응용통계연구 제 7권 1호
1994년 3월 pp. 19 - 34

k개의 회귀직선에서 기울기들의 우산형 대립가설에 대한 평행성의 비모수 검정법에 관한 연구[1)]

김동희[2)], 임동훈[3)]

요약

본 논문에서는 k개의 회귀직선에서 기울기들의 우산형 대립가설에 대한 평행성을 검정 하는 비모수 검정법을 제안하고 제안된 검정법의 점근적 성질들을 고찰하고자 한다. 정점을 알고 있는 경우와 정점을 모르고 있는 경우로 구분하여 검정법을 제시하고 몇몇 분포에 대해 모의 실험을 해본 결과 제안된 검정통계량이 꼬리가 두터운 분포에서 우도비 검정법들보다 검정력이 우수함을 보였다.

1. 서론

우산형 대립가설에 대한 문제는 Mack 과 Wolfe(1981)에 의하여 처음으로 위치 문제(location problem)에 대한 비모수 검정법이 소개된 후 Simpson 과 Margolin (1986), Hettmansperger 와 Norton(1987), Shi(1988a, 1988b)등에 의해 연구 발전되었다.

본 논문에서 다루고자 하는 k개의 회귀직선의 평행성을 우산형 대립가설에 대하여 검정하는 문제는 생물학, 경제학, 농학등 많은 실제적인 상황에 적용 가능하다.

우리는 다음과 같은 회귀모형을 생각 한다.

$$Y_{ij} = \alpha_i + \beta_i (X_{ij} - \overline{X}_i) + e_{ij}, \quad i=1,\dots,k\ ;\ j=1,\dots,n_i . \tag{1.1}$$

여기서 α_i 들은 미지의 절편이고, β_i 들은 미지의 기울기이며, $\overline{X}_i$ 는 i번째 직선의 X_{ij} 값들의 평균을 나타내고 e_{ij} 들은 서로 독립이며 항등적으로 분포하는 확률변수로서 연속인 분포함수 F 를 갖는다고 가정한다.

위의 모형 (1.1)에서 귀무가설

$$H_0 : \beta_1 = \cdots = \beta_k \ (=\beta,\ \text{미지})$$

을 우산형 대립가설

$$H_1 : \beta_1 \le \cdots \le \beta_{\ell-1} \le \beta_\ell \ge \beta_{\ell+1} \ge \cdots \ge \beta_k \quad (\text{적어도 하나의 부등호 성립})$$

---

1) 이 논문은 1992년도 교육부 지원 한국 학술 진흥재단의 자유공모 (지방대학 육성) 과제 학술 연구 조성비에 의하여 연구 되었음.

2) (609-735) 부산시 금정구 잔전동 부산대학교 자연과학대학 통계학과
3) (609-735) 부산시 금정구 잔전동 부산대학교 자연과학대학 통계학과

에 대하여 검정하고자 하며, 여기서 ℓ 은 정점 (peak)이라 부르기로 한다. 본 논문에서는 귀무가설 H_0을 대립가설 H_1에 대하여 검정하기 위한 비모수 검정법을 제안하고자 하며, 제안된 검정통계량의 점근적 성질들을 알아보고자 한다. 또한, 제안된 검정법과 비교하기 위하여 우도비 검정법과 우도비 검정법에 기초한 순위검정법을 아울러 소개하고 정점을 아는 경우와 모르는 경우 각각에 대해 살펴 보고자 한다.

2장에서는 우도비 검정법과 우도비 검정법에 기초한 순위검정법을 알아보고자 하며 3장에서는 검정통계량을 제안하고, 제안한 검정통계량의 점근적 성질을 알아보며, 4장에서는 몇몇 분포에 대해 모의 실험을 통한 검정력을 비교하고 5장에서는 참고사항 및 앞으로의 연구과제를 다루고자 한다.

본 논문을 통하여 $i=1,\dots,k$ 에 대해 다음의 기호를 사용하고자 한다.

$$n = \sum_{i=1}^{k} n_i \ ; \quad \overline{X}_i = \sum_{j=1}^{n_i} X_{ij}/n_i \ ; \quad \overline{Y}_i = \sum_{j=1}^{n_i} Y_{ij}/n_i$$

$$C_{n_i}^2 = \sum_{j=1}^{n_i}(X_{ij}-\overline{X}_i)^2 \ ; \quad C_n^2 = \sum_{i=1}^{k} C_{n_i}^2 \ ; \quad \rho_{n_i} = C_{n_i}^2/C_n^2$$

$$A_\psi^2 = \int \psi^2(u)du - \left(\int \psi(u)du\right)^2 .$$

단 $\psi(u)$, $0<u<1$ 는 점수생성함수(score generating function)이다.

2. 우도비 검정법과 우도비 검정법에 기초한 순위검정법

2.1. 우도비 검정법

2.1.1. 정점을 아는 경우

회귀모형(1.1)에서 오차항 e_{ij} 들의 분포는 평균이 0 이고 미지의 분산 σ^2 을 갖는 정규분포임을 가정하고 귀무가설 H_0를 우산형 대립가설 H_1에 대하여 검정하기 위한 우도비 검정법을 유도해 보자.

