WSPÓŁCZYNNIKI ZBIEŻNOŚCI ALGORYTMÓW GIBBSA

Paweł Kopciuszewski

Instytut Matematyki i Informatyki, Politechnika Częstochowska

Streszczenie. Praca poświęcona jest badaniu łańcuchów Markowa generowanych różnymi modyfikacjami metody Gibbsa. W szczególności obliczane są współczynniki zbieżności dla trzech najbardziej znanych w literaturze modyfikacji algorytmu Gibbsa dla pewnych szczególnych przestrzeni stanów.

1. Modyfikacje metody losowania Gibbsa

Metoda losowania Gibbsa umożliwia generowanie próbki losowej z gęstości $\pi(x)$ o nośniku $S \subset R^n$ na podstawie gęstości warunkowych $\pi(x_i|x_j, j \neq i)$, $i = 1, \dots, n$, gdzie $x = (x_1, \dots, x_n) \in S$. Łańcuch Markowa wygenerowany za pomocą tej metody nazywamy łańcuchem Gibbsa. Aby wygenerować próbkę losową z gęstości π, stosujemy następujący algorytm:
Najpierw wybieramy stan początkowy $x^{(0)} = (x_1^{(0)}, \dots, x_n^{(0)}) \in S$ w łańcuchu Markowa $\{X^{(k)}\}$. W kroku algorytmu $k = 1, 2, \dots$, generujemy stan $x^{(k)} = (x_1^{(k)}, \dots, x_n^{(k)}) \in S$ w następujący sposób:

$$\begin{array}{l}
x_1^{(k)} \text{ jest realizacją z } \pi(x_1|x_2^{(k-1)}, \dots, x_n^{(k-1)}) \\
x_2^{(k)} \text{ jest realizacją z } \pi(x_2|x_1^{(k)}, x_3^{(k-1)}, \dots, x_n^{(k-1)}) \\
\vdots \\
x_{n-1}^{(k)} \text{ jest realizacją z } \pi(x_{n-1}|x_1^{(k)}, \dots, x_{n-2}^{(k)}, x_n^{(k-1)}) \\
x_n^{(k)} \text{ jest realizacją z } \pi(x_n|x_1^{(k)}, \dots, x_{n-1}^{(k)})
\end{array} \tag{1}$$

Gęstość prawdopodobieństwa przejścia $h(x, y)$ ze stanu $x = (x_1, \dots, x_n) \in S$ do stanu $y = (y_1, \dots, y_n) \in S$ można wyrazić za pomocą wzoru

$$h(x, y) = \prod_{i=1}^{n} \pi(y_i|y_1, \dots, y_{i-1}, x_{i+1}, \dots, x_n) \tag{2}$$

Każdy z powyższych $n!$ wariantów metody Gibbsa nazywany jest algorytmem deterministycznego uaktualniania łańcucha Gibbsa (determining updating Gibbs sampler), w skrócie DUGS.

Przypuśćmy, że k-ty krok w algorytmie generowania łańcucha Markowa $\{X^{(k)}\}$ przebiega według następującego schematu: Losujemy najpierw jedną z liczb ze zbioru $\{1,\ldots,n\}$ z prawdopodobieństwem p_i, $i = 1,\ldots,n$. Wylosowaną liczbę oznaczmy przez i_0. Następnie generujemy wartość $x_{i_0}^{(k)}$ z gęstości $\pi(x_{i_0}|x_j^{(k-1)}, j \neq i_0)$. W k-tym kroku otrzymamy stan łańcucha $x^{(k)} = (x_1^{(k-1)},\ldots,x_{i_0}^{(k)},\ldots,x_n^{(k-1)})$. Gęstość prawdopodobieństwa przejścia ze stanu $x^{(k-1)}$ do stanu $x^{(k)}$ jest równa

$$h(x^{(k-1)}, x^{(k)}) = p_{i_0}\pi(x_{i_0}^{(k)}|x_j^{(k-1)}, j \neq i_0) \tag{3}$$

Metodę tę nazywa się algorytmem strategii losowej dla łańcucha Gibbsa (random sweep strategy Gibbs sampler), w skrócie RSGS.

Opiszemy teraz trzecią modyfikację metody losowania Gibbsa.

