telekomunikacji Uniwersytetu Zielonogórskiego (2003 r.). Obecnie adiunkt w Instytucie Metrologii Elektrycznej Uniwersytetu Zielonogórskiego. Zajmuje się zagadnieniami związanymi z cyfrowym przetwarzaniem sygnałów i oceną niepewności pomiarów.



e-mail: s.sienkowski@ime.uz.zgora.pl

Streszczenie

Przedmiotem badań jest estymator wartości oczekiwanej. Sprawdzano dokładność estymacji wartości oczekiwanej w sytuacji, gdy estymator obliczany jest na podstawie danych z rozkładu Gaussa. Sprawdzanie dokładności estymacji wykonano z zastosowaniem metody Monte Carlo.

Słowa kluczowe: metoda Monte Carlo, wartość oczekiwana, rozkład Gaussa.

Verification of accuracy of mean value estimation using Monte Carlo method

Abstract

The subject of the research is the mean value estimator. The estimator is determined based on data obtained from a Gaussian distribution. The accuracy of the mean value estimator was examined using the Monte Carlo method. Chapter 1 provides basic information on the reasons for use the Monte Carlo method. In Chapter 2 the basic definitions were presented. Eq. (1) describes the expected value of the random variable. Eq. (3) presents the mean value estimator. Eq. (4) it is the error of the estimator (3). In the next part of Chapter 2 the mean value estimator for Gaussian distribution was presented (eq. 6). Eq. (7) describes the error of the mean value estimator (6). Next equation describes coverage factor for Gaussian distribution. In the next part of the paper the Monte Carlo methods were presented. In this article the Crude and Hit-or-Miss Monte Carlo methods have been used. Eq. (13) presents the mean value estimator obtained using the Crude Monte Carlo method. Eq. (14) describes the error of the estimator. Eq. (15) presents the mean value estimator obtained using the Hit-or-Miss Monte Carlo method. Eq. (16) it is the error of the estimator. In Fig. 1 the errors (4), (14) and (16) have been shown. Tab. 1 presents the errors obtained in Matlab, MatCAD and LabWINDOWS. The researches have been summarized in Chapter 3.

Keywords: Monte Carlo method, mean value, Gaussian distribution.

1. Wprowadzenie

W artykule podjęto problematykę sprawdzania dokładności estymacji wartości oczekiwanej metodą Monte Carlo. Zastosowana metoda opiera się na całkowaniu numerycznym funkcji. Istnieje wiele metod całkowania numerycznego funkcji. Niektóre z tych metod, jak na przykład kwadratury adaptacyjne, cechuje duża dokładność obliczeń oraz możliwość kontrolowania błędu całkowania [1]. W porównaniu z tymi metodami metoda Monte Carlo nie daje na ogół wyników z tak dużą dokładnością [2]. Metoda ta ma jednak tę zaletę, że często pozwala na ominięcie skomplikowanego aparatu matematycznego, jest prosta w implementacji i nie wymaga stosowania złożonych technik algorytmicznych i programistycznych. Metodę Monte Carlo stosuje się najczęściej w sytuacji, gdy zawodzą metody analityczne, a inne metody numeryczne nie dają zadowalającego wyniku. Jest tak na przykład podczas wielowymiarowego całkowania wieloparametrowych

funkcji wielu zmiennych [2]. Powstaje zatem pytanie, czy podejście przedstawione w pracy ma sens? Szczególnie, że całkowaniu numerycznemu poddawana jest jednoargumentowa funkcja o dwóch parametrach, której składową jest funkcja gęstości prawdopodobieństwa rozkładu Gaussa. Zdaniem autora podejście przedstawione w pracy ma swoje uzasadnienie. Jest tak z dwóch powodów. Po pierwsze uzyskane dane bardzo często podlegają rozkładowi Gaussa, po drugie szacowanie wartości oczekiwanej to w probabilistyce jedna z najczęściej wykonywanych procedur obróbki danych. O dokładności estymacji tej wielkości decyduje liczba danych. W metodzie Monte Carlo stosuje się zwykle dużą liczbę danych. Sprawdzenie dokładności szacowania wartości oczekiwanej, gdy do obliczeń zastosowana zostanie metoda Monte Carlo jest więc uzasadnione.

2. Estymacja wartości oczekiwanej

Niech X będzie zmienną losową o gęstości $f_X(x)$, $x \in \mathbf{R}$. Wartość oczekiwana θ zmiennej losowej X ma postać [3]:

$$\theta = E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx. \quad (1)$$

Wielkość (1) to moment zwykły pierwszego rzędu, nazywany w technice wartością średnią.

