

## 区間データにもとづく一般化ラページ検定 (2 標本問題)

川崎医科大学名誉教授

仮 谷 太 一

(平成12年 8 月29日受理)

### A Generalized Lepage Test Based on Interval Data Samples

**Taichi KARIYA**

*Professor Emeritus,*

*Kawasaki Medical School,*

*Kurashiki, Okayama, 701-0192, Japan*

*(Received on August 29, 2000)*

#### 概 要

コンピュータの急速な発展普及は、ソフトプログラムの利便性と相まって、ときに統計的解析の誤用を引き起こしている。特に医学データでは、普通例数が少なく、データの分布に関する情報がよく分からない場合が多い。ラページ検定は、こうした状況に対応して、母集団分布について事前情報がない場合、2 標本の間に差があるかないか、中央値またはチラバリなどの間に差があるかないか、を検定するもので、優れた検定法であることが知られている。

ところで、医学データには、手術後再発した時期とか、特定乳歯の萌出日時とか、眼底出血の時点など、正確な生起時点を知ることが困難なものが少なくない。このようなときには、事象の生起した直前と直後の検診日時から定まる区間内に、それが生起したとするのが最も信頼できるデータと言えよう。また、われわれの血圧を1分おきに数回測定してみると、かなり大きく変動していることが分かる。このような場合には、個体の特性値を単一の値で表示するよりも、その値を内に含む区間データとして表示するのが妥当のように考えられる。

この小論では、独立な2 標本の各個体特性値が、それぞれ区間データで与えられている場合に、ラページ検定を一般化した検定法を提案するものである。

#### Abstract

Misapplication of statistical significance tests has been found here and there with the rapid computalization. In medical investigations, the number of data is not usually large and the information of populations is often unknown. Under these circumstances, it is well known that the Lepage test is a good two-sample test for testing the differences between medians or variances.

Now, the observed values cannot be often obtained exactly in the multistage follow-up examination about a period of time from a surgical operation to relapse of the disease, an eruptive age of a specified deciduous tooth, or a point of time when cerebral hemorrhage in one's eyes and so on. And, for example, maximum blood

pressure of any person varies every moment ; therefore, successively observed values of the same person will be different considerably every moment.

Consequently, it is preferable that the observation in these cases should be specified by interval data.

In this paper, a distribution-free two-sample test is proposed that is an extension of the Lepage test to interval data samples.

## 1. 序

$(X_1, X_2, \dots, X_m), (Y_1, Y_2, \dots, Y_n)$  を連続な分布関数  $F(x), G(y)$  をもつ母集団からの、それぞれ大きさ  $m, n$  の独立な無作為標本とする。問題は  $\{X_i\}, \{Y_j\}$  の観測値にもとづいて、対立仮説  $H_a : G(X) = F(ax+b)$ , ただし  $a \neq 1$  ( $a > 0$ ) または  $b \neq 0$ , に対し、帰無仮説  $H_0 : G(X) = F(X)$  を検定することである。

ラページ検定は、 $\{X_i\}, \{Y_j\}$  の観測値が通常の観測値である場合の話で、その検定における統計量  $T$  は、Wilcoxon 統計量 ( $W$ -統計量) と Ansari-Bradley 統計量 ( $AB$ -統計量) の関数である。

周知のように、 $W$ -統計量は、その統計量にもとづく両側検定が、 $G(x) = F(x+b)$  ( $b \neq 0$ ) 型の対立仮説に対する帰無仮説  $H_0 : G(x) = F(x)$  の検定に関し Consistent である。また、 $AB$ -統計量は、その統計量にもとづく両側検定が、 $G(x-\theta) = F(a(x-\theta))$ , ( $a > 0, a \neq 1$ ) 型の対立仮説に対する帰無仮説  $H_0 : G(x) = F(x)$  の検定に関し Consistent である。ここに  $\theta$  は共通の中央値とする。

