

211/2002

**Raport Badawczy**

**RB/4/2002**

**Research Report**

**Diagnostyka zbieżności metod  
MCMC**

**M. Romaniuk**

**Instytut Badań Systemowych  
Polska Akademia Nauk**

**Systems Research Institute  
Polish Academy of Sciences**



# **POLSKA AKADEMIA NAUK**

## **Instytut Badań Systemowych**

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 8373578

fax: (+48) (22) 8372772

Kierownik Pracowni zgłaszający pracę:  
Prof. dr hab. inż. Olgierd Hryniewicz

Warszawa 2002

# Diagnostyka zbieżności metod MCMC

Maciej Romaniuk

Instytut Badań Systemowych PAN,

ul. Newelska 6, 01-447 Warszawa

e-mail: mroman@ibspan.waw.pl

6 listopada 2002 roku

## Streszczenie

Głównym celem poniższej pracy jest ogólne wprowadzenie do metod Monte Carlo (MC) oraz ściśle z nimi związanych metod Markov Chain Monte Carlo (MCMC), powstałych z powiązania metod MC z teorią łańcuchów Markowa. Ponadto zaprezentowany zostanie przegląd występujących w literaturze metod diagnostycznych dla MCMC wraz z zasygnalizowaniem tematyki przyszłych badań autora. Niniejszy raport ma postać przegląдово – badawczą i związany jest z przyszłą pracą doktorską autora.

## 1 Wprowadzenie

Metody Monte Carlo zostały wprowadzone w latach 40-tych ubiegłego wieku przez dwóch wielkich matematyków – Metropolis i Ulama (patrz [39]).

Początkowo wykorzystane w zastosowaniach militarnych przy słynnym projekcie Manhattan, w ciągu kilkadziesiąt lat podbiły świat matematycznych symulacji komputerowych.

Sukces odniesiony przez metody Monte Carlo (w skrócie MC) został zdyskontowany przez ich rozwinięcie – metody Markov Chain Monte Carlo (skrótowo określane jako MCMC, patrz [16]). Rozwinięcie to zostało zaproponowane przez Metropolisa i innych (patrz [37]).

Obecnie lista dziedzin matematyki stosowanej, w których wykorzystywane są metody MC i MCMC jest imponująca — poczynając od zadań znajdowania całek i zagadnień optymalizacyjnych, poprzez generowanie zmiennych losowych z analitycznie skomplikowanych rozkładów prawdopodobieństwa, estymację złożonych statystyk testowych, zagadnienia z teorii ruiny dla matematyki ubezpieczeniowej, symulacyjne wyznaczanie cen instrumentów finansowych, odszumianie obrazów, problemy związane z wnioskowaniem bayesowskim, aż po zagadnienia z fizyki, psychologii i medycyny (patrz np. [7], [8], [10], [13], [14], [18], [19], [20], [23], [24], [25], [27], [28], [31]).

Opracowane zostały również specjalne programy komputerowe wspomagające użycie tych metod w praktyce, jak np. BUGS (patrz np. [4]).

Mimo ogromnego rozwoju metod MC i MCMC, ich stosowanie wiąże się nadal z uciążliwym problemem zwanym *diagnostyką zbieżności* (patrz rozdział 4), z którym autor niniejszej prezentacji zamierza zmierzyć się w swojej pracy doktoranckiej.

Niniejsza praca podzielona została w następujący sposób. Rozdział 2 zawiera krótkie wprowadzenie do metod Monte Carlo wraz z omówieniem stosowanych w nich algorytmów. Rozdział 3 omawia w skrócie metody Markov

Chain Monte Carlo wraz z odpowiednimi algorytmami i zasygnalizowaniem pewnych twierdzeń i definicji z nimi związanych. Rozdział 4 przybliży problem monitorowania zbieżności metod MCMC.

## 2 Metody Monte Carlo

W swej najbardziej klasycznej formie metody MC związane są z zagadnieniami *całkowania i optymalizacji* funkcji (patrz np. [15], [25]). Przyjrzyjmy się pokrótce tym dwóm kwestiom.

W przypadku, gdy całki z danej funkcji nie da się wyrazić przy pomocy funkcji elementarnych, do jej obliczenia konieczne jest skorzystanie z metod numerycznych. Oprócz deterministycznych metod numerycznych, takich jak np. całka Riemanna, zasada trapezoidu, zasada Simpsona, splajny itp., można posłużyć się także metodą MC dla całki o ogólnej postaci

$$\mathbb{E}_f h(x) = \int h(x)f(x)dx, \quad (1)$$

gdzie  $f(x)$  jest gęstością pewnego rozkładu prawdopodobieństwa.

W dalszej części pracy zakładamy, iż  $\mathbb{E}_f h(x) < \infty$ .

Obliczenie wartości wyrażenia (1) wymaga w metodzie MC wygenerowania losowej próby  $X_1, X_2, \dots, X_n$  *iid* z rozkładu o gęstości  $f(x)$  przy pomocy odpowiednich komputerowych algorytmów pseudolosowych (szersze informacje na ten temat znaleźć można np. w [11]). Średnia o postaci

$$\tilde{h}_f = \frac{1}{n} \sum_{i=1}^n h(X_i) \quad (2)$$

przybliża, zgodnie z Mocnym Prawem Wielkich Liczb, wyrażenie  $\mathbb{E}_f h(x)$ , przy czym zbieżność ta jest zbieżnością prawie na pewno dla  $n \rightarrow \infty$ . Metodę

tą określa się czasami w literaturze jako *crude Monte Carlo* i pochodzi ona z lat 40-tych (patrz [39]).

Obecnie istnieje wiele różnorodnych rozwinięć powyższej metody, których głównym celem jest minimalizacja wariancji estymatora  $\tilde{h}_f$ . Podejście takie wynika z faktu, iż do  $\tilde{h}_f$  możemy zastosować Centralne Twierdzenie Graniczne, otrzymując

$$\tilde{h}_f - \mathbb{E}_f h(x) \xrightarrow{D} \mathcal{N}(0, \text{Var}(\tilde{h}_f)), \quad (3)$$

gdzie

$$\text{Var}(\tilde{h}_f) = \frac{1}{n^2} \sum_{i=1}^n (h(X_i) - \tilde{h}_f)^2 \quad (4)$$

jest estymatorem wariancji  $\tilde{h}_f$  (o ile wariancja ta ma wartość skończoną).

Wynika z tego, iż redukcja wariancji wielkości  $\tilde{h}_f$  zwiększa dokładność uzyskiwanych wyników, pozwalając na zmniejszenie ilości losowań  $n$  (a więc i skrócenie nieraz bardzo czasochłonnych symulacji). Przykładami metod pozwalających uzyskać redukcję wariancji jest *importance sampling*, metoda zmiennych antytetycznych (w j. ang. *antithetic variables*), czy zmiennych kontrolnych (w j. ang. *control variates*) (dokładniejsze informacje na ten temat można znaleźć np. w [25]).

