数理解析研究所講究録
第150巻 1972年 37-48

確率的大小とその応用

統計数理研　柳本　武美
日本IBM　渋谷　政昭

確率的大小の概念はノンパラメトリック検定の理論においてきわめて基本的であり、単調尤度比，正（負）の相関性，increasing hazard rate，対称性の検定（一変数および二変数）と密接な関係をもっている。

S を $\mathbb{R}^2$ の半平面 $\mathbb{R}_x=\{(x,y);\ x>y\}$ のボレル部分集合，${}^tS$ を直線 $x=y$ に関して S に対称な集合とする。$\mathbb{R}_x$ のボレル集合のある族 $\mathcal{R}=\{S,\ S\subset\mathbb{R}_x\}$ が与えられたとき，確率ベクトル（X, Y）が

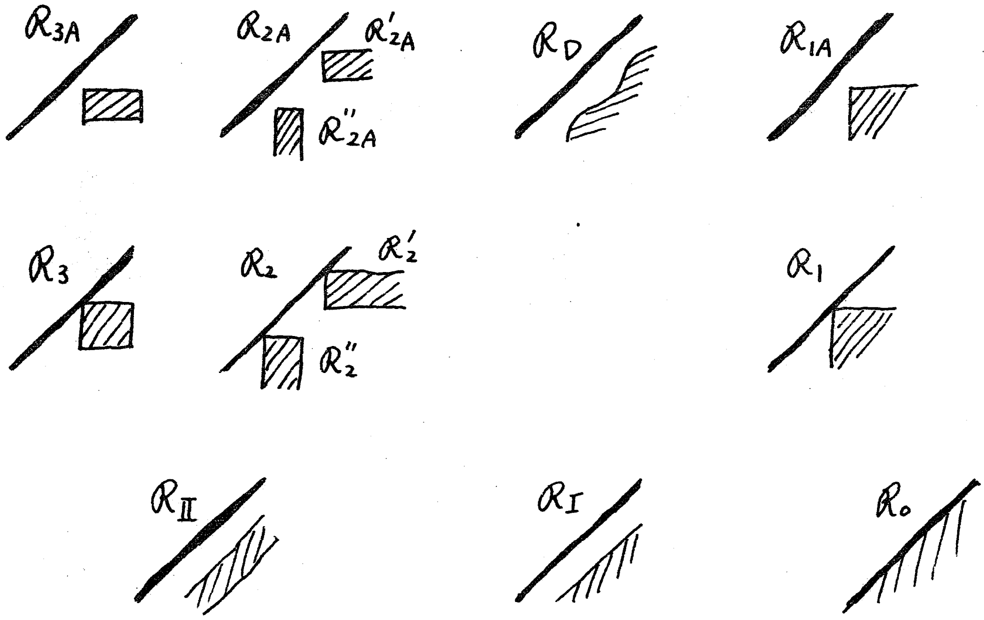
$$P\{(X,Y)\in S\}\ge P\{(X,Y)\in {}^tS\} \quad \text{for all } S\in\mathcal{R}$$