これらは共に、優れた検定であるが、上述の対立仮説から明らかなように、 $W$ -検定は、両母集団分布の分散が等しいことが分かっている場合の検定であり、 $AB$ -検定は、両母集団分布の中央値が等しいことが分かっている場合の検定である。ところが、両母集団分布について、分散が等しいとか、中央値が等しいとかの情報はよく分かっていないのが通例である。

そこで、ラページは、このような母集団分布についての事前情報がない場合にも威力を発揮する統計量  $T$  を

$$T = \frac{(W - E(W))^2}{V(W)} + \frac{(AB - E(AB))^2}{V(AB)}$$

と定義し、この  $T$  が大きいとき、 $G(x) = F(ax+b)$  ( $a > 0$ ) 型の対立仮説  $H_a$  に対し、帰無仮説  $H_0 : G(x) = F(x)$  を棄却するというラページ検定を提案した。 $W$ -統計量と、 $AB$ -統計量とは無相関であるから、統計量  $T$  は、 $m, n$  が大きいとき、 $H_0$  の下で、自由度 2 の  $\chi^2$  分布で近似することができる。また、統計量  $T$  は、分布の中央値の違い、分散の違いに関する広い範囲の対立仮説  $H_a$  に対し、優れた検出力を持つことが知られている。

本論文では、 $\{X_i\}, \{Y_j\}$  の観測値が区間データで与えられている場合に、ラページ検定の一般化を行っている。

## 2. 区間データにもとづく一般化ラページ検定

2 標本  $(X_1, X_2, \dots, X_m)$ ,  $(Y_1, Y_2, \dots, Y_n)$  を、連続な分布関数  $F(x)$ ,  $G(y)$  をもつ母集団からの、それぞれ大きさ  $m$ ,  $n$  の独立な無作為標本とし、しかも  $X_i$ ,  $Y_j$  の実現値は、それぞれ、区間データ  $[x_{iL}, x_{iU}]$ ,  $[y_{jL}, y_{jU}]$  によって特性づけられているとする。すなわち、 $X_i$  の実現値は区間  $[x_{iL}, x_{iU}]$  の 1 要素、 $Y_j$  の実現値は区間  $[y_{jL}, y_{jU}]$  の 1 要素であることだけが知られているとする。ここで、 $m$ ,  $n$  は 2 より大きい正の整数、 $x_{iL}$ ,  $x_{iU}$ ,  $y_{jL}$ ,  $y_{jU}$  は有限な実数とする。

分布関数	標本の大きさ	区間データ
$F(x)$	$m$	$[x_{1L}, x_{1U}], [x_{2L}, x_{2U}], \dots, [x_{mL}, x_{mU}]$
$G(y)$	$n$	$[y_{1L}, y_{1U}], [y_{2L}, y_{2U}], \dots, [y_{nL}, y_{nU}]$

問題は、対立仮説  $H_a: G(x)=F(ax+b)$ ,  $a \neq 1$  または  $b \neq 0$  ( $a > 0$ ) に対し、帰無仮説  $H_0: G(x)=F(x)$  を検定することである。

提案する統計量  $GT$  は、ラページ統計量  $T^{(1)}$  にならい、次のように定義する。

まず、区間データ  $[x_{iL}, x_{iU}]$ ,  $[y_{jL}, y_{jU}]$  にもとづく確率変数  $X_i - Y_j$  の一般化符号  $U_{ij}^{(2,3,4,5)}$  を次のように定義する。

$$U_{ij} = E(\text{sgn}(x_i - y_j)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{sgn}(x - y) f_i(x; x_{iL}, x_{iU}) g_j(y; y_{jL}, y_{jU}) dx dy$$

ここに、 $f_i(x; x_{iL}, x_{iU})$ ,  $g_j(y; y_{jL}, y_{jU})$  はそれぞれ  $X_i$ ,  $Y_j$  の確率密度関数であり、また、

$$\text{sgn}(x - y) = \begin{cases} 1 & \text{もし } x > y \text{ ならば} \\ 0 & \text{もし } x = y \text{ ならば} \\ -1 & \text{もし } x < y \text{ ならば} \end{cases}$$