귀무가설 H_o 와 대립가설 H_1에서 σ^2의 최우추정량은 각각

$$\widehat{\sigma^2_{H_0}} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\{Y_{ij} - \overline{\alpha}_i - \overline{\beta}(X_{ij}-\overline{X}_i)\}^2/n \ ,$$

$$\widehat{\sigma^2_{H_1}} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\{Y_{ij} - \widetilde{\alpha}_i - \widetilde{\beta}_i(X_{ij}-\overline{X}_i)\}^2/n$$

이며 $\overline{\alpha}_i$ 와 $\overline{\beta}$ 는 H_0에서 최우추정량이고, $\widetilde{\alpha}_i$ 와 $\widetilde{\beta}_i$는 H_1에서의 최우추정량이다. 여기서 $\overline{\alpha}_i = \widetilde{\alpha}_i = \overline{Y}_i$ 이고, β_i 들에 대해 아무런 제한이 없다는 가설하에서 β_i 의 최우추정량은 $\overline{\beta}_i = \sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)Y_{ij} / C^2_{n_i}$ 이므로 $\overline{\beta}$는 다음과 같다.

$$\overline{\beta} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)Y_{ij} / C^2_n = \sum_{i=1}^{k}\rho_{n_i}\overline{\beta}_i.$$

또한 우산형 대립가설 H_1에서 최우추정량 $\widetilde{\beta}_i$는 $\overline{\beta}_i$의 동위회귀(isotonic regression)를 사용함으로써 얻어진다. 즉, $\widetilde{\beta}_i$는 제약조건 $\beta_1 \le \cdots \le \beta_\ell \ge \cdots \ge \beta_k$하에서 $\sum_{i=1}^{k} C^2_{n_i}(\overline{\beta}_i - \beta_i)^2$ 를 최소화 하는 해이다. 동위회귀 $\widetilde{\beta}_i$를 계산하는 알고리즘에 대한 자세한 사항은 Shi(1988a)를 참고 하기 바란다.

따라서 정점 ℓ 를 아는 경우 H_1에 대한 H_0를 검정하는 우도비 검정통계량은

$$\overline{E}^2_{k,\ell} = \sum_{i=1}^{k} C^2_{n_i}(\widetilde{\beta}_i - \overline{\beta})^2 / \sum_{i=1}^{k}\sum_{j=1}^{n_i}\{Y_{ij} - \overline{Y}_i - \overline{\beta}(X_{ij} - \overline{X}_i)\}^2 \qquad (2.1)$$

이 되며 이때 $\overline{E}^2_{k,\ell}$ 의 값이크면 귀무가설을 기각하게 된다.

Barlow et al.(1972) 과 Robertson et al.(1988)으로부터 H_0에서 $\overline{E}^2_{k,\ell}$ 의 분포는 다음과 같다.

정 리 2.1 귀무가설이 참일때,

$$P_0(\overline{E}^2_{k,\ell} \ge c) = \sum_{h=2}^{k} P_\ell(h,k) P\{B_{(h-1)/2,\,(n-h)/2} \ge c\}, \quad c>0$$

$$P_0(\overline{E}^2_{k,\ell} = 0) = P_\ell(1,k).$$

단, $P_\ell(h,k)$는 H_0에서 동위회귀 $\widetilde{\beta}_i$가 h개의 서로다른 값을 가질때의 확률이며 $B_{a,b}$는 모수가 a, b인 베타분포를 나타낸다.

증 명 위 정리의 증명은 β_i 들에 대한 모든 형태의 대립가설를 다룬 Barlow et al. (1972)의 정리 3.2 와 Robertson et al. (1988)의 정리 2.3.1의 증명의 특수한 경우이다.

H_0에서 $\overline{E}^2_{k,\ell}$의 분포는 $P_\ell(h,k)$에 의해서 결정된다. Shi(1988a)는 $k \leq 10$ 이고 똑같은 가중치 $C^2_{n_1} = \cdots = C^2_{n_k}$를 갖는 경우 $P_\ell(h,k)$의 값을 계산하였으나 k 값이 크고, 가중치들이 서로 다른 경우 $P_\ell(h,k)$의 정확한 계산은 현실적으로 불가능하다. 따라서 $\overline{E}^2_{k,\ell}$ 의 p-value 계산은 근사적인 방법에 의존할 수 밖에 없다.

2.1.2. 정점을 모르는 경우

정점을 모르는 경우 Shi(1988a)의 연구결과를 회귀분석문제에 적용시키고자 한다. $\nu = 1, \ldots, k$에 대해 $H_1^{(\nu)}$를 정점이 ν인 우산형 대립가설이라 하고 $\tilde{\beta}^{(\nu)} = (\tilde{\beta}_1^{(\nu)}, \ldots, \tilde{\beta}_k^{(\nu)})$를 최우추정량 $\overline{\beta} = (\overline{\beta}_1, \ldots, \overline{\beta}_k)$의 동위회귀라고 하면 정점을 모를 때 우산형 대립가설 H_1은 다음과 같이 표현할 수 있다.

$$H_1 = H_1^{(1)} \cup \cdots \cup H_1^{(k)}.$$

따라서 σ^2를 모를 때 우도비 검정통계량은

$$\widetilde{E}_k^2 = \sum_{i=1}^{k} C^2_{n_i} (\tilde{\beta}_i^{(\tau)} - \overline{\beta})^2 / \sum_{i=1}^{k} \sum_{j=1}^{n_i} \{ Y_{ij} - \overline{Y}_i - \overline{\beta}(X_{ij} - \overline{X}_i) \}^2 \qquad (2.2)$$

이 되며 $\overline{\beta} = \sum_{i=1}^{k} \rho_{n_i} \overline{\beta}_i$ 이고 $\tilde{\beta}^{(\tau)} = (\tilde{\beta}_1^{(\tau)}, \ldots, \tilde{\beta}_k^{(\tau)})$은

$$\sum_{i=1}^{k} C^2_{n_i} (\tilde{\beta}_i^{(\tau)} - \overline{\beta}_i)^2 = \min_{1 \leq \nu \leq k} \sum_{i=1}^{k} C^2_{n_i} (\tilde{\beta}_i^{(\nu)} - \overline{\beta}_i)^2$$

를 만족하는 벡터이다.

$\widetilde{E}_k^2$의 분포를 구하는 것은 매우 어렵지만 Geng 과 Shi(1990)가 개발한 프로그램을 이용하여 모의 실험을 통한 경험적 유의확률을 구할 수 있다.

