

Piotr Dzierża

Algorytm Expectation - Maximization w uproszczonym problemie znajdowania motywu w ciągach DNA

1. Wstęp

Założmy, że mamy ciąg $\underline{x}_1 = (x_{11}, x_{12}, \dots, x_{1w})$ długości $w \in \mathbb{N}$, gdzie każdy element $x_{1i} \in \Sigma = \{A, C, G, T\}$. Dla wygody często utożsamiamy $\{A, C, G, T\} \equiv \{1, 2, 3, 4\}$. Oznaczmy $d = |\Sigma|$, czyli w naszym przypadku $d = 4$. Mamy dane rozkłady $\underline{\theta}_j = (\theta_{1,j}, \dots, \theta_{d,j})^T$, $j = 1 \dots w$. Ciąg \underline{x}_1 jest realizacją zmiennej losowej $\underline{X} = (X_1, \dots, X_w)$, gdzie

$$P(X_j = a) = \theta_{a,j}, \quad a = 1, \dots, d, \quad (\equiv a \in \{A, C, G, T\}).$$

Innymi słowy, $\theta_{a,j}$ jest prawdopodobieństwem, że na j -tej pozycji występuje literka a . W naukach biologicznych macierz

$$\underline{\theta} = (\underline{\theta}_1, \dots, \underline{\theta}_w)$$

nazywana jest macierzą "position weight matrix of a motif".

Założmy teraz, że mamy k obserwacji, mianowicie $\underline{x}_1, \dots, \underline{x}_k$:

$$\underline{x}_1 = (x_{11}, x_{12}, \dots, x_{1w}),$$

$$\underline{x}_2 = (x_{21}, x_{22}, \dots, x_{2w}),$$

...

$$\underline{x}_k = (x_{k1}, x_{k2}, \dots, x_{kw}).$$

Dodatkowo ustalmy liczbę $\alpha \in (0, 1)$, która określa sposób otrzymywania ciągów \underline{x}_i , $i = 1, \dots, k$:

- Dla każdego i losujemy Z_i z rozkładu zero-jedynkowego z parametrem α .
- Jeśli wylosujemy wartość 1, to \underline{x}_i generujemy z użyciem macierzy $\underline{\theta}$.
- Jeśli wylosujemy wartość 0, to zamiast $\underline{\theta}$ korzystamy z rozkładu tła $\underline{\theta}^b$, który definiujemy przez

$$\underline{\theta}^b = (\theta_1^b, \theta_2^b, \theta_3^b, \theta_4^b),$$

gdzie θ_m^b , $m = 1, 2, 3, 4$, opisuje prawdopodobieństwo wylosowania wartości 1, 2, 3, 4, odpowiednio, niezależnie od pozycji.

Rozpatrywać będziemy problem, w którym na podstawie tylko macierzy danych $\underline{x} = (\underline{x}_1^T, \dots, \underline{x}_k^T)$ oraz opcjonalnie przy znanej wartości α , będziemy chcieli wyestymować wektor tła $\underline{\theta}^b$ oraz macierz $\underline{\theta}$. W przypadku braku informacji o α , będziemy też estymować wartość tego parametru. W tym celu wykorzystamy algorytm EM.

2. Algorytm EM

Niech $\Theta = (\underline{\theta}, \underline{\theta}^b, \alpha)$. Korzystając z postaci funkcji prawdopodobieństwa rozkładu zero-jedynkowego otrzymujemy, że interesującą nas funkcją wiarygodności będzie rozkład brzegowy postaci

$$L(\Theta) = \mathbf{P}(x_1, \dots, x_k; (\underline{\theta}, \underline{\theta}^b, \alpha)) = \prod_{i=1}^k (\alpha \mathbf{P}(x_i, \underline{\theta}) + (1 - \alpha) \mathbf{P}(x_i, \underline{\theta}^b)).$$

Naszym celem jest otrzymanie

$$\arg \max_{(\underline{\theta}, \underline{\theta}^b, \alpha)} L[(\underline{\theta}, \underline{\theta}^b, \alpha)] = \arg \max_{(\underline{\theta}, \underline{\theta}^b, \alpha)} \log(L[(\underline{\theta}, \underline{\theta}^b, \alpha)]).$$

Algorytm zaczynamy od ustalenia momentu początkowego $t = 0$. Generujemy początkową losową macierz $\underline{\theta}^{(t)}$ rozmiaru $4 \times w$ oraz wektor $\underline{\theta}^{(t),b}$ długości 4 tak, by każda kolumna oraz wektor $\underline{\theta}^{(t),b}$ były rozkładami prawdopodobieństwa (sumowanie się elementów do 1 oraz wartości z przedziału $[0,1]$). W naszych oznaczeniach indeks górny (t) oznacza, że dany parametr został wyznaczony w czasie t . W przypadku braku informacji o α , losujemy również $\alpha^{(t)}$ z rozkładu jednostajnego na przedziale $[0,1]$. Następnie dla każdego $i = 1, \dots, k$, liczymy prawdopodobieństwa określone wzorami

