It is unlikely that the behaviour of time series would be the same for a longer periods. Due to governmental inventions, different economic condition, natural disasters the behaviour of the time series might change but the time is usually unknown. It is important to check the stability of our model. Neglecting changes and use our data as a stationary sequence could case misleading conclusions and false predictions. We wish to investigate how to segment a data set.

We start with a simple example. Let $X_1, X_2, \ldots, X_N$ be independent normal random defined as

$$X_i = \left\{ \begin{array}{ll} 2 + \epsilon_i, & \text{if} \quad 1 \le i \le \lfloor N/3 \rfloor, \\ 1 + \epsilon_i, & \text{if} \quad \lfloor N/3 \rfloor + 1 \le i \le \lfloor 2N/3 \rfloor \\ \epsilon_i, \quad \text{if} \quad \lfloor 2N/3 \rfloor + 1 \le i \le N, & \end{array} \right\}$$

where $\lfloor x \rfloor$ denotes the integer part of $x$. We assume that $\epsilon_1, \epsilon_2, \ldots, X_N$ are independent and identically distributed standard normal random variables. According to our model we start with mean 2, this changes to 1 at $\lfloor N/3 \rfloor + 1$ and to 0 at $\lfloor 2N/3 \rfloor + 1$. We wish to estimate to estimate the number of changes in the sequence. Let $c(\alpha)$ the critical value for the supremum of the absolute value of a Brownian bridge, i.e.

$$P\{\sup_{0 \le t \le 1} |B(t)| \ge c(\alpha)\} = \alpha.$$

These values are widely available, these are the asymptotic values for the Kolmogorov–Smirnov statistic

We suggest the following algorithm: we look at the function

$$T^{(1)}(k, X_1, X_2, \ldots, X_N)) = \frac{1}{N^{1/2}} \left| \sum_{i=1}^{k} X_i - \frac{k}{N} \sum_{i=1}^{N} X_i \right|$$

and we check if

$$\max_{1 \le k \le N} T(k, X_1, X_2, \ldots, X_N)) \ge c(\alpha).$$

If it is true than the sequence changed mean and the location of the maximum is our estimator for the time of change. This is denoted by $\bar{k}_1$. If the maximum is reached at more than one point, take the smallest one. Cut the data into to subsets $X_1, X_2, \ldots, X_{\bar{k}_1}$ and $X_{\bar{k}_1+1}, X_{\bar{k}_1+2}, \ldots, X_N$ and again we compute the CUSUM sequence from each subset

$$T^{(1,1)}(k, X_1, X_2, \ldots, X_{\bar{k}_1}) = \frac{1}{\bar{k}_1^{1/2}} \left| \sum_{i=1}^{k} X_i - \frac{k}{\bar{k}_1} \sum_{i=1}^{\bar{k}_1} X_i \right|.$$

If the maximum of this sequence is less than $c(\alpha)$, then there is no change in this subset and nothing more is done with this subsequence. If it is above the critical value we found a change, and the location of the maximum $\bar{k}_{1,1}$. Now we cut the subset $X_1, X_2, \ldots, X_{\bar{k}_1}$ into two subsets at $\bar{k}_{1,1}$ and continue the segmentation on each subset.

From $X_{\bar{k}_1+1}, X_{\bar{k}_1+2}, \ldots, X_N$ we compute

$$T^{(1,2)}(k, X_{\bar{k}_1+1}, X_{\bar{k}_1+2}, \ldots, X_N) = \frac{1}{(N - \bar{k}_1)^{1/2}} \left| \sum_{i=\bar{k}_1+1}^{k} X_i - \frac{k - \bar{k}_1 + 1}{N - \bar{k}_1} \sum_{i=\bar{k}_1+1}^{N} X_i \right|.$$

If the maximum of this sequence is less than $c(\alpha)$, then there is no change in this subset and nothing more is done with this subsequence. If it is above the critical value we found a change, and the location of the maximum $\bar{k}_{1,2}$. We cut $X_{\bar{k}_1+1}, X_{\bar{k}_1+2}, \ldots, X_N$ into two subsets at $\bar{k}_{1,2}$ and continue the segmentation.

If it goes well you should have 0, 1,2,3,4 estimators for the change point, 3 should be the most common

Hence you have the times of changes $\bar{k}_1$, $\bar{k}_{1,1}$, $\bar{k}_{1,2}, \ldots$. If there is no change the location of the time of change is zero (this is when you stop at the first step). You repeat this experiment $M$ times and you will get $M$ vectors of the times of changes. Compute the histogram of the times of changes. Display the histogram.

You need to choose $N$ and $M$. I'd start small values like $N = 50$ and $M = 200$. Then we could check if the simulations are working. If looks fine, we start changing $N$ and $M$.