2.2. 우도비 검정법에 기초한 순위 검정법

2.2.1. 정점을 아는 경우

귀무가설 H_0에서 기울기 β의 Hodges-Lehmann 형태의 추정량을 구하기 위해 R_{ij}를 $Y_{i1}, \ldots, Y_{in_i}$ 들 중 Y_{ij}의 순위라 하고 다음의 통계량을 정의한다.

$$T_n(Y) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)\psi(R_{ij}/(n_i+1))/C_n^2 . \tag{2.3}$$

그러면 β의 Hodges-Lehmann 형태의 추정량은

$$\beta^* = (\beta_1^* + \beta_2^*) / 2$$

이며 $\beta_1^* = \sup\{b : T_n(Y-bX) > 0\}$, $\beta_2^* = \inf\{b : T_n(Y-bX) < 0\}$이고 $T_n(Y-bX)$는 통계량(2.3)에 측정치 Y_{ij} 대신 $Y_{ij} - b(X_{ij} - \overline{X}_i)$를 대입시킨 것이다. $\overline{R}_{ij}$를 $Y_{i1} - \beta^*(X_{i1} - \overline{X}_i)$,…, $Y_{in_i} - \beta^*(X_{in_i} - \overline{X}_i)$들 중 $Y_{ij} - \beta^*(X_{ij} - \overline{X}_i)$의 순위라 할 때

$$\overline{T}_{n_i} = \sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)\psi(\overline{R}_{ij} / (n_i+1)) / C_{n_i}^2$$

라 두고 $\widetilde{T_{n_1}} \le \cdots \le \widetilde{T_{n_l}} \ge \cdots \ge \widetilde{T_{n_k}}$를 우산형 대립가설 $\beta_1 \le \cdots \le \beta_\ell \ge \cdots \ge \beta_k$ 에서 $\overline{T}_{n_1}, \ldots, \overline{T}_{n_k}$의 동위회귀라 하면 정점을 아는 경우의 H_1에 대한 H_0의 우도비 검정법에 기초한 순위 검정통계량은

$$\overline{\chi}^2_{k,\ell}(R) = \sum_{i=1}^{k} C_{n_i}^2 (\hat{T}_{n_i} - \overline{T}_n)^2 / A_\psi^2 \tag{2.4}$$

로 주어지며, $\overline{T}_n = \sum_{i=1}^{k}\rho_{n_i}\overline{T}_{n_i}$이다.

Barlow et al.(1972)의 3.3절에 따르면 $\overline{\chi}^2_{k,\ell}(R)$의 분포는 σ^2를 아는 경우의 우도비 검정통계량(5장의 식(5.1))에 대한 분포와 같은 극한분포를 가짐을 알 수 있다.

2.2.2. 정점을 모르는 경우

$H_1^{(\nu)}$를 정점이 ν인 경우의 우산형 대립가설이라 하고 $\widetilde{T_n^{(\nu)}} = (\widetilde{T_{n_1}^{(\nu)}}, \ldots, \widetilde{T_{n_k}^{(\nu)}})$ 를 $H_1^{(\nu)}$에 따르는 대립가설에서 $\overline{T}_n = (\overline{T}_{n_1}, \ldots, \overline{T}_{n_k})$의 동위회귀라고 하자. 그러면 대립가설은 2.1.2절에서와 같이 $H_1 = H_1^{(1)} \cup \cdots \cup H_1^{(k)}$로 둘 수 있으며, Shi(1988b)의 결과를 이용하면 H_0를 검정하기 위한 검정통계량은

$$\widetilde{\chi}_k^2(R) = \sum_{i=1}^{k} C_{n_i}^2 (\widetilde{T_{n_i}}^{(\tau)} - \overline{T}_n)^2 / A_\psi^2 \tag{2.5}$$

이고, 여기서 $\widetilde{T_n}^{(\tau)} = (\widetilde{T_{n_1}}^{(\tau)}, \ldots, \widetilde{T_{n_k}}^{(\tau)})$는

$$\sum_{i=1}^{k} C_{n_i}^2 (\widetilde{T_{n_i}}^{(\tau)} - \overline{T}_{n_i})^2 = \min_{1 \le v \le k} \sum_{i=1}^{k} C_{n_i}^2 (\widetilde{T_{n_i}^{(v)}} - \overline{T}_{n_i})^2$$

를 만족한다.

$\widetilde{\chi}_k^2(R)$의 분포는 Barlow et al.(1972)의 4.4절에서 논의된 바와같이 σ^2을 아는 경우의 우도비 검정통계량(5장의 식(5.2))과 똑같은 점근분포를 갖는다.

3. 제안된 검정통계량

3.1. 검정통계량 및 점근적 성질 (정점을 아는 경우)

이 절에서는 정점 ℓ 을 아는 경우 H_1에 대한 H_0를 검정하기 위한 검정통계량을 유도해 보자. 회귀 모형(1.1)에서 일반성을 잃지않고 $X_{i_1} \le \cdots \le X_{n_i}$라 하고 절편 $\alpha_i = \alpha$일때 i번째 데이타들로 부터 얻은 잔차 $Z_{ij}(\hat{\beta})$를 다음과 같이 정의한다.

$$Z_{ij}(\hat{\beta}) = (Y_{ij} - \hat{\beta}(X_{ij} - \overline{X}_i)) sgn(X_{ij} - \overline{X}_i), \quad i = 1, \ldots, k; \quad j = 1, \ldots, n_i.$$

여기서 $sgn(t)$는 $t \ge 0$이면 1, $t < 0$이면 -1이고 $\hat{\beta}$는 H_0에서 β의 Hodges-Lehmann 형태의 추정량으로 다음과 같이 구한다.