Niech $p_{z_1,\ldots,z_n}$ będzie prawdopodobieństwem określonym na zbiorze $\mathcal{Z} = \{(z_1,\ldots,z_n) : z_i \in \{1,\ldots,n\}, i = 1,\ldots,n\}$ wszystkich permutacji zbioru $\{1,\ldots,n\}$. Wówczas k-ty krok w algorytmie przebiega według następującego schematu: Ze zbioru $\mathcal{Z}$ losujemy permutację $(z_1,\ldots,z_n)$ z prawdopodobieństwem $p_{z_1,\ldots,z_n}$. Wartości $x_{z_i}^{(k)}$, $i = 1,\ldots,n$, w k-tym kroku generować będziemy w porządku $x_{z_1}^{(k)},\ldots,x_{z_n}^{(k)}$. Gęstość prawdopodobieństwa przejścia ze stanu $x \in S$ do stanu $y \in S$ przy ustalonej permutacji $(z_1,\ldots,z_n)$ wynosi

$$h(x,y|(z_1,\ldots,z_n)) = \prod_{i=1}^{n} \pi(y_{z_i}|y_{z_1},\ldots,y_{z_{i-1}},x_{z_{i+1}},\ldots,x_{z_n}) \tag{4}$$

Gęstość prawdopodobieństwa przejścia ze stanu x do stanu y wyraża się wzorem

$$h(x,y) = \sum_{\{(z_1,\ldots,z_n)\in\mathcal{Z}\}} p_{z_1,\ldots,z_n} \prod_{i=1}^{n} \pi(y_{z_i}|y_{z_1},\ldots,y_{z_{i-1}},x_{z_{i+1}},\ldots,x_{z_n}) \tag{5}$$

Jeśli nośnik S gęstości π jest zbiorem skończonym, to macierz prawdopodobieństw przejścia P dla łańcucha Markowa wygenerowanego tą metodą może być zapisana w postaci

$$P = \sum_{\{(z_1,\ldots,z_n)\in\mathcal{Z}\}} p_{z_1,\ldots,z_n} R_{z_1,\ldots,z_n} \tag{6}$$

gdzie $R_{z_1,\ldots,z_n}$ jest macierzą prawdopodobieństw przejścia dla łańcucha wygenerowanego metodą DUGS, określoną przez permutację $(z_1, \ldots, z_n)$.

Metodę opisaną powyżej nazywamy algorytmem losowania permutacji dla łańcucha Gibbsa (random permutation Gibbs sampler), w skrócie RPGS.

2. Współczynniki zbieżności

Niech S będzie zbiorem N-elementowym. Niech $\{X^{(k)}\}$ będzie łańcuchem Markowa z przestrzenią stanów S oraz macierzą prawdopodobieństw przejścia H.

Zajmiemy się obecnie oceną współczynników zbieżności $\lambda(H)$ dla łańcuchów Markowa $\{X^{(k)}\}$, generowanych za pomocą modyfikacji metody losowania Gibbsa.

Twierdzenie 1 *Łańcuchy Markowa $\{X^{(k)}\}$ z niezmienniczym prawdopodobieństwem π oraz skończoną przestrzenią stanów $S \subset \mathcal{R}^2$, generowane metodami DUGS, posiadają ten sam współczynnik zbieżności.*

Dowód. Bez straty ogólności możemy założyć, że $S = \{(u_i, v_j) : \ i, j = 1, 2, \ldots, s\}$. Liczba wszystkich stanów jest wtedy równa $N = s^2$. Aby wygenerować łańcuch Markowa z niezmienniczym prawdopodobieństwem π, można zastosować dwa różne algorytmy DUGS. Dla pierwszego z nich prawdopodobieństwo przejścia $h(x, y)$ z punktu $x = (u_i, v_j)$, $i, j = 1, \ldots, s$, do punktu $y = (u_m, v_l)$, $m, l = 1, \ldots, s$, spełnia warunek

