Z. Wahrscheinlichkeitstheorie verw. Geb. 14, 269–274 (1970)

Zur Unverfälschtheit des Pitman-Tests auf positive stochastische Abhängigkeit

G. NEUHAUS und G. NÖLLE*

Summary. Let $\mathfrak{F}_p(\mathfrak{F}_n)$ be the class of all two dimensional continuous distributions with positive (negative) likelihood ratio dependence. It is shown that the Pitman permutation test for testing independence against positive likelihood ratio dependence at level α is unbiased and of level α for $\mathfrak{F}_n$. Furthermore, it is optimal against the class of normal distributions with positive correlation coefficient ρ.

1. Terminologie und Sätze

Wir benutzen die folgenden Bezeichnungen

$\mathfrak{F}_c$ sei die Menge der zweidimensionalen stetigen Verteilungsfunktionen F;

$\mathfrak{F}_u$ sei die Menge der Verteilungsfunktionen $F \in \mathfrak{F}_c$, die das Produkt ihrer Randverteilungen F_1, F_2 sind;

$\mathfrak{F}_p$ sei die Menge der Verteilungsfunktionen $F \in \mathfrak{F}_c - \mathfrak{F}_u$, die zu Verteilungen mit positiver Likelihood-Abhängigkeit gehören (Lehmann [2], S. 1150), d.h. für die ein dominierendes Produktmaß $\mu \times \nu$ existiert (μ, ν σ-endlich) derart, daß es Festlegungen der zugehörigen Dichten $p_F(x, y)$ und Borelsche Mengen M, N gibt mit $\mu(M) = \nu(N) = 0$ und

$$p_F(x, y')\, p_F(x', y) \leqq p_F(x, y)\, p_F(x', y') \qquad \forall\, x < x',\ y < y'\colon\ x, x' \notin M,\ y, y' \notin N. \quad (1)$$

$\mathfrak{F}_n$ werde analog erklärt, wobei nur (1) ersetzt ist durch

$$p_F(x, y')\, p_F(x', y) \geqq p_F(x, y)\, p_F(x', y') \qquad \forall\, x < x',\ y < y'\colon\ x, x' \notin M,\ y, y' \notin N. \quad (2)$$

Man beachte, daß μ, ν und M, N von dem jeweiligen F abhängen können. (1) ist eine hinreichende Bedingung dafür, daß X und Y positiv abhängig sind in dem Sinne, daß gilt

$$F^{Y|x'}(y) \leqq F^{Y|x}(y) \qquad \forall\, y \in \mathbb{R}_1, \qquad \forall\, x < x'\colon x, x' \notin M$$

für die bedingte Verteilungsfunktion $F^{Y|x}$ der Verteilung von Y bei gegebenem x und entsprechend für $F^{X|y}$ (Regressions-Abhängigkeit). Verteilungen die (1) bzw. (2) genügen, findet man in [2].

Gegeben seien n stochastisch unabhängige zweidimensionale Zufallsgrößen $(X_1, Y_1), \ldots, (X_n, Y_n)$ mit der gleichen Verteilungsfunktion $F \in \mathfrak{F}_c$, so daß insbesondere alle Komponenten mit Wahrscheinlichkeit 1 verschiedene Werte annehmen. Zu einem n-Tupel $(x_1, \ldots, x_n)$ sei $x_{[\,]} = (x_{[1]}, \ldots, x_{[n]})$ das nach wachsender Größe der Komponenten geordnete n-Tupel; entsprechend sei $(x_{[\,]}, y_{([\,])}) = (x_{[1]}, y_{([1])}; \ldots; x_{[n]}, y_{([n])})$ dasjenige $2n$-Tupel, das aus $(x_1, y_1; \ldots; x_n, y_n)$ entsteht durch Anordnen der Paare (x_i, y_i) nach wachsender Größe der x-Komponenten. $(x_{[\,]}, y_{([\,])})$ ist f.s. eindeutig bestimmt. Entsprechend werden die Zufallsgrößen $X_{[\,]}$ und $(X_{[\,]}, Y_{([\,])})$ gebildet.

---

* Herr Dr. G. Nölle ist am 20. 7. 1968 verstorben.

Bezeichnen wir mit $\mathfrak{S}_n = \{s = (s_1, \ldots, s_n)\colon s$ Permutation von $1, \ldots, n\}$ die Permutationsgruppe von $1, \ldots, n$, so wird die Anordnung in $(x_{[\,]}, y_{([\,])})$ bestimmt durch genau eine Permutation $s \in \mathfrak{S}_n$, d.h. $(x_{[\,]}, y_{([\,])}) = (x_{[1]}, y_{[s_1]}; \ldots; x_{[n]}, y_{[s_n]}) =: (x_{[\,]}, y_{[s]})$. Weiter sei $(x_{[\,]}, y_{[\,]}) := (x_{[\,]}, y_{[e]})$, $e := (1, \ldots, n)$.