を満たす時，X が Y よりも $\mathcal{R}$ の意味で確率的に大きいと言い，

$$X \succ Y \quad (\mathcal{R})$$

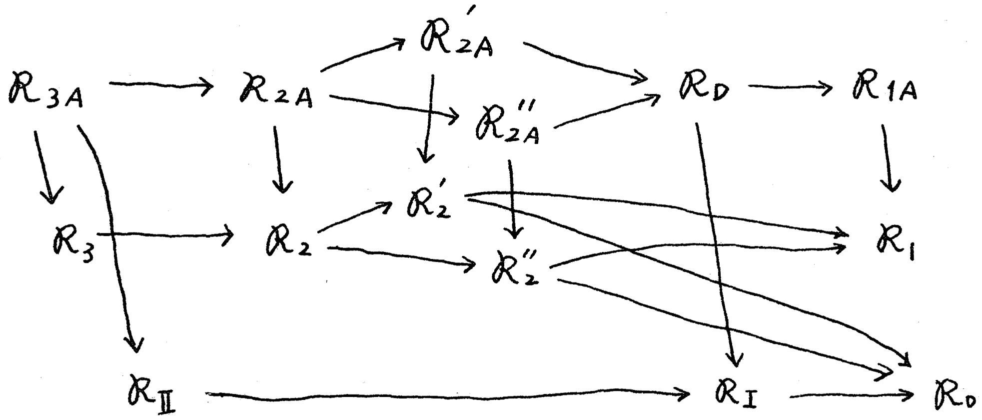
と書く。

$\mathcal{R}$ として意味のあるのは，次の図で典型的な集合 S が表わされるような族である。斜の太線は，$x=y$ である。



ただし$\mathcal{R}_D$は　$x<x'$, $y>y'$のとき　$(x,y)\in S$なら$(x',y')\in S$であるようなボレル集合Sの族である。　また、
$\mathcal{R}_{2A}=\mathcal{R}'_{2A}\cup\mathcal{R}''_{2A}$,　$\mathcal{R}_2=\mathcal{R}'_2\cup\mathcal{R}''_2$　である。

このような集合族で定義される確率的大小について次のような包含関係が成立する。　また矢印のない集合族間では、包含関係は成り立たない。



包含関係の不成立を確かめることは，比較的少数の反例でできる。たとえば右のような9点上の分布を考えると $X \succ Y$ ($\mathcal{R}_3$) だが，$X \not\succ Y$ ($\mathcal{R}_I$) であり，これから直ちに $X \succ Y$ ($\mathcal{R}_3, \mathcal{R}_2, \mathcal{R}'_2, \mathcal{R}''_2, \mathcal{R}_1, \mathcal{R}_0$) のいずれも $X \succ Y$ ($\mathcal{R}_{3A}, \mathcal{R}_{2A}, \mathcal{R}'_{2A}, \mathcal{R}''_{2A}, \mathcal{R}_D, \mathcal{R}_{II}, \mathcal{R}_I$) を含まぬことが言える。

$y \backslash x$	1	2	3
1	2/9	1/9	0
2	1/9	0	2/9
3	0	2/9	1/9

$X \perp\!\!\!\perp Y$ のときは，上の諸定義のいくつかが同等となり，図式はずっと簡単となる。

$$
\begin{tikzcd}
\mathcal{R}_{3A} \arrow[d, equal] & \mathcal{R}_{2A} \arrow[d, equal] \arrow[r] & \mathcal{R}'_{2A}=\mathcal{R}'_2 \arrow[r] & \mathcal{R}_D=\mathcal{R}_{1A} \arrow[d, equal] & \\
\mathcal{R}_3 \arrow[r] \arrow[drr] & \mathcal{R}_2 \arrow[r] & \mathcal{R}''_{2A}=\mathcal{R}''_2 \arrow[ur] & \mathcal{R}_1 \arrow[r] \arrow[d] & \mathcal{R}_E \\
 & & \mathcal{R}_{II} \arrow[r] & \mathcal{R}_I \arrow[r] & \mathcal{R}_0
\end{tikzcd}
$$

$\mathcal{R}_3$ は単調尤度比，$\mathcal{R}_1$ は「確率的に大」，$\mathcal{R}_2$ は Pfanzagl が導入した概念である。上で新しく導入した $\mathcal{R}_E$ は X，Y の分布関数をそれぞれ $G(x)$，$H(y)$ とするとき，$\int (H(t)-G(t))\,dt \geq 0$ で $X \succ Y$ ($\mathcal{R}_E$) を定義する。ただし，積分は $\pm\infty$ を許すものとし，$+\infty$ は正，$-\infty$ は負とみなす。

