28 października 2024

1 O metodzie

- metoda została opublikowana w artykule **Gajda Janusz, Sikora Grzegorz, Wyłomańska Agnieszka: Regime variance testing - a quantile approach, Acta Phys. Polon B 44(5), 1015-1035, 2013.**,
- metoda wykorzystuje statystykę, jaką jest skumulowana suma kwadratów,
- jest metodą nieparametryczną, wykorzystującą jedynie empiryczne momenty badanego szeregu danych.

1.1 Założenia i statystyka bazowa dla metody opartej na podejściu kwantylowym

Metoda zakłada, że wprowadzony zbiór danych będzie się składał z ciągu niezależnych zmiennych losowych, pochodzących z dowolnego rozkładu o skończonej wariancji.

Niech $X_1, X_2, ..., X_N$ będą nieskorelowanymi zmiennymi losowymi z rozkładu normalnego o średniej 0 i wariancji σ_i^2, $i = 1, 2, ..., N$. Wykorzystaną statystyką testową jest skumulowana suma kwadratów C_n dana wzorem:

$$C_n = \sum_{i=1}^{n} X_i^2, \quad n = 1, 2, ..., N. \tag{1}$$

Zakładając, że dane mogą zawierać jeden punkt zmiany reżimu, można dokonać analizy każdego z możliwych podziałów danych. Metoda bazuje na minimalizacji sumy kwadratów błędów, otrzymanych poprzez dopasowywanie prostej regresji liniowej do danych z dwóch ustalonych segmentów pochodzących z oryginalnej próbki. W celu usystematyzowania przyjętych konwencji warto wprowadzić następujące pojęcia:

Załóżmy, że mamy zbiór danych $\{Y_i\}_{i=1,2,...,N}$, $N \in \mathbb{N}$. W teoretycznym modelu regresji liniowej zmienna objaśniana Y_i jest następującej postaci:

$$Y_i = a + bx_i + \epsilon_i, \quad i = 1, 2, ..., N,$$

przy czym a oraz b są stałe, x_i to zmienna objaśniająca (deterministyczna), a $\{\epsilon_i\}_{i=1}^N$ to niezależne zmienne losowe o następujących własnościach:

$$E[\epsilon_i] = 0,$$
$$Var[\epsilon_i] = \sigma^2 < \infty.$$

Aby dopasować prostą regresji liniowej $\hat{y}_i = \alpha + \beta x_i$ w taki sposób, by odległość chmary punktów od prostej była jak najmniejsza stosuje się metodę najmniejszych kwadratów, która w niniejszym zastosowaniu sprowadza się do minimalizacji funkcji $S(\alpha, \beta)$, danej wzorem:

$$S(\alpha, \beta) = \sum_{i=1}^{N} (Y_i - (\alpha + \beta x_i))^2$$

Następuje to wówczas, gdy spełniony jest układ równań:

$$\begin{cases} \dfrac{\partial S(\alpha, \beta)}{\partial \alpha} = 0 \\ \dfrac{\partial S(\alpha, \beta)}{\partial \beta} = 0. \end{cases}$$

1.2 Procedura wyznaczania punktu zmiany reżimu za pomocą metody kwantylowej

Procedurę wyznaczania zmiany strukturalnej za pomocą metody kwantylowej można przedstawić na schemacie. Na początku dla każdego z możliwych punktów zmiany obliczana jest wartość statystyki i na tej podstawie wyznacza się punkt, będący potencjalną zmianą strukturalną. Następnie przeprowadzony zostaje test, którego celem jest określenie, czy uzyskaną wartość można faktycznie uznać jako punkt zmiany lub czy wynik należy odrzucić ze względu na to, że dane pozbawione są punktu zmian. Procedurę można przedstawić za pomocą następujących kroków:

1. Znajdowanie potencjalnego punktu zmiany reżimu:

   (a) Oblicz wartość

   $$C_n = \sum_{i=1}^{n} X_i^2, \quad n = 1, 2, ..., N.$$

   $\forall j \in \{1, 2, ..., N\}$, gdzie $N \in \mathbb{N}$ jest długością próby.

