trend: Non-Parametric Trend Tests and Change-Point Detection, R package version 0.0.1

Technical Report · January 2015

DOI: 10.13140/RG.2.1.2633.4243

1 author:

Thorsten Pohlert

Non-Parametric Trend Tests and Change-Point Detection

Thorsten Pohlert

March 24, 2015

Contents

1 Trend detection

1.1 Mann-Kendall Test

The non-parametric Mann-Kendall test is commonly employed to detect monotonic trends in series of environmental data, climate data or hydrological data. The null hypothesis, H_0, is that the data come from a population with independent realizations and are identically distributed. The alternative hypothesis, H_A, is that the data follow a monotonic trend. The Mann-Kendall test statistic is calculated according to :

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \operatorname{sgn}\left(X_j - X_k\right) \tag{1}$$

with

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x = 0 \\ -1 & \text{if} \quad x < 0 \end{cases} \tag{2}$$

The mean of S is $E[S] = 0$ and the variance σ^2 is

$$\sigma^2 = \left\{ n\left(n-1\right)\left(2n+5\right) - \sum_{j=1}^{p} t_j\left(t_j - 1\right)\left(2t_j + 5\right) \right\} / 18 \tag{3}$$

where p is the number of the tied groups in the data set and t_j is the number of data points in the jth tied group. The statistic S is approximately normal distributed provided that the following Z-transformation is employed:

$$Z = \begin{cases} \frac{S-1}{\sigma} & \text{if} \quad S > 0 \\ 0 & \text{if} \quad S = 0 \\ \frac{S+1}{\sigma} & \text{if} \quad S > 0 \end{cases} \tag{4}$$

The statistic S is closely related to Kendall's τ as given by:

$$\tau = \frac{S}{D} \tag{5}$$

where

$$D = \left[\frac{1}{2} n\left(n-1\right) - \frac{1}{2} \sum_{j=1}^{p} t_j\left(t_j - 1\right) \right]^{1/2} \left[\frac{1}{2} n\left(n-1\right) \right]^{1/2} \tag{6}$$

The univariate Mann-Kendall test is envoked as folllows:

```
> require(trend)
> data(maxau)
> Q <- maxau[,"Q"]
> mk.test(Q)
```

```
Mann-Kendall Test


two-sided homogeinity test
H0: S = 0 (no trend)
HA: S != 0 (monotonic trend)


Statistics for total series
     S  varS    Z    tau  pvalue
1 -144 10450 -1.4 -0.145 0.16185
```

1.2 Seasonal Mann-Kendall Test

The Mann-Kendall statistic for the gth season is calculated as:

$$S_g = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \operatorname{sgn}\left(X_{jg} - X_{ig}\right), \quad g = 1, 2, \ldots, m \tag{7}$$

According to Hirsch *et al.* (1982), the seasonal Mann-Kendall statistic, $\hat{S}$, for the entire series is calculated according to

$$\hat{S} = \sum_{g=1}^{m} S_g \tag{8}$$

For further information, the reader is referred to Hipel and McLoed (1994, p. 866-869) and Hirsch *et al.* (1982). The seasonal Mann-Kendall test ist conducted as follows:

```
> require(trend)
> smk.test(nottem)

Seasonal Mann-Kendall Test without correlation

two-sided homogeinity test
H0: S = 0 (no trend)
HA: S != 0 (monotonic trend)

Statistics for individual seasons
     S  varS    Z    tau    pvalue
1   -7 944.3 -0.2 -0.037 0.8198092
2    3 949.0  0.1  0.016 0.9224214
3    1 949.0  0.0  0.005 0.9741041
4   31 947.0  1.0  0.163 0.3137596
5  -23 944.3 -0.7 -0.121 0.4541863
6   45 949.0  1.5  0.237 0.1440808
7   -9 949.0 -0.3 -0.047 0.7701701
8   80 946.0  2.6  0.421 0.0092946
9   67 944.3  2.2  0.353 0.0292368
10  -2 946.0 -0.1 -0.011 0.9481536
11  59 947.0  1.9  0.311 0.0552071
12 -21 949.0 -0.7 -0.111 0.4954357

Statistics for total series
    S  varS   Z   tau   pvalue
1 224 11364 2.1 0.098 0.035617
```

Only the temperature data in Nottingham for August ($S = 80$, $p = 0.009$) as well as for September ($S = 67$, $p = 0.029$) show a significant ($p < 0.05$) positive trend according to the seasonal Mann-Kendall test. Thus, the global trend for the entire series is significant ($S = 224$, $p = 0.036$).

1.3 Correlated Seasonal Mann-Kendall Test

The correlated seasonal Mann-Kendall test can be employed, if the data are coreelated with e.g. the pre-ceeding months. For further information the reader is referred to Hipel and McLoed (1994, p. 869-871).

```
> require(trend)
> csmk.test(nottem)

Seasonal Mann-Kendall Test with correlations

two-sided homogeinity test
H0: S = 0 (no trend, seasons are correlated)
HA: S != 0 (monotonic trend)

Statistics for total series
    S     varS   Z  pvalue
1 224 19663.33 1.6 0.11017
```

1.4 Partial Mann-Kendall Test

This test can be conducted in the presence of co-variates. For full information, the reader is referred to Libiseller and Grimvall (2002).

