Testing for a Change in Mean of a Weakly Stationary Time Series

Shreyas Kowshik (17MA20039)

Indian Institute of Technology Kharagpur

shreyaskowshik@iitkgp.ac.in

29/11/2021

Outline

Background

Time Series

Definition

Time series is a collection of random variables $\{X_t | t \in T\}$ over a time index set T, which might be a finite, countably infinite or an uncountable set.

- What we observe are the realized values of the time series i.e. the data set is $\{X_1 = x_1, \cdots, X_n = x_n\}$, where the x_is are some numeric or categorical values.

For example : **Population of India**, **Stock Prices**, **Rainfall in a city**, etc.

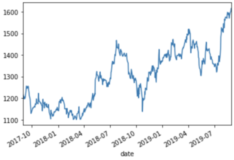


Figure: Time Series Data : Stock Prices

Time Series

Mean and Covariance

- The mean $\mu_X(t)$ of a series $\{X_t\}$ is : $\mu_X(t) = \mathbb{E}[X_t]$
- Covariance (autocovariance) function of $\{X_t\}$:
  $\gamma_X(r, s) = \mathrm{Cov}(X_r, X_s) = \mathbb{E}\left[(X_r - \mu_X(r))\,(X_s - \mu_X(s))\right]$

Weak Stationarity

A time series $\{X_t\}$ is said to be weakly stationary if :

- $\mu_X(t)$ is independent of t
- For every $h \in \mathbb{Z}$, $\gamma_X(t + h, t)$ is independent of t

Strong Stationarity

A time series $\{X_t\}$ is said to be strongly stationary if for all $k, h, t_1 \cdots t_k, x_1 \cdots x_k$, shift of the time axis does not affect the distribution i.e. $P\left(X_{t_1} \leq x_1, \cdots, X_{t_k} \leq x_k\right) = P\left(X_{t_1+h} \leq x_1, \cdots, X_{t_k+h} \leq x_k\right)$

Time Series

AR(p) Process

An $AR(p)$ (autoregressive) process of order p is defined as :

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + W_t$$

where $W_t \sim WN(0, \sigma^2)$ is white noise.

- The series $\{W_t\}$ is a white noise process.
- For the rest of this work, we will specifically consider AR processes of order 1 i.e. $AR(1)$ processes i.e. :

$$X_t = \rho X_{t-1} + W_t$$

where $|\rho| < 1$, $\rho \neq 0$

Changepoint Detection

Changepoint Detection

- Detection of the existence of an abrupt change in the distribution of a time series
- Can be change in mean, variance, parameter value, etc.
- We focus our attention on detecting a **change in mean** of a **weakly stationary** time series

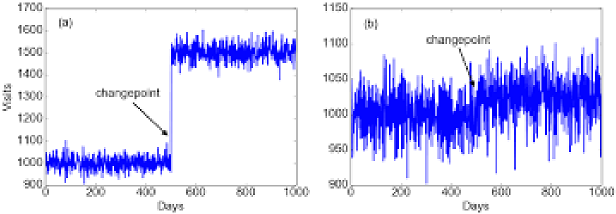


Figure: Example changepoint in a time series

Change in Mean Detection

Problem Statement

Given a sample $\mathbf{X} = \{X_1, \cdots, X_n\}$ from a time-series $\{X_t\}$, we are interested in testing the following hypothesis :

$$H_0 : \mathbb{E}[X_1] = \cdots = \mathbb{E}[X_n]$$

versus

$$H_1 : \mathbb{E}[X_1] = \cdots = \mathbb{E}[X_{k^*}] \neq \mathbb{E}[X_{k^*+1}] = \cdots = \mathbb{E}[X_n]$$

where, $1 \leq k^* < n$ is the location of the changepoint and is unknown.