$\phi(t)$를 $t \ge 0$ 이면 1, $t < 0$ 이면 0인 지시함수라 하고 가중 순위 통계량(weighted rank statistic) T_β를 다음과 같이 정의하자.

$$T_\beta = \sum_{i=1}^{k} \sum_{s<t}^{n_i} w_{ist} \, \phi(Y_{it} - Y_{is} - \beta(X_{it} - X_{is})), \quad i = 1, \ldots, k \; ; \quad s, t = 1, \ldots, n_i \; ,$$

여기서 가중치 w_{ist}는 $X_{is} = X_{it}$이면 $w_{ist} = 0$, $X_{is} \ne X_{it}$ 이면 $w_{ist} \ge 0$이다. 그러면 β의 Hodges-Lehmann 형태의 추정량은

$$\hat{\beta} = (\hat{\beta}_1 + \hat{\beta}_2)/2, \tag{3.1}$$

$\hat{\beta}_1 = \sup \{\beta : T_\beta \ge w_{\ldots}/2\}$; $\hat{\beta}_2 = \inf \{\beta : T_\beta \le w_{\ldots}/2\}$이고, $w_{\ldots} = \sum_{i=1}^{k} \sum_{s<t}^{n_i} w_{ist}$ 이다.

식 (3.1)의 $\hat{\beta}$는 가중치 w_{ist}를 갖는 $(Y_{it}-Y_{is})/(X_{it}-X_{is})$, $1 \le s < t \le n_i$, 의 가중 중앙값(weighted median)이며 최적의 가중치는 $w_{ist}=X_{it}-X_{is}$이다. 가중 중앙 회귀 추정량(weighted median regression estimator)에 대한 좀더 자세한 사항은 Scholz (1978) 와 Sievers (1978)를 참조 하기 바란다.

위에서 정의한 u번째 데이타와 v번째 데이타로 부터 얻은 잔차들로 부터 맨-휘트니 통계량 $U_{uv}(\hat{\beta})$는 다음과 같다.

$$U_{uv}(\hat{\beta})=\sum_{s=1}^{n_u}\sum_{t=1}^{n_v}\phi(Z_{vt}(\hat{\beta})-Z_{us}(\hat{\beta})) .$$

그러면 정점 ℓ 를 아는 경우에 H_0를 검정하는 검정통계량은 다음과 같다.

$$\overline{V}_{\ell}(\hat{\beta})=\sum_{1\le u<v\le \ell}U_{uv}(\hat{\beta})+\sum_{\ell\le u<v\le k}U_{vu}(\hat{\beta}). \tag{3.2}$$

$\overline{V}_{\ell}(\hat{\beta})$은 정점 ℓ 를 중심으로 Jonckheere 통계량과 역 Jonckheere 통계량들의 합의 형태이다. $\overline{V}_{\ell}(\hat{\beta})$의 값이 크게되면 귀무가설을 기각한다.

참고로, 회귀계수들의 순서대립가설(ordered alternative)에 대한 평행성의 검정법을 연구한 Jee(1989)의 검정통계량은 본 논문에서 제시한 검정통계량 $\overline{V}_{\ell}(\hat{\beta})$에서 정점이 $\ell=k$인 특수한 경우이다.

제안된 검정통계량 $\overline{V}_{\ell}(\hat{\beta})$의 점근적 성질에 대하여 알아보자. Randles(1984)의 정리 A.9에 있는 조건들을 만족하므로 다음의 정리를 얻을 수 있다.

정 리 3.1 귀무가설이 참이고, $\hat{\beta}$은 β의 $\sqrt{n}$ -일치추정량일때,

$$n^{-3/2}\,[\,\overline{V}_{\ell}(\hat{\beta})-\overline{V}_{\ell}(\beta)] \rightarrow 0$$

이다.

Archambault, Mack 그리고 Wolfe(1977)의 결과를 이용하여 검정통계량 $\overline{V}_{\ell}(\hat{\beta})$의 점근분포를 다음정리에서 얻는다.

정 리 3.2 $i=1,\ldots,k$에 대해 $n_i/n \rightarrow \lambda_i$, $0<\lambda_i<1$일때 귀무가설이 참이면

$$[\,\overline{V}_{\ell}(\beta)-E_0(\overline{V}_{\ell}(\beta))]\,/\,[Var_0(\overline{V}_{\ell}(\beta))]^{1/2}$$

는 극한 정규 분포를 가지며, $\overline{V}_{\ell}(\beta)$의 평균과 분산은 각각 다음과 같다.

$$E_0(\overline{V}_\ell(\beta)) = (N_1^2 + N_2^2 - \sum_{i=1}^{k} n_i^2 - n^2_\ell)/4,$$
$$Var_0(\overline{V}_\ell(\beta)) = \{2(N_1^3+N_2^3) + 3(N_1^2+N_2^2) - \sum_{i=1}^{k} n_i^2(2n_i+3) - n_\ell(2n_\ell+3) + 12n_\ell N_1N_2 - 12n^2_\ell n\}/72,$$

단, $N_1 = \sum_{i=1}^{\ell} n_i$, $N_2 = \sum_{i=\ell}^{k} n_i$ 임.

따라서 유의수준 α에서 $\overline{V}_\ell(\hat{\beta}) \geq E_0(\overline{V}_\ell(\beta)) + z_\alpha (Var_0(\overline{V}_\ell(\beta)))^{1/2}$ 이면 귀무가설을 기각한다.

3.2. 정점을 모르는 경우

이 절에서는 먼저 정점을 추정하고 추정된 정점을 제안된 검정통계량에 적용시키고자 한다.