Aufgrund der n vorliegenden Beobachtungen soll getestet werden

$$H\colon\ F \in \mathfrak{F}_n + \mathfrak{F}_u \quad \text{gegen} \quad K\colon\ F \in \mathfrak{F}_p. \tag{3}$$

Der Pitmansche Permutationstest (vgl. [3]) für (3) lautet bekanntlich (Lehmann [1], S. 199)

$$\varphi(x_1, y_1; \ldots; x_n, y_n) = \begin{cases} 1 \\ \gamma(x_{[\,]}, y_{[\,]}) \\ 0 \end{cases} \quad \text{für} \quad \sum_{i=1}^{n} x_i y_i \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} c(x_{[\,]}, y_{[\,]}), \tag{4}$$

wobei die nur von $(x_{[\,]}, y_{[\,]})$ abhängenden Zahlen c und γ, $0 < \gamma < 1$, bestimmt werden aus der Gleichung

$$\frac{1}{n!} \sum_{s \in \mathfrak{S}_n} \varphi(x_{[\,]}, y_{[s]}) = \alpha \tag{5}$$

bei vorgegebenem Niveau α, $0 < \alpha < 1$.

φ ist also ein bedingter Test, der auf jeder der Mengen $\{(x_{[s]}, y_{[s']})\colon s, s' \in \mathfrak{S}_n\}$ gesondert festgelegt wird und nur über $(x_{[\,]}, y_{([\,])})$ von $(x_1, y_1; \ldots; x_n, y_n)$ abhängt. Wir können also schreiben $\varphi(x_1, y_1; \ldots; x_n, y_n) = \psi(x_{[\,]}, y_{([\,])})$. Nach diesen Vorbereitungen formulieren wir das Hauptergebnis dieser Note.

Satz 1. *Der Test* (4)–(5) *hat die folgenden Eigenschaften*

a) $E_F\, \varphi = \alpha$ *für alle* $F \in J := \mathfrak{F}_u$, *d.h.* φ *ist ähnlich auf* J.

b) $E_F\, \varphi \leqq \alpha$ *für alle* $F \in \mathfrak{F}_n + \mathfrak{F}_u$, *d.h.* φ *ist ein Test zum Niveau* α *für* H.

c) $E_F\, \varphi \geqq \alpha$ *für alle* $F \in \mathfrak{F}_p$, *d.h.* φ *ist unverfälscht für* K.

Dabei soll ein unterer Index F stets die Verteilungsfunktion der zugrundeliegenden Zufallsgrößen (X_i, Y_i) kennzeichnen.

Da nun, wie man leicht zeigt, jeder unverfälschte Test zum Niveau α für (3) ähnlich ist auf J und (4)–(5) nach Konstruktion (Lehmann [1]) unter allen auf J ähnlichen Tests optimal ist gegen Normalverteilungsalternativen mit positiver Korrelation, erhält man

Korollar 2. *Der Test* (4)–(5) *ist ein unverfälschter Test zum Niveau* α *für das Problem* (3) *und unter allen diesen Tests optimal gegen Normalverteilungsalternativen mit positiver Korrelation.*

Statt der Aussage des Satzes 1 beweisen wir nun etwas schärfer den

Satz 3. *Bezeichnet* $w^{(X_{[\,]}, Y_{([\,])})|(x_{[\,]}, y_{[\,]})}$ *die bedingte Verteilung von* $(X_{[\,]}, Y_{([\,])})$ *bei gegebenem* $(x_{[\,]}, y_{[\,]})$, *so existiert ein* N_F *mit* $w_F^{(X_{[\,]}, Y_{[\,]})}(N_F) = 0$ *derart, daß für* $(x_{[\,]}, y_{[\,]}) \notin N_F$ *gilt*

$$\int \psi\, dw_F^{(X_{[\,]}, Y_{([\,])})|(x_{[\,]}, y_{[\,]})} \begin{cases} = \alpha, & F \in \mathfrak{F}_u \quad \text{(a')} \\ \leqq \alpha, & F \in \mathfrak{F}_n \quad \text{(b')} \\ \geqq \alpha, & F \in \mathfrak{F}_p. \quad \text{(c')} \end{cases}$$