$$h(x, y) = \pi(v_l|u_m)\pi(u_m|v_j) \tag{7}$$

Dla drugiego algorytmu prawdopodobieństwo przejścia spełnia warunek

$$h(x, y) = \pi(u_m|v_l)\pi(v_l|u_i) \tag{8}$$

Niech D_1, D_2 będą macierzami prawdopodobieństw przejścia dla łańcuchów Markowa generowanych odpowiednio pierwszym oraz drugim z algorytmów. Niech T_m, $m = 1, \ldots, s$, będą macierzami o elementach $t_{i,j}^{(m)}$, spełniających warunki:

$$t_{i,j}^{(m)} = \pi(y_j|x_m), \ i, j = 1, \ldots, s$$

Niech U_m, $m = 1, \dots, s$, będą macierzami o elementach $u_{i,j}^{(m)}$, $i,j = 1, \dots, s$, spełniających warunki:

$$u_{i,j}^{(m)} = \begin{cases} \pi(x_m | y_i) & \text{dla } j = i \\ 0 & \text{dla } j \neq i \end{cases}$$

Niech macierz T będzie zdefiniowana następująco:

$$T = \begin{pmatrix} T_1 & 0 & \dots & 0 \\ 0 & T_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & T_s \end{pmatrix} \tag{9}$$

Niech macierz U będzie zdefiniowana następująco:

$$U = \begin{pmatrix} U_1 & U_2 & \dots & U_s \\ U_1 & U_2 & \dots & U_s \\ \vdots & \vdots & & \vdots \\ U_1 & U_2 & \dots & U_s \end{pmatrix} \tag{10}$$

Wtedy macierze prawdopodobieństw przejścia D_1 oraz D_2 spełniają warunek:

$$\begin{aligned} D_1 &= UT \\ D_2 &= TU \end{aligned} \tag{11}$$

Wartości własne macierzy UT oraz TU są identyczne. Zatem współczynniki zbieżności dla tych dwóch algorytmów są sobie równe

$$\lambda(D_1) = \lambda(D_2) \tag{12}$$

W trzech poniższych przykładach obliczę współczynniki zbieżności dla łańcuchów Markowa, generowanych metodami DUGS, RSGS oraz RPGS, dla szczególnego przypadku przestrzeni stanów $S = \{(u_i, v_j),\ i,j = 1,2\}$.

Przykład 1 jest ilustracją twierdzenia 1. W przykładzie 2, spośród klasy algorytmów RSGS, wyznaczę jeden algorytm RSGS, dla którego łańcuch Markowa jest najszybciej zbieżny (posiada najmniejszy współczynnik zbieżności). W przykładzie 3 pokażę, że łańcuch Markowa, generowany za pomocą algorytmu RPGS, jest najwolniej zbieżny, gdy permutacje zbioru $\{1,2\}$ generuje się z jednakowym prawdopodobieństwem. Przykład ten sugeruje wybór innego prawdopodobieństwa do generowania permutacji niż prawdopodobieństwo jednostajne.

Niech $p_1 = \pi(u_1|v_1) > 0$, $p_2 = \pi(u_1|v_2) > 0$, $q_1 = \pi(v_1|u_1) > 0$, $q_2 = \pi(v_1|u_2) > 0$ będą prawdopodobieństwami warunkowymi oraz $\delta = (p_2 - p_1)(q_2 - q_1)$.

Przykład 1

Istnieją dwa algorytmy DUGS, za pomocą których można generować łańcuchy Markowa z przestrzenią stanów S. Niech D_1 oraz D_2 oznaczają macierze prawdopodobieństw przejścia dla łańcuchów Markowa generowanych za pomocą algorytmów DUGS.

Z twierdzenia 1 wynika, że łańcuchy Markowa generowane za pomocą tych algorytmów posiadają ten sam współczynnik zbieżności. W związku z tym przeanalizujemy zbieżność tylko jednego z tych dwóch algorytmów.