Drugą klasą problemów, które można rozwiązać poprzez użycie metod MC są zagadnienia optymalizacji, ze szczególnym uwzględnieniem kwestii poszukiwania ekstremów danej funkcji, np. maksimum globalnego. Oczywiście także i dla tych zagadnień możliwe jest zastosowanie deterministycznych metod numerycznych (choćby np. metody Newtona, czy jej uogólnień – patrz np. [38]). Jednak jeśli rozpatrywana funkcja  $h(x)$  posiada kilka maksimumów, zastosowanie metody deterministycznej może zaowocować odnalezieniem tylko jednego z ekstremów lokalnych, a nie maksimum globalnego. Podobnie

częstokroć konieczne są dodatkowe założenia co do szczególnych własności funkcji  $h(x)$  (np. jej różniczkowalność).

Remedium na takie problemy może być zastosowanie metody MC zwanej symulowanym wyżarzaniem (w j. ang. *simulated annealing*) z modyfikacją zaproponowaną przez Metropolisa (patrz np. [18], [25], [37]). Metoda ta wymaga wprowadzenia dodatkowej zmiennej  $T > 0$ , tradycyjnie określanej jako temperatura ze względu na historyczne zastosowanie w fizyce. Startując z pewnej wartości  $X_0$ , kolejne kroki symulowanego wyżarzania polegają na generacji wartości  $X_n$  według następującego algorytmu:

1. Wylosowanie wartości  $Y_n$  z pewnego rozkładu prawdopodobieństwa na otoczeniu  $x_{n-1}$  (np. z rozkładu jednostajnego na tym otoczeniu, w ogólności z dowolnego rozkładu o gęstości postaci  $g(|x_{n-1} - Y_n|)$ ).
2. Nowy punkt  $X_n$  losowany jest według przepisu:

$$X_n = \begin{cases} Y_n & \text{z prawd. } p = \min \{ \exp(\Delta h/T), 1 \} \\ x_{n-1} & \text{z prawd. } 1 - p \end{cases} \quad (5)$$

gdzie  $\Delta h = h(Y_n) - h(x_{n-1})$ . Ponadto dla kolejnych kroków powyższego algorytmu wprowadza się założenie o monotonicznej zbieżności  $T \rightarrow 0$ .

Jak widzimy, obecność warunku (5) umożliwia procedurze „ucieczkę” z lokalnego maksimum z pewnym dodatnim prawdopodobieństwem  $p$  (o ile tylko  $\exp(\Delta h/T) < 1$ ). Ponadto stałe zmniejszanie się wartości  $T$  („zamrażanie”) założone w powyższym algorytmie umożliwia „zatrzymanie” łańcucha w pewnym maksimum. Wynika z tego, iż na odpowiednie zachowanie się procedury (a więc i znalezienie z dużym prawdopodobieństwem globalnego ekstremum)

olbrzymi wpływ ma sposób zbiegania temperatury  $T$  do zera. Dodatkową bibliografię dotyczącą tego zagadnienia znaleźć można w odsyłaczach w [25].

Obecnie jako metody Monte Carlo określane są również metody związane z przeprowadzaniem symulacji dla szerokiej klasy procesów stochastycznych, np. symulowaniem ruchów Browna. Tego typu zagadnienia spotykane są w matematyce finansowej i ubezpieczeniowej. Przykładem może być tutaj problematyka wyceny tzw. opcji egzotycznych (patrz np. [7], [14]), czy rozwiązania problemu ruiny dla procesu ryzyka firmy ubezpieczeniowej. Należy podkreślić, iż w powyższych zastosowaniach w symulacjach stosuje się często zamiast algorytmów (pseudo)losowych, deterministyczne algorytmy numeryczne. Metoda taka określana jest często jako *quasi Monte Carlo* (dalsze szczegóły dotyczące tej problematyki znaleźć można np. w [13] i tamtejszych odsyłaczach).

### 3 Metody Markov Chain Monte Carlo

W największym skrócie rzecz ujmując, metody Markov Chain Monte Carlo (w skrócie MCMC) powstają z połączenia symulacji metodami MC z teorią łańcuchów Markowa.

Najważniejszym wynikiem teoretycznym dla jednorodnych łańcuchów Markowa wykorzystywanym w metodach MCMC są tzw. twierdzenia ergodyczne. Są one odpowiednikiem MPWL sformułowanego dla ciągu *iid* zmiennych losowych.

W przypadku łańcucha Markowa o przeliczalnej przestrzeni stanów  $\mathcal{S}$  twierdzenie ergodyczne przedstawić można w następującej postaci:



**Twierdzenie 1 (Ergodyczne dla przeliczalnej przestrzeni stanów).** *Jeżeli łańcuch Markowa  $(X_n)$  jest nieprzywiedlny i istnieje rozkład stacjonarny  $\pi$ , a funkcja  $g : S \rightarrow \mathbb{R}$  jest taka, że wartość oczekiwana  $\mathbb{E}_\pi g(X)$  istnieje, to zachodzi zbieżność*

$$\frac{1}{n} \sum_{i=1}^n g(X_n) \xrightarrow[p.n.]{n \rightarrow \infty} \mathbb{E}_\pi g(X) \quad (6)$$

dla dowolnego rozkładu początkowego  $\pi_0$ .

Przypomnijmy, że łańcuch nazywamy nieprzywiedlnym, jeśli wszystkie jego stany komunikują się ze sobą (tzn. dla każdej pary  $i, j \in S$  istnieją takie  $k, l \in \mathbb{N}$ , że  $P(X_k = i | X_0 = j) > 0$  oraz  $P(X_l = j | X_0 = i) > 0$ ).

Odpowiednikiem powyższego twierdzenia dla łańcuchów Markowa o nieprzeliczalnej przestrzeni stanów jest

**Twierdzenie 2 (Ergodyczne dla nieprzeliczalnej przestrzeni stanów).** *Jeżeli łańcuch Markowa  $(X_n)$  jest nieprzywiedlny, ma  $\sigma$ -skończoną miarę niezmienniczą (w j. ang. invariant measure)  $\pi$  oraz jest powracający w sensie Harrisa (w j. ang. Harris recurrent), a funkcja  $g(x)$  ma skończoną wartość oczekiwaną względem  $\pi$ , to zachodzi zbieżność p. n.*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n g(X_n) = \mathbb{E}_\pi g(X) . \quad (7)$$

Łańcuch Markowa nazywamy powracającym w sensie Harrisa, jeśli spełnia następującą definicję:

**Definicja 1 (Powracalność w sensie Harrisa).** *Zbiór  $\mathcal{A}$  nazywamy powracalnym w sensie Harrisa, jeśli  $P(\eta_{\mathcal{A}} = \infty | X_0 = x) = 1$  dla każdego  $x \in \mathcal{A}$ , gdzie  $\eta_{\mathcal{A}}$  jest ilością powrotów łańcucha do zbioru  $\mathcal{A}$ .*

*Łańcuch Markowa*  $(X_n)$  nazywamy *powracającym w sensie Harrisa*, jeśli istnieje miara  $\psi$ , względem której jest on nieprzywiedlny, oraz każdy zbiór  $A$ , spełniający warunek  $\psi(A) > 0$  jest zbiorem powracalnym w sensie Harrisa.