$$Q_i^{(t)}(0) = \mathbf{P}(Z_i = 0 | x_i; (\underline{\theta}^{(t)}, \underline{\theta}^{(t),b}, \alpha^{(t)})) = \frac{(1 - \alpha^{(t)}) \prod_{j=1}^w \theta_{x_{i,j}}^{(t),b}}{(1 - \alpha^{(t)}) \prod_{j=1}^w \theta_{x_{i,j}}^{(t),b} + \alpha^{(t)} \prod_{j=1}^w \theta_{x_{i,j},j}^{(t)}},$$

$$Q_i^{(t)}(1) = \mathbf{P}(Z_i = 1 | x_i; (\underline{\theta}^{(t)}, \underline{\theta}^{(t),b})) = 1 - Q_i^{(t)}(0).$$

W ten sposób wykonaliśmy tzw. krok *Expectation* w momencie czasowym t .

Krok *Maximization* zaczynamy od zdefiniowania

$$\mathfrak{Q}(\Theta, \Theta^{(t)}) = \sum_{i=1}^k \sum_{h=0}^1 Q_i^{(t)}(h) \log(\mathbf{P}(x_i, Z_i = h; \Theta)),$$

gdzie $\Theta^{(t)} = (\underline{\theta}^{(t)}, \underline{\theta}^{(t),b}, \alpha^{(t)})$.

W tym kroku naszym celem jest uzyskanie

$$\Theta^{(t+1)} = \arg \max_{\Theta} \mathfrak{Q}(\Theta, \Theta^{(t)}).$$

Korzystając z metody mnożników Lagrange'a, jesteśmy w stanie podać wzory estymatorów $\underline{\theta}^{(t+1)}$ i $\underline{\theta}^{(t+1),b}$. Zakładając, że elementy są z przedziału $[0,1]$ i odpowiednie sumy są równe 1 (z warunków bycia rozkładem prawdopodobieństwa), dla $i = 1, \dots, k$ oraz dla $j = 1, \dots, w$, otrzymujemy:

$$\theta_1^{(t+1),b} = \frac{\sum_{i=1}^k Q_i^{(t)}(0) \cdot \#\{j : x_{i,j} = 1\}}{\sum_{i=1}^k Q_i^{(t)}(0)w},$$

$$\theta_2^{(t+1),b} = \frac{\sum_{i=1}^k Q_i^{(t)}(0) \cdot \#\{j : x_{i,j} = 2\}}{\sum_{i=1}^k Q_i^{(t)}(0)w},$$

$$\theta_3^{(t+1),b} = \frac{\sum_{i=1}^k Q_i^{(t)}(0) \cdot \#\{j : x_{i,j} = 3\}}{\sum_{i=1}^k Q_i^{(t)}(0)w},$$

$$\theta_4^{(t+1),b} = \frac{\sum_{i=1}^k Q_i^{(t)}(0) \cdot \#\{j : x_{i,j} = 4\}}{\sum_{i=1}^k Q_i^{(t)}(0)w},$$

gdzie symbol # oznacza moc zbioru. Ponadto

$$\theta_{1,j}^{(t+1)} = \frac{\sum_{i: x_{i,j}=1} Q_i^{(t)}(1)}{\sum_{i=1}^k Q_i^{(t)}(1)},$$

$$\theta_{2,j}^{(t+1)} = \frac{\sum_{i: x_{i,j}=2} Q_i^{(t)}(1)}{\sum_{i=1}^k Q_i^{(t)}(1)},$$

$$\theta_{3,j}^{(t+1)} = \frac{\sum_{i: x_{i,j}=3} Q_i^{(t)}(1)}{\sum_{i=1}^k Q_i^{(t)}(1)},$$

$$\theta_{4,j}^{(t+1)} = \frac{\sum_{i: x_{i,j}=4} Q_i^{(t)}(1)}{\sum_{i=1}^k Q_i^{(t)}(1)}.$$

Jeśli nie znamy wartości α , możemy ją estymować za pomocą

$$\alpha^{(t+1)} = \frac{\sum_{i=1}^k Q_i^{(t)}(1)}{\sum_{i=1}^k Q_i^{(t)}(0) + Q_i^{(t)}(1)}.$$

W przypadku znanej wartości α , kontynuując przyjętą tutaj konwencję, przyjmujemy, że niezależnie od czasu t , $\alpha = \alpha^{(0)} = \alpha^{(t)}$.

W ten sposób otrzymamy kolejne przybliżenia interesujących nas wielkości. Należy iteracyjnie obliczać kolejne wartości $Q_i^{(t)}(0)$, $Q_i^{(t)}(1)$, $\theta^{(t),b}$, $\theta^{(t)}$ oraz ewentualnie $\alpha^{(t)}$ (bez ponownego ich losowania). Algorytm należy zakończyć, gdy spełniony zostanie określony warunek zbieżności lub po pewnej liczbie iteracji.