$J_t(\hat{\beta}) = \sum_{i \neq t}^{k} U_{it}(\hat{\beta})$, $i = 1, \ldots, k$ 라 두면 $J_t(\hat{\beta})$는 t번째 표본 $Z_{tj}(\hat{\beta})$의 값이 다른 $k-1$개의 표본의 값보다 크게되는 경우의 갯수이며 $J_t(\hat{\beta})$의 극한분포는 $J_t(\beta)$의 극한분포와 같고 평균과 분산은 다음과 같다.

$$E_0\{J_t(\beta)\} = \{n_t(n-n_t)\}/2,$$
$$Var_0\{J_t(\beta)\} = \{n_t(n-n_t)(n+1)\}/12.$$

$J_t(\hat{\beta})$의 표준화된 형태는

$$J_t^*(\hat{\beta}) = [J_t(\hat{\beta}) - E_0(J_t(\beta))] / [Var_0(J_t(\beta))]^{1/2}$$

이며 r $(1 \leq r \leq k)$을 최대값 $J_i^*(\hat{\beta})$을 갖는 회귀직선의 갯수라 하면

$$\chi_t = \begin{cases} 1/r, & t \text{ 번째 직선이 최대값 } J_i^*(\hat{\beta}) \text{을 갖는 경우} \\ 0, & \text{그외} \end{cases}$$

라 둘 수 있다. 보통은 $r=1$ 이다. 즉, 최대값 $J_i^*(\hat{\beta})$를 갖는 회귀직선은 하나 뿐이라고 간주하며, 이로써 정점을 추정할 수 있다. 이 추정된 정점을 이용하여 검정통계량은 앞절에서와 같이

$$\hat{V}(\hat{\beta}) = \sum_{t=1}^{k} \chi_t [\overline{V}_t(\hat{\beta}) - E_0(\overline{V}_t(\beta))] / [Var_0(\overline{V}_t(\beta))]^{1/2} \tag{3.3}$$

로 표현되며 $\hat{V}(\hat{\beta})$의 값이 크면 귀무가설을 기각하면 된다.

다음은 $\widetilde{V}(\widehat{\beta})$의 기각치를 구하기 위해 H_0하에서 $\widetilde{V}(\widehat{\beta})$의 극한분포에 대해 알아보자. 임의의 실수 d 에 대해 $\lim_{n\to\infty} P_0(\widetilde{V}(\widehat{\beta}) \le d)$은 평균이 0이고 서로 다른 공분산 행렬를 갖는 k 개의 정규밀도함수의 k 다중적분의 합으로 표현되므로 실제적인 상황에 적용시키는데는 어려움이 있다. 따라서 본 논문에서는 10,000 번의 모의실험을 통하여 각각의 경우에 대해서 근사적인 기각치를 얻었다.

4. 모의 실험

4.1. 실험의 계획

이 절에서는 정점을 아는 경우와 모르는 경우에 대하여 각각 식 (2.1)과 (2.2)에 주어진 우도비 검정통계량과 식(2.4)과 (2.5)에 주어진 우도비 검정법에 기초한 순위검정통계량 및 식(3.2)과 (3.3)에 주어진 제안된 검정통계량의 검정력을 비교하고자 한다. 검정력에 사용되는 분포는 균등분포, 정규분포, 코쉬분포, 이중지수분포, 오염정규분포를 사용하고자 하며 오염정규분포의 분포함수는

$$F(x) = (1-\varepsilon)\Phi(x) + \varepsilon\Phi(x/\sigma)$$

이며 Φ는 표준정규분포 함수이다.

이 실험에서 $k=3$, 4로 두었고 각 직선에서의 자료의 크기는 같게 두었으며 X_{ij}의 값들은 IMSL의 부프로그램 GGUBS를 이용하여 생성하였다. $k=3$일때는 정점 $\ell=2$, $k=4$일때는 정점 $\ell=2$, 3 인 경우에 대하여 실험을 하였다. 오염정규분포의 경우에는 $\varepsilon=0.05$, $\sigma=5$ 와 $\varepsilon=0.10$, $\sigma=10.0$인 두가지를 고려하였다. 균등분포와 정규분포에서의 난수 생성은 IMSL의 GGUBS 와 GGNML을 각각 이용하였고 코쉬분포와 이중지수분포의 난수는 역적분변환(inverse integral transfomation)을 이용하여 생성하였으며 오염정규분포는 GGUBS와 GGNML을 이용하였다. 모든 분포는 0에 대해 대칭이고 코쉬분포를 제외하고는 분산 1을 갖도록 하였으며 코쉬분포의 경우는 2차적률이 존재하지 않으므로 $F^{-1}(0.84) - F^{-1}(0.5)=1.8326$ 의 값을 사용하였다. 절편 α_i의 값들은 편의상 0으로 놓고 실험을 하였다. β_i의 값은 다음과 같이 등간격으로 두었다.

$$\beta_i = 1 + c_i m \Delta, \quad i=1,\dots,k$$

여기서 $i=1,\dots,\ell$ 일때 $c_i=i$ 이고 $i=\ell+1,\dots,k$일때는 $c_i=2\ell-i$ 이며 Δ는 결합 표본에서 β의 최소제곱추정량의 표준편차로 두었다. m의 값은 $k=3$일때는 0.0, 2.0, 4.0, 6.0 으로 두었고 $k=4$일때는 0.0, 1.5, 3.0, 4.5로 두었다. $k=3$, $m=2.0$; $k=4$, $m=1.5$인 경우는 β_i들이 귀무가설에서 조금 벗어나고 $k=3$, $m=4.0$, 6.0; $k=4$, $m=3.0$, 4.5인 경우는 귀무가설에

서 많이 벗어나는 우산형 대립가설을 나타내기 위해 m의 범위를 설정하였다. 이 실험은 유의수준 α=0.05에서 수행하였으며 1000번 반복하였다.

4.2. 실험의 결과

실험검정력은 귀무가설을 기각하는 횟수에 1000을 나눈 값이며 m=0일때 실험검정력은 실험유의수준이 된다. 정점을 아는 경우의 실험검정력이 표4.1, 표4.2, 표4.3에 나타나 있고 정점을 모르는 경우의 실험검정력이 표4.4, 표4.5, 표4.6에 나타나 있다. 표에서 알 수 있듯이 정점을 아는 경우 기대한 대로 균등분포와 정규분포에서는 유의수준과 검정력 측면에서 볼때 우도비 검정통계량 $\overline{E}^2_{k,\ell}$가 보편적으로 우수하게 나타났으며 꼬리가 두터운 분포에서는 $\overline{E}^2_{k,\ell}$의 검정력은 현저하게 증가세가 둔화되고 있음을 알 수 있으며 제안된 검정통계량 $\overline{V}_\ell(\hat{\beta})$의 실험 검정력이 한층 우수하게 나타났다. 균등분포와 정규분포하에서도 $\overline{V}_\ell(\hat{\beta})$의 검정력은 $\overline{\chi}^2_{k,\ell}(R)$보다 대체로 좋은 검정력을 갖고 있음을 알 수 있었다. 또한 정점을 모르는 경우에도 유사한 결론을 얻을 수 있으며 예측한대로 정점을 아는 경우보다 전반적으로 실험 검정력이 조금 낮게 나타나 있음을 알수 있다.