である。

実際の積分計算においては

$$x < x_{iL} \text{ または } x > x_{iU} \text{ のとき } f_i(x; x_{iL}, x_{iU}) = 0, \int_{x_{iL}}^{x_{iU}} f_i(x; x_{iL}, x_{iU}) dx = 1$$

$$y < y_{jL} \text{ または } y > y_{jU} \text{ のとき } g_j(y; y_{jL}, y_{jU}) = 0, \int_{y_{jL}}^{y_{jU}} g_j(y; y_{jL}, y_{jU}) dy = 1$$

とする。

$$U_{ij} = \begin{cases} -1 & (x_{iU} < y_{jL} \text{ のとき}) \\ +1 & (y_{jU} < x_{iL} \text{ のとき}) \\ -1 \text{ と } +1 \text{ の間の実数} & (\text{その他のとき}) \end{cases} \quad (1)$$

である。

$U_{ij} > 0$  ならば、 $X_i$  が  $Y_j$  より大きい確率、

$U_{ij} < 0$  ならば、 $Y_j$  が  $X_i$  より大きい確率と考えてもよい。

つぎに、 $W$ —統計量、 $AB$ —統計量に対応する一般化統計量  $GW^{(4)}$ ,  $GAB$  をそれぞれ

$$GW = \sum_{i \in (X)}^m (N + 1 + U_{i.}) / 2, \quad (2)$$

$$GAB = \sum_{i \in (X)}^m (N+1 - |U_{i.}|) / 2 \quad (3)$$

により定義する。ここに  $N=m+n$ ,  $U_{i.} = \sum_{j=1}^N (j \neq i) U_{ij} = \sum_{j=1}^N U_{ij}$  ( $\because U_{ii}=0$ ) である。また  $\sum_{i \in (X)}^m$  は  $X$  群に属する  $m$  個の  $i$  について和をとることを意味する。

かくして、区間データにもとづく一般化ラページ統計量  $GT$  を、次式により定義する。

$$GT = \frac{(GW - E(GW))^2}{V(GW)} + \frac{(GAB - E(GAB))^2}{V(GAB)} \quad (4)$$

ここで、 $E(GW)$ ,  $E(GAB)$  は、 $N=m+n$  個のデータを合併し、 $X$ ,  $Y$  群へのすべての可能な分割を考えると、帰無仮説  $H_0$  の下で得られる、それぞれ  $\binom{N}{m}$  個の  $GW$ ,  $GAB$  の平均である。また、 $V(GW)$ ,  $V(GAB)$  は、それらの分散を表す。 $E(GW)$  は、後で述べる  $E(GW|P, H_0)$  を略記したもので、他も同様である。

この  $GT$  が、大きいとき、帰無仮説  $H_0$  を対立仮説  $H_a$  に対して棄却する。そして、この統計量  $GT$  は、区間データの間に重なりがなく、普通の順序づけが可能で、同順位がない場合には、本来のラページ統計量  $T^{(1)}$  に一致する。

また、 $m$ ,  $n$  が大きいとき、 $GT$  は帰無仮説  $H_0$  の下で自由度 2 の  $\chi^2$  の分布で近似できる。

### 3. 帰無仮説 $H_0$ の下における $GT$ -統計量の精密分布

帰無仮説  $H_0$  の下では、 $(X_1, X_2, \dots, X_m)$ ,  $(Y_1, Y_2, \dots, Y_n)$  は、同じ母集団からの独立な標本であるから、合併して、大きさ  $N$  の独立標本  $(XY_1, XY_2, \dots, XY_N)$ ,  $(N=m+n)$  として考察する。このとき、前述した  $U_{ij}$  は、確率変数  $XY_i - XY_j$  の一般化符号を表し、 $U_{ij} = -U_{ji}$ ,  $U_{ii}=0$  である。さて、 $(XY_1, XY_2, \dots, XY_N)$  を、それぞれ大きさ  $m$ ,  $n$  の  $X$  群,  $Y$  群に分割する方法は、 $\binom{N}{m}$  通りあり、 $H_0$  の下ではすべて同じ生起確率をもつ。従って、我々は統計量  $GT$  の、 $H_0$  の下における精密分布を求めることが出来る。このとき、観測された 2 区間データ標本の有意確率 ( $P$ -値) は次式で与えられる。

$$P\text{-値} = (GT \text{ の値が } GT_0 \text{ より小さくない分割の数}) / \binom{N}{m} \quad (5)$$