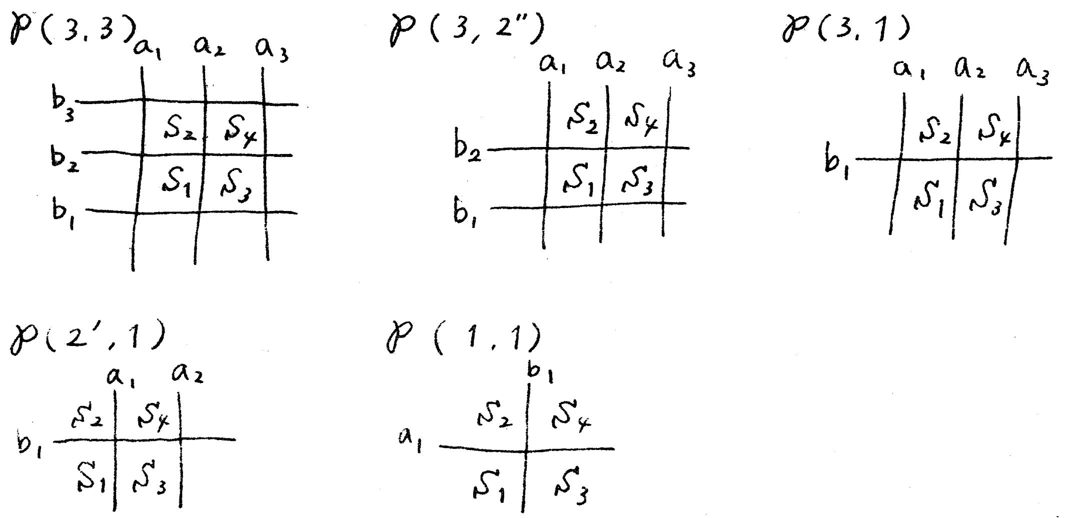
2次元確率分布が正の相関性 $\mathcal{P}(i,j)$ ~~をもつとは~~ をもつとは，図のような任意の $a_1 < a_2 < a_3$; $b_1 < b_2 < b_3$ によって作られる

2次元区間 S_1, S_2, S_3, S_4 に対して常に

$$P(S_1)\cdot P(S_4) \geq P(S_2)\cdot P(S_3)$$

が成り立つことである。逆の不等号が常に成り立つならば負の相関性 $N(i,j)$ と呼ぶ。

図は必ずしもすべての場合を含まない。それは $P(3,2'')$ など ″ で正の方向への無限の2次元区間を含むこと，$P(2',1)$ など ′ で負の方向への無限区間を含むことを示し，また $P(i,j)$ で $P(i,j')$ および $P(i,j'')$ を意味する便法をとっている。



包含関係も $P(i,j)$ が $P(i,j')$，$P(i,j'')$ 等を含み，これらがより弱い関係を含むことになるが それを省略している。

$$\begin{tikzcd}
P(3,3) \arrow[r] & P(3,2) \arrow[r] \arrow[dr] & P(3,1) \arrow[rr] \arrow[dr] & & P(3,E) \arrow[ddd] & \\
& & P(2,2) \arrow[r] & P(2,1) \arrow[d] & & P(1,0) \\
& & & P(1,1) \arrow[urr] \arrow[dr] & & \\
& & & & P(1,E) \arrow[d] & \\
& & & & P(E) &
\end{tikzcd}$$

$P(3,3)$ は positively likelihood ratio dependent とよばれているものである。 $P(3,1)$ は positively regression dependent とよばれているもの, $P(1,1)$ は positively quadrant dependent とよばれているものである。このような定義はまた条件付分布の確率的大小によって定義することもできる。 たとえば $P(3,1)$ は

$$Y|_{X=x_2} > Y|_{X=x_1} \quad (R_1) \qquad \forall x_2 > x_1$$

$$Y|_{X>x} > Y|_{X\leq x} \quad (R_3) \qquad \forall x$$

と同等である。 $P(3,E)$ などは

$$Y|_{X=x_2} > Y|_{X=x_1} \quad (R_E) \qquad \forall x_2 > x_1$$

で定義される。 $P(E)$ は

$$\int \{F(x,y)-F(x,\infty)F(\infty,y)\}\,dx\,dy > 0$$

で定義される。 積分は $\pm\infty$ を許す。

上の包含関係により, 正の相関性についてのいくつかの命題はより簡単に, より本質の明確な形で証明することができる。 たとえば $X_{(1)} \leq \cdots\cdots \leq X_{(n)}$ を任意の確率分布から

