

Dariusz Biskup

Akademia Ekonomiczna we Wrocławiu

BAYESOWSKA ESTYMACJA PARAMETRÓW MIESZANEK ROZKŁADÓW O NIEZNANEJ LICZBIE SKŁADNIKÓW

1. Wstęp

Algorytm Reversible Jump Markov Chain Monte Carlo (RJMCMC) służy do generowania próbek losowych z przestrzeni parametrów o zmiennej liczbie wymiarów (por. [Green 1995; *Highly Structured...* 2003]). Sytuacja taka występuje np. wówczas, gdy celem jest znalezienie prawdopodobieństwa *a posteriori* jednego z modeli. Typowym przykładem takiego zagadnienia jest wybór liczby składników mieszanki ustalonej rodziny rozkładów prawdopodobieństwa. Inne obszary zastosowań to wybór zmiennych do modelu liniowego, wybór rzędu procesu autoregresji i wiele innych.

W przypadku rozpatrywania modeli o różnym wymiarze przestrzeni parametrów algorytm RJMCMC cechuje znacznie większa efektywność niż np. algorytm Gibbsa bądź Metropolisa-Hastingsa.

2. Zagadnienie estymacji parametrów mieszanki rozkładów

W artykule opisana zostanie metoda estymacji parametrów mieszanki jednowymiarowych rozkładów normalnych z założeniem nieznaności liczby jej składników (por. [Richardson, Green 1997]). Wybór liczby składników mieszanki dokonany będzie na podstawie kryterium prawdopodobieństwa *a posteriori*.

Założmy, że dany jest ciąg n -niezależnych zmiennych losowych Y_1, \dots, Y_n , których gęstość prawdopodobieństwa dana jest wzorem

$$Y_i \sim \sum_{j=1}^k w_j f(y_i | \theta_j), \quad i = 1, \dots, n, \quad \sum_{j=1}^k w_j = 1, \quad (1)$$

gdzie $f(y_i|\theta_j)$ oznacza dowolną funkcję gęstości zależną od wektora parametrów θ_j , parametry w_j oznaczają wagi poszczególnych składników mieszanki (interpretowane jako prawdopodobieństwa *a priori*, że dana obserwacja pochodzi z j -tego składnika), k oznacza liczbę składników mieszanki. Parametry modelu (1) traktowane będą jako zmienne losowe. W szczególności zmienną losową jest liczba składników mieszanki k .

Do każdej obserwacji y_i można przypisać nieobserwowaną zmienną losową Z_i , która opisywać będzie numer składnika, z którego pochodzi i -ta obserwacja. Oznacza to, że:

$$p(Z_i = j) = w_j, \quad \text{dla } j = 1, \dots, k, i = 1, \dots, n. \quad (2)$$

3. Metoda Reversible Jump Markov Chain Monte Carlo

Algorytm RJMCMC stanowi uogólnienie algorytmu Metropolisa-Hastingsa. Gdy nie jest zmieniany wymiar przestrzeni parametrów modelu, odpowiada on dokładnie algorytmowi Metropolisa-Hastingsa (z kolei szczególnym przypadkiem algorytmu Metropolisa-Hastingsa jest algorytm Gibbsa). W sytuacji gdy występuje zmiana wymiaru przestrzeni parametrów, algorytm RJMCMC składa się z następujących kroków (por. [Han, Carlin 2001]):

1. Niech (j, θ_j) oznacza aktualnie wylosowane wartości wektora parametrów θ_j modelu o numerze j , którego wymiar wynosi k_j .
2. Wykonanie przejścia do stanu $(j', \theta_{j'})$ ze stanu (j, θ_j) z prawdopodobieństwem $h(j, j')$.
3. Wylosowanie wektora \mathbf{u} z zadanego rozkładu $q(\mathbf{u}|\theta_j, j, j')$.
4. Podstawienie $(\theta_{j'}) = g_{j,j'}(\theta_j, \mathbf{u})$, gdzie $g_{j,j'}$ oznacza wzajemnie jednoznaczna, deterministyczną funkcję. Podstawienie to ma na celu dopasowanie wymiarów modelu aktualnego do modelu docelowego ($k_j + \dim(\mathbf{u}) = k_{j'}$).
5. Akceptacja zmiany modelu z prawdopodobieństwem $\min(1, A)$, gdzie:

$$A = \frac{f(\mathbf{y}|\theta_{j'}, j') p(\theta_{j'}|j') p(j') h(j', j)}{f(\mathbf{y}|\theta_j, j) p(\theta_j|j) p(j) h(j, j') q(\mathbf{u}|\theta_j, j, j')} \left| \frac{\partial g_{j,j'}(\theta_j, \mathbf{u})}{\partial(\theta_j, \mathbf{u})} \right|. \quad (3)$$

Jeżeli nie następuje zmiana modelu, nowe wartości wektora parametrów mogą być generowane z wykorzystaniem algorytmu Gibbsa.

4. Zastosowanie algorytmu RJMCMC do estymacji parametrów mieszanek

4.1. Wprowadzenie

Rozpatrzony zostanie przypadek, w którym składnikami mieszanki będą jednowymiarowe rozkłady normalne, co oznacza, że:

$$f(y_i | \theta_j) = f(y_i | \mu_j, \sigma_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(y_i - \mu_j)^2}{2\sigma_j^2}\right). \quad (4)$$

Przyjęte będą następujące nieinformacyjne rozkłady *a priori* dla parametrów μ_j oraz σ_j^{-2} (por. [Richardson, Green 1997]):

$$\mu_j \sim N\left(\xi, \sqrt{\kappa^{-1}}\right), \quad \sigma_j^{-2} \sim \Gamma(\alpha, \beta). \quad (5)$$

Ustalona zostanie jednakowa wartość parametru ξ dla każdego składnika mieszanki. Wyznaczona ona będzie na poziomie środkowej wartości zakresu danych. Wartość parametru κ ustalona zostanie na poziomie $\frac{1}{R^2}$, gdzie R oznacza różnicę pomiędzy wartością maksymalną a minimalną zbioru danych. Ponadto przyjmie się, iż $\alpha = 2$ oraz $\beta \sim \Gamma(g, h)$, gdzie $g = 0,2$ i $h = 10$ (por. [Richardson, Green 1997]).

Przyjmie się ponadto, że rozkład wektora wag $\mathbf{w} = (w_1, \dots, w_k)$ jest rozkładem Dirichleta: $\mathbf{w} \sim D(\delta, \delta, \dots, \delta)$. Brak informacji *a priori* o wartościach wag można opisać przez przyjęcie wartości parametru $\delta = 1$. Liczba składników mieszanki k będzie miała *a priori* rozkład jednostajny, przy ustalonej maksymalnej liczbie składników mieszanki na poziomie $k_{\max} = 30$.