Niech dana będzie ciągła funkcja $g(x)$, $x \in \mathbf{R}$, całkowana na przedziale $[a, b]$, $a, b \in \mathbf{R}$, $-\infty < a \leq b < \infty$, taka, że:

$$g(x) = x f_X(x). \quad (2)$$

Estymator parametru (1) przyjmuje postać:

$$\tilde{\theta}_1 = \int_a^b g(x) dx. \quad (3)$$

Błąd szacowania parametru (1) można opisać wzorem:

$$e_{\tilde{\theta}_1} = \frac{|\tilde{\theta}_1 - \theta|}{|\theta|}. \quad (4)$$

Niech X będzie zmienną losową o rozkładzie Gaussa ze znanymi, niezerowymi parametrami $\sigma_X \in \mathbf{R}_+ \setminus \{0\}$, $\mu_X \in \mathbf{R} \setminus \{0\}$. Gęstość $f_X(x)$ rozkładu Gaussa jest funkcją o niezerowych wartościach. Powstaje zatem problem doboru wartości końców przedziału $[a, b]$. Najłatwiej zastosować wzory, do opisu których stosujemy parametry rozkładu oraz reguły związane z jego probabilistycznymi własnościami. Przyjmijmy, że:

$$a = -k\sigma_X + \mu_X, \quad b = k\sigma_X + \mu_X, \quad (5)$$

gdzie $k \in \mathbf{R} \setminus \{0\}$.

Ponieważ dla rozkładu Gaussa $\theta = \mu_X$, to na podstawie (2)-(5) oraz gęstości $f_X(x)$ otrzymujemy:

$$\begin{aligned} \tilde{\theta}_1 &= \int_a^b g(x) dx = \int_{-k\sigma_X + \mu_X}^{k\sigma_X + \mu_X} x \frac{1}{\sigma_X \sqrt{2\pi}} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} dx \\ &= \frac{1}{\sigma_X \sqrt{2\pi}} \int_{-k\sigma_X + \mu_X}^{k\sigma_X + \mu_X} x e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} dx = \mu_X \operatorname{erf}\left(\frac{k}{\sqrt{2}}\right), \end{aligned} \quad (6)$$

gdzie $\operatorname{erf}(\cdot)$ to funkcja błędu Gaussa [4].

Ponadto:

$$e_{\theta_1} = 1 - \operatorname{erf}\left(\frac{k}{\sqrt{2}}\right). \quad (7)$$

Zwróćmy uwagę, że:

$$k = \sqrt{2} \operatorname{ierf}(1 - e_{\theta_1}), \quad (8)$$

gdzie $\operatorname{ierf}(\cdot)$ to odwrotna funkcja błędu Gaussa [4].

Należy wziąć pod uwagę, że zastosowane we wzorach (6)-(8) funkcje $\operatorname{erf}(\cdot)$ i $\operatorname{ierf}(\cdot)$ w istocie dalej są wyrażeniami całkowymi. W dodatku funkcje pierwotne tych całek nie dadzą się wyrazić za pomocą funkcji elementarnych. Oznacza to, że nie jest możliwe obliczenie dokładnych wartości tych wielkości. W literaturze znaleźć można wzory, które umożliwiają wyznaczenie ich przybliżonych wartości. Zdaniem autora na uwagę zasługują następujące zależności [5]:

$$\operatorname{erf}(x) \approx \operatorname{sgn}(x) \sqrt{1 - e^{-\frac{4 + \xi x^2}{-x^2 \frac{\pi}{1 + \xi x^2}}}}, \quad (9)$$

$$\operatorname{ierf}(x) \approx \operatorname{sgn}(x) \left[\sqrt{\left(\frac{2}{\pi \xi} + \frac{\ln(1-x^2)}{2}\right)^2 - \frac{\ln(1-x^2)}{\xi}} - \left(\frac{2}{\pi \xi} + \frac{\ln(1-x^2)}{2}\right) \right]^{\frac{1}{2}}, \quad (10)$$

gdzie:

$$\xi = \frac{8(\pi - 3)}{3\pi(4 - \pi)} \approx 0.14, \quad (11)$$

oraz:

$$\operatorname{sgn}(x) = \begin{cases} 1, & x > 0, \\ 0, & x = 0, \\ -1, & x < 0. \end{cases} \quad (12)$$

W [5] podano, że względne dokładności wyznaczania za pomocą (9) i (10) wartości funkcji $\operatorname{erf}(\cdot)$ i $\operatorname{ierf}(\cdot)$ są na poziomie odpowiednio 10^{-4} i 10^{-3} . W [6] przedstawiono modyfikację wzoru (9) zwiększającą o rząd dokładność wyznaczania wartości funkcji $\operatorname{erf}(\cdot)$.

