統計学入門

安田正實

2008 年 11 月 24 日

目次

0.1 ノンパラメトリックな方法 2

図目次

表目次

0.1 ノンパラメトリックな方法

これまでの述べてきた母数 μ, σ^2 の推定および検定では母集団の分布型が既知であることを仮定し多 (普通は正規分布)。分布型が未知の場合や既知であってもその標本理論が確立していない場合には、これまでの方法に変わる統計的方法が必要になる。ここではノンパラメトリックな方法 (または母数によらない方法) とよばれ、いろいろな場面で実際に使われる標準的方法を述べる。

基本事項　連続型確率密度関数 $f(x)$ をもつ母集団から抽出した大きさ n の標本を $x_1, x_2, \cdots, x_n$ とする。これら n 個の標本を小さいものから大きいものへと順に並べ替えたものを

$$x_{(1)} < x_{(2)} < \cdots < x_{(n-1)} < x_{(n)}$$

を大きさ n の順序統計量という。

0.1.1 中央値の信頼区間

母集団中央値は $\int_{\xi}^{\infty} f(x)dx = 1/2$ を満たす ξ を分布の中央値として定義される。したがって一つの観測値が ξ より大きい確率、また小さい確率はともに $1/2$ である。いま

$$x_{(i)} < \xi < x_{(n-i+1)}, i = 1, 2, \cdots, n-1, i < n/2$$

を考えると、この事象は n 個の観測値のうち少なくとも i 個が ξ より小さく、少なくとも i 個が ξ より大きいときに起っている。いいかえると n 個のうち少なくと i 個が ξ より小で、高々 $n-i$ 個がより大きい。

$$x_{(1)} < \cdots < x_{(i)} < \cdots < \xi < \cdots < x_{(n-i+1)} < x_{(n-i+2)} < \cdots < x_{(n)}$$

したがってこの事象の確率は、コインの n 回投げにおいて表が少なくとも i から高々 $n-i$ 回起こる確率と同じである。それゆえこれを観測値が中央値を含む確率、つまり ξ の信頼区間と考えれば、その信頼係数 C は 2 項分布の計算から

$$\begin{aligned} c &= P(x_{(i)} < \xi < x_{(n-i+1)}) \\ &= \sum_{k=i}^{n-i} \binom{n}{k} (1/2)^k \\ &= 1 - 2\sum_{k=0}^{i-1} \binom{n}{k} (1/2)^k \qquad (1 < i < n/2). \end{aligned}$$

たとえば

信頼係数	ξ の信頼区間
$1 - 2(1/2)^n$	$(x_{(1)}, x_{(n)})$
$1 - 2(1/2)^n\{1 + n\}$	$(x_{(2)}, x_{(n-1)})$
$1 - 2(1/2)^n\{1 + n + n(n-1)/2\}$	$(x_{(3)}, x_{(n-2)})$
$1 - 2(1/2)^n\{1 + n + n(n-1)/2 + n(n-1)(n-2)/6\}$	$(x_{(4)}, x_{(n-3)})$

0.1.2 中央値の符号検定

中央値の符号検定は、正規分布の平均値における t 検定に比較できるもので最もその分布によらない性質が表れている。与えられた値 ξ_0 として大きさ n の標本 $x_1, x_2, \cdots, x_n$

(1) 片側検定 ( )　帰無仮説　$H_0 : \xi = \xi_0$
　　対立仮説　$H_1 : \xi < \xi_0$

(2) 片側検定　帰無仮説　$H_0 : \xi = \xi_0$
　　対立仮説　$H_1 : \xi > \xi_0$

(3) 両側検定　帰無仮説　$H_0 : \xi = \xi_0$
　　対立仮説　$H_1 : \xi \neq \xi_0$

の検定を考える。検定統計量としては標本値 x_i と ξ_0 の差を計算し、その正負の符号数に注目する。$x_i - \xi_0 (i = 1, 2, \cdots, n)$ の中で正となるものの数を X、負のものは、$n - X$ である。帰無仮説のもとでは X はパラメータ $n, p = 1/2$ の 2 項分布にしたがう。これから検定の棄却域を定められる。すなわち (1) の場合には、2 項分布の左裾を、(2) の場合には右裾を棄却域に選べ良い。しかし 2 項分布の確率を厳密に求めることは一般には面倒であるから、個数が大きいときには、正規分布で近似する。とくに確率が $1/2$ のときの計算であるから n がそう大きくなくても比較的よい近似が得られる。棄却域は

(1) $X < n/2 - z_{2\alpha}\sqrt{n}/2$

(2) $X > n/2 + z_{2\alpha}\sqrt{n}/2$

(3) $X < n/2 - z_{\alpha}\sqrt{n}/2, X > n/2 + z_{\alpha}\sqrt{n}/2$

とする。ただし z_α は有意水準 α に対する標準正規分布のパーセント点。

また符号検定は標本データ値が対応のある場合にも適用可能である。n 対の組みデータとして与えられた標本を $(x_i, y_i), i = 1, 2, \cdots, n$ とするとき、差 $d_i = x_i - y_i$ が正となる対の数に対して、上記の符号検定を行えばよい。

0.1.3 Wilcoxson の順位和検定

独立な 2 標本に対する検定に用いられる。$x_1, x_2, \cdots, x_n,\ y_1, y_2, \cdots, y_n$ をそれぞれ連続型密度関数 $f(x), g(x)$ をもつ母集団から大きさ m, n の独立な 2 標本で、$m \leq n$ とする。これら 2 組の標本値を各組ごとに大きさの順に並び替えたもの、すなわち各組の順序統計量を

m 個： $x_{(1)} < x_{(2)} < \cdots < x_{(m)}$

n 個： $y_{(1)} < y_{(2)} < \cdots < y_{(n)}$

さらにこれらを合併したもの, たとえば $m = 3, n = 4$ 個

$$\text{合併した：} y_{(1)} < x_{(1)} < y_{(2)} < y_{(3)} < y_{(4)} < x_{(2)} < x_{(3)}$$

を考える。

このときつぎの仮説 (1) あるいは (2) を

(1) 帰無仮説 H_0 ： $f(x) = g(x)$
　　対立仮説 H_1 ： $f(x)$ は $g(x)$ の右側に位置する

(2) 帰無仮説 H_0 ： $f(x) = g(x)$
　　対立仮説 H_1 ： $f(x)$ は $g(x)$ の左側に位置する

としたとき、合併した順位数から求める和から検定する方法を Wilcoxson の順位和検定という。

もし 2 データの合併順序が上記のようであるならば、

連番	1	2	3	4	5	6	7
合併順序	$y_{(1)}$	$x_{(1)}$	$y_{(2)}$	$y_{(3)}$	$y_{(4)}$	$x_{(2)}$	$x_{(3)}$
$f(x)$ の対応順位	-	2	-	-	-	6	7
順位和 $T =$	$0 +$	$2 +$	$0 +$	$0 +$	$0+$	$6+$	$7 = 15$