Z równania (2) wynika, że prawdopodobieństwo przejścia z punktu $x = (u_i, v_j)$, $i, j = 1, 2$, do punktu $y = (u_m, v_l)$, $m, l = 1, 2$, dla pierwszego z algorytmów DUGS spełnia następujący warunek:

$$h(x, y) = \pi(v_l|u_m)\pi(u_m|v_j)$$

Wtedy macierz prawdopodobieństw przejścia D_1 może być zapisana w następującej postaci:

$$D_1 = \begin{pmatrix} p_1q_1 & p_1(1-q_1) & (1-p_1)q_2 & (1-p_1)(1-q_2) \\ p_2q_1 & p_2(1-q_1) & (1-p_2)q_2 & (1-p_2)(1-q_2) \\ p_1q_1 & p_1(1-q_1) & (1-p_1)q_2 & (1-p_1)(1-q_2) \\ p_2q_1 & p_2(1-q_1) & (1-p_2)q_2 & (1-p_2)(1-q_2) \end{pmatrix}$$

Wartości własne macierzy D_1 są równe

$$\lambda_1 = \lambda_2 = 0,\ \lambda_3 = 1,\ \lambda_4 = \delta$$

Wynika stąd, że współczynnik zbieżności $\lambda(D_1) = |\delta|$.

Zauważmy, że im mniejsze są różnice pomiędzy prawdopodobieństwami p_1, p_2 lub q_1, q_2, tym mniejszy jest współczynnik zbieżności $|\delta|$.

Przykład 2

W przykładzie tym obliczę współczynniki zbieżności dla łańcuchów Markowa, generowanych za pomocą metod RSGS.

Niech prawdopodobieństwa wylosowania liczb z dwuelementowego zbioru

$\{1,2\}$ wynoszą a_1 oraz $a_2 = 1 - a_1$. Z równania (3) wynika, że macierz prawdopodobieństw przejścia $R_{(a_1,a_2)}$ spełnia warunek

$$R_{(a_1,a_2)} =$$

$$\begin{pmatrix} a_1p_1 + a_2q_1 & a_2(1-q_1) & a_1(1-p_1) & 0 \\ a_2q_1 & a_2(1-q_1) + a_1p_2 & 0 & a_1(1-p_2) \\ a_1p_1 & 0 & a_2q_2 + a_1(1-p_1) & a_2(1-q_2) \\ 0 & a_1p_2 & a_2q_2 & a_2(1-q_2) + a_1(1-p_2) \end{pmatrix}$$

Wartości własne macierzy $R_{(a_1,a_2)}$ są równe

$$\lambda_1 = 0,\ \lambda_2 = 1,\ \lambda_3 = \tfrac{1-\sqrt{\Delta}}{2},\ \lambda_4 = \tfrac{1+\sqrt{\Delta}}{2}$$

gdzie $\Delta = 1 - 4a_1a_2(1-\delta)$.

Wynika stąd, że współczynniki zbieżności dla łańcuchów Markowa wygenerowanych za pomocą metod RSGS przyjmują wartość

$$\lambda(R_{(a_1,a_2)}) = \left|\frac{1+\sqrt{\Delta}}{2}\right| \tag{13}$$

Zauważmy, że współczynnik zbieżności $\lambda(R_{(a_1,a_2)})$ osiąga wartość najmniejszą dla $a_1 = \frac{1}{2}$ oraz $a_2 = \frac{1}{2}$. Niech R będzie macierzą prawdopodobieństw przejścia, $R = R_{(\frac{1}{2},\frac{1}{2})}$. Wartości własne macierzy R są równe

$$\lambda_1 = 0,\ \lambda_2 = 1,\ \lambda_3 = \tfrac{1}{2} + \tfrac{1}{2}\sqrt{\delta},\ \lambda_4 = \tfrac{1}{2} - \tfrac{1}{2}\sqrt{\delta}$$

Wynika stąd, że współczynnik zbieżności $\lambda(R)$ wynosi

$$\lambda(R) = \begin{cases} \frac{1}{2} + \frac{1}{2}\sqrt{\delta}, & \delta \geq 0 \\ \frac{1}{2}\sqrt{1-\delta}, & \delta < 0 \end{cases} \tag{14}$$