この  $P$ -値が大きいとき、対立仮説  $H_a$  に対し帰無仮説  $H_0$  を棄却することになる。ここに、 $GT_0$  は観測された 2 区間データ標本の  $GT$  の値を表す。

次に、 $H_0$  の下で合併標本に関する  $GW$ ,  $GAB$  の期待値、分散をそれぞれ  $E(GW|P, H_0)$ ,  $V(GW|P, H_0)$ ,  $E(GAB|P, H_0)$ ,  $V(GAB|P, H_0)$  とする。ここに  $P$  は、観測された区間データのパターンを表す。

期待値および分散は、観測されたデータ・パターン  $P$  に帰着する全部で  $\binom{N}{m}$  通りの等確率標本にわたって計算される。次の式が成り立つ。

$$E(GW|P, H_0) = m(N+1)/2,$$

$$E(GAB|P, H_0) = m(N+1)/2 - (m/N) \sum_{i=1}^N |U_{i.}|/2,$$

$$V(GW|P, H_0) = \left( \frac{m \cdot n}{N(N-1)} \right) \sum_{i=1}^N \left( \frac{U_{i.}}{2} \right)^2,$$

$$V(\text{GAB}|\text{P}, \text{H}_0) = \left( \frac{m \cdot n}{N(N-1)} \right) \left\{ \sum_{i=1}^N \left| \frac{U_{i\cdot}}{2} \right|^2 - \left( \frac{1}{N} \right) \cdot \left( \sum_{i=1}^N \left| \frac{U_{i\cdot}}{2} \right| \right)^2 \right\}$$

#### 4. 数値例

この節では、一般化ラページ検定 (GT 検定) を、男女学生の最高血圧に関する標本に適用する。各人の血圧測定は、自動計測装置により連続して 5 回行われた。被験者はほぼ 20 歳の男女学生で、クラスは 30 人の男子学生と 22 人の女子学生で構成されていた。

最高血圧値の性差が主題である。一般に男子の方が女子より高いように考えられているが、両母集団の分散が等しいという確証を得ていないので、直ちにウィルコクソン検定を適用するのは妥当とはいえない。そこで、母集団分布に関する仮定を一切必要としない GT 検定を適用することにした。

各人の区間データは、5 回の測定値の [最小値, 最大値] をもって構成し、一般化符号  $U_{ij}$  の計算では、 $f_i(x)$ ,  $g_j(y)$  にそれぞれ一様分布を仮定した。性別の区間データ標本は、普通例数の少ない医学データを想定して、それぞれ 30, 22 のデータから無作為に 8 個ずつを選んだ (2 組)<sup>4)</sup>。

##### データ (I)

男子： [115, 124], [106, 114], [108, 112], [117, 128],  
 [100, 119], [118, 134], [124, 130], [116, 131]  
 女子： [105, 112], [110, 113], [110, 116], [ 90, 108],  
 [ 99, 108], [112, 127], [108, 142], [103, 106]

##### データ (II)

男子： [111, 118], [107, 110], [117, 130], [114, 120],  
 [121, 130], [116, 122], [106, 114], [ 99, 108]  
 女子： [ 92, 101], [ 84, 95], [ 98, 102], [ 99, 109],  
 [110, 113], [ 90, 101], [104, 118], [105, 112]

検定結果は次の通りである。

表 1 P-値

組	標本の大きさ	P-値	GW-検定による P-値
(I)	8, 8	0.1927	0.0341*
(II)	8, 8	0.0242*	0.0033**
全	30, 22	(0.0005)***	(0.0003)***