Można pokazać, że warunkowe rozkłady parametrów modelu (wykorzystywane w sytuacji, gdy stosowany jest algorytm Gibbsa) są następujące (por. [Richardson, Green 1997]):

$$\mathbf{w} | \dots \sim D(\delta + n_1, \dots, \delta + n_k), \quad (6)$$

gdzie $n_j = \text{card}\{i : z_i = j\}$,

$$\mu_j | \dots \sim N\left(\frac{\sigma_j^{-2} \sum_{i:z_i=j} y_i + \kappa \xi}{\sigma_j^{-2} n_j + \kappa}, \sqrt{(\sigma_j^{-2} n_j + \kappa)^{-1}}\right), \quad (7)$$

$$\sigma_j^{-2} | \dots \sim \Gamma \left(\alpha + \frac{1}{2} n_j, \beta + \frac{1}{2} \sum_{i: z_i = j} (y_i - \mu_j)^2 \right), \quad (8)$$

$$p(z_i = j | \dots) \propto \frac{w_j}{\sigma_j} \exp \left(-\frac{(y_i - \mu_j)^2}{2\sigma_j^2} \right), \quad (9)$$

$$\beta | \dots \sim \Gamma \left(g + k\alpha, h + \sum_j \sigma_j^{-2} \right), \quad (10)$$

Losowanie dla parametrów danych wzorami (6)-(10) wykonywane będzie z użyciem algorytmu Gibbsa.

Pojedynca iteracja algorytmu RJMCMC zaadaptowanego do estymacji parametrów mieszanki składa się z następujących kroków (por. [Richardson, Green 1997]):

- a) aktualizacja wag w_j ($j = 1, \dots, k$),
- b) aktualizacja parametrów μ_j i σ_j ($j = 1, \dots, k$),
- c) aktualizacja przypisań do klas z_i ($i = 1, \dots, n$),
- d) aktualizacja parametru β ,
- e) podział pojedynczego składnika mieszanki na dwa składniki lub połączenie dwóch składników w jeden,
- f) dołączenie lub usunięcie pustego składnika.

W kolejnych dwóch podpunktach opisane zostaną szczegółowo kroki e) oraz f) algorytmu.

4.2. Podział lub połączenie składnika

W kroku e) dokonywany jest losowy wybór pomiędzy podziałem i połączeniem składników z prawdopodobieństwami b_k (dla podziału) oraz $d_k = 1 - b_k$ (dla połączenia). Prawdopodobieństwa te zależą od parametru k . Przyjęto, że: $d_1 = 0$, $b_{k_{\max}} = 0$ oraz $b_k = d_k = 0,5$ dla $k = 2, 3, \dots, k_{\max} - 1$. Połączenie składników polega na losowym wyborze pary składników o numerach j_1 i j_2 , które sąsiadują ze sobą ze względu na wartość parametru μ . Składniki te są łączone i jest tworzony nowy składnik o numerze j^* (numery istniejących składników są następnie odpowiednio przenumerowywane). Charakterystyki dla nowego składnika j^* wyznaczone są zgodnie z następującymi wzorami (por. [Richardson, Green 1997]):

$$\begin{aligned} w_{j^*} &= w_{j_1} + w_{j_2}, \\ w_{j^*} \mu_{j^*} &= w_{j_1} \mu_{j_1} + w_{j_2} \mu_{j_2}, \\ w_{j^*} (\mu_{j^*}^2 + \sigma_{j^*}^2) &= w_{j_1} (\mu_{j_1}^2 + \sigma_{j_1}^2) + w_{j_2} (\mu_{j_2}^2 + \sigma_{j_2}^2). \end{aligned} \quad (11)$$

Podział składnika polega na losowym wyborze komponentu j^* , który będzie podzielony na składniki j_1 i j_2 . Parametry oraz wagi składników j_1 i j_2 spełniają zależności dane równaniami (11). Ponieważ równania te nie wyznaczają jednoznacznie nowych parametrów, losowane są trzy wartości u_1, u_2 i u_3 , zgodnie z następującymi rozkładami (por. [Richardson, Green 1997]):

$$u_1 \sim Be(2,2), \quad u_2 \sim Be(2,2), \quad u_3 \sim Be(1,1). \quad (12)$$

Wybór rozkładów podanych we wzorze (12) jest w zasadzie dowolny. Celem jest uzyskanie jak najwyższego prawdopodobieństwa akceptacji (wzór (3)), co powoduje, że zmiany modelu występują stosunkowo częściej. Pozwala to zredukować ogólną liczbę iteracji algorytmu. Wagi oraz parametry dla nowych składników wyznaczone są następnie zgodnie z równaniami (wynikają one z zależności (11)):

$$\begin{aligned} w_{j_1} &= w_{j^*} u_1, \\ w_{j_2} &= w_{j^*} (1 - u_1), \\ \mu_{j_1} &= \mu_{j^*} - u_2 \sigma_{j^*} \sqrt{\frac{w_{j_2}}{w_{j_1}}} = \mu_{j^*} - u_2 \sigma_{j^*} \sqrt{\frac{1}{u_1} - 1}, \\ \mu_{j_2} &= \mu_{j^*} + u_2 \sigma_{j^*} \sqrt{\frac{w_{j_1}}{w_{j_2}}} = \mu_{j^*} + u_2 \sigma_{j^*} \sqrt{\frac{u_1}{1 - u_1}}, \\ \sigma_{j_1}^2 &= u_3 (1 - u_2^2) \sigma_{j^*}^2 \frac{w_{j^*}}{w_{j_1}} = u_3 (1 - u_2^2) \sigma_{j^*}^2 \frac{1}{u_1}, \\ \sigma_{j_2}^2 &= (1 - u_3) (1 - u_2^2) \sigma_{j^*}^2 \frac{w_{j^*}}{w_{j_2}} = (1 - u_3) (1 - u_2^2) \sigma_{j^*}^2 \frac{1}{1 - u_1}. \end{aligned} \quad (13)$$

Gdy nowe składniki nie spełniają warunku sąsiedztwa (czyli wartość oczekiwana innego składnika należy do przedziału (μ_{j_1}, μ_{j_2})), wówczas krok algorytmu, polegający na podziale składnika, nie jest wykonywany. Obserwacje y_i należące do dzielonego składnika przypisywane są do nowych składników, zgodnie ze wzorem (9).