3. Kryteria stopu

Rozpatrywać będziemy 3 możliwe kryteria stopu algorytmu EM. Celem każdego z nich jest zatrzymanie działania algorytmu w momencie, gdy kolejne iteracje nie wnoszą już znaczącej

poprawy wyników estymacji. Pierwsze kryterium to wartość RMSE (*Root Mean Squared Error*) między wynikami dwóch kolejnych iteracji. Algorytm przerywa się, gdy jednocześnie zachodzą następujące warunki:

- RMSE między macierzami $\theta^{(t)}$ i $\theta^{(t+1)}$ jest mniejsze od 0.0001. Innymi słowy, pierwiastek ze średniej wartości kwadratów różnic wartości o tych samych współrzędnych macierzy $\theta^{(t)}$ i $\theta^{(t+1)}$ musi być mniejsza niż 0.0001.
- RMSE między wektorami $\theta^{(t),b}$ i $\theta^{(t+1),b}$ jest mniejsze od 0.0001.
- Wartość bezwzględna różnicy $\alpha^{(t)}$ i $\alpha^{(t+1)}$ jest mniejsza od 0.0001.

Drugie kryterium to maksimum z modułów różnic współrzędnych estymowanych wielkości. Algorytm przerywa się, gdy jednocześnie zachodzą następujące warunki:

- Maksimum z modułów wszystkich różnic wartości o tych samych współrzędnych w macierzach $\theta^{(t)}$ i $\theta^{(t+1)}$ jest mniejsze od 0.0001. Innymi słowy, obliczymy macierz, która jest różnicą $\theta^{(t)}$ i $\theta^{(t+1)}$. Na każdą współrzędną tak obliczonej macierzy nakładamy moduł i sprawdzamy, czy największy z nich jest mniejszy od 0.0001.
- Maksimum z modułów wszystkich różnic wartości o tych samych współrzędnych w wektorach $\theta^{(t),b}$ i $\theta^{(t+1),b}$ jest mniejsze od 0.0001.
- Wartość bezwzględna różnicy $\alpha^{(t)}$ i $\alpha^{(t+1)}$ jest mniejsza od 0.0001.

Trzecim rozpatrywanym kryterium stopu jest wartość *loglikelihood*. Dla każdej iteracji oblicza się wartość

$$l(\Theta^{(t)}) = \sum_{i=1}^k \left(Q_i^{(t)}(0) \sum_{j=1}^w \log(\theta_{x_{i,j}}^{(t),b} (1 - \alpha^{(t)})) + Q_i^{(t)}(1) \sum_{j=1}^w \log(\theta_{x_{i,j},j}^{(t)} \alpha) \right).$$

Algorytm przerywa się, gdy zachodzi $l(\Theta^{(t+1)}) - l(\Theta^{(t)}) < 0.0001$.

Oprócz podanych kryteriów stopu, wykonywanie algorytmu przerywano po wykonaniu 1000 iteracji.

Wszystkie zaprezentowane tutaj kryteria stopu wraz z progami zatrzymania algorytmu nie są jedynymi możliwymi rozwiązaniami tego zagadnienia. Podane tutaj wartości progów zostały wybrane w wyniku symulacji, które opisane będą w następnych częściach tego raportu.

4. Rozważane wartości parametrów

W symulacjach zostały wykorzystane następujące wartości parametrów.

- Długość motywu to $w = 5$.
- Liczba obserwacji to $k = 20$ albo $k = 1000$.
- Wartości $\theta^{(0),b}$ zawsze losowane były tak, że liczby na pierwszych trzech współrzędnych były liczbami pseudolosowymi wygenerowanymi z rozkładu jednostajnego na przedziale (0.1,0.3), a czwarta współrzędna była różnicą 1 i sumy tych wylosowanych liczb.
- Macierz $\theta^{(0)}$ zawsze otrzymywano przez niezależne losowanie wartości każdej współrzędnej z rozkładu jednostajnego na przedziale (0.1,0.3), a następnie normowanie kolumn.

5. Wyniki symulacji

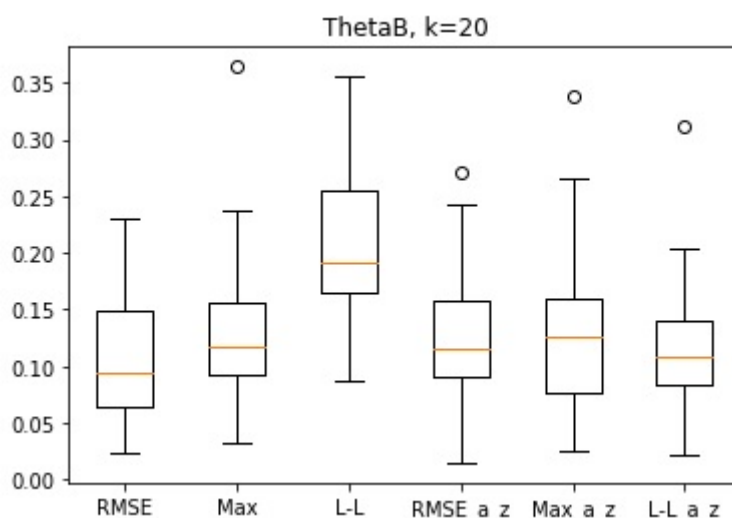
W celu znalezienia warunków dających najlepsze rezultaty postanowiono wykonać symulacje porównujące wszystkie rozważane tu metody stopu. Osobno rozważano przypadki, gdzie wartość parametru α nie była estymowana.