Z kolei miarę  $\pi$  nazywamy *niezmienniczą*, jeśli spełnia ona warunek

$$\pi(x) = \int_{\mathcal{S}} P(X \in \mathcal{B} | Y = y) d\pi(y) \quad (8)$$

dla każdego  $\mathcal{B} \in \mathbb{B}(\mathcal{S})$ , gdzie  $\mathbb{B}(\mathcal{S})$  jest  $\sigma$ -ciałem zbiorów borelowskich na  $\mathcal{S}$ .

Pełniejsze wprowadzenie do teorii łańcuchów Markowa o przeliczalnej przestrzeni stanów znaleźć można np. w [36], a dla nieprzeliczalnej przestrzeni stanów np. w [15], [25].

Należy zwrócić uwagę, iż w założeniach twierdzeń ergodycznych nie jest wymagana nieokresowość łańcucha Markowa. Jednak w zastosowaniach, a więc w metodach MCMC, generuje się zazwyczaj tylko łańcuchy spełniające taki postulat, gdyż ułatwia to dokonywanie obliczeń i upraszcza zagadnienie diagnostyki zbieżności (patrz paragraf 4). Ponadto nieokresowość wymagana jest w tzw. mocnych twierdzeniach ergodycznych.

Najważniejsza idea metody MCMC polega na tym, aby do obliczania wyrażeń typu (1) nie generować próby *iid*  $X_1, X_2, \dots, X_n$  bezpośrednio z gęstości  $f(x)$ . Zamiast tego w metodzie MCMC generuje się ciąg, który jest jednorodnym łańcuchem Markowa o *rozkładzie stacjonarnym*  $f(x)$ . Podejście takie umożliwia uniknięcie odwoływania się w algorytmach pseudolosowych do bezpośredniego generowania z gęstości  $f(x)$ . Gęstość ta może mieć bowiem zbyt skomplikowaną postać, aby wykorzystanie jej było efektywne lub nawet możliwe. Może także być znana jedynie z dokładnością do pewnej stałej normującej, a stała ta częstokroć jest zbyt skomplikowana do dostatecznie

szybkiego obliczenia. Oba te przypadki szczególnie często spotykane są przy bayesowskim wnioskowaniu statystycznym (patrz np. [16], [18], [20], [25]).

Do stworzenia próby  $X_1, X_2, \dots, X_n$ , która będzie łańcuchem Markowa o z góry zadany rozkładzie stacjonarnym  $f(x)$  służą dwa bardzo ogólne algorytmy, zwane algorytmem Metropolisa – Hastingsa i próbnikiem Gibbsa (w j. ang. *Gibbs sampler*). Jeśli próba taka będzie ponadto ergodycznym łańcuchem Markowa, tzn. spełniać będzie założenia twierdzenia 1 lub 2, to, podobnie jak dla metody MC średnia

$$\bar{h}_{MC} = \frac{1}{n} \sum_{i=1}^n h(X_i) \quad (9)$$

będzie zbieżna p. n. do  $E_f h(x)$ . Jednocześnie w obu tych algorytmach warunków ergodyczności łańcucha Markowa jest zazwyczaj, oprócz pewnych zdegenerowanych przypadków, dostatecznie łatwy do sprawdzenia w praktyce.

Przyjrzymy się teraz dokładniej konstrukcji każdego z obu wspomnianych powyżej algorytmów. Szersze informacje na ich temat znaleźć można w [9], [10], [15], [18], [20], [22], [25], [35].

W algorytmie Metropolisa – Hastingsa (w skrócie MH) wykorzystywana jest dodatkowa gęstość  $g(y|x)$ , zwana gęstością warunkową (w j. ang. *conditional density*). Gęstość ta może być wybrana dość dowolnie, tzn. niezależnie od gęstości  $f(x)$ . Jednak żeby zapewnić poprawne działanie algorytmu MH, tzn. ergodyczność powstałego łańcucha Markowa, funkcja  $g(y|x)$  musi spełniać następujące warunki:

1. Gęstość  $g(y|x)$  musi być łatwa do generowania z niej próbek lub symetryczna (tzn.  $g(y|x) = g(x|y)$ ).
2. Musi istnieć możliwość wyliczenia wielkości  $f(y)/g(y|x)$  (ewentualnie z dokładnością do stałej niezależnej od  $x$ ).

3. Nośnik funkcji  $g(y|x)$  (względem zmiennej  $y$ ) powinien zawierać nośnik  $f(x)$ .
4. Powinny być spełnione dwa warunki w postaci nierówności. Związane są one z zapewnieniem nieprzywiedlności oraz nieokresowości powstałego łańcucha (patrz (12) i (13)).

Startując z dowolnej, początkowej wartości  $X_0$ , algorytm MH generuje kolejne wyrazy łańcucha Markowa poprzez wykonanie następujących kroków:

1. Wylosowanie zmiennej  $Y_i$  z gęstości  $g(\cdot|x_{i-1})$
2. Nowy punkt  $X_i$  losowany jest według przepisu:

$$X_i = \begin{cases} Y_i & \text{z prawd. } p = m(x_{i-1}, Y_i) \\ x_{i-1} & \text{z prawd. } 1 - p \end{cases} \quad (10)$$

gdzie

$$m(x, y) = \min \left\{ \frac{f(y)}{g(y|x)} \frac{g(x|y)}{f(x)}, 1 \right\}. \quad (11)$$

Gęstość  $m(x, y)$  nazywana jest czasami gęstością pomocniczą lub instrumentalną (w j. ang. *proposal, instrumental distribution*). W języku powyższego algorytmu nierówności o których mowa w punkcie 4 przy omawianiu warunków dotyczących gęstości warunkowej mają postać:

$$g(y|x) > 0 \text{ dla dowolnych } x, y \in \mathcal{S} \times \mathcal{S} \quad (12)$$

(co implikuje nieprzywiedlność i powracalność w sensie HARRISA),

$$P \left( \frac{f(Y)}{g(Y|X)} \frac{g(X|Y)}{f(X)} \leq 1 \right) < 1 \quad (13)$$

(co implikuje zachodzenie warunku  $P(X_i = X_{i-1}) > 0$ , a więc i nieokresowość).