Jakobian transformacji wektora $(w_{j^*}, \mu_{j^*}, \sigma_{j^*}^2, u_1, u_2, u_3)$ na wektor $(w_{j_1}, \mu_{j_1}, \sigma_{j_1}^2, w_{j_2}, \mu_{j_2}, \sigma_{j_2}^2)$, wyznaczany na podstawie funkcji opisanych wzorem (13), ma postać:

$$\begin{vmatrix}
 u_1 & 0 & 0 & w_{j'} & 0 & 0 \\
 0 & 1 & -u_2 \sqrt{\frac{1}{u_1}-1} \frac{1}{2\sigma_{j'}} & u_2 \sigma_{j'} \frac{1}{2u_1^2 \sqrt{\frac{1}{u_1}-1}} & -\sigma_{j'} \sqrt{\frac{1}{u_1}-1} & 0 \\
 0 & 0 & u_3 (1-u_2^2) \frac{1}{u_1} & \frac{-u_3 (1-u_2^2) \sigma_{j'}^2}{u_1^2} & -2u_2 u_3 \sigma_{j'}^2 \frac{1}{u_1} & (1-u_2^2) \sigma_{j'}^2 \frac{1}{u_1} \\
 1-u_1 & 0 & 0 & -w_{j'} & 0 & 0 \\
 0 & 1 & u_2 \sqrt{\frac{u_1}{1-u_1}} \frac{1}{2\sigma_{j'}} & u_2 \sigma_{j'} \frac{\frac{1}{1-u_1} + \frac{u_1}{(1-u_1)^2}}{2\sqrt{\frac{u_1}{1-u_1}}} & \sigma_{j'} \sqrt{\frac{u_1}{1-u_1}} & 0 \\
 0 & 0 & (1-u_3)(1-u_2^2) \frac{1}{1-u_1} & \frac{(1-u_3)(1-u_2^2) \sigma_{j'}^2}{(1-u_1)^2} & -2u_2 (1-u_3) \sigma_{j'}^2 \frac{1}{1-u_1} & -(1-u_2^2) \sigma_{j'}^2 \frac{1}{1-u_1}
 \end{vmatrix} = (14)$$

$$= \frac{\sigma_{j'}^3 w_{j'} (u_2^2 - 1) \left(\sqrt{\frac{1}{u_1}-1} + \sqrt{\frac{u_1}{1-u_1}} \right)}{u_1 (u_1 - 1)}.$$

Prawdopodobieństwo akceptacji wyrażone wzorem (3) w odniesieniu do zagadnienia estymacji parametrów mieszanki, gdy liczba składników mieszanki powiększana jest o jeden, przybiera następującą postać:

$$A = \frac{p(k') p\left(y \mid \mu', (\sigma^2)', z'\right) p(w' \mid k', \delta) p(z' \mid w', k') p\left(\mu', (\sigma^2)' \mid k'\right)}{p(k) p(y \mid \mu, \sigma^2, z) p(w \mid k, \delta) p(z \mid w, k) p(\mu, \sigma^2 \mid k)} \times (15)$$

$$\times \frac{h(k', k)}{h(k, k')} \times \frac{1}{q(\mathbf{u} \mid \mu, \sigma^2, j, j')} \times |J|,$$

gdzie: $k' = k + 1$,

$\mathbf{z} = (z_1, \dots, z_n)$,

$\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)$,

$\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_k^2)$,

$\mathbf{y} = (y_1, \dots, y_n)$,

$|J|$ – jacobian transformacji.

W dalszej części pracy przedstawione zostanie wyprowadzenie wzoru na prawdopodobieństwo akceptacji dla podziału składnika mieszanki.

Założono jednostajny rozkład *a priori* dla liczby składników mieszanki, co oznacza, że $p(k+1) = p(k) = \frac{1}{k_{\max}}$. Składniki $p\left(\mathbf{y} \mid \boldsymbol{\mu}', (\boldsymbol{\sigma}^2)', \mathbf{z}'\right)$ oraz $p(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}^2, \mathbf{z})$ opisują funkcje wiarygodności dla modelu o odpowiednio $k+1$ i k składnikami. Funkcja dla modelu o k składnikami ma następującą postać (przyjęto uproszczone oznaczenie $L(k)$):

$$L(k) = \prod_{j=1}^k \prod_{i: z_i=j} \frac{1}{\sigma_j \sqrt{2\pi}} \exp\left(-\frac{(y_i - \mu_j)^2}{2\sigma_j^2}\right). \quad (16)$$

Analogiczny wzór można podać dla funkcji $L(k+1)$.

Składniki $p(\mathbf{w}' \mid k', \delta)$ oraz $p(\mathbf{w} \mid k, \delta)$ wzoru (15) oznaczają rozkłady *a priori* wag mieszanki o $k+1$ i k składnikami. Wagi te mają *a priori* rozkład $D(\delta, \dots, \delta)$, co oznacza, że

$$\begin{aligned} \frac{p(\mathbf{w}' \mid k', \delta)}{p(\mathbf{w} \mid k, \delta)} &= \frac{\frac{\Gamma((k+1)\delta)}{\Gamma(\delta) \cdots \Gamma(\delta)}}{\frac{\Gamma(k\delta)}{\Gamma(\delta) \cdots \Gamma(\delta)}} \times \frac{w_1^{\delta-1} \cdots w_{j_1}^{\delta-1} w_{j_2}^{\delta-1} \cdots w_{k+1}^{\delta-1}}{w_1^{\delta-1} \cdots w_{j_*}^{\delta-1} \cdots w_k^{\delta-1}} = \frac{\Gamma(k\delta + \delta)}{\Gamma(\delta)\Gamma(k\delta)} \times \frac{w_{j_1}^{\delta-1} w_{j_2}^{\delta-1}}{w_{j_*}^{\delta-1}} = \\ &= \frac{1}{B(\delta, k\delta)} \times \frac{w_{j_1}^{\delta-1} w_{j_2}^{\delta-1}}{w_{j_*}^{\delta-1}}, \end{aligned} \quad (17)$$

gdzie $B(\cdot, \cdot)$ oznacza funkcję beta.