Z przeprowadzonych obliczeń wynika, że w celu uzyskania błędu (7) na poziomie jednego promila należy dobrać $k=3.29$, natomiast jeżeli przyjmiemy $k=2.57$, to błąd ten będzie równy w przybliżeniu 1%. Podane wartości mnożnika k wskazują, że wyznaczenie k na podstawie (8) równoważne jest obliczeniu przybliżonej wartości współczynnika rozszerzenia [7, 8]. Interesujące jest, że współczynnik ten nie zależy od parametrów rozkładu oraz od wartości końców przedziału $[a, b]$, ale jedynie od wartości błędu (4). W literaturze spotkać można najczęściej tabele z wartościami współczynnika rozszerzenia. Wartości te zależą od poziomu istotności, który odpowiada wartościom błędu (4).

Rozważmy teraz zagadnienie całkowania numerycznego funkcji $g(x)$ metodą Monte Carlo. Taka operacja odbywa się również na przedziale $[a, b]$. Przyjmijmy, że a i b wyznaczane są na podstawie (5). Uzasadnione jest więc sprawdzenie, jakie dokładności szacowania parametru (1) otrzymamy stosując metodę Monte Carlo. Rozważmy dwie popularne metody Monte Carlo całkowania numerycznego funkcji, metodę Crude i Hit-or-Miss [2]. Dokładność obu metod zależy od liczby danych oraz od okresu generatora liczb pseudolosowych. Może również zależeć od postaci funkcji $g(x)$, dlatego zakładamy dla uproszczenia, że $g(x)$ nie jest funkcją tzw. wysokiego pik.

Metoda Crude opiera się na generowaniu $N \in \mathbb{N} \setminus \{0\}$ liczb pseudolosowych x_i , $i=0, 1, \dots, N-1$ z rozkładu równomiernego na przedziale $[a, b]$. Estymator parametru (1) uzyskany tą metodą ma postać [2]:

$$\tilde{\theta}_2 = \frac{b-a}{N} \sum_{i=0}^{N-1} g(x_i). \quad (13)$$

Błąd szacowania parametru (1) wyznaczamy na podstawie wzoru:

$$e_{\tilde{\theta}_2} = \frac{|\tilde{\theta}_2 - \theta|}{|\theta|}. \quad (14)$$

Załóżmy, że wartości funkcji $g(x)$ znajdują się w obszarze $\Omega = \{(x, y) \in \mathbb{R} : a \leq x \leq b, c \leq y \leq d\}$, gdzie $-\infty < c \leq 0$, $0 \leq d < \infty$, $c, d \in \mathbb{R}$. Metoda Hit-or-Miss opiera się na generowaniu N liczb pseudolosowych x_i i y_i , $i=0, 1, \dots, N-1$ z rozkładu równomiernego odpowiednio na przedziałach $[a, b]$ i $[c, d]$. Estymator parametru (1) uzyskany metodą Hit-or-Miss ma postać [2]:

$$\tilde{\theta}_3 = \frac{k_N}{N} (b-a)(d-c), \quad (15)$$

gdzie k_N to moc zbioru $\{i \in \mathbb{N} : 0 < y_i \leq g(x_i)\}$ pomniejszona o moc zbioru $\{i \in \mathbb{N} : g(x_i) \leq y_i < 0\}$. Jeżeli $y_i=0$, to k_N nie zmienia swojej wartości.

Błąd szacowania parametru (1) wyznaczamy na podstawie wzoru:

$$e_{\tilde{\theta}_3} = \frac{|\tilde{\theta}_3 - \theta|}{|\theta|}. \quad (16)$$

