Change point problems for multi-indexed random variables

Vaidotas Zemlys

Vilnius University
Universite des Sciences et Technologies de Lille

17th of June, 2008

The change point problem

- Having a sample (one dimensional index) we want to test whether there is a point where the mean changes.
- Null and alternative hypotheses:

$$H_0 : EX_i = \mu_0$$

against

$$H_A : \exists k^* \text{ such that } EX_i = \mu_0 + (\mu_1 - \mu_0)\mathbf{1}\left(k^* < i \leq n\right)$$

Change point is known

- Simple idea (Csörgo and Horvath (1997)): compare the means.

$$\frac{1}{k^*}S_{k^*} - \frac{1}{n-k^*}(S_n - S_{k^*}) = \frac{n}{k^*(n-k^*)}\left(S_{k^*} - \frac{k^*}{n}S_n\right)$$

- Denote

$$R = (S_{k^*} - \frac{k^*}{n}S_n)$$

- Assume $k^*/n \to c$. Then under null due to CLT we have

$$n^{-1/2}R \to N(0, \sigma^2 c(1-c))$$

- Under alternative we have

$$n^{-1/2}R = n^{1/2}\frac{k^*}{n}\left(1 - \frac{k^*}{n}\right)(\mu_1 - \mu_0) + O_P(1) \to \infty$$

Change point is unknown

- If the change point is unknown, “look” for it:

$$Q = \max_{1<k<n} |S_k - \frac{k}{n} S_n|$$

- Invariance principle and continuous mapping gives us :

$$n^{-1/2} Q \xrightarrow{D} \sup_{0<t<1} |W(t) - tW(1)|$$

Epidemic alternatives

- We want to test the epidemic: the mean changes, then reverts to previous value.
- Alternative hypothesis: $EX_i = \mu_0 + (\mu_1 - \mu_0)\mathbf{1}\,(k^* < i \leq m^*)$
- The same argumentation gives test statistic and its limit:

$$Q_E = \max_{1 \leq i < j \leq \ n} |S_j - S_i - (j-i)/nS_n|,$$

$$n^{-1/2} Q_E \to \sup_{0 < |t-s| < 1} |W(t) - W(s) - (t-s)W(1)|$$

Short epidemics

- Division by the length of epidemic leads to detection of shorter epidemics:

$$UI(n,\alpha) = \max_{1\le i<j\le n} \frac{|S_j - S_i - (j-i)/nS_n|}{[(j-i)/n]^{\alpha}}$$

- Under alternative

$$n^{-1/2}UI(n,\alpha) \ge \frac{l^{*(1-\alpha)}}{n^{1/2-\alpha}}\left(1-\frac{l^*}{n}\right)|\mu_1-\mu_0| + O_P(1)$$

- If $l^* = n^{\gamma}$ then

$$n^{-1/2}UI(n,\alpha) \to \infty \text{ for } \gamma > \frac{1/2-\alpha}{1-\alpha}$$

Hölder space

- Under null hypothesis the FCLT in Hölder space (Račkauskas and Suquet (2004)) gives

$$n^{-1/2} UI(n, \alpha) \to \sup_{0<s<t<1} \frac{|W(t) - W(s) - (t-s)W(1)|}{[t-s]^{\alpha}}$$

- Note $0 < \alpha < 1/2$, since W "lives only" in $\mathrm{H}_{\alpha}^{o}([0,1])$ for $\alpha < 1/2$.

Examples of multi-dimensional index data

- Usual situation: we observe some random variable at given time points (price of stock), or we observe group of variables at certain given time (Inflation rate of EU countries in 2007). In both cases our observation have one index.
- Panel data - observations of group of random variables at certain time points. Yearly inflation rate history of EU countries. Observed variables have two indexes: X_{it}, i - individuals, t - time.

- (H_0) : X_{ij} have all the same mean μ_0.
- (H_A) : There is a set $D^* \subset [0,1]^2$ such that

$$EX_{ij} = \mu_0 + \mu_1 \mathbf{1}\left(\left(\frac{i}{n}, \frac{j}{m}\right) \in D^*\right)$$

- The analog of R statistic, when the change set is known.

$$R = \sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}\mathbf{1}\left(\left(\frac{i}{n}, \frac{j}{m}\right) \in D^*\right) - \frac{N_{D^*}}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij},$$

where

$$N_{D^*} = \sum_{i=1}^{n}\sum_{j=1}^{m} \mathbf{1}\left(\left(\frac{i}{n}, \frac{j}{m}\right) \in D^*\right)$$

The limiting distribution

- If $N_{D^*}/(nm) \to |D^*| > 0$, then

$$(nm)^{-1/2} R \to N(0, \sigma^2 |D^*|(1 - |D^*|))$$

- What to do when D^* is unknown? Restrict the class of possible sets:

$$(nm)^{-1/2} \sup_{D \in \mathcal{D}} \left|S_{nm}(D) - \frac{N_D}{nm} S_{nm}(T)\right| \to \sup_{D \in \mathcal{D}} |W(D) - |D| W(T)|$$

due to resuls by Bass (1985), Alexander, Pyke (1986). Here $T = [0,1]^2$, $\mathcal{D}$ is suitable class of subsets.