Jeśli gęstość  $g(y|x)$  spełnia założenia z punktów 1–4, to ciąg  $X_1, X_2, \dots, X_n$  powstały w wyniku działania algorytmu MH jest ergodycznym łańcuchem Markowa zbieżnym do ustalonej gęstości  $f(x)$  oraz spełnia równanie (9). Ponadto, jak łatwo zauważyć, w algorytmie MH nie wykonuje się bezpośrednio losowania z gęstości  $f(x)$ , a znajomość postaci tej funkcji ogranicza się jedynie do możliwości obliczenia wartości  $f(y)/g(y|x)$  lub równoważnie  $f(y)/f(x)$ , a więc bez określania stałej normującej.

Drugi z algorytmów – próbnik Gibbsa – jest, dzięki specyficznemu sposobowi swej konstrukcji, szczególnie łatwy w implementacji przy bayesowskim podejściu statystycznym (patrz np. [18], [20], [25], [34]).

W próbniku Gibbsa konstruowany jest ciąg wektorów losowych  $\mathbb{X}_1, \mathbb{X}_2, \dots, \mathbb{X}_n$ . Przez  $(X_i^1, \dots, X_i^m)$  będziemy oznaczać wektor współrzędnych zmiennej losowej  $\mathbb{X}_i$ , przy czym każda ze współrzędnych może być jedno- lub wielowymiarowa. Przez  $\mathbb{X}_i^{-j}$  oznaczymy wektor losowy  $(X_i^1, \dots, X_i^{j-1}, X_i^{j+1}, \dots, X_i^m)$ . Ponadto założymy, iż możliwe jest efektywne generowanie zmiennych losowych o gęstościach warunkowych  $g(X^j|\mathbb{X}^{-j})$  dla  $j = 1, \dots, m$ . Takie założenie jest zazwyczaj spełnione przy praktycznych aplikacjach bayesowskich.

Niech ponadto dana będzie wartość startowa  $\mathbb{X}_0$ . Wtedy, przy powyższych założeniach, próbnik Gibbsa generuje kolejne elementy ciągu według następującego algorytmu:

1. Generuj  $Y_i^j$  z gęstości  $g(\cdot|y_i^1, \dots, y_i^{j-1}, x_i^{j+1}, \dots, x_i^m)$  kolejno dla  $j = 1, \dots, m$
2.  $\mathbb{X}_{i+1} = (Y_i^1, \dots, Y_i^m)$

W literaturze spotykana jest modyfikacja powyższego algorytmu, polegająca na wykorzystaniu dodatkowej, niezależnej gęstości pomocniczej  $v(z)$ . Przy

takiej modyfikacji, losowania elementów w punkcie 1 powyższego algorytmu nie dokonuje się dla  $j$  przebiegającego po kolei elementy zbioru  $1, \dots, m$ . Zamiast tego generuje się zmienną losową  $J$  zgodnie z gęstością  $v(\cdot)$  i dla  $J = j$  losuje  $Y_i^j$  z gęstości  $g(\cdot | \mathbb{X}_i^{-j})$ .

Istnieją rozmaite twierdzenia, które formułując warunki dotyczące rodziny gęstości  $g(X^j | \mathbb{X}^{-j})$ , są równoważne odpowiednim twierdzeniom ergodycznym (patrz tw. 1 i 2). Zapewniając ergodyczność łańcucha Markowa powstałego w wyniku działania próbnika Gibbsa, są jednocześnie łatwiejsze do sprawdzenia w praktycznych zastosowaniach. Jednym z nich jest

**Twierdzenie 3.** *Ciąg Markowa wygenerowany próbnikiem Gibbsa jest nieprzywiedlny i nieokresowy, jeśli gęstości warunkowe  $g(X^j | \mathbb{X}^{-j})$  spełniają następujące warunki:*

1. Niech  $\mathbb{Y} = (y_1, \dots, y_m)$  i  $\mathbb{Y}' = (y'_1, \dots, y'_m)$  oraz istnieje  $\delta > 0$  dla której  $\mathbb{Y}, \mathbb{Y}' \in \text{supp}(g)$ ,  $|\mathbb{Y} - \mathbb{Y}'| < \delta$  i

$$g(y_i | y_1, \dots, y_{i-1}, y'_{i+1}, \dots, y'_m) > 0 \text{ dla } i = 1, \dots, m \quad (14)$$

2. Niech  $\delta' < \delta$ , taka, że każda para  $\mathbb{Y}, \mathbb{Y}' \in \text{supp}(g)$  może być połączona skończoną ilością kul o promieniu  $\delta'$  mających parami przecięcia o dodatniej mierze.

Mimo skomplikowanego sformułowania twierdzenia, jego zasadniczym elementem jest żądanie, aby próbnik Gibbsa poruszał się po spójnej przestrzeni (tzn. miał spójny w sensie topologicznym nośnik).

Próbnik Gibbsa, oprócz wykorzystywania go w rozmaitych problemach związanych z podejściem bayesowskim, używany jest również dla tzw. pól

Markowowskich (rozszerzenia warunku Markowa na przestrzeń wielowymiarową). Przykładem zastosowania próbnika Gibbsa dla takich pól może być model Isinga zachowania się ferromagnetyków, czy też zagadnienia odszumiania obrazów.

## 4 Diagnostyka zbieżności metod Markov Chain Monte Carlo

Zaprezentowane w paragrafie 3 twierdzenia ergodyczne, choć bardzo istotne z teoretycznego punktu widzenia, mają jednak poważną wadę aplikacyjną. Są one bowiem twierdzeniami *granicznymi*. Tymczasem w zastosowaniach najważniejszą informacją związaną z przeprowadzaniem symulacji jest udzielenie odpowiedzi na pytanie, jak wcześnie można taką symulację zatrzymać. Innymi słowy – przy którym kroku symulacji osiągnięte za jej pomocą wyniki są dostatecznie bliskie prawdziwym, poszukiwanym przez nas wartościom.

Zagadnienia związane z tak specyficznym sformułowaniem *problemem stopu* nazywane są diagnostyką (lub monitoringiem) metod MCMC.

Należy przy tym od razu zauważyć, iż zastosowanie klasycznych statystycznych twierdzeń granicznych (np. CTG) jest dla metod MCMC znacznie utrudnione. Choć bowiem możliwość zastosowania przybliżeń np. rozkładem normalnym w celu stworzenia przedziałów ufności, jest bardzo pociągająca, równocześnie w metodach MCMC jest szczególnie mało efektywna. Problem ten wynika z dwóch przyczyn. Po pierwsze, próbka wygenerowana przez metodę MCMC nie jest *iid*. Ponadto w sformułowaniu CTG (lub jego odpowiednika dla łańcuchów Markowa) występuje wariancja. Zaś wielkość ta

dla problemów rozwiązywanych metodami MCMC, jest po pierwsze nieznaną (należałoby ją zatem najpierw wyestymować), a po drugie i znowu – w łańcuchu Markowa kolejne wyrazy nie są oczywiście niezależne i nie tworzą ciągu *iid*. Stąd też należałoby w estymatorze wariancji wziąć pod uwagę odpowiednie współczynniki kowariancji, co niesamowicie komplikuje postać tych estymatorów i czyni cały proces bardzo mało efektywnym. Wynika to oczywiście z konieczności przeprowadzenia w takim przypadku dodatkowych symulacji estymujących nieznaną wartość.