の順序統計量とすると　$(X_{(r)}, X_{(s)}) \in \mathcal{P}(3,3)$，したがって相関係数が正である，というもっとも弱い意味を含めた種々の正相関性をもつ。

一変量対称性の検定

一変量の対称性の検定は R^1 上の確率変数 X，その分布関数 $F(x)$，が点 μ に関して対称であるか否かを扱う。以下一般性を失う事なしに，$\mu = 0$ とする。

対立仮説として分布が"正に偏よっている"を採用する。$\mathcal{R}$ を一つの確率的大小とするとき，

1、$X \succ -X \quad (\mathcal{R})$

2、$X \succ -X' \quad (\mathcal{R})$

3、$X \succ 0 \quad (\mathcal{R})$

4　$X|_{X \geq 0} \succ -X|_{X \leq 0} \quad (\mathcal{R})$

により定義される。ただし X' は X と独立に同一分布に従う確率変数とする。

上を調べる事に依り5種類の"正の偏り"が定義される。以下簡単のためにのみ密度関数 $f(x)$ が存在するとする。

1、$X \succ 0 \quad (\mathcal{P}_0) \iff F(0) \leq \frac{1}{2}$

2、$X \succ 0 \quad (\mathcal{P}_1) \iff F(x) \leq 1 - F(-x)$

3. $X \succ 0$ $(\mathcal{P}_2)$ $\Longleftrightarrow$ $f(-x) \leq f(x)$ ただし $x \geq 0$

4. $X \succ 0$ $(\mathcal{P}_3)$ $\Longleftrightarrow$ $f(x)/f(-x)$ は $x \geq 0$ で増加.

5. $X \succ 0$ $(\mathcal{P}_4)$ $\Longleftrightarrow$ $f(x)/f(-x)$ は増加.

上の定義を拡張して、二つの確率変数 X，Y に於て、"X の方が Y より正に偏よっている" を定義出来、以下と同様の議論が成り立つ。しかしはんざつになるのでここでは省略する。

定義の内には次の包含関係が成り立つ。

(i)
$$\begin{tikzcd} \mathcal{P}_4 \arrow[r] \arrow[dr] & \mathcal{P}_2 \arrow[r] & \mathcal{P}_1 \arrow[r] & \mathcal{P}_0 \\ & \mathcal{P}_3 & & \end{tikzcd}$$

(ii) $X \succ 0$ $(\mathcal{P}_2)$ かつ $(\mathcal{P}_3)$ $\Rightarrow$ $X \succ 0$ $(\mathcal{P}_4)$

(iii) $X \succ 0$ $(\mathcal{P}_0)$ かつ $(\mathcal{P}_3)$ $\Rightarrow$ $X \succ 0$ $(\mathcal{P}_1)$

手始に次の性質を述べる。

(i) $X \succ 0$ $(\mathcal{P})$, $-X \succ 0$ $(\mathcal{P})$ であれば X は対称である。但し. $\mathcal{P}$ は $\mathcal{P}_4$, $\mathcal{P}_3$, $\mathcal{P}_2$ 及び $\mathcal{P}_1$ のいずれかとする。

(ii) $\gamma(x)$ が増加な奇関数とする。この時勝手な $\mathcal{P}$ について $X \succ 0$ $(\mathcal{P})$ ならば $\gamma(X) \succ 0$ $(\mathcal{P})$ が成り立つ。

次に $x_1, \cdots\cdots, x_n$ を母集団，その連続な分布関数 $F(x)$ からの size n の標本とする。$|x_i|$ の $\{|x_1|, \cdots, |x_n|\}$ でのランクを r_i とし，$z_{r_i} = 1$ $x_i > 0$，$z_{r_i} = 0$ $x_i < 0$ とし，$z = (z_1, \cdots, z_n)$ を考える。それらを確率変数とみなす時大文字で

記す。 z の集合 Π に半順序を導入する。

1. $z=(z_1,\cdots\cdots,z_n) \geq z'=(z_1',\cdots,z_n)$ (O_1')

$\Longleftrightarrow$ $i<j$ があって $z_i=z_j'=1,$ $z_i'=z_j=0$.