Każde badanie składało się z 50 powtórzeń i wyglądało następująco. Najpierw generowano obserwacje przy ustalonych losowo wybranych wartościach Θ . Następnie wykonywano algorytm EM zgodnie z opisem w rozdziale 2. Na koniec każdego powtórzenia liczono błąd RMSE wyrażony wzorem

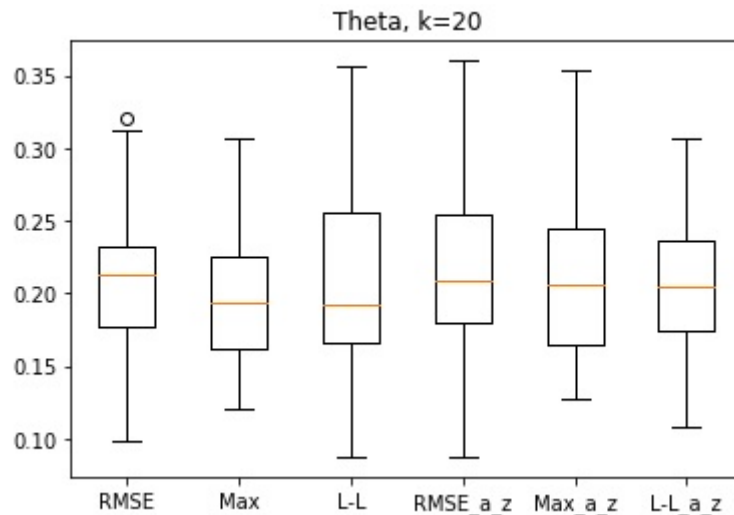
$$RMSE = \sqrt{\frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m (\underline{A}'_{i,j} - \underline{A}_{i,j})^2},$$

gdzie $\underline{A}_{i,j}$ to prawdziwa wartość, a $\underline{A}'_{i,j}$ to wartość wyestymowana. Za \underline{A} bierzemy zatem $\underline{\theta}$, $\underline{\theta}^b$ lub α , a za \underline{A}' ich ostateczne przybliżenie. Należy zauważyć, że dla α takie podejście sprowadza się do zwykłego modułu z różnicy wartości prawdziwej i wyestymowanej.

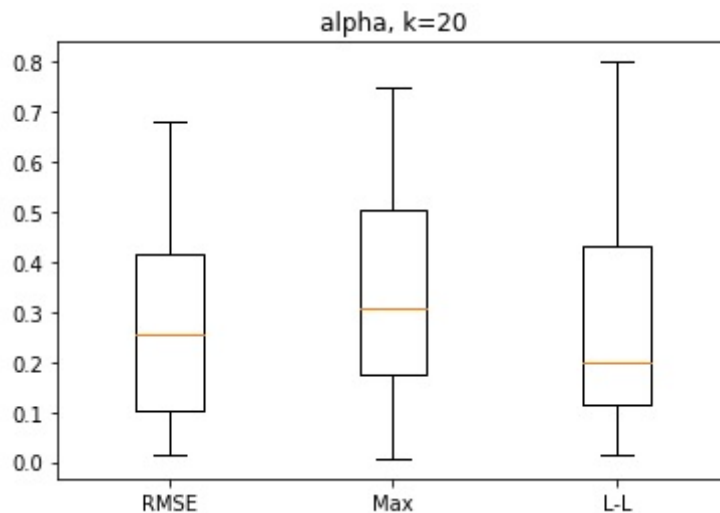
Przedstawione zostaną teraz wykresy pudełkowe wartości RMSE, gdzie za warunek stopu brano określoną w rozdziale 3 wartość: RMSE dwóch kolejnych estymowanych parametrów (skrót: RMSE), maksimum modułu różnic współrzędnych dwóch kolejnych estymowanych parametrów (skrót: Max) oraz zdefiniowaną wcześniej wartość *loglikelihood* dwóch kolejnych iteracji (skrót: L-L). Na początek przyjrzymy się sytuacji, gdy $k = 20$. Końcówka „_a_z” przy nazwie oznacza, że wykres dotyczy sytuacji, gdzie α było znane i go nie estymowano.



Rysunek 1: Boxploty wartości RMSE dla estymatorów $\underline{\theta}^b$ dla różnych kryteriów stopu przy $k = 20$.