표 4.1. 실험검정력 (k=3, n_1 = n_2 = n_3 =10 , ℓ =2, 정점을 아는 경우, α=0.05)

검정통계량	m=0.0	2.0	4.0	6.0	m=0.0	2.0	4.0	6.0
	균등 분포				정규 분포			
$\overline{E}^2_{k,\ell}$	0.049	0.151	0.383	0.648	0.055	0.174	0.382	0.683
$\overline{\chi}^2_{k,\ell}(R)$	0.042	0.107	0.264	0.467	0.041	0.122	0.277	0.531
$\overline{V}_\ell(\hat{\beta})$	0.038	0.156	0.337	0.553	0.046	0.170	0.372	0.635
	이중 지수 분포				코쉬 분포			
$\overline{E}^2_{k,\ell}$	0.045	0.148	0.408	0.704	0.051	0.081	0.124	0.179
$\overline{\chi}^2_{k,\ell}(R)$	0.031	0.130	0.371	0.657	0.034	0.102	0.235	0.368
$\overline{V}_\ell(\hat{\beta})$	0.047	0.177	0.459	0.720	0.033	0.145	0.313	0.530
	오염 정규 분포 (ε=0.05, σ=5)				오염 정규 분포 (ε=0.10, σ=10)			
$\overline{E}^2_{k,\ell}$	0.060	0.180	0.466	0.746	0.063	0.248	0.545	0.732
$\overline{\chi}^2_{k,\ell}(R)$	0.056	0.171	0.446	0.774	0.031	0.422	0.775	0.863
$\overline{V}_\ell(\hat{\beta})$	0.045	0.214	0.504	0.815	0.040	0.504	0.880	0.963

표 4.2. 실험검정력 (k=4, n_1 = … = n_4 =10, ℓ =2, 정점을 아는 경우 α=0.05)

검정통계량	m=0.0	1.5	3.0	4.5	m=0.0	1.5	3.0	4.5
	균등 분포				정규 분포			
$\overline{E}^2_{k,\ell}$	0.049	0.201	0.485	0.826	0.053	0.200	0.499	0.821
$\overline{\chi}^2_{k,\ell}(R)$	0.037	0.138	0.342	0.636	0.040	0.142	0.374	0.692
$\overline{V}_\ell(\hat{\beta})$	0.046	0.177	0.379	0.638	0.048	0.169	0.428	0.686
	이중 지수 분포				코쉬 분포			
$\overline{E}^2_{k,\ell}$	0.056	0.215	0.519	0.801	0.071	0.094	0.115	0.174
$\overline{\chi}^2_{k,\ell}(R)$	0.037	0.185	0.500	0.758	0.038	0.130	0.323	0.474
$\overline{V}_\ell(\hat{\beta})$	0.048	0.232	0.508	0.765	0.040	0.153	0.400	0.543
	오염 정규 분포 (ε=0.05, σ=5)				오염 정규 분포 (ε=0.10, σ=10)			
$\overline{E}^2_{k,\ell}$	0.055	0.264	0.585	0.840	0.064	0.288	0.600	0.808
$\overline{\chi}^2_{k,\ell}(R)$	0.037	0.230	0.632	0.883	0.042	0.571	0.917	0.971
$\overline{V}_\ell(\hat{\beta})$	0.066	0.241	0.603	0.872	0.052	0.600	0.950	0.990

표 4.3. 실험검정력 (k=4, n_1 = … = n_4 =10, ℓ =3, 정점을 아는 경우 α=0.05)

검정통계량	m=0.0	1.5	3.0	4.5	m=0.0	1.5	3.0	4.5
	균등 분포				정규 분포			
$\overline{E}^2_{k,\ell}$	0.052	0.212	0.594	0.892	0.052	0.240	0.573	0.865
$\overline{\chi}^2_{k,\ell}(R)$	0.044	0.176	0.461	0.792	0.052	0.193	0.471	0.793
$\overline{V}_\ell(\hat{\beta})$	0.041	0.196	0.460	0.721	0.048	0.213	0.449	0.775
	이중 지수 분포				코쉬 분포			
$\overline{E}^2_{k,\ell}$	0.068	0.252	0.606	0.877	0.071	0.095	0.143	0.221
$\overline{\chi}^2_{k,\ell}(R)$	0.051	0.227	0.585	0.859	0.059	0.165	0.373	0.546
$\overline{V}_\ell(\hat{\beta})$	0.041	0.254	0.573	0.831	0.060	0.198	0.431	0.666
	오염 정규 분포 (ε=0.05, σ=5)				오염 정규 분포 (ε=0.10, σ=10)			
$\overline{E}^2_{k,\ell}$	0.052	0.300	0.682	0.873	0.066	0.340	0.677	0.862
$\overline{\chi}^2_{k,\ell}(R)$	0.039	0.308	0.731	0.940	0.048	0.684	0.947	0.975
$\overline{V}_\ell(\hat{\beta})$	0.044	0.300	0.699	0.923	0.050	0.675	0.968	0.996

표 4.4. 실험검정력(k=3, n_1 = n_2 = n_3 =10, ℓ =2, 정점을 모르는 경우, α=0.05)