Przykład 3

W przykładzie tym obliczę współczynniki zbieżności dla łańcuchów Markowa generowanych za pomocą metod RPGS. Zbiór $\mathcal{Z}$ składa się w tym przypadku z dwóch permutacji $(1,2)$ oraz $(2,1)$. Niech p oznacza prawdopodobieństwo wylosowania permutacji $(1,2)$ ze zbioru $\mathcal{Z}$. Wtedy $1-p$ oznacza prawdopodobieństwo wylosowania permutacji $(2,1)$. Niech P_p, $0 < p < 1$, oznacza macierz prawdopodobieństw przejścia dla łańcucha Markowa generowanego za pomocą metody RPGS. Z równania (6) wynika, że macierz P_p spełnia następujący warunek:

$$P_p = pD_1 + (1-p)D_2$$

Wartości własne macierzy P_p są równe

$$\lambda_1 = 1,\ \lambda_2 = 0,\ \lambda_3 = \tfrac{\delta}{2} + \tfrac{1}{2}\sqrt{\delta^2 + 4\delta p(1-p)},\ \lambda_4 = \tfrac{\delta}{2} - \tfrac{1}{2}\sqrt{\delta^2 + 4\delta p(1-p)}$$

Jeśli $\delta > 0$, to współczynnik zbieżności $\lambda(P) = |\lambda_3|$ osiąga wartość najmniejszą dla $p = 0$ lub $p = 1$. Wtedy macierz P_p równa jest macierzy D_1 lub macierzy D_2, a przypadek ten rozważany już był w przykładzie 1.

Jeśli $\delta < 0$, to współczynnik zbieżności $\lambda(P) = |\lambda_3|$ osiąga wartość najmniejszą dla

$$p = \frac{1}{2} + \frac{1}{2}\sqrt{1+\delta} \ \text{ lub } \ p = \frac{1}{2} - \frac{1}{2}\sqrt{1+\delta} \tag{15}$$

Niech P oznacza macierz prawdopodobieństw przejścia dla p spełniającego warunek (6). Wówczas współczynnik zbieżności

$$\lambda(P) = \frac{|\delta|}{2} \tag{16}$$

W literaturze rozważa się algorytmy RPGS, dla których permutacje ze zbioru $\mathcal{Z}$ generuje się z jednakowym prawdopodobieństwem [6]. W naszym przykładzie odpowiada to sytuacji, gdy $p = \frac{1}{2}$. Okazuje się jednak, że współczynnik zbieżności $\lambda(P_p)$ osiąga wartość największą dla $p = \frac{1}{2}$. Zatem łańcuch Markowa, generowany za pomocą algorytmu RPGS dla $p = \frac{1}{2}$, jest najwolniej zbieżny.

Dokonamy teraz porównania współczynników zbieżności $\lambda(D_1)$ dla metody DUGS, $\lambda(R)$ dla metody RSGS oraz $\lambda(P)$ dla metody RPGS. Z obliczeń wynika, że

$$\lambda(P) < \lambda(D_1) \text{ oraz } \lambda(P) < \lambda(R) \tag{17}$$

oraz

$$\lambda(D_1) > \lambda(R) \Leftrightarrow -1 < \delta < \frac{-1-\sqrt{17}}{8} \tag{18}$$

Współczynnik zbieżności $\lambda(P)$ jest najmniejszym współczynnikiem z wszystkich analizowanych w powyższych przykładach. Zatem łańcuch Markowa, generowany za pomocą algorytmu RPGS z macierzą prawdopodobieństw przejścia P, jest najszybciej zbieżny.

Literatura

[1] Amit Y., Convergence properties of the Gibbs sampler for perturbations Gaussians, Ann. Statist. 1996, 24, 122-140.

[2] Besag J.E., Spatial interaction and the statistical analysis of lattice systems (with discussion), J. Roy. Statist. Soc. Ser. B 1974, 36, 192-236.

[3] Brémaud P., Markov Chains, Springer-Verlag, New York 1999.

[4] Kopciuszewski P., The class of joint densities uniquely determined by conditionals used in Gibbs sampling, Demonstratio Math. 1999, 32, 847-860.

[5] Kopciuszewski P., Convergence of Hastings-Metropolis algorithm, Proc. of the Third International Conference on Parallel Processing and Applied Mathematics, Kazimierz Dolny 1999, 543-549.

[6] Roberts G.O., Sahu S.K., Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler, J. Roy. Statist. Soc. *Ser.* B 1997, 59, 291-317.

[7] Tierney L., Markov chains for exploring posterior distributions, Ann. Statist. 1994, 22, 1701-1762.

[8] Zellner A., Min C.K., Gibbs sampler convergence criteria, J. Amer. Statist. Assoc. 1995, 90, 921-927.