

DESCRIPTION OF GALEANO AND PEÑA'S COVARIANCE CHANGEPOINT DETECTION  
ALGORITHM BASED ON CUSUM STATISTIC

Christopher Natoli

## 1 Formulating the test statistic

The following changepoint detection algorithm, which focuses on changes in the covariance matrix, was designed by Galeano and Peña [1].

Suppose the  $k$ -dimensional time series  $\mathbf{y}_t$  can be represented as a VARIMA process

$$\Phi(B)\mathbf{y}_t = \mathbf{c} + \Theta(B)\mathbf{a}_t,$$

where  $B$  is the backshift operator and the error or innovation  $\mathbf{a}_t$  is an iid Gaussian random variable with mean  $\mathbf{0}$  and covariance matrix  $\Sigma_i$ , where  $i$  indicates the particular regime.

The cusum test statistic is derived from accumulating the sum of squared errors. Since we do not know the true errors  $\mathbf{a}_t$ , we first fit a VARIMA process  $\hat{\mathbf{y}}_t$  to the time series  $\mathbf{y}_t$  and use the residuals  $\mathbf{e}_t = \mathbf{y}_t - \hat{\mathbf{y}}_t$  as estimates for the errors. Consider the interval  $\{\ell, \dots, r\}$ . We can estimate the covariance matrix of the time series in this interval by

$$\hat{\Sigma}_\ell^r = \frac{1}{r - \ell} \sum_{t=\ell}^r \mathbf{e}_t \mathbf{e}_t^\top.$$

Denote the sum of squares accumulated up to time  $h$  by

$$A_\ell^r(h) = \sum_{t=\ell}^h \mathbf{e}_t^\top \left( \hat{\Sigma}_\ell^r \right)^{-1} \mathbf{e}_t,$$

where the squared errors are in some sense normalized by the empirical covariance  $\hat{\Sigma}_\ell^r$  of the entire interval.

The cusum test statistic for a changepoint at time  $h + 1$  is then

$$C_\ell^r(h) = \frac{h}{\sqrt{2k(r - \ell + 1)}} \left( \frac{A_\ell^r(h)}{h} - \frac{A_\ell^r(r - \ell + 1)}{r - \ell + 1} \right).$$

The term outside the parentheses is another normalization factor. The left-hand term inside the parentheses is the cumulative sum of squares up to time  $h$ . The right-hand term inside the parentheses is the cumulative sum of squares for the entire interval  $\{\ell, \dots, r\}$ . The test statistic  $C_\ell^r(h)$  compares these latter two terms. For example, if there is no changepoint in  $\{\ell, \dots, r\}$ , then  $\frac{A_\ell^r(h)}{h}$  and  $\frac{A_\ell^r(r - \ell + 1)}{r - \ell + 1}$  will be approximately equal. However, if there is a single changepoint at time  $t = h + 1$ , then the left-hand term is the sum of squares for all of the

first regime (and only the first regime), while the right-hand term is the sum of squares over the entirety of both regimes. This suggests that the cumulative sums of squares are most different, and thus  $C_\ell^r(h)$  is greatest in magnitude, at the changepoint  $t = h + 1$ .

This observation is confirmed by plotting  $|C_\ell^r(h)|$  vs.  $h$  for simulated data. Therefore, let

$$\begin{aligned}\Gamma_\ell^r &= \max_{h \in \{\ell, \dots, r\}} |C_\ell^r(h)| \\ \bar{h}_\ell^r &= \operatorname{argmax}_{h \in \{\ell, \dots, r\}} |C_\ell^r(h)|.\end{aligned}$$

Galeano and Peña prove that  $\Gamma_\ell^r$  asymptotically has the distribution of the supremum over  $[0, 1]$  of a Brownian bridge. This is a known distribution, allowing us to calculate the critical value  $C_\alpha$  for any given significance level  $\alpha$ .

## 2 Algorithm for finding multiple changepoints

The approach to finding multiple changepoints is similar to binary segmentation. Let  $d = k(p + q + 1) + \frac{k(k+1)}{2} + 1$  be the resolution of the changepoint detection algorithm. First fit a VARIMA model to the data and computing the residuals. Then let  $h_{\text{first}} = 1 + d$  and  $h_{\text{last}} = T - d$ , the final time step. Search for a changepoint in  $\{h_{\text{first}}, \dots, h_{\text{last}}\}$  by checking if  $\Gamma_{h_{\text{first}}}^{h_{\text{last}}}$  exceeds the critical value. Let  $\bar{h}_{\text{old}} = \bar{h}_{h_{\text{first}}}^{h_{\text{last}}}$ .

If a candidate is found, split the entire time series at the time  $t_2 = \bar{h}_{\text{old}}$  of the candidate changepoint. Check  $\Gamma_{h_{\text{first}}}^{t_2}$  if greater than the critical value; if so, redefine  $t_2$  as  $\bar{h}_{h_{\text{first}}}^{t_2}$ . Repeat until  $\Gamma_{h_{\text{first}}}^{t_2}$  is no longer significant. This procedure thus finds the earliest point in the time series at which a changepoint could occur. Redefine  $h_{\text{first}}$  as the last significant value of  $\bar{h}_{h_{\text{first}}}^{t_2}$ . Perform the same procedure over the interval  $\{\bar{h}_{\text{old}}, \dots, h_{\text{last}}\}$ , acquiring a new  $h_{\text{last}}$  as the latest point in the time series at which a changepoint could occur. If  $|h_{\text{first}} - h_{\text{last}}| < d$ , i.e., the resolution of the algorithm is not high enough to distinguish between  $h_{\text{first}}$  and  $h_{\text{last}}$ , then record  $\bar{h}_{\text{old}}$  as a candidate changepoint. Otherwise, record both  $h_{\text{first}}$  and  $h_{\text{last}}$  as candidate changepoints. Then repeat this procedure in the narrower interval  $\{h_{\text{first}}, \dots, h_{\text{last}}\}$ . Thus, rather than continually cutting the interval in two and repeating the procedure in each part, the algorithm instead narrows down the interval in which changepoints can occur, using previous changepoints as the new endpoints of the narrower interval.

The algorithm ends up find excess candidates. To solve this retrospectively, let  $x_1 = 1$ ,  $x_s = T$ , and  $\{x_2, \dots, x_{s-1}\}$  be the sorted list of candidate changepoints. In each interval  $\{x_i + 1, \dots, x_{i+2} - 1\}$ , drop  $x_i$  from the list of candidates if  $\Gamma_{x_i+1}^{x_{i+2}-1}$  is insignificant. Repeat this procedure until convergence. Then remove  $x_1$  and  $x_s$  from the list of candidate changepoints. Denote the winnowed list of candidates by  $X$ . Then the final changepoints detected by the algorithm are  $\{x + 1 : x \in X\}$ .

The entire procedure is detailed more explicitly in Algorithm 1.