검정통계량	m=0.0	2.0	4.0	6.0	m=0.0	2.0	4.0	6.0
	균등 분포				정규 분포			
$\widetilde{E}^2_k$	0.046	0.114	0.297	0.546	0.051	0.125	0.301	0.574
$\widetilde{\chi}^2_k(R)$	0.037	0.072	0.177	0.331	0.036	0.082	0.179	0.398
$\widetilde{V}(\widehat{\beta})$	0.038	0.093	0.174	0.345	0.041	0.088	0.195	0.396
	이중 지수 분포				코쉬 분포			
$\widetilde{E}^2_k$	0.048	0.107	0.329	0.632	0.050	0.077	0.092	0.139
$\widetilde{\chi}^2_k(R)$	0.032	0.078	0.252	0.533	0.028	0.068	0.164	0.290
$\widetilde{V}(\widehat{\beta})$	0.048	0.089	0.259	0.501	0.027	0.082	0.180	0.317
	오염 정규 분포 (ε=0.05, σ=5)				오염 정규 분포 (ε=0.10, σ=10)			
$\widetilde{E}^2_k$	0.063	0.144	0.376	0.688	0.054	0.193	0.468	0.677
$\widetilde{\chi}^2_k(R)$	0.044	0.108	0.323	0.653	0.027	0.297	0.714	0.824
$\widetilde{V}(\widehat{\beta})$	0.056	0.127	0.292	0.605	0.038	0.310	0.740	0.904

표4.5. 실험검정력 (k=4, n_1 = ⋯ = n_4 =10, ℓ =2, 정점을 모르는 경우 α=0.05)

검정통계량	m=0.0	1.5	3.0	4.5	m=0.0	1.5	3.0	4.5
	균등 분포				정규 분포			
$\widetilde{E}^2_k$	0.047	0.129	0.373	0.697	0.041	0.135	0.346	0.727
$\widetilde{\chi}^2_k(R)$	0.031	0.077	0.209	0.462	0.024	0.080	0.231	0.549
$\widetilde{V}(\widehat{\beta})$	0.043	0.111	0.266	0.480	0.051	0.125	0.274	0.513
	이중 지수 분포				코쉬 분포			
$\widetilde{E}^2_k$	0.051	0.140	0.412	0.709	0.077	0.085	0.084	0.150
$\widetilde{\chi}^2_k(R)$	0.031	0.100	0.370	0.622	0.034	0.084	0.215	0.340
$\widetilde{V}(\widehat{\beta})$	0.056	0.149	0.359	0.591	0.052	0.098	0.252	0.399
	오염 정규 분포 (ε=0.05, σ=5)				오염 정규 분포 (ε=0.10, σ=10)			
$\widetilde{E}^2_k$	0.050	0.169	0.468	0.747	0.069	0.213	0.488	0.725
$\widetilde{\chi}^2_k(R)$	0.024	0.124	0.433	0.781	0.031	0.420	0.848	0.938
$\widetilde{V}(\widehat{\beta})$	0.050	0.172	0.445	0.752	0.041	0.431	0.882	0.970

표 4.6. 실험검정력(k=4, n_1 = … = n_4 =10, ℓ =3,정점을 모르는 경우 α=0.05)

검정통계량	m=0.0	1.5	3.0	4.5	m=0.0	1.5	3.0	4.5
	균등 분포				정규 분포			
$\widetilde{E}^2_k$	0.047	0.139	0.424	0.796	0.041	0.137	0.388	0.781
$\widetilde{\chi}^2_k(R)$	0.031	0.074	0.261	0.576	0.024	0.088	0.259	0.618
$\widetilde{V}(\widehat{\beta})$	0.043	0.120	0.307	0.539	0.051	0.139	0.285	0.587
	이중 지수 분포				코쉬 분포			
$\widetilde{E}^2_k$	0.051	0.146	0.459	0.780	0.077	0.083	0.105	0.176
$\widetilde{\chi}^2_k(R)$	0.031	0.114	0.380	0.717	0.034	0.084	0.231	0.399
$\widetilde{V}(\widehat{\beta})$	0.056	0.151	0.443	0.685	0.052	0.104	0.302	0.506
	오염 정규 분포 (ε=0.05, σ=5)				오염 정규 분포 (ε=0.10, σ=10)			
$\widetilde{E}^2_k$	0.050	0.185	0.539	0.790	0.069	0.241	0.570	0.769
$\widetilde{\chi}^2_k(R)$	0.024	0.151	0.487	0.833	0.031	0.496	0.869	0.939
$\widetilde{V}(\widehat{\beta})$	0.050	0.189	0.501	0.800	0.041	0.512	0.916	0.981

5. 참고사항 및 연구과제

본 논문에서 우도비 검정통계량을 유도하는데 σ^2를 알고 있는 경우는 현실적으로 거의 없다고 생각하여 σ^2를 모르는 경우에 대해서만 통계량을 유도하였으나 참고로 σ^2를 아는 경우의 통계량은 다음과 같다.

참고사항 1 정점을 아는 경우, σ^2도 아는 경우의 우도비 검정통계량은

$$\overline{\chi}^2_{k,\ell} = \sum_{i=1}^{k} C^2_{n_i}(\widetilde{\beta_i} - \overline{\beta})^2 / \sigma^2 \qquad (5.1)$$

으로 표현되고 이의 분포는

$$P_0(\overline{\chi}^2_{k,\ell} \geq c) = \sum_{h=2}^{k} P_\ell(h, k)\ P(\chi^2_{h-1} \geq c), \quad c>0$$
$$P_0(\overline{\chi}^2_{k,\ell} = 0) = P_\ell(1, k)$$

이다. 여기서 χ^2_a는 자유도가 a인 카이 제곱 분포를 나타낸다.

