統計数理(2012)
第 60 巻 第 1 号 215–218

［研究ノート］

ノンパラメトリック変化点問題に対する順位統計量について

西山 陽一†

（受付 2012 年 1 月 27 日；改訂 3 月 7 日；採択 3 月 8 日）

要　　旨

Nishiyama (2011, *Journal of the Japan Statistical Society*) において提案された，変化点の存在の有無を検定するための統計量についての新しい解釈を述べる．また，同論文においては，対立仮説として変化点が複数個存在するという命題を立てて記述しているので，かえってわかりにくくなっている面もあるので，本ノートにおいて 1 個の場合にどうなるかをはっきり書くことによって，普及のための一助としたい．また，変化点が複数個ある対立仮説の場合についての考察を深め，同論文よりも良い形の対立仮説を提案する．

キーワード： 変化点問題，順位統計量，検定．

1. はじめに

$X_1,\ldots,X_n$ は独立な 1 次元実数値確率変数の列であるとする．これに対する次の仮説検定問題を考える：

H_0: 全ての $q=1,\ldots,n$ について，X_q の分布は同一の連続分布である；

H_1: ある $u_1\in(0,1)$ および連続分布 F_1, F_2 (ただし $\sup_{x\in\mathbb{R}}|F_1(x)-F_2(x)|>0$) が存在して，$q\le u_1 n$ については X_q は F_1 に従い，$q>u_1 n$ については F_2 に従う．

この設定は，ノンパラメトリック変化点問題の中で最も基本的なものである．

この問題のための検定統計量として，Nishiyama (2011) は，$X_1,\ldots,X_n$ の順位統計量 $R_1,\ldots,R_n$ に基づいて

$$D_n=\frac{1}{\sqrt{n}}\max_{1\le i,j\le n}\left|\sum_{q=1}^{i}1\{R_q\le j\}-\frac{ij}{n}\right| \tag{1.1}$$

を提案し，H_0 のもとで，$n\to\infty$ とするとき，D_n が $\sup_{(s,t)\in[0,1]\times[0,1]}|W^\circ(s,t)|$ に分布収束することを示した．ただし W° は平均がゼロで共分散構造が

$$\mathrm{E}[W^\circ(s_1,t_1)W^\circ(s_2,t_2)]=(s_1\wedge s_2-s_1s_2)(t_1\wedge t_2-t_1t_2)$$

であるような正規確率場であり，standard Brownian pillow と呼ばれる．また，H_1 のもとで，$n\to\infty$ とするとき，

$$D_n\ge\sqrt{n}u_1(1-u_1)\sup_{x\in\mathbb{R}}|F_1(x)-F_2(x)|-o(\sqrt{n})-O_P(1) \tag{1.2}$$

† 統計数理研究所：〒190–8562 東京都立川市緑町 10–3

が成り立つことを示した．後者からは，任意の $K>0$ に対し $P(D_n>K)\to 1$ であることが分かるので，この検定は一致性をもつ．

Nishiyama (2011) は，H_1 としては変化点が κ 個（ただし κ は任意の正整数）である場合について考察したが，この節では敢えてわかりやすい $\kappa=1$ の場合の結果を紹介した．$\kappa\geq 2$ に対しても，検定統計量 D_n は修正する必要はないので，当然 H_0 のもとでの極限分布も変更はなく，従って κ が未知であっても棄却域は構成できる．ただし H_1 のもとでの漸近的不等式 (1.2) は一般化した形に変更される．これについては第 3 節で復習する．

ノンパラメトリック変化点問題についての先行研究としては Pettitt (1979)，Lombard (1987) および Murakami (2010) らがあるが，我々が考えている D_n を最初に導入したのは Nishiyama (2011) であった．この統計量は，H_0 のもとで漸近的に分布不変であるばかりでなく，n を固定しても分布不変である（なぜなら，連続分布からのデータに基づく順位統計量は分布不変であるから）という利点があり，また上述したように κ が未知である場合にも使えるので，実用的に重要であると考えられる．本ノートでは，この統計量の新しい解釈の仕方を報告する．

2. 検定統計量の新しい解釈

(1.1) で定義した統計量は

$$D_n=\sqrt{n}\max_{1\leq i,j\leq n}\left|\frac{1}{n}\sum_{q=1}^{n}1\{q\leq i,\ R_q\leq j\}-\frac{ij}{n^2}\right| \tag{2.1}$$