かつ $z_k=z_k'$ $k \neq i,j$

$z \geq z'$ (O_1) は (O_1') の連なりで結ばれる時.

2. $z \geq z'$ $(O_2) \Longleftrightarrow z_i \geq z_i'$ $i=1,2,\cdots,n$

3. $z \geq z'$ $(O_3) \Longleftrightarrow \sum_{i=1}^{k} z_i \geq \sum_{i=1}^{k} z_i'$ $k=1,2,\cdots,n$

これらの半順序の間に次の関係が存在する。

$z > z'$ $(O_3) \Longleftrightarrow z \geq z'$ $(O_1 \vee O_2)$

但し右の式は z と z' が (O_1) と (O_2) の連なりで結ばれることを意味する。

母集団の分布と統計量の分布の間に次の命題が成り立つ。

$F > 0$ $(\mathscr{P})$ は母集団と分布関数を同一視して得る。

(i) $F > 0$ $(\mathscr{P}_2)$ $z \geq z'$ (O_2) ならば、$P_F(Z=z) \geq P_F(Z=z')$

(ii) $F > 0$ $(\mathscr{P}_3)$, $z \geq z'$ (O_1) ならば $P_F(Z=z) \geq P_F(Z=z')$

(i)(ii)をあわせて(iii)を得る。

(iii) $F > 0$ $(\mathscr{P}_k)$, $z \geq z'$ (O_3) ならば $P_F(Z=z) \geq P_F(Z=z')$

(iv) $T(Z)$ が非減少であるとは. $z \geq z'$ (O_3) ならば $T(Z)$

$\geq T(z')$ が成り立つことを云う。 $F \succ 0$ ($\mathcal{P}_1$) で $T(z)$ が非減少であれば，任意の c に対して、 $P_F(T(z) \geq c) \geq P_{F_0}(T(z) \geq c)$。 但し. F_0 は対称な分布関数とする。

二変量対称性の検定

(X, Y) を R^2 上の確率変数，その分布関数を $F(x, y)$，周辺分布関数を $F_1(x)$，$F_2(y)$ とする。 帰無仮説として直線 $x = y$ に関して対称，対立仮説として $X \succ Y$ ($\mathcal{R}$) を採用する。実際には. $X \succ Y$ ($\mathcal{R}_S$)，　$X \succ Y$ ($\mathcal{R}_3$) が重要である。更にこの場合. $X \succ Y$ ($\mathcal{R}_4$) を導入する。

$$X \succ Y\ (\mathcal{R}_4) \Leftrightarrow P_F(S(a_1, b_1; a_2, b_2))\, P_F({}^tS(a_2, b_2; a_3, b_3))$$
$$\geq P_F({}^tS(a_1, b_1; a_2, b_2)) \cdot P_F(S(a_2, b_2; a_3, b_3)) \quad a_3 > a_2 > a_1 > b_1 > b_2 > b_3$$

$\mathcal{R}_{\mathrm{I}}$, $\mathcal{R}_{\mathrm{II}}$ を用いる事も興味あるが. これは $X - Y$ ($\mathcal{P}_1$)，$X - Y$ ($\mathcal{P}_2$) と同等になり. 一変量の対称性の検定の問題に帰着する。