참고사항 2 정점을 모르는 경우, σ^2를 알고 있는 경우의 우도비 검정통계량은

$$\widetilde{\chi}_k^2 = \sum_{i=1}^{k} C_{n_i}^2 (\tilde{\beta}_i^{(\mathcal{T})} - \overline{\beta})^2 / \sigma^2 \tag{5.2}$$

으로 표현되며 $\overline{\beta} = \sum_{i=1}^{k} \rho_{n_i} \overline{\beta}_i$ 이고 $\tilde{\beta}^{(\mathcal{T})} = (\tilde{\beta}_i^{(\mathcal{T})}, \ldots, \tilde{\beta}_k^{(\mathcal{T})})$은 $\sum_{i=1}^{k} C_{n_i}^2 (\tilde{\beta}_i^{(\mathcal{T})} - \overline{\beta}_i)^2 = \min_{1 \le v \le k} \sum_{i=1}^{k} C_{n_i}^2 (\tilde{\beta}_i^{(v)} - \overline{\beta}_i)^2$을 충족시킨다.

Shi(1988a)의 논문에서 k=3이고 가중치가 같을 경우 $\widetilde{\chi}_3^2$의 분포는

$$P_0(\widetilde{\chi}_3^2 \ge c^2) = \frac{1}{2} + \frac{2}{3} P(\chi_2^2 \le c^2) + 2 \int_0^c \Phi(\sqrt{3} z) \phi(z) dz - \Phi(c), \quad c > 0$$

으로 주어지며 Φ 와 ϕ는 표준정규분포의 분포함수 및 밀도함수이다.

본 논문에서 제안하는 검정통계량과 우도비 검정통계량 및 우도비 검정법에 기초한 순위통계량들과의 ARE(Asymptotic Relative Efficiency)로써 efficiency 비교는 이들의 극한분포가 서로 다르기 때문에 불가능하나 제안된 검정법은 다른 검정법들보다 다음의 몇가지의 이점을 갖고 있다.

1. $\overline{V}_\ell(\hat{\beta})$는 H_0 하에서 극한정규분포를 갖는 반면에 $\overline{E}^2_{k,\ell}$ 와 $\overline{\chi}^2_{k,\ell}(R)$의 분포는 $P_\ell(h,k)$ 와 각각의 베타분포 및 카이 제곱 분포와의 혼합된 형태이므로 p-value 계산하는데 어려움이 있다.
2. 모의실험 결과에서 보듯이 제안된 검정법은 이 논문에서 고려되는 모든 분포하에서 우도비 검정법에 기초한 순위 검정법보다는 전반적으로 좋은 검정력을 갖고 있으며 꼬리가 두터운 분포하에서 우도비 검정법보다 우수하다.

앞으로 더 연구가 진행될 수 있다고 생각되는 부분은 검정통계량에 가중치를 부여함으로써 효율을 극대화 시킬 수 있는 가중치를 찾는 것과 척도문제에 있어서 우산형 대립가설의 검정문제에 적용시키는 것등이라 생각한다.

주) 본 논문을 심사해주시고 조언을 해주신 심사위원과 편집위원께 깊은 감사를 드립니다.

참 고 문 헌

[1] Archambault, W.A.T., Tr., Mack, G.A. and Wolfe, D.A.(1977), K-Sample Rank Tests Using Pair-Specific Scoring Functions, *The Canadian journal of Statistics,* Vol. 5, 195-207.

[2] Barlow, R.E., Bartholomew, D.J., Bremner, J.H. and Brunk, H.D.(1972) *Statistical*

Inference under Ordered Restriction, John Wiley and Sons, New York.
[3] Geng, Z., and Shi, N-Z.(1990), Isotonic Regression for Umbrella Orderings, *Applied Statistics,* Vol. 39, 397-424.
[4] Hettmansperger, T.P. and Norton, R.M.(1987), Tests for Patterned Alternatives in K-Sample Problem, *Journal of the American Statistical* Associations, Vol. 82, 292-299.
[5] Jee, E.S.(1989), A Nonparametric Test for the Parallelism of Regression Lines against Ordered Alternatives, Ph.D. Dissertation, Department of Computer Science and Statistics, Seoul National University, Seoul, Korea.
[6] Mack, G.A. and Wolfe, D.A.(1981),K-Sample Rank Tests for Umbrella Alternatives, *Journal of the American Statistical Associations*, Vol. 76, 175-181.
[7] Randles, R.H.(1984), On Tests Applied to Residuals, *Journal of the American Statistical Associations*, Vol. 79, 349-354.
[8] Robertson, T., Wright, F.T. and Dykstra, R.L.(1988), *Order Restricted Statistical Inference,* John Wiley and Sons, New York.
[9] Scholz, F.W.(1978), Weighted Median Regression Estimates, *The Annals of Statistics,* Vol. 6, 603-609.
[10] Shi, N.-Z.(1988a), A Test of Homogeneity for Umbrella Alternatives and Tables of the Level Probabilities, *Communications in Statistics-Theory and Methods*, A17(3), 657-670.
[11] Shi, N.-Z.(1988b), Rank Test Statistics for Umbrella Alternatives, *Communications in Statistics-Theory and Methods*, A17(6), 2059-2073.
[12] Sievers, G.L.(1978), Weighted Rank Statistics for Simple Linear Regression, *Journal of the American Statistical Associations,* Vol. 73, 628-631.
[13] Simpson, D.G. and Margolin, B.H.(1986), Recursive Nonparametric Testing for Dose-Response Relationships Subject to Downturns at High Doses, *Biometrika,* Vol. 73, 589-596.

Nonparametric Tests of Parallelism Against Umbrella Alternatives of Slopes in k-Regression Lines[4]

Dong Hee Kim[5], Dong Hoon Lim[6]

Abstract

In this paper we propose nonparametric tests of parallelism against umbrella alternatives of slopes in k-regression lines and investigate the asymptotic properties of the proposed test statistics. For the known peak and unknown peak, we suggest the test statistics and show that, from Monte carlo study, the proposed test statistics have good empirical powers for heavy tailed distributions than the likelihood ratio tests.

4) This paper was supported(in part) by NON DIRECTED RESEARCH FUND, Korea Research Foundation, 1992.

5) Department of Statistics, Pusan National University, Pusan 609-735, Korea.

6) Department of Statistics, Pusan National University, Pusan 609-735, Korea.