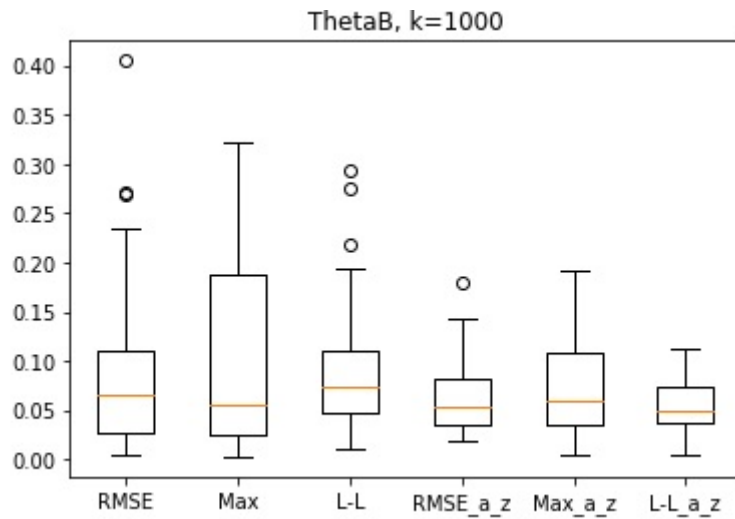
Rysunek 2: Boxploty wartości RMSE dla estymatorów θ dla różnych kryteriów stopu przy $k = 20$.



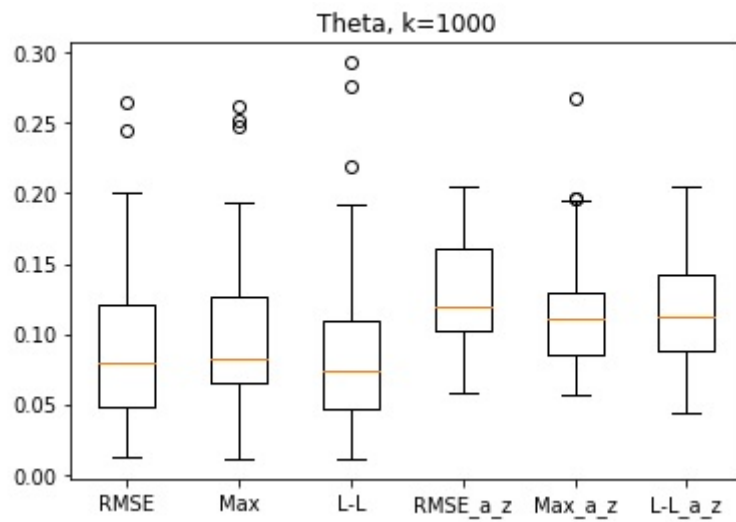
Rysunek 3: Boxploty wartości RMSE dla estymatorów α dla różnych kryteriów stopu przy $k = 20$.

Z wykresów 1, 2 i 3 możemy odczytać, że dla tak małego k , zastosowana metoda działa bardzo słabo bez względu na dobór rozważanych kryteriów stopu. Należy zaznaczyć, że zwiększanie lub zmniejszanie tolerancji błędu nie zmieniało znacznie postaci i ułożenia wykresów, więc nie ma sensu rozpatrywać innych wartości niż 0.0001. Trzeba zauważyć, że estymacja wartości α daje wyniki wręcz absurdalne, gdyż błąd estymacji potrafi wynieść wartości bliskie 0.8. Kolejną obserwacją, jaką można poczynić, jest niewielkie znaczenie samego parametru α przy estymowaniu innych parametrów. Jeśli chodzi o wybór najlepszego kryterium stopu, to wydaje się, że najrozsądniej będzie korzystać z kryterium korzystającego z RMSE.

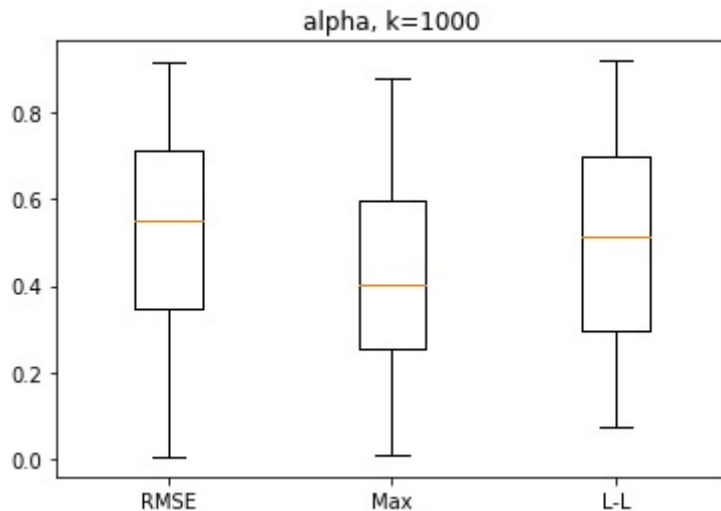
Bardzo słabe wyniki przy niskich wartościach k inspirują, by sprawdzić, co się stanie, gdy rozpatrywać będziemy duże k , na przykład równe 1000. Symulacje wykonano podobnie jak wcześniej.