Z kolei

$$\frac{p(\mathbf{z}' \mid \mathbf{w}', k')}{p(\mathbf{z} \mid \mathbf{w}, k)} = \frac{w_1^{n_1} \cdots w_{j_1}^{l_1} w_{j_2}^{l_2} \cdots w_{k+1}^{n_{k+1}}}{w_1^{n_1} \cdots w_{j_*}^{l_1+l_2} \cdots w_k^{n_k}} = \frac{w_{j_1}^{l_1} w_{j_2}^{l_2}}{w_{j_*}^{l_1+l_2}}, \quad (18)$$

gdzie l_1, l_2 jest to liczba obserwacji przypisanych do składników j_1 i j_2 . Oznacza to, że

$$\frac{p(\mathbf{w}' \mid k', \delta)}{p(\mathbf{w} \mid k, \delta)} \frac{p(\mathbf{z}' \mid \mathbf{w}', k')}{p(\mathbf{z} \mid \mathbf{w}, k)} = \frac{w_{j_1}^{\delta-1+l_1} w_{j_2}^{\delta-1+l_2}}{w_{j_*}^{\delta-1+l_1+l_2} B(\delta, k\delta)}. \quad (19)$$

Ponieważ parametry μ_j oraz σ_j^2 są *a priori* niezależne, $p(\boldsymbol{\mu}, \boldsymbol{\sigma}^2 \mid k) = p(\boldsymbol{\mu} \mid k) p(\boldsymbol{\sigma}^2 \mid k)$. Ze wzoru (5) otrzymujemy, że

$$\frac{p(\mu'|k')}{p(\mu|k)} = \frac{(k+1)! \sqrt{\frac{\kappa}{2\pi}} \sqrt{\frac{\kappa}{2\pi}} \exp\left(-\frac{1}{2}\kappa\left((\mu_{j_1}-\xi)^2 + (\mu_{j_2}-\xi)^2\right)\right)}{k! \sqrt{\frac{\kappa}{2\pi}} \exp\left(-\frac{1}{2}\kappa(\mu_{j^*}-\xi)^2\right)} = \quad (20)$$

$$= (k+1) \sqrt{\frac{\kappa}{2\pi}} \exp\left(-\frac{1}{2}\kappa\left((\mu_{j_1}-\xi)^2 + (\mu_{j_2}-\xi)^2 - (\mu_{j^*}-\xi)^2\right)\right)$$

oraz

$$\frac{p\left((\sigma^2)'\right)}{p(\sigma^2)} = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\beta^\alpha}{\Gamma(\alpha)} \times \frac{(\sigma_{j_1}^{-2})^{\alpha-1} \exp(-\beta\sigma_{j_1}^{-2}) \times (\sigma_{j_2}^{-2})^{\alpha-1} \exp(-\beta\sigma_{j_2}^{-2})}{(\sigma_{j^*}^{-2})^{\alpha-1} \exp(-\beta\sigma_{j^*}^{-2})} = \quad (21)$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{\sigma_{j_1}^2 \sigma_{j_2}^2}{\sigma_{j^*}^2}\right)^{-\alpha+1} \exp(-\beta(\sigma_{j_1}^{-2} + \sigma_{j_2}^{-2} - \sigma_{j^*}^{-2})).$$

Do wyznaczenia prawdopodobieństwa akceptacji niezbędne jest również obliczenie prawdopodobieństwa uzyskania określonej alokacji elementów składnika podlegającego podziałowi na składniki docelowe. Prawdopodobieństwo to można obliczyć zgodnie z następującym wzorem:

$$P_{alloc} = \prod_{i:z_i=j_1} \frac{\frac{w_{j_1}}{\sigma_{j_1}} \exp\left(-\frac{(y_i - \mu_{j_1})^2}{2\sigma_{j_1}^2}\right)}{\frac{w_{j_1}}{\sigma_{j_1}} \exp\left(-\frac{(y_i - \mu_{j_1})^2}{2\sigma_{j_1}^2}\right) + \frac{w_{j_2}}{\sigma_{j_2}} \exp\left(-\frac{(y_i - \mu_{j_2})^2}{2\sigma_{j_2}^2}\right)} \times \quad (22)$$

$$\times \prod_{i:z_i=j_2} \frac{\frac{w_{j_2}}{\sigma_{j_2}} \exp\left(-\frac{(y_i - \mu_{j_2})^2}{2\sigma_{j_2}^2}\right)}{\frac{w_{j_1}}{\sigma_{j_1}} \exp\left(-\frac{(y_i - \mu_{j_1})^2}{2\sigma_{j_1}^2}\right) + \frac{w_{j_2}}{\sigma_{j_2}} \exp\left(-\frac{(y_i - \mu_{j_2})^2}{2\sigma_{j_2}^2}\right)}.$$

Uwzględniając ponadto, że $h(j', j) = d_{k+1}$, $h(j, j') = b_k P_{alloc}$, wyznaczając funkcję $q(\cdot|\cdot)$ na podstawie rozkładów (12) oraz korzystając ze wzoru (14), otrzymujemy, że prawdopodobieństwo akceptacji dla podziału wynosi $\min(1, A)$, przy czym

$$\begin{aligned}
A &= \frac{L(k+1)}{L(k)} \frac{p(k+1)}{p(k)} (k+1) \frac{w_{j1}^{\delta-1+l_1} w_{j2}^{\delta-1+l_2}}{w_{j^*}^{\delta-1+l_1+l_2} B(\delta, k\delta)} \times \\
&\times \sqrt{\frac{\kappa}{2\pi}} \exp\left(-\frac{1}{2} \kappa ((\mu_{j1} - \xi)^2 + (\mu_{j2} - \xi)^2 - (\mu_{j^*} - \xi)^2)\right) \times \\
&\times \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{\sigma_{j1}^2 \sigma_{j2}^2}{\sigma_{j^*}^2}\right)^{-\alpha+1} \exp(-\beta(\sigma_{j1}^{-2} + \sigma_{j2}^{-2} - \sigma_{j^*}^{-2})) \times \\
&\times \frac{d_{k+1}}{b_k P_{alloc}} (g_{2,2}(u_1) g_{2,2}(u_2) g_{1,1}(u_3))^{-1} \times \\
&\times \frac{\sigma_{j^*}^3 w_{j^*} (u_2^2 - 1) \left(\sqrt{\frac{1}{u_1} - 1} + \sqrt{\frac{u_1}{1 - u_1}}\right)}{u_1 (u_1 - 1)},
\end{aligned}$$

gdzie: k – liczba składników przed podziałem,

$B(\cdot, \cdot)$ – funkcja beta,

$L(\cdot)$ – funkcja wiarygodności,

P_{alloc} – prawdopodobieństwo dokonania zadanego podziału,

$g_{p,q}$ – gęstość rozkładu beta o parametrach p i q .