We assume a correlation between concentration of suspended sediments (s) and flow at Maxau.

```
> data(maxau)
> s <- maxau[,"s"]; Q <- maxau[,"Q"]
> cor.test(s,Q, meth="spearman")

        Spearman's rank correlation rho

data:  s and Q
S = 10564, p-value = 0.0427
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.3040843
```

As s is significantly positive related to flow, the partial Mann-Kendall test can be employed as follows.

```
> require(trend)
> data(maxau)
> s <- maxau[,"s"]; Q <- maxau[,"Q"]
> partial.mk.test(s,Q)

        Partial Mann-Kendall trend test

data:  t AND  s  .  Q
Z = -3.597, p-value = 0.0003218
alternative hypothesis: true trend exists in series
sample estimates:
        S      varS
-350.6576 9503.2898
```

The test indicates a highly significant decreasing trend ($S = -350.7$, $p < 0.001$) of s, when Q is partialled out.

1.5 Partial correlation trend test

This test performs a partial correlation trend test with either the Pearson's or the Spearman's correlation coefficients (r(tx.z)). The magnitude of the linear / monotonic trend with time is computed while the impact of the co-variate is partialled out.

```
> require(trend)
> data(maxau)
> s <- maxau[,"s"]; Q <- maxau[,"Q"]
> partial.cor.trend.test(s,Q, "spearman")

        Spearman's partial correlation trend test

data:  t AND  s  .  Q
t = -4.158, df = 43, p-value = 0.0001503
alternative hypothesis: true correlation is not equal to 0
sample estimates:
   r(ts.Q)
-0.5355055
```

Likewise to the partial Mann-Kendall test, the partial correlation trend test using Spearman's correlation coefficient indicates a highly significant decreasing trend ($r_{S(ts.Q)} = -0.536$, $n = 45$, $p < 0.001$) of s when Q is partialled out.

2 Magnitude of trend

2.1 Sen's slope

This test computes both the slope (i.e. linear rate of change) and intercept according to Sen's method. First, a set of linear slopes is calculated as follows:

$$d_k = \frac{X_j - X_i}{j - i} \quad (9)$$

for $(1 \leq i < j \leq n)$, where d is the slope, X denotes the variable, n is the number of data, and i, j are indices.

Sen's slope is then calculated as the median from all slopes: $b = \text{Median}\, d_k$. The intercepts are computed for each timestep t as given by

$$a_t = X_t - b * t \quad (10)$$

and the corresponding intercept is as well the median of all intercepts.

This function also computes the upper and lower confidence limits for sens slope.

```
> require(trend)
> s <- maxau[,"s"]
> sens.slope(s)

Sen's slope and intercept


slope:  -0.2876
95 percent confidence intervall for slope
-0.1519 -0.4196

intercept: 31.8574
nr. of observations: 45
```

2.2 Seasonal Sen's slope

Acccording to Hirsch *et al.* (1982) the seasonal Sen's slope is calculated as follows:

$$d_{ijk} = \frac{X_{ij} - xik}{j - k} \quad (11)$$

for each $(x_{ij}, xik$ pair $i = 1, 2, \ldots, m$, where $1 \leq k < j \leq n_i$ and n_i is the number of known values in the ith season. The seasonal slope estimator is the median of the d_{ijk} values.

```
> require(trend)
> sea.sens.slope(nottem)

Seasonal Sen's slope and intercept


slope:  0.05
intercept: 42.1
nr. of observations: 240
```

3 Change-point detection

3.1 Pettitt's test

The approach after Pettitt (1979) is commonly applied to detect a single change-point in hydrological series or climate series with continuous data. It tests the H_0: The T variables follow one or more distributions that have the same location parameter (no change), against the alternative: a change point exists. The non-parametric statistic is defined as:

$$K_T = \max |U_{t,T}| , \tag{12}$$

where

$$U_{t,T} = \sum_{i=1}^{t} \sum_{j=t+1}^{T} \operatorname{sgn}(X_i - X_j) \tag{13}$$

The change-point of the series is located at K_T, provided that the statistic is significant. The significance probability of K_T is approximated for $p \leq 0.05$ with

$$p \simeq 2 \ \exp \left(\frac{-6 \ K_T^2}{T^3 + T^2} \right) \tag{14}$$

The Pettitt-test is conducted in such a way:

```
> require(trend)
> data(PagesData)
> pettitt.test(PagesData)

        Pettitt's test for single change-point detection

data:  PagesData
K = 232, p-value = 0.01456
alternative hypothesis: true change point is present in the series
sample estimates:
probable change point at tau
                          17
```

As given in the publication of Pettitt (1979) the change-point of Page's data is located at $t = 17$, with $K_T = 232$ and $p = 0.014$.

References

Hipel KW, McLoed AI (1994). *Time series modelling of water resources and environmental systems*. Elsevier, Amsterdam.

Hirsch R, Slack J, Smith R (1982). "Techniques of Trend Analysis for Monthly Water Quality Data." *Water Resour. Res.*, **18**, 107–121.

Libiseller C, Grimvall A (2002). "Performance of partial Mann-Kendall tests for trend detection in the presence of covariates." *Environmetrics*, **13**, 71–84. doi:10.1002/env.507.

Pettitt AN (1979). "A non-parametric approach to the change-point problem." *Appl. Statist.*, **28**, 126–135.