Rysunek 4: Boxploty wartości RMSE dla estymatorów $\underline{\theta}^b$ dla różnych kryteriów stopu przy $k = 1000$.



Rysunek 5: Boxploty wartości RMSE dla estymatorów $\underline{\theta}$ dla różnych kryteriów stopu przy $k = 1000$.



Rysunek 6: Boxploty wartości RMSE dla estymatorów α dla różnych kryteriów stopu przy $k = 1000$.

Na wykresach 4, 5 i 6 widzimy zmiany w porównaniu z poprzednim zestawem wykresów. Estymowanie rozkładów prawdopodobieństw wykazało się raczej niższymi wartościami RMSE niż dla $k = 20$. Niezmienna pozostaje jednak kompletnie błędna estymacja α . Zaskakuje jednak fakt, że znajomość wartości α okazała się być czynnikiem, który pogorszył estymację $\underline{\theta}$. Znowu ciężko jest wybrać najlepsze kryterium, wszystkie trzy dalej działają podobnie i niestety, dalej dość słabo. Wydaje się, że ponownie najlepiej wypadło kryterium oparte na RMSE.

6. Podsumowanie

Podsumowując, zaimplementowane zostało kryterium stopu oparte na liczeniu RMSE. Choć nieznacznie, wypadło ono najlepiej podczas symulacji. Wydaje się, że przedstawiony tutaj i zaimplementowany algorytm EM nie poradził sobie dobrze w tak postawionym problemie i warunkach. Szczególnie martwi tutaj estymacja α , która kategorycznie nie powinna być używana. Bardzo zastanawiające jest również to, że mimo tak słabej estymacji α , przedstawione na rysunku 5 wykresy błędów są niższe, gdy estymacja $\underline{\theta}$ opiera się na źle wyestymowanych wartościach α .

7. Załączone programy

Zaimplementowane zostały dwie funkcje. Pierwszą z nich, generowanie przykładowych danych według zadanych parametrów, wykonuje program `motif_generate.py`. Drugą funkcją jest estymacja parametrów na podstawie zadanych obserwacji i opcjonalnie wartości α , którą wykonuje program `motif_estimate.py`.

7.1. Program generujący dane

Program `motif_generate.py` ma dwa argumenty:

- `--params`, czyli plik wejściowy formatu `.json` z parametrami do generowania,
- `--output`, czyli nazwa pliku wyjściowego, również formatu `.json`.

Program generuje losową próbkę według zawartych w pliku wejściowym parametrów. Tymi parametrami są:

- długość motywu w ,
- wartość liczby α ,
- liczbę wierszy do symulowania k ,
- macierz rozkładów prawdopodobieństw $\underline{\theta}$,
- rozkład tła $\underline{\theta}^b$.

Rezultatem programu jest plik formatu *.json* o nazwie zadanej w argumencie `--output`, który zawiera wartość wykorzystanego parametru α oraz wygenerowaną próbkę. Próbka jest generowana według schematu przedstawionego w rozdziale 1. Losowana jest liczba z rozkładu zero-jedynkowego, a następnie:

- jeśli wylosowano 1, to wartości na kolejnych współrzędnych są losowane zgodnie z macierzą prawdopodobieństw $\underline{\theta}$.
- jeśli wylosowano 0, to wartości na kolejnych współrzędnych są losowane zgodnie z wektorem prawdopodobieństw $\underline{\theta}^b$.

Całość powtarza się do momentu uzyskania zadanej liczby obserwacji.

Załączone zostały dodatkowo 3 pomocnicze pliki:

- Program `json_save_params.py` służy do łatwego zapisania parametrów do pliku o rozszerzeniu *.json*. Nie przyjmuje on argumentów, zmiana parametrów wymaga edycji części kodu.
- Plik `params_set1.json` to przykładowy plik, który może być wykorzystany jako argument `--params` programu generującego dane `motif_generate.py`. Należy zadbać, by każdy plik wykorzystany jako argument `--params` miał strukturę taką samą jak `params_set1.json`, to znaczy musi zawierać w sobie te same nazwy parametrów.
- Plik `generated_data.json` to przykładowy rezultat programu `motif_generate.py`.

7.2. Program estymujący wartości parametrów

Program `motif_estimate.py` przyjmuje następujące parametry.

- `--input`, czyli plik rozszerzenia *.json*, który zawiera dane oraz wartość parametru α .
- `--output`, czyli plik formatu *.json*, który zostanie utworzony i zawierać będzie wyestymowane parametry: $\underline{\theta}$, $\underline{\theta}^b$ oraz opcjonalnie wartość wyestymowanego α .
- `--estimate-alpha`, który przyjmuje dwie możliwe wartości. Gdy wpisane zostanie *yes*, to program zignoruje informację o wartości α i będzie ją estymować. Gdy zostanie wpisane *no*, to program nie będzie obliczał α , tylko wykorzysta wartość podaną w pliku z argumentu `--input`.