#### 4.1 Podział metod diagnostycznych

Zanim przejdziemy do bardziej szczegółowego przyjrzenia się kilku wybranym metodom diagnostyki MCMC, najpierw dokonamy ogólnego ich podziału. Klasyfikacji takiej dokonać można ze względu na rozmaite kryteria. Jednym z owych kryteriów może być rodzaj funkcji występujący w granicy rozpatrywanej zbieżności.

Jeśli granicą taką jest sam rozkład stacjonarny  $\pi$ , to mówimy o diagnostyce zbieżności MCMC do rozkładu stacjonarnego. Oczywiście, samo istnienie takiego typu zbieżności dla danej metody MCMC jest warunkiem niezbędnym, zarówno w teorii, jak i w zastosowaniach.

Z kwestią powyższą związane jest również istotne zagadnienie eksplorowania przez metodę MCMC całej przestrzeni stanów oraz szybkość przebiegu tego procesu. Pozostawanie łańcucha Markowa przez dłuższy okres w jednym stanie (określane jako wolno mieszejący łańcuch – w j. ang. *slow mixing*) może prowadzić do konieczności zwiększenia ilości symulacji i / lub wyciągnięcia złych wniosków z przeprowadzonych dotąd obserwacji (np. co do istniejących



ekstremów funkcji). Diagnostyka wskazująca na występowanie takich sytuacji jest więc w wielu zagadnieniach bardzo potrzebna.

Drugim typem granicy w badaniu zbieżności może być średnia  $E_{\pi}g(X)$  (patrz (9) oraz twierdzenia 1 i 2). Diagnostyka związana z tak postawionym zagadnieniem jest dla praktycznych aplikacji bardzo istotna, w szczególności dlatego, iż jak już zostało to wspomniane, twierdzenia ergodyczne zapewniają jedynie zbieżność asymptotyczną. Tymczasem w symulacjach dysponujemy oczywiście tylko skończoną ilością prób (choć liczbę ich możemy zwiększyć). Podobnie jak w poprzednim przypadku diagnostyka tego typu związana jest również z rozpoznawaniem zjawiska *slow mixing*. Ponadto istotnym może być sprawdzenie, czy łańcuch Markowa zdążył już „zapomnieć” swoją wartość startową  $X_0$ , tzn. czy wybór wartości początkowej nie będzie miał wpływu na wyciągane przez nas wnioski.

Trzecie zagadnienie jest z kolei związane ze sprawdzeniem, jak bardzo ciąg  $X_1, \dots, X_n$  zbliżony jest do próby *iid* w sensie pewnej, ustalonej normy. Należy od razu zwrócić uwagę, iż próba bezpośredniego zastosowania wielu łańcuchów Markowa, czyli jednoczesnego symulowania wielu trajektorii, i brania pod uwagę jedynie ostatnich wartości każdego z ciągów powoduje dużą stratę efektywności i niewykorzystanie uzyskanych już informacji. Podobnie, choć na mniejszą skalę, dzieje się przy tzw. *subsamplingu* (patrz paragraf 4.2 i [35]). Diagnostyka tego typu problemu związana jest z możliwością wykorzystania wniosków płynących z CTG, co umożliwia np. budowę przedziałów ufności, prostsze oszacowanie prędkości zbieżności itp.

Drugim kryterium podziału dla metod diagnostyki MCMC jest wykorzystywanie w ich implementacji pewnych postaw teoretycznych (mniej lub

bardziej uzasadnionych) lub jedynie obserwacji *stricte* heurystycznych.

Trzecie kryterium związane jest z wydzieleniem metod graficznych, tzn. bazujących nie tylko na wzorach matematycznych, ale i na pewnych graficznych wykresach otrzymanych z symulacji metodą MCMC.

Należy ponadto zauważyć, iż metody diagnozowania zbieżności odwołują się do różnych sposobów implementacji MCMC. Związane jest to z zauważalnym w literaturze sporem pomiędzy zwolennikami wykorzystywania tylko jednej trajektorii ( $X_n$ ), a autorami preferującymi jednoczesne użycie wielu łańcuchów ( $X_n^k$ ) (dla  $k > 1$ ) (patrz np. [26]). Choć kwestia ta w ogólności zdaje się być nierozstrzygnięta, warto jednak zwrócić uwagę na wady i zalety każdego z tych podejść.

Porównując wykorzystanie jednego łańcucha o długości  $KN$  z jednoczesnym zasymulowaniem  $K$  trajektorii o długości  $N$  każda, widzimy, iż istnieje większe prawdopodobieństwo zbliżenia się pojedynczego, dłuższego łańcucha do oczekiwanego rozkładu stacjonarnego. Ponadto w symulacjach MCMC część wyrazów początkowych każdego wygenerowanego ciągu należy odrzucić, aby początkowy stan nie wpłynął negatywnie na wyestymowaną informację (tzw. *burning* lub *pre burning* trajektorii). Tak więc procentowe straty danych wynikające z zastosowania wielu łańcuchów są oczywiście o wiele większe niż przy pojedynczym ciągu, co obniża efektywność symulacji. Jednocześnie optymalne wykorzystanie metody wielołańcuchowej wydaje się wymagać choć częściowej *a priorycznej* informacji na temat gęstości stacjonarnej. Wynika to z potrzeby jak najlepszego „rozzrucenia” wartości początkowych po całym jej nośniku.

Mimo przedstawionych powyżej wad, wygenerowanie tylko pojedyncze-

go łańcucha związane jest z problemem określanym jako „*you've only seen, where you've been*”. Oznacza to, iż ciąg taki mógł nie znaleźć się we wszystkich możliwych stanach, co prowadzić może do błędnych wniosków diagnostycznych, a w konsekwencji i złych wyników. Problem taki jest szczególnie istotny dla łańcuchów wykazujących *slow mixing* – np. długotrwałe pozostawanie ciągu w jednym stanie może zostać niepoprawnie zinterpretowane jako osiągnięcie zbieżności, czy ekstremum funkcji.

Problem taki nie ujawnia się tak mocno przy użyciu metod wielołańcuchowych, o ile oczywiście zostaną one poprawnie zaimplementowane. Możliwość porównania kilku estymatorów osiągniętych z różnych trajektorii stanowi w takim przypadku bardzo silne narzędzie diagnostyczne. Niemniej należy pamiętać, iż podejście takie jest bardzo konserwatywne, tzn. znaczna różnica wartości pomiędzy choć jednym estymatorem a pozostałymi może wymusić niepotrzebne kontynuowanie symulacji.