W celu znalezienia wartości końców przedziału $[c, d]$ wyznaczyć należy ekstrema funkcji $g(x)$. Znanych jest wiele efektywnych sposobów wyznaczania ekstremów funkcji. Oparte są one między innymi na metodach optymalizacji, w tym na programowaniu liniowym i nieliniowym [9]. Istnieją również metody znajdowania ekstremum metodą Monte Carlo [10]. Ponieważ w przypadku $g(x)$ można znaleźć wszystkie punkty podejrzane o istnienie ekstremum, to do wyznaczenia wartości końców przedziału $[c, d]$ wystarczy zastosować twierdzenie Weierstrassa. Wtedy:

$$c = \min \{x_1 f_X(x_1), x_2 f_X(x_2), a f_X(a), b f_X(b)\}, \\ d = \max \{x_1 f_X(x_1), x_2 f_X(x_2), a f_X(a), b f_X(b)\},$$

$$x_1 = \begin{cases} x_{0_1}, & a \leq x_{0_1} \leq b, \\ 0, & x_{0_1} < a \vee x_{0_1} > b, \end{cases} \quad x_2 = \begin{cases} x_{0_2}, & a \leq x_{0_2} \leq b, \\ 0, & x_{0_2} < a \vee x_{0_2} > b, \end{cases} \quad (17)$$

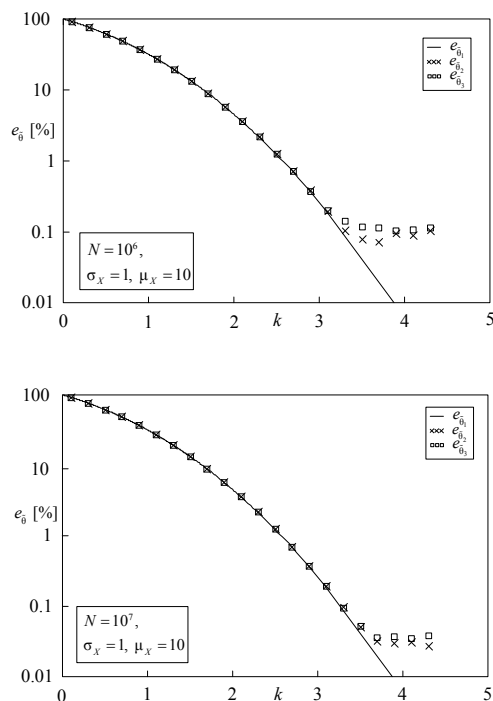
$$x_{0_1} = \frac{\mu_X - \sqrt{4\sigma_X^2 + \mu_X^2}}{2}, \quad x_{0_2} = \frac{\mu_X + \sqrt{4\sigma_X^2 + \mu_X^2}}{2}.$$

Jeżeli $k > 3$, to do obliczenia wartości c i d można zastosować wzór [11]:

$$c \approx x_{0_1} f_X(x_{0_1}), \quad d \approx x_{0_2} f_X(x_{0_2}). \quad (18)$$

Probabilistyczne własności estymatorów (13) i (15) zostały opisane w [2]. W pracy przyjęto założenie, że rzeczywiste wartości parametru (1) są znane. Należy jednak podkreślić, że na ogół rzeczywista wartość szacowanego parametru nie jest znana. Wówczas, w celu znalezienia wartości błędów estymacji parametrów szacowanych metodą Monte Carlo stosuje się podejście oparte na przedziałach ufności. Opis takiego podejścia przedstawiono w [2]. W celu zwiększenia ogólności rozważań podanych w [2] należy również wziąć pod uwagę informacje przedstawione w [12].

Sporządzono wykres błędów (7) oraz wykresy uśrednionych 100-krotnie procentowych wartości błędów (14) i (16) w funkcji wartości mnożnika (8).



Rys. 1. Wykresy błędów (7), (14) i (16) w funkcji mnożnika (8)
Fig. 1. The graphs of the errors (7), (14) and (16) as a function of the multiplier (8)

Tab. 1. Błędy estymacji wartości oczekiwanej
Tab. 1. The errors of the mean value estimator

$N=10^6, k=2.57$	MathCAD	Matlab	LabWINDOWS
$e_{\hat{\theta}_1}$ [%]	1		
$e_{\hat{\theta}_2}$ [%]	1.0201	1.0137	1.0169
$e_{\hat{\theta}_3}$ [%]	1.0219	1.0157	1.0175
$N=10^6, k=3.29$	MathCAD	Matlab	LabWINDOWS
$e_{\hat{\theta}_1}$ [%]	1		
$e_{\hat{\theta}_2}$ [%]	1.1297	1.1469	1.1119
$e_{\hat{\theta}_3}$ [%]	1.3762	1.3211	1.2899
$N=10^7, k=2.57$	MathCAD	Matlab	LabWINDOWS
$e_{\hat{\theta}_1}$ [%]	1		
$e_{\hat{\theta}_2}$ [%]	1.0164	1.0162	1.0174
$e_{\hat{\theta}_3}$ [%]	1.0172	1.0164	1.0186
$N=10^7, k=3.29$	MathCAD	Matlab	LabWINDOWS
$e_{\hat{\theta}_1}$ [%]	1		
$e_{\hat{\theta}_2}$ [%]	0.9990	0.9941	1.0223
$e_{\hat{\theta}_3}$ [%]	1.0062	1.0274	1.0311