- This framework is usually considered in retrospective changepoint study
- The other paradigm is online changepoint analysis

Approaches to Change in Mean Detection

KS Statistic

Construction

Given a sample $\mathbf{X} = \{X_1, \cdots, X_n\}$ from a weakly stationary time-series $\{X_t\}$ and defining $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, construct :

$$T_n(\lfloor nt \rfloor) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nt \rfloor} (X_t - \overline{X}_n)$$

Convergence

Under weak dependence and moment conditions ([Phi87]) :

$$T_n(\lfloor nt \rfloor) \implies \sigma(B(t) - tB(1))$$

where $\implies$ denotes weak convergence in $D[0,1]$ endowed with Skorokhod topology, $\sigma^2 = lim_{n\to\infty} n\text{Var}(\overline{X}_n) = \Sigma_{k\in\mathbb{Z}}\gamma(k)$ is the long run variance and $B(t)$ is the one-dimensional standard Brownian motion.

KS Statistic

Kolmogorov-Smirnov Statistic

KS_n statistic is defined as : $KS_n = \sup_{t \in [0,1]} \left| \frac{T_n(\lfloor nt \rfloor)}{\hat{\sigma}_n} \right|$

- $\hat{\sigma}_n$ is a consistent estimator of σ
- Estimating $\hat{\sigma}_n$ generally requires using a kernel-based estimate : $\hat{\sigma}_n^2 = \sum_{k=-l_n}^{l_n} \hat{\gamma}(k) \mathbb{K}\left(\frac{k}{l_n}\right)$
- Here l_n is a bandwidth parameter which can be a function of sample size n or be chosen from the data
- l_n as function of sample size : **not adaptive to presence of changepoint**
- l_n that is data-dependent : **can introduce bias in estimation of σ^2 under alternative hypothesis**

Self-Normalizing Statistic

- Construct a statistic which is pointwise scaled with its estimated pointwise standard deviation
- This construction can help avoid direct estimation of σ^2.

$\widetilde{KS}_n$ Statistic

$\widetilde{KS}_n$ statistic is defined as (Shao'10) :

$$\widetilde{KS}_n = \sup_{t\in[0,1]} \left| \frac{T_n(\lfloor nt \rfloor)}{D_n} \right|$$

where $D_n^2 = n^{-2} \sum_{t=1}^{n} (\sum_{j=1}^{t} (X_j - \bar{X}_n)^2)$.

- No need to estimate σ^2 : Avoids bandwidth selection

$\widetilde{KS}_n$ Statistic

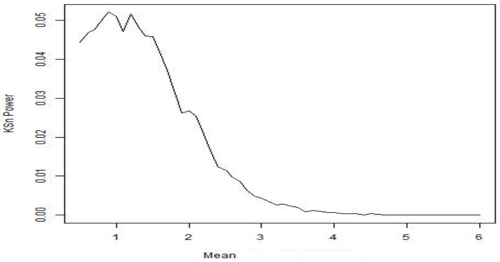


Figure: Power of $\widetilde{KS}_n$

- The mean of data before changepoint is fixed at 1 and the mean after the changepoint is varied above.
- As one moves away from the null hypothesis, the power decreases
- Reason : D_n does not take the alternative into account i.e. the presence and location of a changepoint

G_n Statistic

G_n statistic is defined as :

$$T_n(\lfloor nk \rfloor) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nk]} (X_t - \overline{X}_n)$$

$$S_{t_1,t_2} = \sum_{j=t_1}^{t_2} X_j \text{ if } t_1 \leq t_2, 0 \text{ otherwise}$$

$$V_n(k) = n^{-2}[\sum_{t=1}^{k}(S_{1,t} - \frac{t}{k}S_{1,k})^2 + \sum_{t=k+1}^{n}(S_{t,n} - \frac{n-t-1}{n-k}S_{k+1,n})^2]$$

