Assessing Change-Points in Surface Air Temperature Over Alaska

Elena A. Khapalova[1], Venkata K. Jandhyala[2*], Stergios B. Fotopoulos[3], James E. Overland[4]

[1]Schneider School of Business & Economics, St. Norbert College, De Pere, WI, USA
[2*]Department of Mathematics and Statistics, Washington State University, Pullman, WA, USA
[3]Department of Finance and Management Science, Washington State University, Pullman, WA, USA
[4]NOAA Pacific Marine Environmental Laboratory, Seattle, WA, USA

[*]Correspondence:
Dr. Venkata K. Jandhyala
e-mail: jandhyala@wsu.edu

Appendix 1

PELT Method

The computational steps associated with PELT developed by Killick et al. (2012) are presented below. One may see Killick et al. (2012) for clarity of notations used in the following steps.
Required input:

1. A time series, $(y_1, y_2, ..., y_n)$, where $y_i \in R$.
2. Fit measure (cost function), $C(.)$, that depends on the data.
3. A constant penalty value β (to prevent over-fitting) that is not dependent on the change-point location or number of change-points.
4. A constant K that satisfies $C(y_{(t+1):s}) + C(y_{(s+1):T}) + K \le C(y_{(t+1):T})$, $t < s < T$.

We let $n =$ length of time series and set $F(0) = -\beta, cp(0) = NULL, R_1 = \{0\}$ and iterate for $\tau^* = 1,...,n$, where τ denotes the change-point position.

1. Calculate $F(\tau^*) = \min_{\tau \in R_{\tau^*}} \left[F(\tau) + C(y_{(\tau+1):\tau^*}) + \beta \right]$.
2. Let $\tau^1 = \arg\left\{ \min_{\tau \in R_{\tau^*}} \left[F(\tau) + C(y_{(\tau+1):\tau^*}) + \beta \right] \right\}$.
3. Set $cp(\tau^*) = \left[cp(\tau^1), \tau^1 \right]$.
4. Set $R_{\tau^*+1} = \left\{ \tau^* \cap \left\{ \tau \in R_{\tau^*} : F(\tau) + C(y_{(\tau+1):\tau^*}) + K \le F(\tau^*) \right\} \right\}$

Resulting output: the change-points stored in $cp(n)$.

Appendix 2

Distribution of Change Detection Statistic and Change-Point Estimate

The likelihood ratio method (Csörgő and Horváth, 1997) of detecting an unknown change-point in the mean of a Gaussian time series begins by letting the data be $Y_1, Y_2, \cdots, Y_n$, $n \geq 1$. Here, let the Gaussian parameters be μ and σ^2. The process begins with (μ_1, σ^2) and then changes to (μ_2, σ^2) with $\mu_1 \neq \mu_2$ at an unknown time-point τ_n where $\tau_n \in \{1, 2, \cdots, n-1\}$. The corresponding twice log-likelihood ratio statistic may be computed as: $U_n = \max_{1 \leq t \leq n-1} n \log\left(\hat{\sigma}_n^2 / \hat{\sigma}_t^2\right)$, where $\hat{\sigma}_t^2 = n^{-1}\left\{\sum_{i=1}^{t}\left(Y_i - \hat{\mu}_{1,t}\right)^2 + \sum_{i=t+1}^{n}\left(Y_i - \hat{\mu}_{2,t}\right)^2\right\}$, $\hat{\mu}_{1,t} = t^{-1}\sum_{i=1}^{t} Y_i$, and $\hat{\mu}_{2,t} = (n-t)^{-1}\sum_{i=t+1}^{n} Y_i$, $t = 1, \cdots, n$.

Letting $W_n = (2 \log\log n U_n)^{1/2} - \left(2 \log\log n + \frac{1}{2} \log\log\log n - \log\Gamma(1/2)\right)$, one then utilizes the limiting distribution of W_n given by $\lim_{n\to\infty} P[W_n \leq t] = \exp\left(-2\, e^{-t}\right)$ to compute the p-value associated with W_n. When the test is significant, the mle $\hat{\tau}_n$ of τ_n is the argument at which W_n attains its maximum.

It is also of interest to test for change in the variance of the data series. The generalized log-likelihood ratio statistic for the constancy of the variance over time against the alternative that the variance has changed at an unknown time is given by $U_n^* = \max_{1 \leq t \leq n-1} \log\left\{\hat{\sigma}_{1:n}^{n} / \left(\hat{\sigma}_{1:t}^{t} \hat{\sigma}_{t+1:n}^{(n-t)}\right)\right\}$, where $\hat{\sigma}_{1:t}$ and $\hat{\sigma}_{t+1:n}$ are the usual estimators of the variance based on the first t and last $n-t$ deviations, respectively. The limiting distribution of U_n^* is obtained through the distribution of W_n^*, where W_n^* is defined upon U_n^* in an analogous manner.

Asymptotic distribution of the mle $\hat{\tau}_n$ as derived by Fotopoulos et al. (2010) can be computed through the centered estimator given by $\xi_n = \hat{\tau}_n - \tau_n$. Computing the limiting distribution of ξ_n denoted by ξ_∞ involves the following:

$$P(\xi_\infty = k) = \begin{cases} \left(1 - \|G_+\|\right)\left(q_{|k|} - \|G_+\| \tilde{q}_{|k|}\right) & k = \pm 1, \pm 2, \cdots \\ \left(1 - \|G_+\|\right)^2 & k = 0, \end{cases}$$

where $1 - \|G_+\| = \exp\left\{-\sum_{j=1}^{\infty} \frac{1}{j} \overline{\Phi}\left(\eta\sqrt{j}/2\right)\right\}$, and $q_k = E\left\{I\left(T_1^- > k\right)\right\}$, $\tilde{q}_k = E\left\{e^{-S_k} I\left(T_1^- > k\right)\right\}$, $k = 1, 2, \cdots$

, with T_1^- representing the first time that a random walk with negative drift becomes negative. Also, $q_0 = \tilde{q}_0 = 1$, $\eta^2 = (\mu_1 - \mu_2)^2 / \sigma^2$, and $\overline{\Phi}(\cdot)$ is the survival function of the standard normal distribution.