2. Beweis der Sätze

Die Aussagen a), b), c) von Satz 1 folgen aus a′), b′), c′) von Satz 3 wegen

$$E_F\,\varphi=\int \varphi\, dw_F^{(X_1,Y_1,\ldots,X_n,Y_n)}=\int \psi\, dw_F^{(X_{[\,]},Y_{([\,])})}$$
$$=\iint \psi\, dw_F^{(X_{[\,]},Y_{([\,])})|(x_{[\,]},y_{[\,]})}\, dw_F^{(X_{[\,]},Y_{[\,]})}.$$

Das Wahrscheinlichkeitsmaß $w_F(\circ|(x_{[\,]},y_{[\,]})):=w_F^{(X_{[\,]},Y_{([\,])})|(x_{[\,]},y_{[\,]})}$ konzentriert sich für $(x_{[\,]},y_{[\,]})\notin N_F$ auf die $n!$ Punkte der Menge $\{(x_{[\,]},y_{[s]}): s\in\mathfrak{S}_n\}$, und sein Bildungsgesetz ist, wie man leicht sieht

$$w_F(\{(x_{[\,]},y_{[s]})\}|(x_{[\,]},y_{[\,]}))=\frac{\prod_{i=1}^{n} p_F(x_{[i]},y_{[s_i]})}{\sum_{s'\in\mathfrak{S}_n}\prod_{i=1}^{n} p_F(x_{[i]},y_{[s'_i]})},\qquad s\in\mathfrak{S}_n, \tag{6}$$

wobei $(x_{[\,]},y_{[\,]})\notin N_F$ und N_F die Menge der $(x_{[\,]},y_{[\,]})$ ist, die mindestens eine der vier folgenden Bedingungen erfüllen:

1. Der Nenner in (6) verschwindet.
2. Mindestens zwei Komponenten von $x_{[\,]}$ oder $y_{[\,]}$ sind gleich.
3. Eine x-Komponente liegt in M.
4. Eine y-Komponente liegt in N.

Dabei stammen M, N aus der Definition von $\mathfrak{F}_p$ bzw. $\mathfrak{F}_n$; für $F\in\mathfrak{F}_u$ sei $M=N=\emptyset$ gewählt. Dann gilt $w_F^{(X_{[\,]},Y_{[\,]})}(N_F)=0\ \ \forall F\in\mathfrak{F}_c$.

Bei von nun ab festgehaltenem $(x_{[\,]},y_{[\,]})\notin N_F$ ist (6) eine Funktion der Permutation $s\in\mathfrak{S}_n$ allein, die wir $u_F(s)$ nennen. Ebenso ist ψ bei festem $(x_{[\,]},y_{[\,]})$ eine Funktion $\chi(s)$. Für $F\in\mathfrak{F}_u$ ist $p_F(x,y)=p_{F_1}(x)\,p_{F_2}(y)$ bei geeigneter Fixierung von p_F, so daß in diesem Fall $u_F(s)=1/n!$ für alle $s\in\mathfrak{S}_n$ gilt. Dann aber ist a′) äquivalent mit der nach Voraussetzung erfüllten Bedingung (5).

Die Aussagen a′)–c′) schreiben sich nun mit Hilfe von u_F und χ zu

a″) $\frac{1}{n!}\sum_{s\in\mathfrak{S}_n}\chi(s)=\alpha$,

b″) $\sum_{s\in\mathfrak{S}_n} u_F(s)\,\chi(s)\leqq\alpha$ für $F\in\mathfrak{F}_n$,

c″) $\sum_{s\in\mathfrak{S}_n} u_F(s)\,\chi(s)\geqq\alpha$ für $F\in\mathfrak{F}_p$.

Der Beweis von b″) reduziert sich also wegen $\sum_{s\in\mathfrak{S}_n} u_F(s)=1$ auf den Nachweis von

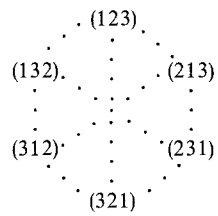
$$\sum_{s\in\mathfrak{S}_n} u_F(s)\,\chi(s)\leqq\frac{1}{n!}\sum_{s\in\mathfrak{S}_n}\chi(s)\sum_{s\in\mathfrak{S}_n} u_F(s)\qquad \text{für } F\in\mathfrak{F}_n; \tag{7}$$

c″) beweist man dann analog.

Um (7) zu zeigen führen wir in $\mathfrak{S}_n$ eine Ordnung ein und nennen s' *besser geordnet* als s, $s'\succeq s$, wenn s' aus s entsteht durch sukzessive Beseitigung von Fehlständen. Dabei ist ein Fehlstand in $s=(s_1,\ldots,s_n)$ definiert durch $s_i>s_j$ für $i<j$.