   (b) Ustal punkt $k \in \{2, ..., N-2\}$, który będzie dzielić próbkę danych na dwa segmenty $\{C_i\}_{i=1}^{k}$ oraz $\{C_j\}_{j=k+1}^{N}$.

   (c) Wyestymuj współczynniki prostych regresji do obu z wyznaczonych segmentów, korzystając z metody najmniejszych kwadratów ($\alpha_{1,k}, \beta_{1,k}$ - dla pierwszego segmentu oraz $\alpha_{2,k}, \beta_{2,k}$ dla drugiego z segmentów).

   (d) Powtórz kroki 1. (b), 1. (c) dla każdego możliwego punktu podziału k (tzn. $\forall k \in \{2, ..., N-2\}$).

   (e) Niech

   $$V(k) = \sum_{j=1}^{k} (C_j - (\alpha_{1,k} + \beta_{1,k} \cdot j))^2 + \sum_{j=k+1}^{N} (C_j - (\alpha_{2,k} + \beta_{2,k} \cdot j))^2. \qquad (2)$$

   Potencjalny punkt zmiany reżimu, to k^*, dla którego:

   $$k^* = arg \min_{2 \leqslant k \leqslant N-2} [V(k)].$$

2. Sprawdzenie, czy k^* wskazane przez algorytm można uznać za punkt zmiany reżimu:

(a) Dokonaj podziału kwadratowych wartości szeregu czasowego na dwa segmenty względem wyznaczonego punktu k^*: $\mathcal{X}_1 = [X_1^2, ..., X_{k^*}^2]$ oraz $\mathcal{X}_2 = [X_{k^*+1}^2, ..., X_N^2]$.

(b) Oblicz empiryczne wartości odchyleń standardowych w obu segmentach: $\hat{\sigma}_1$, $\hat{\sigma}_2$.

(c) Niech A będzie segmentem o mniejszym odchyleniu standardowym, a B niech będzie drugim z segmentów (spośród $\mathcal{X}_1$ oraz $\mathcal{X}_2$). Wyznacz empiryczne kwantyle na zadanym poziomie α z rozkładu o mniejszym empirycznym odchyleniu standardowym (czyli $q_{\alpha/2}$ oraz $q_{1-\alpha/2}$)). Kwantyle spełniają poniższą zależność:

$$P(q_{\alpha/2} < X_i^2 < q_{1-\alpha/2}) = 1 - \alpha.$$

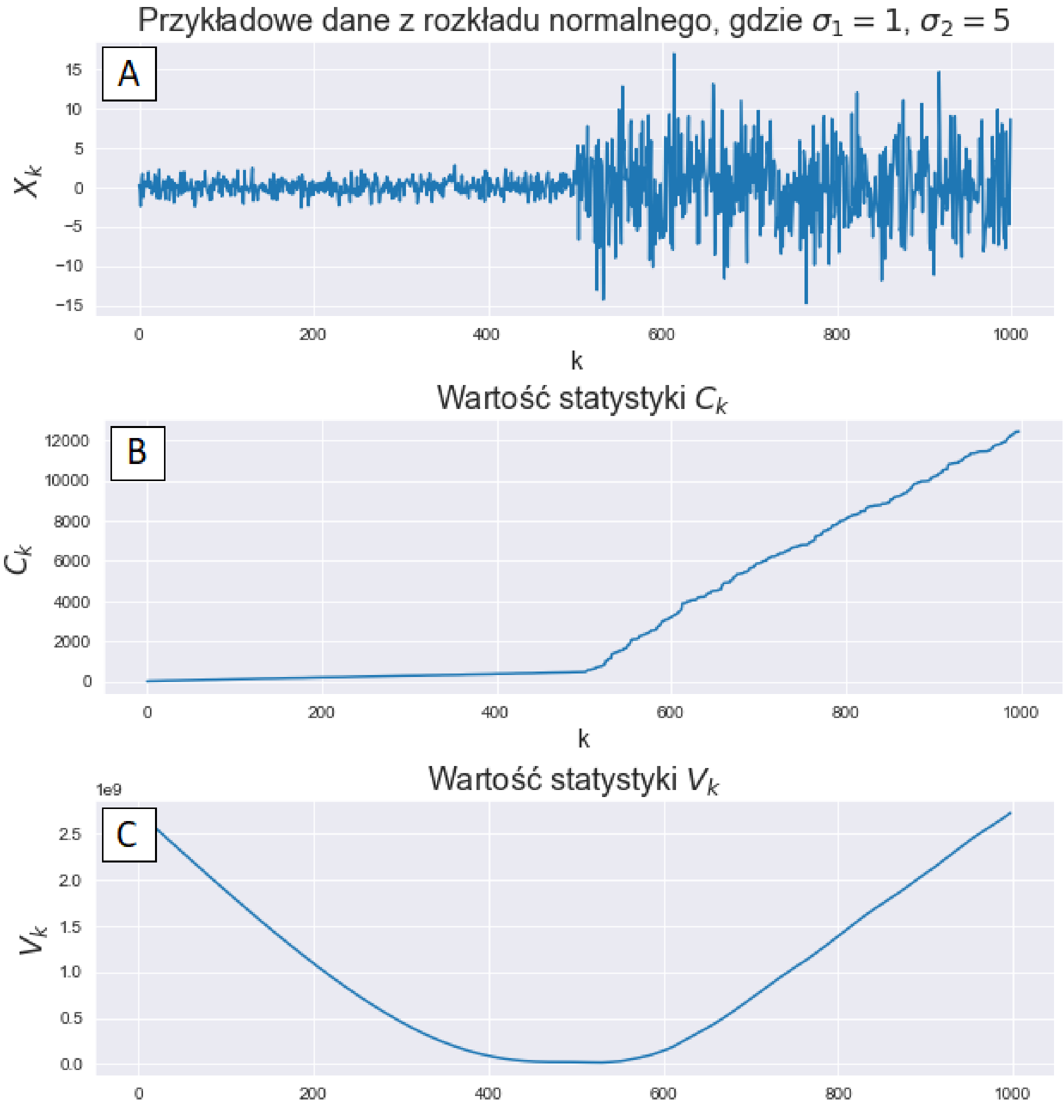
Oblicz ilość obserwacji z B, które znajdują się pomiędzy wyznaczonymi wartościmi kwantyli:

$$D = \#\{q_{\alpha/2} \leqslant B \leqslant q_{1-\alpha/2}\}. \tag{3}$$

Statystyka D ma rozkład dwumianowy z parametrami $n-$ ilość obserwacji w B oraz $p = 1 - \alpha$, p-wartość dla testu jest obliczana jako $P(Z < D)$.

(d) Jeśli obliczona p-wartość jest większa niż α, wówczas nie ma podstaw do odrzucenia hipotezy zerowej (punkt k^* nie może być traktowany jako punkt zmiany reżimu). W przeciwnym razie hipoteza zerowa zostaje odrzucowna na rzecz hipotezy alternatywnej, co oznacza, że wyznaczony punkt k^* jest punktem, w którym zmienia się warancja badanych danych.

Statystyką, która podlega minimalizacji jest statystyka $V(k)$, której postać dla przykładowych danych (wraz z pośrednio obliczoną statystyką C_k) przedstawiono na Rysunku 1.



Rysunek 1: Wizualizacja statystyk testowych używanych w metodzie kwantylowej. A: Przykład danych, wśród których można wydzielić dwa segmenty - pierwszy o rozkładzie $N(0,1)$, a drugi o rozkładzie $N(0,5)$. B: Statystyka C_j obliczona na podstawie danych z A. C: Statystyka $D(k)$ obliczona na podstawie danych z A.