他の統計の分野と次のつながりを持つ。

(i) $X \perp\!\!\!\perp Y$ の場合には結局. 等 size の 2 標本問題になる。

(ii) $y = -x$ に退化している場合. 一変量対称性の検定と同じになる。

(iii) $x > y$，$x < y$ に落ちる標本の個数が一定という条件を

つけると $(X,Y)|_{X>Y}$ と $(Y,X)|_{X<Y}$ の二変量二標本問題とみなされる。

確率的大小の不変性を調べる。

(i) $\mathcal{R}$ が $\mathcal{R}_1, \mathcal{R}_D, \mathcal{R}_2, \mathcal{R}_3, \mathcal{R}_4$ のいずれかとする。$f(x)$ が単調増加ならば $X>Y\ (\mathcal{R})$ の時、$f(X)>f(Y)\ (\mathcal{R})$。逆に $\mathcal{R}$ が $\mathcal{R}_1$ か $\mathcal{R}_D$ ならば、任意の $X>Y\ (\mathcal{R})$ なる (X,Y) に対して $f(X)>f(Y)$ ならば f は単調増加である。

(ii) $\mathcal{R}$ が $\mathcal{R}_D$ か $\mathcal{R}_3$ とする。$\gamma=(\gamma_1(x,y), \gamma_1(y,x))$ について $\gamma_1(x,y)$ が x について単調増加、y について単調減少であれば $X>Y\ (\mathcal{R})$ の時、$\gamma_1(X,Y)>\gamma_1(Y,X)\ (\mathcal{R})$。逆に任意の $X>Y\ (\mathcal{R}_D)$ なる (X,Y) に対して $\gamma_1(X,Y)>\gamma_1(Y,X)\ (\mathcal{R}_D)$ が成立てば $\gamma_1(x,y)$ は上の条件を満たす。

(ii) は関連して (iii) を与える。$\mathcal{R}_D$ の例を得る。

(iii) $\gamma=(\gamma_1(x,y), \gamma_2(y,x))$ で γ_1, γ_2 が (ii) の条件を満たす。$X>Y\ (\mathcal{R}_D)$ ならば $\gamma_1(X,Y)>\gamma_2(Y,X)\ (\mathcal{R}_D)$。特に (X,Y) が対称、$f(x)\geq g(x)$ で f, g 共に増加ならば $f(X)>g(Y)\ (\mathcal{R}_D)$。

$\binom{x_1}{y_1},\cdots\cdots,\binom{x_n}{y_n}$ を母集団からの size n のサンプルとする。分布は連続でかつ $x=y$ 上には正の確率を持たないとする。$\max\{x_1,y_1\}>\cdots>\max\{x_n,y_n\}$ なるよう添字をつけかえて、$x_i'=\max\{x_i,y_i\}$, $y_i'=\min\{x_i,y_i\}$ と記す。$z_i=$

$sgn(x_i - y_i)$, r_i, s_i を x_i', y_i' の $\{x_1, y_1, \cdots, x_n, y_n\}$ でのランクとする。 $m = (m_{ij})$ を $m_{ij} = c(x_i' - x_j')\,c(y_j' - y_i')$ で定義する。 ここで $c(x) = 1$ for $x > 0$, $= 0$ otherwise とする。

(i) $P_F\left(T\left(\binom{x_1}{y_1}, \cdots, \binom{x_n}{y_n}\right) \geq c\right)$ がすべての対称な F について一定であれば, $P_F\left(T\left(\binom{x_1}{y_1}, \cdots, \binom{x_n}{y_n}\right) \geq c \,\middle|\, ((x_1', y_1'), \cdots, (x_n', y_n'))\right)$ $= P_F\left(T(z_1, \cdots, z_n) \geq c \,|\, ((x_1', y_1'), \cdots, (x_n', y_n'))\right) = P_F(T \geq c)$ が a.e. な $((x_1', y_1'), \cdots, (x_n', y_n'))$ について成り立つ。 即ち、条件付き符号検定になる。