Podobnie można wyprowadzić prawdopodobieństwo akceptacji dla połączenia. Wynosi ono $\min(1, B)$, przy czym

$$\begin{aligned}
B &= \frac{L(k-1)}{L(k)} \frac{p(k-1)}{p(k)} \frac{1}{k} \frac{w_{j^*}^{\delta-1+l_1+l_2} B(\delta, (k-1)\delta)}{w_{j1}^{\delta-1+l_1} w_{j2}^{\delta-1+l_2}} \times \\
&\times \sqrt{\frac{2\pi}{\kappa}} \exp\left(-\frac{1}{2} \kappa ((\mu_{j^*} - \xi)^2 - (\mu_{j1} - \xi)^2 - (\mu_{j2} - \xi)^2)\right) \times \\
&\times \frac{\Gamma(\alpha)}{\beta^\alpha} \left(\frac{\sigma_{j^*}^2}{\sigma_{j1}^2 \sigma_{j2}^2}\right)^{-\alpha+1} \exp(-\beta(\sigma_{j^*}^{-2} - \sigma_{j1}^{-2} - \sigma_{j2}^{-2})) \times \\
&\times \frac{b_{k-1} P_{alloc}}{d_k} g_{2,2}(u_1) g_{2,2}(u_2) g_{1,1}(u_3) \times \\
&\times \frac{u_1 (u_1 - 1)}{\sigma_{j^*}^3 w_{j^*} (u_2^2 - 1) \left(\sqrt{\frac{1}{u_1} - 1} + \sqrt{\frac{u_1}{1 - u_1}}\right)}.
\end{aligned}$$

4.3. Dołączenie lub usunięcie pustego składnika

Krok f) algorytmu opisuje drugą możliwość modyfikacji wymiaru przestrzeni parametrów modelu. Polega on na dołączeniu bądź usunięciu pustego składnika mieszanki, a więc takiego, do którego w danym momencie nie jest przypisana żadna obserwacja. W pierwszej kolejności dokonywany jest wybór pomiędzy usunięciem a dołączeniem składnika – wykorzystywane są do tego prawdopodobieństwa b_k oraz d_k . Gdy wybrane zostanie dołączenie nowego składnika, wówczas jego parametry losowane są z następujących rozkładów:

$$w_{j^*} \sim B(1, k), \quad \mu_{j^*} \sim N(\xi, \kappa^{-1}), \quad \sigma_{j^*}^{-2} \sim \Gamma(\alpha, \beta). \quad (23)$$

Oznacza to, że parametry μ_{j^*} oraz $\sigma_{j^*}^{-2}$ losowane są z ich rozkładów *a priori*. Po dołączeniu nowego składnika niezbędne jest zmodyfikowanie wag dla istniejących składników. Wykonywane jest one zgodnie ze wzorem $w'_j = w_j(1 - w_{j^*})$. Usunięcie składnika polega na losowym wybraniu pustego składnika oraz skorygowanie wag, które zapewni ich sumowanie do jedności.

Ogólna postać wzoru na prawdopodobieństwo akceptacji dla dołączenia składnika jest identyczna jak w przypadku opisanym w punkcie 0 (por. wzór (15)). Iloraz funkcji wiarygodności $p(\mathbf{y} | \boldsymbol{\mu}', (\boldsymbol{\sigma}^2)', \mathbf{z}') / p(\mathbf{y} | \boldsymbol{\mu}, \boldsymbol{\sigma}^2, \mathbf{z})$ jest równy 1 (dołączenie pustego składnika nie zmienia wartości funkcji wiarygodności). Iloraz rozkładów wag ma postać (wykorzystany zostanie warunek $w'_j = w_j(1 - w_{j^*})$):

$$\begin{aligned} \frac{p(\mathbf{w}' | k', \delta)}{p(\mathbf{w} | k, \delta)} &= \frac{\Gamma((k+1)\delta)}{\Gamma(\delta) \cdots \Gamma(\delta)} \times \frac{(w'_1)^{\delta-1} \cdots w_{j^*}^{\delta-1} \cdots (w'_k)^{\delta-1}}{w_1^{\delta-1} \cdots w_k^{\delta-1}} = \\ &= \frac{\Gamma(k\delta + \delta)}{\Gamma(\delta)\Gamma(k\delta)} \times (1 - w_{j^*})^{k(\delta-1)} w_{j^*}^{\delta-1} = \frac{1}{B(\delta, k\delta)} \times (1 - w_{j^*})^{k(\delta-1)} w_{j^*}^{\delta-1}. \end{aligned} \quad (24)$$

Z kolei

$$\frac{p(\mathbf{z}' | \mathbf{w}', k')}{p(\mathbf{z} | \mathbf{w}, k)} = \frac{(w'_1)^{n_1} \cdots (w_{j^*})^0 \cdots (w'_{k+1})^{n_{k+1}}}{w_1^{n_1} \cdots w_k^{n_k}} = (1 - w_{j^*})^{n_1} \cdots (1 - w_{j^*})^{n_k} = (1 - w_{j^*})^n, \quad (25)$$

zatem

$$\frac{p(\mathbf{w}'|k')}{p(\mathbf{w}|k)} \frac{p(\mathbf{z}'|\mathbf{w}',k')}{p(\mathbf{z}|\mathbf{w},k)} = \frac{1}{B(\delta, k\delta)} \times w_{j^*}^{\delta-1} (1-w_{j^*})^{n+k\delta-1}. \quad (26)$$

Ponadto

$$\frac{p(\boldsymbol{\mu}'|k')}{p(\boldsymbol{\mu}|k)} = \frac{(k+1)!}{k!} \sqrt{\frac{\kappa}{2\pi}} \exp\left(-\frac{1}{2}\kappa(\boldsymbol{\mu}_{j^*} - \boldsymbol{\xi})^2\right) = (k+1) \sqrt{\frac{\kappa}{2\pi}} \exp\left(-\frac{1}{2}\kappa(\boldsymbol{\mu}_{j^*} - \boldsymbol{\xi})^2\right)$$

oraz

$$\frac{p\left((\boldsymbol{\sigma}^2)'\right)}{p(\boldsymbol{\sigma}^2)} = \frac{\beta^\alpha}{\Gamma(\alpha)} \times (\boldsymbol{\sigma}_{j^*}^{-2})^{\alpha-1} \exp(-\beta\boldsymbol{\sigma}_{j^*}^{-2}).$$

Ponieważ wartości parametrów nowego składnika są losowane z rozkładu *a priori* (wektor $\mathbf{u} = (\boldsymbol{\mu}_{j^*}, \boldsymbol{\sigma}_{j^*}^{-2}, w_{j^*})$):

$$q(\mathbf{u}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2, j, j') = g_{1,k}(w_{j^*}) \sqrt{\frac{\kappa}{2\pi}} \exp\left(-\frac{1}{2}\kappa(\boldsymbol{\mu}_{j^*} - \boldsymbol{\xi})^2\right) \frac{\beta^\alpha}{\Gamma(\alpha)} \times (\boldsymbol{\sigma}_{j^*}^{-2})^{\alpha-1} \exp(-\beta\boldsymbol{\sigma}_{j^*}^{-2}),$$

co oznacza, że

$$\frac{p(\boldsymbol{\mu}'|k')}{p(\boldsymbol{\mu}|k)} \frac{p\left((\boldsymbol{\sigma}^2)'\right)}{p(\boldsymbol{\sigma}^2)} \frac{1}{q(\mathbf{u}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2, j, j')} = \frac{1}{g_{j,k}(w_{j^*})} (k+1).$$