$$G_n = \sup_{k=1,\cdots,n-1} T_n(k) V_n^{-1}(k) T_n(k)$$

G_n Statistic

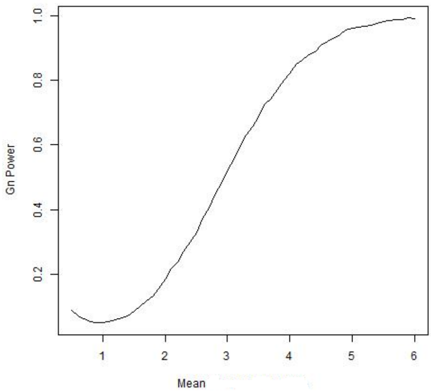


Figure: Power of G_n, $0 \leq \mu \leq 6$

G_n Statistic

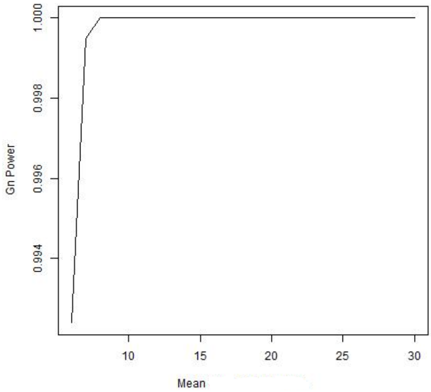


Figure: Power of G_n, $6 \leq \mu \leq 30$

Proposed Statistic

Proposed Statistic

H_n Statistic

Given a sample $\mathbf{X} = \{X_1, \cdots, X_n\}$, the H_n statistic is defined as :

$$T_r(\mathbf{X}) = \frac{1}{n}\left[\left(1 - \frac{r}{n}\right)\sum_{i=1}^{r} X_i + \left(-\frac{r}{n}\right)\sum_{i=r+1}^{n} X_i\right]$$

$$H_n = \sup_{r=1,\cdots,n-1} \frac{\sqrt{n}T_r(\mathbf{X})}{\sqrt{\sum_{|h|<n} w(r,h,n)\gamma(h)}}$$

where $w(r, h, n)$ is a weighting function the details of which we will soon derive and $r \in \{1, \cdots, n-1\}$.

- To obtain a normalized form of H_n, we need to compute $\mathrm{Var}(\sqrt{n}T_r(\mathbf{X}))$

Proposed Statistic

Variance Computation

$$\mathrm{Var}\left[\sqrt{n}T_r(\mathbf{X})\right] = \frac{1}{n}Var\left[\mathbf{a}^T\mathbf{X}\right] = \frac{1}{n}\mathbf{a}^T\Sigma\mathbf{a}$$

$$\text{where : } \mathbf{a} = \left[\underbrace{\left(1-\frac{r}{n}\right),\cdots,\left(1-\frac{r}{n}\right)}_{\text{r times}}\underbrace{\left(,-\frac{r}{n}\right),\cdots,\left(-\frac{r}{n}\right)}_{\text{n-r times}}\right]^T$$

$$\Sigma = \mathrm{Cov}(\mathbf{X})$$

Definining $\alpha = \frac{r}{n}, \beta = \frac{l}{n}$ for lag l, it can be shown that :
$\mathbf{a}^T\Sigma\mathbf{a} = \sum_{l=-(n-1)}^{n-1}\gamma(l)w(r,l,n)$, where :

$$w(r,l,n) = \begin{cases} n\left[\alpha\left(1-\alpha\right)-\beta\left(1-\alpha+\alpha^2\right)\right], & \text{if } \alpha \geq \beta \\ -n\beta\alpha^2, & \text{if } \alpha < \beta \end{cases}$$

Variance Computation in H_n

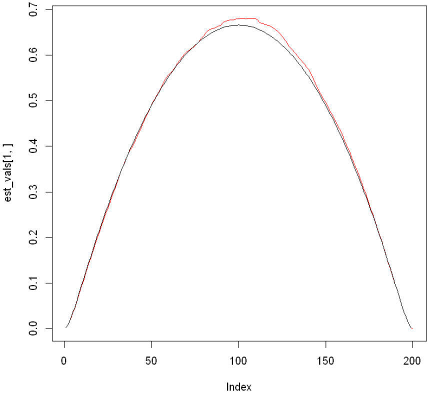
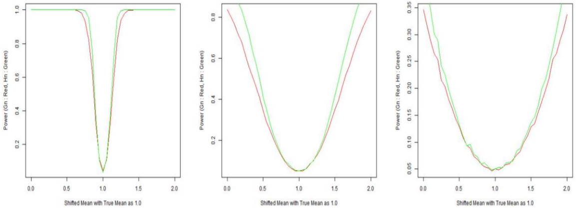


Figure: Predicted Variance (Black) v/s Sample variance of $T_r(\mathbf{X})$ (Red)

Variance Computation in H_n

- Black curve denotes variance at a particular index r predicted from the above formula for an $AR(1)$ process
- Red curve denotes sample variance at a particular index r obtained by simulating multiple $AR(1)$ processes and computing $T_r(\mathbf{X})$
- Very significant overlap!