Rectangles

- Let us analyze the class of the rectangles $\mathcal{D} = \{[\boldsymbol{s}, \boldsymbol{t}] \subset [0,1]^2\}$. If the rectangle is known:

$$D^* = \left[\frac{a^*}{n}, \frac{b^*}{n}\right] \times \left[\frac{c^*}{m}, \frac{d^*}{m}\right].$$

- Under alternative

$$(nm)^{-1/2} R = (nm)^{1/2} \frac{k^* l^*}{nm} \left(1 - \frac{k^* l^*}{nm}\right) (\mu_1 - \mu_0) + O_P(1),$$

  where $k^* = b^* - a^*$, $l = d^* - c^*$.

- The divergence occurs when $k^* = O(n^\gamma)$ and $l^* = O(m^\delta)$ with $\gamma, \delta > 1/2$.

Epidemic rectangle

- As in one dimensional case it is possible to improve the rates.
- Test statistic:

$$DUI(nm,\alpha) = \max_{\substack{1\le a<b\le n\\ 1\le c<d\le m}} \frac{|\Delta^1_{b-a}\Delta^2_{d-c}S_{bd} - (s_b - s_a)(t_d - t_c)S_{nm}|}{\max\{s_b - s_a, t_d - t_c\}^{\alpha}}$$

where $s_i = i/n$, $t_j = j/m$, for $i = 1,\dots,n$, $j = 1,\dots,m$ and

$$\Delta^1_{b-a}\Delta^2_{d-c}S_{bd} = S_{bd} - S_{ad} - S_{bc} + S_{ac}$$

Alternative hypothesis

- Condition for divergence under alternative hypothesis

$$\lim_{\mathbf{n}\to\infty} (nm)^{1/2} \frac{h_{nm}}{d_{nm}^{\alpha}} |\mu_1 - \mu_0| \to \infty,$$

where

$$h_{nm} = \frac{k^* l^*}{nm}\left(1 - \frac{k^* l^*}{nm}\right) \text{ and } d_{nm} = \max\left\{\frac{k^*}{n}, \frac{l^*}{m}\right\},$$

- Taking $k^* = O(n^\gamma)$ and $l^* = O(m^\delta)$ we get

$$\frac{n^{\gamma-1/2} m^{\delta-1/2}}{[n^{\gamma-1} \vee m^{\delta-1}]^{\alpha}} \to \infty$$

If $n^{\gamma-1} > m^{\delta-1}$ then we have

$$n^{\gamma(1-\alpha)+\alpha-1/2} m^{\delta-1/2} \to \infty$$

Null hypothesis

- Given the FCLT by Račkauskas, Suquet, Zemlys (2007)

$$(nm)^{-1/2} DUI(nm, \alpha) \to T_\alpha(W),$$

where

$$T_\alpha(x) = \sup_{\mathbf{0} \le \boldsymbol{s} < \boldsymbol{t} \le \mathbf{1}} \frac{|\Delta_{[\boldsymbol{s},\boldsymbol{t}]} W(\boldsymbol{t}) - (t_1 - s_1)(t_2 - s_2) W(1,1)|}{|\boldsymbol{t} - \boldsymbol{s}|^\alpha}$$

- Division by volume is open question. Hard to find a suitable functional space.

Panel regression

- Ordinary least squares

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + u_{it}, \quad i = 1, \dots, n; t = 1, \dots, T$$

- Fixed effects

$$y_{it} = \mu_i + \mathbf{x}_{it}\boldsymbol{\beta} + u_{it}, \quad i = 1, \dots, n; t = 1, \dots, T$$

CUSUM type statistics

- Generalisation of Ploberger, Kramer (1992) results:
- "Plug-in" residuals of panel regression to statistic $DUI(nm, \alpha)$.
- For the ordinary least squares regression the limiting statistic is the same.
- For the fixed effects regression:

$$DUI^{FE}(\alpha) = \sup_{\mathbf{0} \leq \boldsymbol{s} < \boldsymbol{t} \leq \mathbf{1}} \frac{|\Delta_{[\boldsymbol{s},\boldsymbol{t}]} W - (t_2 - s_2)[W(t_1, 1) - W(s_1, 1)]|}{|\boldsymbol{t} - \boldsymbol{s}|^{\alpha}}.$$

- The results hold as $n \wedge m \to \infty$!

Remaining problems

- Calculation of critical values.
- Power calculations.
- Small sample properties.