Mimo tego, zdaniem autora tej pracy, generowanie tylko pojedynczego łańcucha w dobie szybkich komputerów niesie ze sobą zbyt dużo ewentualnych niebezpieczeństw, aby rewanżowały to korzyści osiągnięte z takiego podejścia (patrz np. [2]).

W następnych podrozdziałach omówiono wybrane metody diagnozujące zbieżność symulacji MCMC. Szerszy ich przegląd oraz dyskusję dotyczącą wad i zalet poszczególnych metod znaleźć można np. w [6], [17], [21], [25].

## 4.2 Wybrane metody diagnostyki zbieżności do rozkładu stacjonarnego

Jedną z prostych metod diagnostycznych mających na celu sprawdzenie, czy symulacja powstała przy wykorzystaniu MCMC osiągnęła zbieżność do rozkładu stacjonarnego, jest wykorzystanie modyfikacji dowolnego nieparametrycznego testu zgodności, np. testu Kołmogorowa-Smirnowa (patrz np. [25]). Zauważmy mianowicie, iż jeśli wyraz  $X_i$  łańcucha Markowa pochodzi z rozkładu stacjonarnego  $\pi$ , element  $X_{i+1}$  również będzie mieć rozkład  $\pi$ .

Zastosowanie testu Kołmogorowa-Smirnowa polega więc na podziale łańcucha Markowa na dwie połowy  $X_1, \dots, X_{n/2}$  oraz  $X_{n/2+1}, \dots, X_n$ . Oczywiście bezpośrednie zastosowanie odpowiednika statystyki testowej Kołmogorowa-Smirnowa nie jest możliwe ze względu na *własność Markowa* łańcucha, czyli zależność pomiędzy kolejnymi jego wyrazami. To niekorzystne zjawisko można próbować wyeliminować dokonując tzw. *subsamplingu*, tzn. wykorzystując tylko co  $k$ -wyraz w obu podłańcuchach. Otrzymujemy dzięki temu podciągi postaci  $X_1, X_k, \dots$  (oznaczany dalej jako  $Y_1, Y_2, \dots$ ) oraz  $X_{n/2+1}, X_{n/2+k}$  (oznaczany jako  $Y'_1, Y'_2, \dots$ ). Choć sposób ten zmniejsza oczywiście stopień zależności pomiędzy próbkami, nie eliminuje całkowicie tego problemu, a ponadto prowadzi do częściowej straty informacji. Interesującym wydaje się sprawdzenie, czy losowy subsampling (a więc losowy wybór wyrazów do podciągów) dałby w tym przypadku lepsze rezultaty.

Następnym krokiem jest zastosowanie zmodyfikowanej statystyki Kołmogorowa-Smirnowa

$$KS = \frac{1}{M} \sup_x \left| \sum_{i=1}^M \mathbb{1}_{(0,x)} Y_i - \mathbb{1}_{(0,x)} Y'_i \right|, \quad (15)$$

gdzie  $M$  jest największą liczbą spełniającą warunki:  $M < n/2$  i  $M = kp$  dla  $p \in \mathbb{N}$ .

Możliwe jest przy tym otrzymanie pełnej postaci gęstości dla statystyki (15), ale tylko asymptotycznie dla  $M \rightarrow \infty$ . Dokładne wyliczenie jej wartości (dla dowolnego  $M$  skończonego) jest niestety znacznie utrudnione ze względu na korelacje występujące pomiędzy poszczególnymi elementami ciągu. W dodatku trudność problemu wzrasta przy zastosowaniu metody subsamplingu z powodu skomplikowania postaci zależności pomiędzy wyrazami.

Innego typu metodą wykorzystywaną przy badaniu zbieżności do rozkładu stacjonarnego jest np. szacowanie odległości (w j. ang. *distance evaluations*, patrz np. [3]).

#### 4.3 Wybrane metody diagnostyki zbieżności do wartości oczekiwanej

Prostą, ale o zaskakująco dobrych własnościach, graficzną metodą diagnozującą zbieżność do wartości oczekiwanej jest CUSUM (z j. ang. *cumulative sums*, patrz [12], [32]). W metodzie tej badany jest wykres wartości

$$C_n(i) = \sum_{k=1}^i (h(X_k) - S_n) \quad (16)$$

dla  $i = 1, \dots, n$ , gdzie

$$S_n = \frac{1}{n} \sum_{k=1}^n h(X_k). \quad (17)$$

Jak wykazano, wykres  $C_n(\cdot)$  jest narzędziem wskazującym na stopień prędkości mieszania łańcucha. Dla szybko mieszających łańcuchów wykres ów jest bardzo nieregularny i położony w pobliżu zera. Z kolei przy wolnym mieszanii, wykres tworzy długie, gładkie trajektorie, z długimi czasami powrotu do

zerowej wartości.

Niestety, z racji bazowania analizy tylko na pojedynczym łańcuchu, wnioski uzyskane z obserwacji wartości (16) mogą być mylące w niektórych przykładach, wskazując błędnie na osiągnięcie zbieżności nawet przy jej braku. Dlatego też interesującą wydaje się próba uogólnienia metody CUSUM na podejście wielołańcuchowe. Innym pomysłem jest opracowanie nowych wskaźników ilościowych bazujących na tej metodzie.

Innym silnym narzędziem jest stosowanie kilku różnych estymatorów zbieżnych do  $\mathbb{E}_f g(X)$  (patrz [17]). Oprócz klasycznego estymatora postaci (9) możliwe jest zastosowanie dodatkowo warunkowej wartości oczekiwanej, stworzonej przy pomocy metody określanej jako *Rao-Blackwellization*. Estymator taki ma ogólną postać

$$\bar{h}_{RB} = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_f(h(X_i)|Z_i), \quad (18)$$

gdzie  $Z_i$  są pewnymi wartościami losowymi (np. momentem dotarcia do ustalonego stanu  $s$ ).

Kolejnym przybliżeniem wartości średniej może być estymator uzyskany przy pomocy metody *importance sampling* (patrz paragraf 2). Jego obliczenie polega na generowaniu z innej, niż docelowa gęstości  $g(x)$ . Następnie obliczana jest wielkość

$$\bar{h}_{IS} = \frac{1}{n} \sum_{i=1}^n h(X_i) \frac{f(X_i)}{g(X_i)}. \quad (19)$$

Estymator  $\mathbb{E}_f g(X)$  może wreszcie być stworzony przez aproksymację Riemanna (w j. ang. *Riemann approximation*, patrz [33]). Wyraża się on wtedy wzorem

$$\bar{h}_{RA} = \sum_{i=1}^{n-1} (X_{(i+1)} - X_{(i)}) h(X_{(i)}) f(X_{(i)}), \quad (20)$$

gdzie  $(X_{(i)})_i$  (dla  $i = 1, \dots, n$ ) oznacza uporządkowany monotonicznie ciąg wyjściowy  $(X_i)_i$ .