Power of H_n



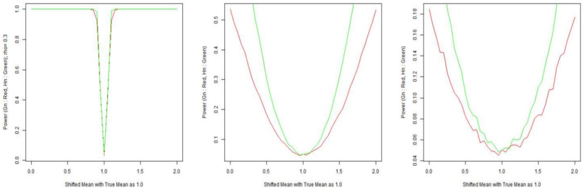
(a) $\rho = 0.7$, $\sigma_{WN} = 0.1, k = 100$

(b) $\rho = 0.7$, $\sigma_{WN} = 0.5, k = 100$

(c) $\rho = 0.7$, $\sigma_{WN} = 1, k = 100$

Figure: Power Curves comparing G_n v/s H_n statistics. Green : H_n, Red : G_n. On X-axis is plotted the new mean after the changepoint with mean before as $\mu = 1$. On Y-axis is the power of the corresponding statistic's test.
ρ:$AR(1)$ coefficient, k:changepoint location, $n = 200$, σ_{WN}:White-Noise Std Dev

Power of H_n



(d) $\rho = 0.3$, $\sigma_{WN} = 0.1, k = 100$

(e) $\rho = 0.7$, $\sigma_{WN} = 0.5, k = 50$

(f) $\rho = 0.7$, $\sigma_{WN} = 1, k = 50$

Figure: Power Curves comparing G_n v/s H_n statistics. Green : H_n, Red : G_n. On X-axis is plotted the new mean after the changepoint with mean before as $\mu = 1$. On Y-axis is the power of the corresponding statistic's test.
ρ:$AR(1)$ coefficient, k:changepoint location, $n = 200$, σ_{WN}:White-Noise Std Dev

- Power of H_n is better (sharper) if not the same as G_n across different values of model parameters ρ, σ_{WN}, k
- H_n shows promise to investigate it further

Normalizing Factor Estimation for Proposed Statistic

Variance Estimation

- Want to estimate : $\mathbf{a}^T\Sigma\mathbf{a} = \sum_{l=-(n-1)}^{n-1} \gamma(l) w(r,l,n)$
- Use a kernel based estimate as they are found to be consistent in the literature :

$$\hat{\sigma}_n^2 = \sum_{k=-l_n}^{l_n} w(r,k,n)\hat{\gamma}(k)\mathbb{K}\left(\frac{k}{l_n}\right)$$

  where l_n is a bandwidth parameter
- The above estimate does not account for the presence of a changepoint i.e. it does nothing special for that
- Consequently, using such an estimate can do good under the null hypothesis, but need not be that good under the alternative hypothesis
- We introduce a variable transformation to address this

Variance Estimation

Transformation of Series

Given a sample $\mathbf{X} = \{X_1, \cdots, X_n\}$, define the following : $\overline{X}_r = \frac{1}{r}\sum_{i=1}^{r} X_i$ and $\overline{\overline{X}}_r = \frac{1}{n-r}\sum_{i=r+1}^{n} X_i$. The series is then transformed as follows :

$$\begin{aligned} Z_1 &= X_1 - \overline{X}_r \\ &\vdots \\ Z_r &= X_r - \overline{X}_r \\ Z_{r+1} &= X_{r+1} - \overline{\overline{X}}_r \\ &\vdots \\ Z_n &= X_n - \overline{\overline{X}}_r \end{aligned}$$

The transformed series $\mathbf{Z} = \{Z_1, \cdots, Z_n\}$ is used for computing the autocovariance estimates $\hat{\gamma}(h)$.