(ii) $T\left(\binom{x_1}{y_1}, \cdots, \binom{x_n}{y_n}\right) = T\left(\binom{f(x_1)}{f(y_1)}, \cdots, \binom{f(x_n)}{f(y_n)}\right)$ が 任意の狭義単調増加な f について a.e. で成り立つ為の必十条件は T が、 $z = (z_1, \cdots, z_n)$ と $((r_1, s_1), \cdots, (r_n, s_n))$ のみに依存することである。

(iii) $T\left(\binom{x_1}{y_1}, \cdots, \binom{x_n}{y_n}\right) = T\left(\binom{r(x_1, y_1)}{r(y_1, x_1)}, \cdots, \binom{r(x_n, y_n)}{r(y_n, x_n)}\right)$ が $r(x, y)$ が x について増加, y について減少, 更に $r = \binom{r(x, y)}{r(y, x)}$ が y について減少である関数について a.e. に成り立つとする。 その為の必十条件は T が m と z にのみ依存することである。

不偏性と(ii)の不変性を仮定して条件付き符号順位統計量について, 統計的性質を与える。

(i) $X \succ Y$ $(\mathcal{R}_3)$, $z \geq z'$ (O_2) ならば $P_F\left(Z = z \,|\, ((r_1, s_1), \cdots \cdots (r_n, s_n))\right) \geq P_F\left(Z = z' \,|\, ((r_1, s_1), \cdots \cdots, (r_n, s_n))\right)$.

(ii) $X > Y$ (R_4) $z \geq z'$ (O_1') $r_i > r_j > s_j > s_i$ ならば、

$P_F(Z=z \mid ((r_1,s_1),\cdots,(r_n,s_n))) \geq P_F(Z=z' \mid ((r_1,s_1),\cdots,(r_n,s_n)))$

(iii) Tを用いて、棄却域Rを $T \geq c(\alpha,((r_1,s_1),\cdots,(r_n,s_n)))$ で構成する。等号はlevelを用いて普通に決める。この時、更に、$\left(\begin{pmatrix}x_1\\y_1\end{pmatrix},\cdots,\begin{pmatrix}x_n\\y_n\end{pmatrix}\right) \in R$ でかつ $x' \geq x$, $y' \leq y$ であれば $\left(\begin{pmatrix}x_1\\y_1\end{pmatrix},\cdots,\begin{pmatrix}x_i'\\y_i'\end{pmatrix},\cdots,\begin{pmatrix}x_n\\y_n\end{pmatrix}\right) \in R$ なる条件を加える。$X > Y$ (R_D) であれば $P_F(R) \geq P_{F_0}(R) = \alpha$。$F_0$はある対称な分布関数とする。

(iv) Tが次の条件を満たすとしよう。

(イ) $z \geq z'$ (O_2) ならば $T(z \mid ((r_1,s_1),\cdots,(r_n,s_n))) \geq T(z' \mid ((r_1,s_1),\cdots,(r_n,s_n)))$.

(ロ) z^0に対して $z_i^0 = 1$ ならば $r_i \geq r_i' > s_i' \geq s_i$, $z_i^0 = 0$ ならば $r_i' \geq r_i > s_i \geq s_i'$ が成立っているとする。この時、$T(z \mid ((r_1',s_1'),\cdots)) < T(z^0 \mid ((r_1',s_1'),\cdots))$ ならば $T(z \mid ((r_1,s_1),\cdots)) < T(z^0 \mid ((r_1,s_1),\cdots))$、亦 $T(z \mid ((r_1,s_1),\cdots)) > T(z^0 \mid ((r_1,s_1),\cdots))$ ならば $T(z \mid ((r_1',s_1'),\cdots)) > T(z^0 \mid ((r_1,s_1),\cdots))$.

更に元の分布(X,Y)について、$(X,Y) \sim (f(U),g(V))$ であると仮定する。但し、(U,V)は適当な対称な確率変数、f, g は $f(x) \geq g(x)$ for $\forall x$ なる適当な実狭義増加関数とする。$P_{(X,Y)}(T \geq c) \geq P_{(U,V)}(T \geq c)$ が成り立つ。