と書き直すことができる．この右辺の絶対値の中の第 1 項は，2 次元のデータ

$$(q,R_q),\quad q=1,\ldots,n \tag{2.2}$$

の経験分布関数である．右辺の絶対値の中の第 2 項は $\frac{i}{n}\times\frac{j}{n}$ であるから，2 次元の格子点 $\{1,\ldots,n\}\times\{1,\ldots,n\}$ 上の一様分布の分布関数であると見なすことができる．(2.1) 式における D_n は Kolmogorov-Smirnov 型の統計量であるが，これはデータ (2.2) の第 1 成分と第 2 成分の独立性の指標となっていると解釈できる．Nishiyama (2011) の原著論文における (1.1) よりも (2.1) の形の方が意味がわかりやすく，記憶しやすいので，普及のためにはこの形で提示しておく方がより適切であろう．

3. 変化点が複数個あるような対立仮説の場合

Nishiyama (2011) は，1 節で考えた対立仮説 H_1 ではなく，実際には次のような，より一般的な対立仮説を考察した（以下，1 節で使った記号 κ に関して言えば $k=\kappa+1$ と見なされたい）：

H_1': ある整数 $k\geq 2$，定数 $0=u_0<u_1<\cdots<u_k=1$，および k 個の線形独立な連続分布 $F_1,\ldots,F_k$ が存在して，$c=1,\ldots,k$ について，$u_{c-1}n<q\leq u_cn$ の範囲にある X_q たちは F_c に従う．

同論文では，この仮説のもとで，次の漸近的不等式が成立することを示している：

$$D_n\geq\sqrt{n}\max_{1\leq c_*<k}\sup_{x\in\mathbb{R}}|B_{c_*}(x)|-o(\sqrt{n})-O_P(1). \tag{3.1}$$

ただし

$$B_{c_*}(x)=\sum_{1\leq c\leq c_*}(u_c-u_{c-1})(1-u_{c_*})F_c(x)-\sum_{c_*<c\leq k}u_{c_*}(u_c-u_{c-1})F_c(x). \tag{3.2}$$

Nishiyama (2011)がこの対立仮説に「$F_1,\ldots,F_k$ が線形独立である」ことを課したのは，その仮定のもとでは，ある c_* (実際には，全ての c_*)に対し B_{c_*} が恒等的にゼロにはならないからであった．しかしながらその仮定は，一旦変化した分布が元に戻る場合(例えば $F_1,F_2,\ldots,F_{k-1},F_1$ の順に分布が変化するような場合)を排除してしまう．そこで，その仮定を次のような形に緩めることを提案する：

H_1'': ある整数 $k\geq k'\geq 2$，定数 $0=u_0<u_1<\cdots<u_k=1$，および k' 個の線形独立な連続分布 $\tilde{F}_1,\ldots,\tilde{F}_{k'}$ が存在して，$c=1,\ldots,k$ について，$u_{c-1}n<q\leq u_cn$ の範囲にある X_q たちの分布 F_c は，$\tilde{F}_1,\ldots,\tilde{F}_{k'}$ のいずれかである．ただし，$F_1=\cdots=F_k$ ではない(つまり，ある $1\leq c<c'\leq k$ が存在して $F_c\neq F_{c'}$)．

この仮定のもとでもやはり(3.1)が成立するが，ある c_* が存在して(3.2)で与えられる B_{c_*} が恒等的にはゼロにならないことがすぐにわかる．実際，$c_*=k-1$ とおき，$F_k=\tilde{F}_l$ によって l を定めると，B_{c_*} を $\tilde{F}_1,\ldots,\tilde{F}_{k'}$ の線形和の形で表した際の係数が正になるような $\tilde{F}_\cdot$ が $\{\tilde{F}_1,\ldots,\tilde{F}_{k'}\}\setminus\{\tilde{F}_l\}$ の中に存在する．よって $\tilde{F}_1,\ldots,\tilde{F}_{k'}$ の線形独立性より，B_{c_*} は恒等的にゼロではない．

謝　辞

このノートの 2 節で新たに紹介した見方は，栗木哲教授からの啓発による．ここに感謝の意を表したい．

参 考 文 献

Lombard, F. (1987). Rank tests for changepoint problems, *Biometrika*, **74**, 615–624.

Murakami, H. (2010). A rank statistic for the change-point problem and its application, *Journal of the Japanese Society of Computational Statistics*, **23**, 27–40.

Nishiyama, Y. (2011). A rank statistic for non-parametric k-sample and change point problems, *Journal of the Japan Statistical Society*, **41**, 67–73.

Pettitt, A.N. (1979). A non-parametric approach to the change-point problem, *Applied Statistics*, **28**, 126–135.

218 Proceedings of the Institute of Statistical Mathematics Vol. 60, No. 1, 215–218 (2012)

Rank Statistic for the Non-parametric Change Point Problem

Yoichi Nishiyama

The Institute of Statistical Mathematics

A new interpretation for the rank statistic for the non-parametric change point problem introduced by Nishiyama (2011, *Journal of the Japan Statistical Society*) is presented. Also, the statement of the alternative is improved.

Key words: The change point problem, rank statisitc, test.