Variance Estimation

- Different kernels exist for the smooth estimation of variance
- We use the Automatic Bandwidth Selection methodology as highlighted in the paper (Andrews'92)
- This is available for direct use in the R-package "cointReg"

Truncated: $$k_{TR}(x) = \begin{cases} 1 & \text{for } |x| \leqslant 1, \\ 0 & \text{otherwise,} \end{cases}$$

Bartlett: $$k_{BT}(x) = \begin{cases} 1 - |x| & \text{for } |x| \leqslant 1, \\ 0 & \text{otherwise,} \end{cases}$$

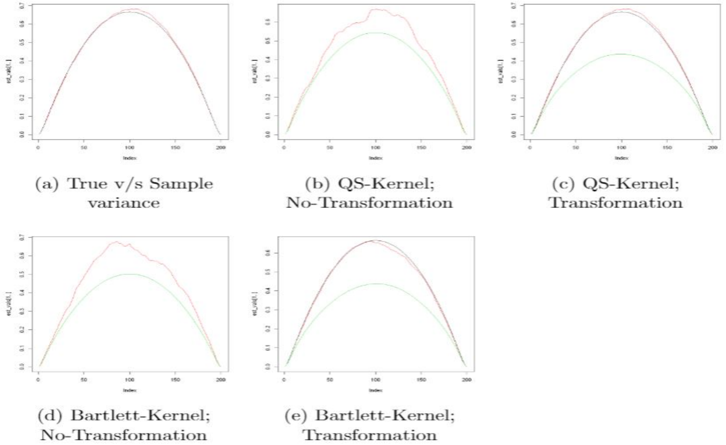
Parzen: $$k_{PR}(x) = \begin{cases} 1 - 6x^2 + 6|x|^3 & \text{for } 0 \leqslant |x| \leqslant 1/2, \\ 2(1 - |x|)^3 & \text{for } 1/2 \leqslant |x| \leqslant 1, \\ 0 & \text{otherwise} \end{cases}$$

Tukey-Hanning: $$k_{TH}(x) = \begin{cases} (1 + \cos(\pi x))/2 & \text{for } |x| \leqslant 1, \\ 0 & \text{otherwise,} \end{cases}$$

Quadratic Spectral: $$k_{QS}(x) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right)$$

Figure: Kernel Formulations

Variance Estimation



(a) True v/s Sample variance

(b) QS-Kernel; No-Transformation

(c) QS-Kernel; Transformation

(d) Bartlett-Kernel; No-Transformation

(e) Bartlett-Kernel; Transformation

Figure: Comparision of variances under null (X-axis : Index r, Y-axis : Variance) Red : Sample variance of $\sqrt{n}T_r(\mathbf{X})$, Black : Predicted Variance ($\mathbf{a}^T\Sigma\mathbf{a}$), Green : Estimated Variance. $n = 200$, $AR(1)$ process, $\rho = 0.7$

Variance Estimation

- It can be observed that without using the transformation, the estimated variance is quite close to the sample variance under the null hypothesis
- Using the transformation may help significantly when one is considering the alternative hypothesis
- This needs to be investigated further and fixed

Conclusion and Future Work

Conclusion and Future Work

- $\widetilde{KS}_n$ statistic suffers from **non-monotonic power** problem due to not incorporating information from alternative hypothesis
- G_n statistic takes alternative hypothesis into account and provides **monotonic power**
- Proposed statistic H_n **is found to outperform** G_n on a wide range of model parameters under exact simulation
- A closed form **normalizer for** H_n **is obtained**
- A **variable transformation was introduced** to estimate the normalizer of H_n. However it requires further investigation

Future Work

- **Investigate discrepancy** between exact and estimated variance using transformation
- **Extensively evaluate** on different processes and parameters
- Study the **asymptotic convergence** of H_n
- Extend research to **multiple changepoints in Bayesian Paradigm**

Thank You

References I

Donald W. K. Andrews, *Heteroskedasticity and autocorrelation consistent covariance matrix estimation*, Econometrica **59** (1991), no. 3, 817–858.

P. C. B. Phillips, *Time series regression with a unit root*, Econometrica **55** (1987), no. 2, 277–301.

Xiaofeng Shao and Xianyang Zhang, *Testing for change points in time series*, Journal of the American Statistical Association **105** (2010), no. 491, 1228–1240.

[SZ10] [And91]