Transformacje wektora $(w_1, \dots, w_k, w_{j^*}, \boldsymbol{\mu}_{j^*}, \boldsymbol{\sigma}_{j^*}^2)$ na wektor $(w'_1, \dots, w'_k, w'_{j^*}, \boldsymbol{\mu}'_{j^*}, (\boldsymbol{\sigma}_{j^*}^2)')$ wyznaczone na podstawie przekształcenia $w'_j = w_j(1-w_{j^*})$ mają następującą postać:

$$\begin{aligned} w'_1 &= w_1(1-w_{j^*}), \\ &\vdots \\ w'_{k-1} &= w_{k-1}(1-w_{j^*}), \\ w'_{j^*} &= w_{j^*}, \\ \boldsymbol{\mu}'_{j^*} &= \boldsymbol{\mu}_{j^*}, \\ (\boldsymbol{\sigma}_{j^*}^2)' &= \boldsymbol{\sigma}_{j^*}^2. \end{aligned} \quad (27)$$

Z kolei jacobian transformacji ma postać:

$$\begin{vmatrix} (1-w_{j^*}) & 0 & 0 & -w_1 & 0 & 0 \\ 0 & (1-w_{j^*}) & \cdots & 0 & -w_2 & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & (1-w_{j^*}) & -w_{k-1} & 0 & 0 \\ 0 & 0 & & 0 & 1 & 0 & 0 \\ 0 & 0 & & 0 & 0 & 1 & 0 \\ 0 & 0 & & 0 & 0 & 0 & 1 \end{vmatrix} = (1-w_{j^*})^{k-1} \quad (28)$$

Uwzględniając ponadto, że $\frac{h(k',k)}{h(k,k')} = \frac{d_{k+1}}{(k_0+1)b_k}$, otrzymujemy:

$$\begin{aligned} A &= \frac{p(k+1)}{p(k)} \frac{1}{B(k\delta, \delta)} w_{j^*}^{\delta-1} (1-w_{j^*})^{n+k\delta-k} (k+1) \times \\ &\times \frac{d_{k+1}}{(k_0+1)b_k} \frac{1}{g_{j,k}(w_{j^*})} (1-w_{j^*})^{k-1}, \end{aligned}$$

gdzie k_0 oznacza liczbę pustych składników.

Analogicznie wyprowadzić można wzór na prawdopodobieństwo akceptacji dla usunięcia składnika:

$$\begin{aligned} B &= \frac{p(k-1)}{p(k)} B((k-1)\delta, \delta) \frac{1}{w_{j^*}^{\delta-1} (1-w_{j^*})^{n+(k-1)(\delta-1)}} \frac{1}{k} \times \\ &\times \frac{b_{k-1}k_0}{d_k} \frac{1}{g_{j,k}(w_{j^*})} \frac{1}{(1-w_{j^*})^{k-1}}. \end{aligned}$$

5. Przykłady

5.1. Estymacja parametrów mieszanki dwóch rozkładów normalnych

Zaprezentowane obliczenia wykonano za pomocą programu NMIX dostępnego pod adresem <http://www.stats.bris.ac.uk/~peter/Nmix/> (por. również [Richardson, Green 1997]). Program ten napisany został w Fortranie i jest dostępny wyłącznie dla systemu operacyjnego UNIX.

W przykładzie wykorzystano dane uzyskane za pomocą generatora liczb losowych. Z rozkładu $N(5, 1)$ wylosowano 200 obserwacji oraz 100 obserwacji z roz-

kładu $N(15, 3)$. Oznacza to, że cały zbiór danych pochodzi z następującej mieszanki rozkładów:

$$Y_i \sim \frac{2}{3} N(5, 1) + \frac{1}{3} N(15, 3) \quad (29)$$

Estymację parametrów mieszanki wykonano na podstawie 100 000 iteracji opisanego algorytmu. Rozkład *a posteriori* liczby składników mieszanki przedstawiony został w tab. 1.

Tabela 1. Rozkład liczby składników mieszanki

k	1	2	3	4	5	6	7	8	9	10
$p(k)$	0	0,50296	0,3224	0,12281	0,03709	0,01049	0,00263	0,00118	0,00037	0,00007

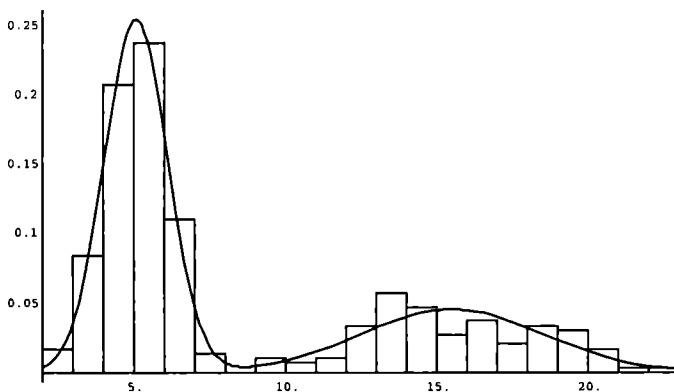
Źródło: opracowanie własne.

Najbardziej prawdopodobna liczba składników mieszanki jest zatem równa 2. Odpowiadające tej liczbie prawdopodobieństwo wynosi 0,50296 (oznacza to, że spośród 100 000 iteracji algorytmu dla 50296 z nich liczba składników mieszanki była równa 2). Stosunkowo wysokie prawdopodobieństwo uzyskano również dla liczby składników równej 3. Wyniosło ono 0,3224.

Dla liczby składników $k = 2$ uzyskano następujące oszacowania parametrów modelu mieszanki:

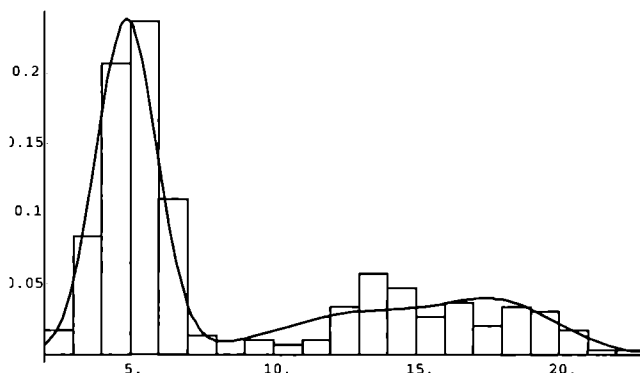
$$Y_i \sim 0,66359 N(5,06318; 1,04394) + 0,33641 N(15,49776; 2,99234) \quad (30)$$

Uzyskane oszacowania są zatem bardzo zbliżone do prawdziwych parametrów zapisanych we wzorze (29), na podstawie których generowane były dane. Podano wyłącznie estymatory punktowe, gdyż wykorzystywany program NMIX nie pozwala na wyznaczanie bayesowskich przedziałów ufności.