\* : 5% 有意, \*\* : 1% 有意, \*\*\* : 0.1% 有意

P-値の ( ) は近似値であることを示す。

参考値として、GW-検定 (一般化ウィルコクソン検定)<sup>4)</sup> による P-値も示した。

表 1 から, 母集団分布に何も仮定しない GT-検定によると, データ (I) では帰無仮説  $H_0$  を棄却することはできず, 性差があるとは言えないということであり, データ (II) では, 危険率 5% で  $H_0$  を棄却できたので, 一応性差があるということになる。こうした食い違いは標本サイズが小さいために起こることで, 標本サイズを大きくするか, さらに研究を重ねなければならない。ここでは, 30, 22 のデータを全部用いれば,  $H_0$  は 0.1% の危険率で棄却でき, 性差ありということが出来る。

なお, 分散については男女差はないとして検定を行う GW-検定では, データ (I) では, 危険率 5% で有意, データ (II) では, 危険率 1% で有意となり, 中央値に性差が認められた。GT-検定にくらべて GW-検定が, 検出力において優るようにはみえるが, これは分散については男女差はないという (ほぼ真に近い) 情報を用いたのであるから当然のことと言えよう。

図 1 に, データ (II) のデータの図表示と, GT 値のヒストグラムおよび GT 値の分布を近似する  $\chi^2$  分布曲線 (自由度 2) を示した。近似はかなり良好であることが分かる。

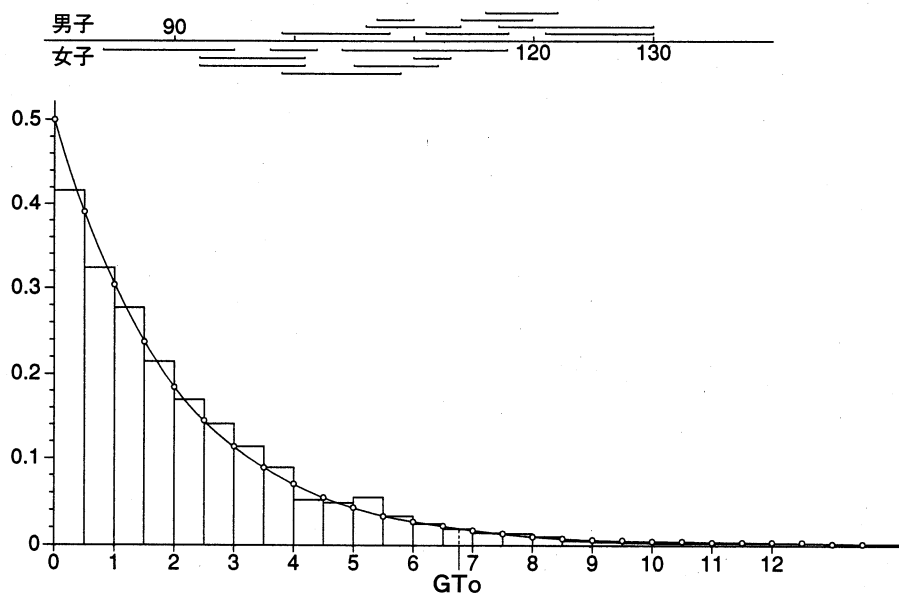


図 1 データ (II) の区間データと GT 値のヒストグラムおよび自由度 2 の  $\chi^2$  曲線

### References

- 1) Lepage Y: A combination of Wilcoxon's and Ansari-Bradley's statistics. *Biometrika* 58: 213-217, 1971
- 2) Kariya T: A generalized sign test based on paired interval data. *The Second Japan-China Symposium on Statistics*: 125-128, 1986
- 3) 仮谷太一: 対応のある 2 標本の区間データに基づく一般化符号検定. *応用統計学* 16: 77-88, 1987
- 4) Kariya T: A generalized Wilcoxon test for comparing interval data samples. *Kawasaki Med J* 14: 187-192, 1988
- 5) Kariya T: A bivariate permutation test for analysis of three interval data samples. *Kawasaki Med J* 18: 25-30, 1992