Program działa wykorzystując opisany w rozdziale 2 algorytm EM, wykorzystuje sposób losowania początkowych parametrów opisany w rozdziale 4 oraz przyjmuje kryterium stopu oparte o wielkość zmiany RMSE.

Do poprawnego działania tego programu, plik użyty jako argument `--input` musi mieć strukturę taką, jak załączony plik `generated_data.json`.

8. Dodatek - wyprowadzenie wzorów

W tej części wyprowadzone zostaną niektóre wzory użyte w rozdziale 2. Zaczniemy od przypomnienia, że zmienna losowa Z_i ma rozkład zero-jedynkowy z parametrem α , co oznacza, że

$$\mathbf{P}(Z_i = 1) = \alpha \quad \text{oraz} \quad \mathbf{P}(Z_i = 0) = 1 - \alpha.$$

Jeśli $Z_i = 1$, to wybór wartości współrzędnych obserwacji \underline{x}_i zależy wyłącznie od macierzy prawdopodobieństw $\underline{\theta}$. Jeśli zaś $Z_i = 0$, to wybór wartości współrzędnych obserwacji \underline{x}_i zależy wyłącznie od wektora prawdopodobieństw $\underline{\theta}^b$. Wzór

$$Q_i^{(t)}(0) = \mathbf{P}(Z_i = 0 | x_i; (\underline{\theta}^{(t)}, \underline{\theta}^{(t),b}, \alpha^{(t)})) = \frac{(1 - \alpha^{(t)}) \prod_{j=1}^w \theta_{x_{i,j}}^{(t),b}}{(1 - \alpha^{(t)}) \prod_{j=1}^w \theta_{x_{i,j}}^{(t),b} + \alpha^{(t)} \prod_{j=1}^w \theta_{x_{i,j},j}^{(t)}}$$

jest konsekwencją użycia własności prawdopodobieństwa warunkowego wraz z (cicho założoną) niezależnością między kolejnymi elementami. Wzór

$$Q_i^{(t)}(1) = \mathbf{P}(Z_i = 1 | x_i; (\underline{\theta}^{(t)}, \underline{\theta}^{(t),b})) = 1 - Q_i^{(t)}(0).$$

wynika z faktu, że $Q_i^{(t)}(0)$ i $Q_i^{(t)}(1)$ opisują prawdopodobieństwa zdarzeń dopełniających się.

Przedstawione teraz zostaną wyprowadzenia wzorów na estymatory $\underline{\theta}^{(t),b}$, $\underline{\theta}^{(t)}$, $\alpha^{(t)}$. Zaczniemy od rozpisania Ω . Zauważmy najpierw, że

$$\mathbf{P}(x_i, Z_i = h; \Theta) = \begin{cases} (1 - \alpha) \prod_{j=1}^w \theta_{x_{i,j}}^b, & \text{gdy } h = 0, \\ \alpha \prod_{j=1}^w \theta_{x_{i,j},j}, & \text{gdy } h = 1. \end{cases}$$

Zatem

$$\begin{aligned} \Omega(\Theta, \Theta^{(t)}) &= \sum_{i=1}^k \sum_{h=0}^1 Q_i^{(t)}(h) \log(\mathbf{P}(x_i, Z_i = h; \Theta)) = \\ &= \sum_{i=1}^k \left(Q_i^{(t)}(0) \log((1 - \alpha) \prod_{j=1}^w \theta_{x_{i,j}}^b) + Q_i^{(t)}(1) \log(\alpha \prod_{j=1}^w \theta_{x_{i,j},j}^b) \right) = \\ &= \sum_{i=1}^k \left(Q_i^{(t)}(0) [\log(1 - \alpha) + \sum_{j=1}^w \log(\theta_{x_{i,j}}^b)] + Q_i^{(t)}(1) [\log(\alpha) + \sum_{j=1}^w \log(\theta_{x_{i,j},j}^b)] \right). \end{aligned}$$

Naszym celem jest uzyskanie

$$\Theta^{(t+1)} = \arg \max_{\Theta} \Omega(\Theta, \Theta^{(t)}).$$

Jego realizację zaczniemy od obliczenia pochodnych cząstkowych po elementach $\underline{\theta}^b$. Dla $m = 1, \dots, 4$ mamy

$$\frac{\partial \Omega}{\partial \theta_m^b} = \sum_{i=1}^k Q_i^{(t)}(0) \sum_{j=1}^w \frac{1}{\theta_{x_{i,j}}^b} \mathbf{1}(x_{i,j} = m) = \frac{1}{\theta_m^b} \sum_{i=1}^k Q_i^{(t)}(0) \cdot \#\{j : x_{i,j} = m\}.$$

Należy pamiętać, że θ_m^b jest prawdopodobieństwem, więc $\sum_{m=1}^4 \theta_m^b = 1$. Warunek ten zapiszemy w formie funkcji

$$g(\theta_1^b, \theta_2^b, \theta_3^b, \theta_4^b) = \theta_1^b + \theta_2^b + \theta_3^b + \theta_4^b - 1 = 0.$$