Uśrednione wartości błędów (13) i (15) są dla $N=10^6$ i $k=2.57$ równe w przybliżeniu odpowiednio 1.02% i 1.04%, natomiast dla $k=3.29$ wynoszą 1.2 i 1.4 promila. Jeżeli błąd (7) jest nie większy niż 1%, to obie metody Monte Carlo dają błędy na porównywalnym poziomie. Zwróćmy uwagę, że dalsze zwiększanie wartości mnożnika k nie powoduje już zmniejszania wartości błędów. Zmniejszenie wartości błędów jest możliwe, ale dopiero po zwiększeniu liczby N danych. Niestety kosztem istotnego wydłużenia czasu obliczeń.

W celu sprawdzenia otrzymanych wyników zwiększono liczbę uśrednień do 1 000 i wykonano analogiczne obliczenia w kilku popularnych programach komputerowych. W zastosowanych programach liczby pseudolosowe generowane są z zastosowaniem liniowego generatora kongruentnego [13]. Szczegółowe wyniki obliczeń przedstawiono w tabeli 1. Wartości parametrów rozkładu przyjęto jak w przypadku wyznaczania wykresów błędów z rys. 1.

Zastosowane programy dają wyniki na zbliżonym poziomie. Uzyskane wyniki potwierdzają wcześniejsze wnioski. Zwiększenie wartości mnożnika k może nie wpłynąć na zwiększenie dokładności szacowania wartości oczekiwanej.

3. Podsumowanie

W artykule podano sposób na wyznaczenie błędów estymacji na określonym poziomie dokładności. Uzyskane tym sposobem wyniki błędów porównano z wynikami błędów estymatorów wyznaczanych metodą Monte Carlo. Dokładność szacowania wartości oczekiwanej metodą Monte Carlo podlega ograniczeniom metody. Ograniczenia te wynikają z zastosowanej liczby danych i długości cyklu generatora liczb pseudolosowych. Generowanie danych z przedziału większego niż trzy sigma bez istotnego zwiększenia liczby danych nie przynosi zwiększenia dokładności szacowania wartości oczekiwanej metodą Monte Carlo. Jednocześnie podczas zwiększania liczby danych należy wziąć pod uwagę ograniczenia generatora liczb pseudolosowych i liczyć się z istotnym wydłużeniem czasu obliczeń.

4. Literatura

- [1] Stoer J., Bulirsch R.: Introduction to numerical analysis (3rd ed.), Springer, 2002.
- [2] Zieliński R.: Metoda Monte Carlo, Wydawnictwo Naukowo Techniczne, 1970.
- [3] Durrett R.: Probability: theory and examples (4th ed.), Cambridge University Press, 2004.
- [4] Greene W. H.: Econometric analysis (5th ed.), Prentice-Hall, 1993.
- [5] Winitzki S.: A handy approximation for the error function and its inverse, 2008, <http://sites.google.com/site/winitzki/sergei-winitzki-files/erf-approx.pdf>.
- [6] Soranzo A., Epure E.: Simply explicitly invertible approximations to 4 decimals of error function and normal cumulative distribution function, 2012, <http://arxiv.org/pdf/1201.1320>.
- [7] Guide to the expression of uncertainty in measurement, 1993-95 ISO.
- [8] Fotowicz P.: Obliczanie niepewności rozszerzonej metodą analityczną opartą na splocie rozkładów wielkości wejściowych, Pomiar Automatyka Robotyka, nr 1, str. 5-9, 2005.
- [9] Avriel M.: Nonlinear programming: analysis and methods, Courier Dover Publ., 2012.
- [10] Zieliński R., Neumann P.: Stochastyczne metody poszukiwania minimum funkcji, WNT, 1986.
- [11] Sienkowski S., Estimation of random variable distribution parameters by Monte Carlo method, Metrology and Measurement System, vol. 20, pp. 249-262, 2013.
- [12] Jain S.: Monte Carlo Simulations of Disordered System, World Scientific Publishing Company, Incorporated, 1992.
- [13] Jaffe R.C., Random Signals for Engineers Using Matlab and MathCAD, New York: Springer, 2000.