Dadurch wird in $\mathfrak{S}_n$ eine reflexive, transitive und antisymmetrische Ordnungsrelation definiert.

Für $n=3$ z.B. hat der Graph dieser Relation die nebenstehende Gestalt, wobei jedoch nur solche Tripel durch eine punktierte Linie verbunden sind, die durch genau eine Inversion auseinander hervorgehen.



Fehlstände
0
1
2
3

Bezüglich dieser Ordnungsrelation haben χ und u_F die folgenden Eigenschaften:

Lemma 4. *Für $s' \succeq s$ gilt*

$$0 \leqq \chi(s) \leqq \chi(s'), \tag{8}$$

$$0 \leqq u_F(s') \leqq u_F(s) \quad \textit{für} \quad F \in \mathfrak{F}_n, \tag{9}$$

$$0 \leqq u_F(s) \leqq u_F(s') \quad \textit{für} \quad F \in \mathfrak{F}_p. \tag{10}$$

Beweis. s' entstehe aus s o.B.d.A. durch die Beseitigung nur eines Fehlstandes in s, d.h. für ein Paar (i,j) mit $i<j$ sei $s_i > s_j$ in

$$s=(\ldots, s_i, \ldots, s_j, \ldots), s'=(\ldots, s_j, \ldots, s_i, \ldots)=(\ldots, s'_i, \ldots, s'_j, \ldots).$$

Dann ist (8) äquivalent mit

$$\sum_{k=1}^{n} x_{[k]}\, y_{[s'_k]} \geqq \sum_{k=1}^{n} x_{[k]}\, y_{[s_k]}$$

oder wegen $s'_i = s_j$, $s'_j = s_i$ mit der (offensichtlich richtigen) Ungleichung

$$(x_{[j]} - x_{[i]})(y_{[s_i]} - y_{[s_j]}) \geqq 0.$$

Entsprechend ist (9) äquivalent mit

$$\prod_{k=1}^{n} p_F(x_{[k]}, y_{[s'_k]}) \leqq \prod_{k=1}^{n} p_F(x_{[k]}, y_{[s_k]});$$

dies wiederum folgt aus der wegen (1) richtigen Beziehung

$$p_F(x_{[i]}, y_{[s'_i]})\, p_F(x_{[j]}, y_{[s'_j]}) = p_F(x_{[i]}, y_{[s_j]})\, p_F(x_{[j]}, y_{[s_i]}) \leqq p_F(x_{[i]}, y_{[s_i]})\, p_F(x_{[j]}, y_{[s_j]}). \quad \lrcorner$$

Zum Nachweis von (7) benötigen wir außerdem das leicht zu verifizierende

Lemma 5. *Für Zahlen $A_1 \geqq A_2 \geqq \cdots \geqq A_n \geqq 0$, $0 \leqq B_1 \leqq \cdots \leqq B_n$ gilt*

$$\sum_{i=1}^{n} A_i B_i \leqq \frac{1}{n} \left(\sum_{i=1}^{n} A_i\right) \left(\sum_{i=1}^{n} B_i\right). \tag{11}$$

In einem letzten Schritt beweisen wir nun die Aussage (7).

Lemma 6. *Für Funktionen* $\chi\colon \mathfrak{S}_n \to \mathbb{R}_+$ *und* $u\colon \mathfrak{S}_n \to \mathbb{R}_+$ *mit der Eigenschaft* (8) *bzw.* (9) *gilt die Aussage* (7).

Beweis. Wir führen den Beweis durch vollständige Induktion über n. Die Behauptung ist trivialerweise richtig für $n=1$. Sie gelte nun für $n-1$ statt n, $n-1 \geqq 1$. Zur Durchführung der Induktion von $n-1$ auf n werde $\mathfrak{S}_n$ zerlegt in n disjunkte „Zellen" Z_k,

$$Z_k := \{s=(s_1, \ldots, s_n) \in \mathfrak{S}_n\colon s_1 = k\}, \qquad k=1, \ldots, n.$$

Jede Zelle Z_k hat genau $(n-1)!$ Elemente. Zwischen jedem Z_k und $\mathfrak{S}_{n-1}$ wird durch

$$g_k(k, s_2, \ldots, s_n) := (s_1', \ldots, s_{n-1}')$$

mit $s_i' := s_{i+1}$ falls $s_{i+1} < k$ und $s_i' := s_{i+1} - 1$ falls $s_{i+1} > k$ eine bijektive Abbildung definiert, und es gilt

$$(k, \tilde{s}_2, \ldots, \tilde{s}_n) \succeq (k, s_2, \ldots, s_n) \Leftrightarrow (\tilde{s}_1', \ldots, \tilde{s}_{n-1}') \succeq (s_1', \ldots, s_{n-1}').$$