Rys. 1. Wykres funkcji gęstości dla mieszanki dwóch rozkładów normalnych

Źródło: opracowanie własne.



Rys. 2. Wykres funkcji gęstości dla mieszanki trzech rozkładów normalnych
Źródło: opracowanie własne.

Dla liczby składników $k = 3$ uzyskano następujące oszacowania parametrów modelu mieszanki:

$$Y_i \sim 0,62951N(4,86238; 1,05866) + 0,21931N(13,18918; 2,99234) + 0,15118N(18,11110; 1,96744) \quad (31)$$

Na rysunkach 1 i 2 przedstawione zostały histogramy danych, dla których estymowano mieszankę rozkładów oraz oszacowane na ich podstawie funkcje gęstości odpowiednio dla dwóch oraz dla trzech składników mieszanki.

5.2. Estymacja rozkładu poziomu zanieczyszczenia środowiska – przykład 1

Dane do przykładu pochodzą ze strony internetowej Wojewódzkiego Inspektoratu Ochrony Środowiska we Wrocławiu (<http://www.wroclaw.pios.gov.pl/>). Opisują one poziomy zanieczyszczenia średniodobowego dwutlenkiem azotu zmierzone we Wrocławiu przy ul. Wierzbowej w roku 2003. Dostępny zbiór danych liczył 261 obserwacji (nie we wszystkie dni dokonywano pomiarów).

Estymację parametrów mieszanki wykonano na podstawie 100 000 iteracji opisanego algorytmu. Rozkład liczby składników mieszanki przedstawiony został w tab. 2.

Najbardziej prawdopodobna liczba składników mieszanki wynosi zatem 3, a odpowiadające temu wariantowi prawdopodobieństwo jest równe 0,38113. Parametry mieszanki dla $k = 3$ są następujące:

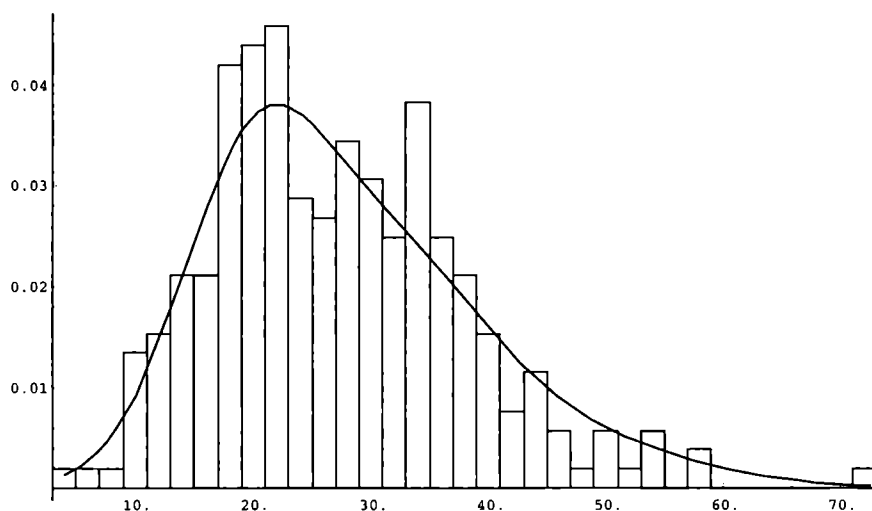
$$Y_i \sim 0,44681N(19,71121; 6,2) + 0,40855N(31,11151; 8,06241) + 0,14464N(44,72662; 10,83801) \quad (32)$$

Tabela 2. Rozkład liczby składników mieszanki

k	1	2	3	4	5	6	7	8	9	10	11
p	0,0031	0,32975	0,38113	0,17856	0,06743	0,02571	0,00923	0,00324	0,00121	0,00047	0,00017

Źródło: opracowanie własne.

Na rysunku 3 przedstawiono histogram oraz funkcję gęstości (32).



Rys. 3. Wykres funkcji gęstości dla trzech składników mieszanki

Źródło: opracowanie własne.

5.3. Estymacja rozkładu poziomu zanieczyszczenia środowiska – przykład 2

Dane wykorzystane w przykładzie pochodzą również ze strony internetowej Wojewódzkiego Inspektoratu Ochrony Środowiska we Wrocławiu. Opisują one poziomy zanieczyszczenia średniodobowego dwutlenkiem azotu, zmierzone w Szczawnie Zdroju w roku 2003. Dostępny zbiór danych liczył 357 obserwacji.

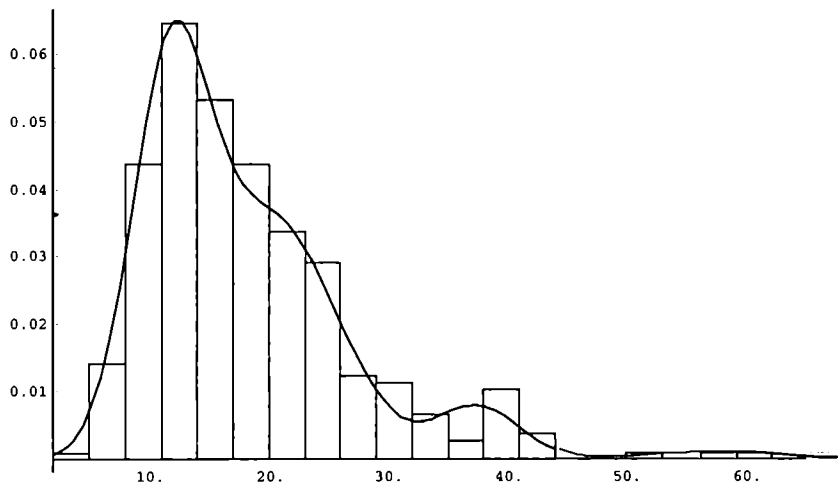
Estymację parametrów mieszanki wykonano na podstawie 100 000 iteracji opisanego algorytmu. Rozkład liczby składników mieszanki przedstawiony został w tab. 3.

