順序統計量による検定の一方法

中　島　達　夫

1　ま　え　が　き

2つの標本群があるとき，その両群が同一の母集団より得られたものと考えるべきか，それとも別の母集団より得られたもので，両者には有意差があると判断すべきか，これは統計学の基本問題の一つであり，また応用的価値の高いものであろう．

このような検定に対しては t-，F-，χ^2-分布等を利用する方法とか，コルモゴロフ・スミルノフの方法など各種のものが用いられるが，ここでは，母集団の分布が不明である場合に，「両群の母集団は同一である」という仮設を，順序統計量を用いて簡単に検定する方法について検討してみる．

この方法は，同数の標本よりなる2群において，各群の標本を数値の小さいものから順に並べ，両群の同順位標本値を比較し，一方の標本群が他方の標本群より小さいもの（あるいは大きいもの）がいくつあるかによって検定しようとするものである．

一般に母集団分布がいかなるものであれ，確率変数の変換により確率密度を［0～1］間の一様分布に帰することができる．したがって順序統計量を用いる下記の方法では母分布として一様分布を採用したが，それは一般性を損なうことなく解析を簡単化するのに有益であろう．

2　理　論　分　布

［0～1］の間に一様に分布する母集団より大きさ n 個の標本を2組ランダムにとり出し，それを G_1，G_2とする．G_1，G_2の各 n 個のデータを大きさの順に

$$G_1 : x_{11} \leqq x_{12} \leqq \cdots\cdots\cdots\cdots\cdots \leqq x_{1n}$$

$$G_2 : x_{21} \leqq x_{22} \leqq \cdots\cdots\cdots\cdots\cdots \leqq x_{2n}$$

とする．両組の同順位のもの x_{1i}と x_{2i}（$i=1,\ 2,\ \cdots,\ n$）を比較し，G_1が G_2より小さいものの数が m である確率 $p(n,\ m)$ の理論値は，m に関係なく $1/(n+1)$ であることが次のように導かれる．

まず $m=0$ である確率 $p(n,\ 0)$ は次のようにして得られる．

$$p(n,\ 0)=(n!)^2\int_0^1 dx_{1n}\int_0^{x_{1n}} dx_{1,n-1}\cdots\int_0^{x_{13}} dx_{12}\int_0^{x_{12}} dx_{11}\int_{x_{1n}}^1 dx_{2n}\int_{x_{1,n-1}}^{x_{2n}} dx_{2,n-1}\cdots\int_{x_{12}}^{x_{23}} dx_{22}\int_{x_{11}}^{x_{22}} dx_{21}$$

$$=(n!)^2\int_0^1 dx_{1n}\int_{x_{1n}}^1 dx_{2n}\int_0^{x_{1n}} dx_{1,n-1}\int_{x_{1,n-1}}^{x_{2n}} dx_{2,n-1}\cdots\int_0^{x_{13}} dx_{12}\int_{x_{12}}^{x_{23}} dx_{22}\int_0^{x_{12}} dx_{11}\int_{x_{11}}^{x_{22}} dx_{21}$$

$$=(n!)^2\int_0^1 dx_{1n}\int_{x_{1n}}^1 \frac{1}{(n-2)!(n-1)!}\left(\frac{1}{n-1}x_{1n}^{n-1}x_{2n}^{n-1}-\frac{1}{n}x_{1n}^{n}x_{2n}^{n-2}\right)dx_{2n}$$

$$=(n!)^2\int_0^1 \frac{1}{(n-1)!n!}(x_{1n}^{n-1}-x_{1n}^{n})\,dx_{1n}$$

$$=\frac{1}{n+1}$$

次に $0\leqq m\leqq n$ の確率 $p(n,\ m)$ は次のように求められる．

$$p(n,\ m)=\binom{n}{m}\left[\frac{1}{n-m+1}-\frac{m!}{(n+1)!}\sum_{i=0}^{m-1}\frac{(n-i)!}{(m-i)!}\right]$$

$$=\binom{n}{m}\left[\frac{1}{n-m+1}-\frac{m!}{(n+1)!}\left\{\frac{(n+1)!}{m!(n-m+1)}-(n-m)!\right\}\right]$$

$$=\frac{1}{n+1}$$

3　シミュレーション

［０～１］の一様乱数より n 個の標本を２組採取し，その一方が同順位の他の組のデータより小さいものの数 m を調べた．

$n=2$，３，４，６，８，10の各ケースについて，それぞれ $N=1000$ 回の繰り返し実測を行い，得られた m の値の分布を図にプロットした．なお図には比較のため

理論分布 $1/(n+1)$

を点線で併記した．

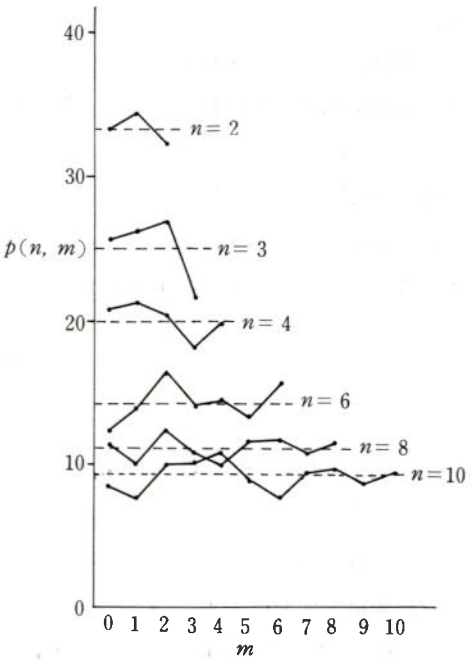

図　シミュレーションの結果と理論分布