Policzymy teraz pochodne cząstkowe pierwszego rzędu funkcji g . Dla $m = 1, \dots, 4$,

$$\frac{\partial g}{\partial \theta_m^b} = 1.$$

Korzystając z metody mnożników Lagrange'a, dla pewnego λ otrzymujemy układ równań

$$\begin{cases} \frac{1}{\theta_1^b} \sum_{i=1}^k Q_i^{(t)}(0) \cdot \#\{j : x_{i,j} = 1\} = \lambda \\ \frac{1}{\theta_2^b} \sum_{i=1}^k Q_i^{(t)}(0) \cdot \#\{j : x_{i,j} = 2\} = \lambda \\ \frac{1}{\theta_3^b} \sum_{i=1}^k Q_i^{(t)}(0) \cdot \#\{j : x_{i,j} = 3\} = \lambda \\ \frac{1}{\theta_4^b} \sum_{i=1}^k Q_i^{(t)}(0) \cdot \#\{j : x_{i,j} = 4\} = \lambda \\ \theta_1^b + \theta_2^b + \theta_3^b + \theta_4^b - 1 = 0. \end{cases}$$

Wyliczając $\theta_1^b, \theta_2^b, \theta_3^b, \theta_4^b$ z pierwszych 4 równań i wstawiając do ostatniego równania, mamy

$$1 = \frac{1}{\lambda} \sum_{m=1}^4 \sum_{i=1}^k Q_i^{(t)}(0) \cdot \#\{j : x_{i,j} = m\} = \frac{w}{\lambda} \sum_{i=1}^k Q_i^{(t)}(0).$$

Zatem

$$\lambda = w \sum_{i=1}^k Q_i^{(t)}(0).$$

Po wstawieniu λ do pierwszych 4 równań otrzymujemy wzory z rozdziału 2.

W celu wyznaczenia wzorów estymatorów dla $\underline{\theta}^{(t)}$, ponownie liczymy odpowiednie pochodne cząstkowe pierwszego rzędu z $\Omega(\Theta, \Theta^{(t)})$. Ponownie, dla $m = 1, \dots, 4$ oraz $j = 1, \dots, w$, mamy

$$\frac{\partial \Omega}{\partial \theta_{m,j}} = \sum_{i=1}^k Q_i^{(t)}(1) \sum_{j=1}^w \frac{1}{\theta_{x_{i,j},j}} \mathbf{1}(x_{i,j} = m) = \sum_{i: x_{i,j}=m} Q_i^{(t)}(1) \frac{1}{\theta_{m,j}}.$$

Wyznaczamy również funkcje opisujące warunki na prawdopodobieństwo. Dla $j = 1, \dots, w$ mamy

$$g_j(\theta_{1,j}, \theta_{2,j}, \theta_{3,j}, \theta_{4,j}) = \theta_{1,j} + \theta_{2,j} + \theta_{3,j} + \theta_{4,j} - 1 = 0.$$

Korzystając z metody mnożników Lagrange'a, dla $j = 1, \dots, w$ oraz pewnego λ otrzymujemy układy równań

$$\begin{cases} \sum_{i: x_{i,j}=1} Q_i^{(t)}(1) \frac{1}{\theta_{1,j}} = \lambda \\ \sum_{i: x_{i,j}=2} Q_i^{(t)}(1) \frac{1}{\theta_{2,j}} = \lambda \\ \sum_{i: x_{i,j}=3} Q_i^{(t)}(1) \frac{1}{\theta_{3,j}} = \lambda \\ \sum_{i: x_{i,j}=4} Q_i^{(t)}(1) \frac{1}{\theta_{4,j}} = \lambda \\ \theta_{1,j} + \theta_{2,j} + \theta_{3,j} + \theta_{4,j} - 1 = 0. \end{cases}$$

Wyznaczenie $\theta_{1,j}, \theta_{2,j}, \theta_{3,j}, \theta_{4,j}$ z pierwszych 4 równań i wstawienie wyników do ostatniego równania pozwala obliczyć λ .

$$\lambda = \sum_{m=1}^4 \sum_{i: x_{i,j}=m} Q_i^{(t)}(1) = \sum_{i=1}^k Q_i^{(t)}(1)$$

Wstawienia obliczonego λ do wyznaczonych równań na $\theta_{1,j}, \theta_{2,j}, \theta_{3,j}, \theta_{4,j}$ pozwala otrzymać wzory z rozdziału 2.

Pozostało jeszcze wyznaczyć wzór na $\alpha^{(t)}$. Liczymy zatem pochodną cząstkową i przyrównujemy ją do 0

$$\frac{\partial \Omega}{\partial \alpha} = \sum_{i=1}^k \left(Q_i^{(t)}(0) \frac{-1}{1-\alpha} + Q_i^{(t)}(1) \frac{1}{\alpha} \right) = 0.$$

Po prostych przekształceniach polegających na wyznaczeniu α , dostajemy wzór z rozdziału 2.