Für die Funktionen $\chi_k(s_1', \ldots, s_{n-1}') := \chi(k, s_2, \ldots, s_n)$ und

$$u_k(s_1', \ldots, s_{n-1}') := u(k, s_2, \ldots, s_n)$$

gilt (8) bzw. (9) für $n-1$ statt n, so daß die Induktionsvoraussetzung liefert

$$\sum_{s' \in \mathfrak{S}_{n-1}} \chi_k(s')\, u_k(s') \leqq \frac{1}{(n-1)!} \Big(\sum_{s' \in \mathfrak{S}_{n-1}} \chi_k(s')\Big)\Big(\sum_{s' \in \mathfrak{S}_{n-1}} u_k(s')\Big),$$

also

$$\sum_{s \in Z_k} \chi(s)\, u(s) \leqq \frac{1}{(n-1)!} \Big(\sum_{s \in Z_k} \chi(s)\Big)\Big(\sum_{s \in Z_k} u(s)\Big) \quad \text{für} \quad k=1, \ldots, n. \tag{12}$$

Zwischen zwei Zellen Z_k und Z_{k+1} wird durch

$$h_k(k, s_2, \ldots, k+1, \ldots, s_n) := (k+1, s_2, \ldots, k, \ldots, s_n)$$

eine bijektive Abbildung definiert mit der Eigenschaft $s \succeq h_k(s)$ für alle $s \in Z_k$, $k=1, \ldots, n-1$. Deshalb gilt für

$$A_k := \sum_{s \in Z_k} u(s), \quad \text{und} \quad B_k := \sum_{s \in Z_k} \chi(s), \qquad k=1, \ldots, n,$$

$0 \leqq A_n \leqq A_{n-1} \leqq \cdots \leqq A_1$ und $0 \leqq B_1 \leqq \cdots \leqq B_n$, so daß schließlich wegen (11) und (12) folgt

$$\begin{aligned}
\sum_{s \in \mathfrak{S}_n} \chi(s)\, u(s) &= \sum_{k=1}^{n} \sum_{s \in Z_k} \chi(s)\, u(s) \leqq \frac{1}{(n-1)!} \sum_{k=1}^{n} \Big(\sum_{s \in Z_k} \chi(s)\Big)\Big(\sum_{s \in Z_k} u(s)\Big) \\
&\leqq \frac{1}{(n-1)!} \frac{1}{n} \Big(\sum_{s \in \mathfrak{S}_n} \chi(s)\Big)\Big(\sum_{s \in \mathfrak{S}_n} u(s)\Big) \\
&= \frac{1}{n!} \Big(\sum_{s \in \mathfrak{S}_n} \chi(s)\Big)\Big(\sum_{s \in \mathfrak{S}_n} u(s)\Big). \qquad \lrcorner
\end{aligned}$$

Bemerkung. Die von Lehmann in [2] gegebene hinreichende Bedingung für die Unverfälschtheit von Permutationstests φ, nämlich $\varphi(x_1, y_1, \ldots, x_n, y_n) \leqq \varphi(x_1, y_1', \ldots, x_n, y_n')$ für alle Paare $(x_1, y_1, \ldots, x_n, y_n)$, $(x_1, y_1', \ldots, x_n, y_n')$ mit

$$x_1 < x_2 < \cdots < x_n$$

und der Eigenschaft: $i < j$ und $y_i < y_j$ impliziert $y_i' < y_j'$, ist hier für $y_{[\,]} \neq y_{[\,]}'$ nicht überall erfüllt, wie man durch einfache Beispiele zeigen kann.

Herrn Professor Dr. H. Witting, Münster, danken wir für Ratschläge und Hinweise bei der Fertigstellung dieser Arbeit.

Literatur

1. Lehmann, E. L.: Testing statistical hypotheses. New York: Wiley 1959.
2. – Some concepts of dependence. Ann. math. Statistics **37**, 1137–1153 (1966).
3. Pitman, E. J. G.: Significance tests which may be applied to samples from any population. J. roy. statist. Soc. Suppl. **4**, 225–232 (1937).

G. Neuhaus
Institut für Mathematische Statistik
der Universität
Münster i. W., Roxeler Str. 62

(Eingegangen am 23. März 1968)