Jeśli cztery powyższe estymatory  $\bar{h}_{MC}, \bar{h}_{RB}, \bar{h}_{IS}, \bar{h}_{RA}$  są w rozważanym przykładzie możliwe do uzyskania, wykorzystać je można jako różne indykatory zbieżności do wartości oczekiwanej. W takim przypadku symulacje kończy się, jeśli wszystkie estymatory uzyskają zgodność do ustalonego miejsca po przecinku.

Choć metoda ta ma silne podstawy teoretyczne, niemniej posiada też kilka istotnych wad. Nie we wszystkich przykładach mianowicie udaje się znaleźć wszystkie ze wspomnianych powyżej estymatorów. Ponadto metoda ta jest bardzo konserwatywna i może prowadzić do znacznego zwiększenia zbędnych obliczeń. Wreszcie, ponieważ wszystkie estymatory bazują tylko na jednym łańcuchu, może zdarzyć się wyciągnięcie na ich podstawie nieprawidłowych wniosków.

Przykładem innej metody diagnostycznej jest badanie estymatorów wariancji jednocześnie pomiędzy wyrazami pojedynczego łańcucha, jak i pomiędzy poszczególnymi łańcuchami (patrz np. [25]). Niech  $K > 1$  oznacza ilość łańcuchów, a  $X_i^k$   $i$ -ty element  $k$ -tego łańcucha (dla  $i = 1, \dots, n$  oraz  $k = 1, \dots, K$ ). W takim przypadku estymator wariancji zewnętrznej (pomiędzy łańcuchami) można przedstawić w postaci

$$\overline{\text{Var}}_E = \frac{1}{K} \sum_{k=1}^K \left( \overline{h(X^k)} - \overline{h(X)} \right)^2, \quad (21)$$

gdzie

$$\overline{h(X^k)} = \frac{1}{n} \sum_{i=1}^n h(X_i^k) \quad \overline{h(X)} = \frac{1}{K} \sum_{k=1}^K \overline{h(X^k)}. \quad (22)$$

Z kolei estymator wariancji wewnętrznej (w pojedynczych łańcuchach) zapi-

sać można jako:

$$\overline{\text{Var}}_I = \frac{1}{K} \sum_{k=1}^K \left( \frac{1}{n} \sum_{i=1}^n (h(X_i^k) - \overline{h(X^k)})^2 \right). \quad (23)$$

Oprócz wielkości  $\overline{\text{Var}}_I$  estymatorem wariancji  $\text{Var}(h(X))$  może też być wartość

$$\overline{\text{Var}}_M = \frac{n-1}{n} \overline{\text{Var}}_I + \overline{\text{Var}}_E. \quad (24)$$

Jak dowiedziono, estymator postaci (24) przybliża od dołu estymowaną wielkość, a estymator (23) przybliża ją od góry (przy zachowaniu pewnych warunków).

Kontrola zbieżności przy tak wprowadzonych oznaczeniach polega na monitorowaniu wielkości

$$R = \frac{\overline{\text{Var}}_M + \frac{\overline{\text{Var}}_E}{K}}{\overline{\text{Var}}_I} \frac{\nu}{\nu - 2}, \quad (25)$$

gdzie

$$\nu = 2 \frac{(\overline{\text{Var}}_M^2 + \frac{\overline{\text{Var}}_E}{K})^2}{\overline{\text{Var}}_I}. \quad (26)$$

Metoda polega wtedy na sprawdzeniu, czy wielkość  $R$  zbliżyła się do jedynki. Inną możliwością jest stworzenie przedziałów ufności przy założeniu przybliżenia rozkładu zmiennej losowej  $R$  pewnym rozkładem Fishera  $\mathcal{F}$ .

#### 4.4 Inne metody diagnostyki zbieżności metod MCMC

Oprócz metod omówionych pokrótce w podrozdziałach 4.2 i 4.3, w literaturze spotkać można jeszcze wiele innych sposobów monitorowania zbieżności. Przykładowo są to: wykorzystanie metody *couplingu* (patrz [34]), estymowanie normy *total variation* dla rozkładu otrzymanego z symulacji (patrz



[1]), wykorzystanie pewnych własności gęstości  $P(X_{i+1}|X_i)$  (patrz [3]), estymatorów uzyskanych heurystycznie z symulacji pomocniczych (patrz [6]) i inne (patrz [5], [6], [17], [21])

Należy też zwrócić uwagę, że przy bezkrytycznym wykorzystaniu metod diagnostycznych bazujących bezpośrednio na wygenerowanej próbce możliwe jest otrzymanie obciążonych estymatorów poszukiwanych wielkości (patrz np. [29]). Obserwacja ta otwiera drogę do nowych, interesujących możliwości badawczych.

## 5 Uwagi końcowe

W świetle powyższego wprowadzenia wydaje się, iż możliwości zaproponowania nowych metod diagnostycznych dla monitorowania zbieżności metod MCMC są bardzo szerokie (patrz np. [16], [19]). Również badania teoretyczne na tym polu mogą być bardzo owocne i interesujące.

Dodatkowym ważkim atutem przemawiającym za kontynuowaniem badań w tej dziedzinie jest unikalna możliwość sprawdzenia działania tych metod na rozmaitych przykładach spotykanych w literaturze przedmiotu, jak i na specjalnych *benchmarkach*.

## Literatura

- [1] Brooks S. P., Dellaportas P., Roberts G. O., „*An Approach to Diagnosing Total Variation Convergence of MCMC Algorithms*”, Journal of Computational and Graphical Statistics, Vol. 6, No. 3, 1997

- [2] Gelman A., Rubin D. B., „*A Single Series from the Gibbs Sampler Provides a False Sense of Security*” w Bernardo J. M., Berger J. O., Dawid A. P., Smith A. F. M. (eds.) *Bayesian Statistics 4*, Oxford University Press 1992
- [3] Kolassa J. E., „*Bounding convergence rates for Markov chains: An example of the use of computer algebra*”, *Statistics and Computing*, 11, 2001
- [4] Thomas A., Spiegelhalter D. J., Gilks W. R., „*BUGS: a Program to Perform Bayesian Inference using Gibbs Sampling*” w Bernardo J. M., Berger J. O., Dawid A. P., Smith A. F. M. (eds.) *Bayesian Statistics 4*, Oxford University Press 1992
- [5] Brooks S. P., Roberts G. O., „*Convergence assesment techniques for Markov chain Monte Carlo*”, *Statistics and Computing*, 8, 1998
- [6] Robert Ch. P., „*Convergence Control Methods for Markov Chain Monte Carlo Algorithms*”, *Statistical Science*, 1995, Vol. 10, No. 3
- [7] Nowak P., Nycz P., Romaniuk M., „*Dobór optymalnego modelu stochastycznego w wycenie opcji metodami Monte-Carlo*”, publikacja na BOS 2002, 2002
- [8] Mehta C. R., Patel N. R., Senchaudhuri P., „*Efficient Monte Carlo Methods for Conditional Logistic Regression*”, *Journal of the American Statistical Association*, March 2000, Vol. 95, No. 449
- [9] Gelman A., Rubin D. B., „*Inference from Iterative Simulation Using Multiple Sequences*”, *Statistical Science*, 1992, Vol. 7, No. 4