Tabela 3. Rozkład liczby składników mieszanki

k	1	2	3	4	5	6	7	8	9	10	>10
p	0	0	0,01515	0,32265	0,32718	0,19203	0,08811	0,03428	0,01346	0,00481	0,00233

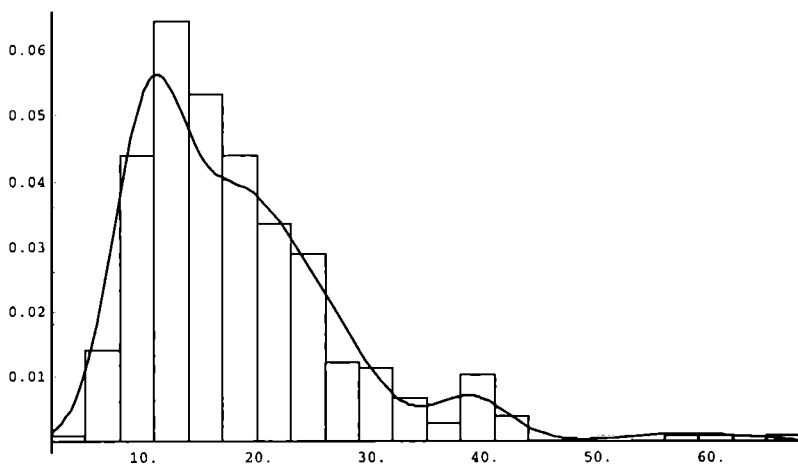
Źródło: opracowanie własne.

Najbardziej prawdopodobna liczba składników mieszanki wynosi zatem 5, niemniej jednak niemal takie samo prawdopodobieństwo występuje dla liczby składników równej 4.



Rys. 4. Wykres funkcji gęstości dla czterech składników mieszanki

Źródło: opracowanie własne.



Rys. 5. Wykres funkcji gęstości dla pięciu składników mieszanki

Źródło: opracowanie własne.

Parametry modelu dla odpowiednio $k = 4$ i $k = 5$ są następujące:

$$Y_i = 0,43604N(11,70417; 3,23724) + 0,47967N(20,26676; 5,41918) + 0,06774N(37,51475; 3,52958) + 0,01654N(57,33292; 5,79742), \quad (33)$$

$$Y_i \sim 0,37146N(10,40209; 3,14174) + 0,31938N(17,47153; 4,27294) + 0,23847N(24,69120; 4,88935) + 0,05552N(38,97253; 3,30936) + 0,01516N(58,62972; 5,50023). \quad (34)$$

Na rysunkach 4 i 5 przedstawiono funkcje gęstości mieszanek odpowiednio dla czterech i pięciu składników. Jak widać, różnica pomiędzy dopasowywanymi funkcjami gęstości jest niewielka w związku z czym również wzrokowa ocena nie pozwala w prosty sposób rozstrzygnąć, który z modeli lepiej opisuje dane empiryczne. Sytuacje, w których kilka modeli posiada podobne prawdopodobieństwa *a posteriori* oznacza, że wykorzystane dane nie dają jednoznacznych przesłanek przesądzających o trafności jednego z modeli. W takiej sytuacji jednym z rozwiązań jest zwiększenie liczebności zbioru danych wykorzystywanych do obliczenia prawdopodobieństwa modelu.

6. Podsumowanie

Algorytm RJMCMC stanowi użyteczne rozwinięcie algorytmu Metropolisa-Hastingsa, które znacznie ułatwia obliczanie prawdopodobieństw *a posteriori* rozpatrywanych modeli statystycznych. Zagadnienie wyznaczania parametrów mieszanek opisano również dla rozkładów Poissona (por. [Viallefont, Richardson, Green 2002]) oraz dla mieszanek wielowymiarowych rozkładów normalnych (por. [Zhang i in. 2004]).

Pewną wadą opisywanego algorytmu jest niewątpliwie jego złożoność. Alternatywą dla opisanej metody jest praca [Stephens 2000], w której problem ustalenia składników mieszanki rozwiązywany jest algorytmem symulacyjnym, wykorzystującym proces urodzeń i śmierci.

Inną trudnością jest kwestia efektywnego wyznaczania rozkładów prawdopodobieństwa, z których generowane są nowe wartości dla parametrów modelu. Pewne propozycje w tym zakresie można znaleźć w pracy [Brooks, Giudici, Roberts 2003].

Mimo iż od pierwszej publikacji dotyczącej algorytmu RJMCMC ([Green 1995]) upłynęło już prawie 10 lat, wiele zagadnień z nim związanych nie zostało do końca rozwiązanych, o czym świadczą wciąż pojawiające się publikacje, w których przedstawione są propozycje jego udoskonaleń.

Literatura

- Brooks S.P., Giudici P., Roberts G.O., *Efficient Construction of Reversible Jump Markov Chain Monte Carlo Proposal Distributions*, „Journal of the Royal Statistical Society: Series B (Statistical Methodology)” 2003, vol. 65, issue 1.
- Green P., *Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination*, „Biometrika” 1995, 82, 711-732.
- Highly Structured Stochastic Systems*, red. P.J. Green, N. Hjort, S. Richardson, University Press, Oxford 2003.
- Han C., Carlin B.P., *MCMC Methods for Computing Bayes Factors: a Comparative Review*, J. Amer. Statist. Assoc., 96, 1122-1132, 2001.

- Richardson S., Green P.J., *On Bayesian Analysis of Mixtures with an Unknown Number of Components*, J.R. Statist. Soc. B (1997) 59, nr 4, s. 731-792.
- Stephens M., *Bayesian Analysis of Mixtures with an Unknown Number of Components – an Alternative to Reversible Jump Methods*, „Annals of Statistics” 2000 nr 28.
- Viallefont V., Richardson S., Green P.J., *Bayesian Analysis of Poisson Mixtures*, „Journal of Non-parametric Statistics” 2002, vol. 14, Issue 1-2, s. 181-202.
- Zhang Z., Chan K.L., Wu Y., Chen Ch., *Learning a Multivariate Gaussian Mixture with the Reversible Jump MCMC Algorithm*, to appear in Statistics and Computing, 2004.

BAYESIAN ESTIMATION OF MIXTURES WITH UNKNOWN NUMBER OF COMPONENTS

Summary

The paper describes in detail the Reversible Jump Markov Chain Monte Carlo algorithm which is used in the Bayesian model choice. One of the frequently met model choice problems involves finding the number of components in a mixture of distributions. The paper describes the case in which the mixture component is univariate normal distribution. The algorithm is illustrated with some simulation and real-life examples. The real-life examples use the data on air pollution in the Lower Silesia region of Poland.