- [10] Gelfand A. E., Hills S. E., Racine-Poon A., Smith A. F. M., „*Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling*”, Journal of the American Statistical Association, December 1990, Vol. 85, No. 412
- [11] Wieczorkowski R., Zieliński R., „*Komputerowe generatory liczb losowych*”, WNT 1997
- [12] Yu B., Mykland P., „*Looking at Markov samplers through cusum path plots: a simple diagnostic idea*”, Statistics and Computing, 8, 1998
- [13] Galanti, Jung, „*Low Discrepancy Sequences: Monte Carlo Simulation of Option Prices*”, Journal of Derivatives, Fall 1997
- [14] Nowak P., Nycz P., Romaniuk M., „*Metody Monte-Carlo w wycenie instrumentów pochodnych*”, raport techniczny IBS PAN, RB/39/2001, 2001
- [15] Brémaud P., „*Markov Chains – Gibbs Fields, Monte Carlo Simulation and Queues*”, Springer Verlag 1999
- [16] Cappé O., Robert Ch. P., „*Markov Chain Monte Carlo: 10 Years and Still Running!*”, Journal of the American Statistical Association, December 2000, Vol. 95, No. 452
- [17] Cowles M. K., Carlin B. P., „*Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review*”, Journal of the American Statistical Association, June 1996, Vol. 91, No. 434
- [18] Gilks W. R., Richardson S., Spiegelhalter D. J., „*Markov Chain Monte Carlo in Practice*”, Chapman & Hall, 1997

- [19] Kass R. E., Carlin B. P., Gelman A., Neal R. M., „*Markov Chain Monte Carlo in Practice: A Roundtable Discussion*”, *The American Statistician*, May 1998, Vol. 52, No. 2
- [20] Brooks S. P., „*Markov chain Monte Carlo method and its application*”, *The Statistician*, 47, 1998
- [21] Mengersen K. L., Robert Ch. P., Guihenneuc-Jouyaux C., „*MCMC Convergence Diagnostics: A Review*” w Bernardo J. M., Berger J. O., Dawid A. P., Smith A. F. M. (eds.) *Bayesian Statistics 6*, Oxford University Press 1999
- [22] Cui L., Tanner M. A., Sinha D., Hall W. J., „*Monitoring Convergence of the Gibbs Sampler: Further Experience with the Gibbs Stopper*”, *Statistical Science*, 1992, Vol. 7, No. 4
- [23] Booth J. G., Sarkar S., „*Monte Carlo Approximation of Bootstrap Variances*”, *The American Statistician*, November 1998, Vol. 52, No. 4
- [24] Boos D. D., Zhang J., „*Monte Carlo Evaluation of Resampling – Based Hypothesis Tests*”, *Journal of the American Statistical Association*, June 2000, Vol. 95, No. 450
- [25] Robert Ch. P., Casella G., „*Monte Carlo Statistical Methods*”, Springer Verlag 1999
- [26] Raftery A. E., Lewis S. M., „*One Long Run with Diagnostics: Implementation Strategies for Markov Chain Monte Carlo*”, *Statistical Science*, 1992, Vol. 7, No. 4

- [27] Doucet A., Godsill S., Andrieu Ch., „*On sequential Monte Carlo sampling methods for Bayesian filtering*”, *Statistics and Computing*, 10, 2000
- [28] Li S., Pearl D. K., Doss H., „*Phylogenetic Tree Construction Using Markov Chain Monte Carlo*”, *Journal of the American Statistical Association*, June 2000, Vol. 95, No. 450
- [29] Cowles M. K., Roberts G. O., Rosenthal J. S., „*Possible Biases Induced by MCMC Convergence Diagnostics*”, *J. Statist. Comput. Simul.*, Vol. 64, 1999
- [30] Romaniuk M., „*Pricing risk-transfer financial instruments via Monte Carlo methods*”, Interim Report, IIASA 2002 (jeszcze nie opublikowany)
- [31] Geyer C. J., „*Practical Markov Chain Monte Carlo*”, *Statistical Science*, 1992, Vol. 7, No. 4
- [32] Brooks S. P., „*Quantitative convergence assessment for Markov chain Monte Carlo via cusums*”, *Statistic and Computing*, 8, 1998
- [33] Philppe A., Robert Ch. P., „*Riemann sums for MCMC estimation and convergence monitoring*”, *Statistics and Computing*, 11, 2001
- [34] Johnson V. E., „*Studying Covergence of Markov Chain Monte Carlo Algorithms Using Coupled Sample Paths*”, *Journal of the American Statistical Association*, March 1996, Vol. 91, No. 433
- [35] MacEachern S. N., Berliner L. M., „*Subsampling the Gibbs Sampler*”, *The American Statistician*, August 1994, Vol. 48, No. 3

- [36] Jakubowski J., Sztencel R., „*Wstęp do teorii prawdopodobieństwa*”,  
SCRIPT, Warszawa 2000
- [37] Metropolis N., Rosenbluth A.W., Rosenbluth M. N., Teller A. H., Teller  
E., „*Equations of state calculations by fast computing machines*”, J.  
Chem. Phys. 21, 1953
- [38] Fortuna Z., Macukow B., Wąsowski J., „*Metody numeryczne*”, WNT  
1998
- [39] Metropolis N., Ulam S., „*The Monte Carlo Method*”, Journal of American  
Statistical Association, 44, 1949

## Spis treści

<b>1</b>	<b>Wprowadzenie</b>	<b>1</b>
<b>2</b>	<b>Metody Monte Carlo</b>	<b>3</b>
<b>3</b>	<b>Metody Markov Chain Monte Carlo</b>	<b>6</b>
<b>4</b>	<b>Diagnostyka zbieżności metod Markov Chain Monte Carlo</b>	<b>13</b>
4.1	Podział metod diagnostycznych . . . . .	14
4.2	Wybrane metody diagnostyki zbieżności do rozkładu stacjo- narnego . . . . .	18
4.3	Wybrane metody diagnostyki zbieżności do wartości oczekiwanej	19
4.4	Inne metody diagnostyki zbieżności metod MCMC . . . . .	22
<b>5</b>	<b>Uwagi końcowe</b>	<b>23</b>



