

평활 적합도 검정에서의 분산 추정의 영향 ¹

윤 용화 ² · 김 종태 ³ · 이 우동 ⁴

요약

본 연구는 Rice 분산추정량을 사용한 기존의 평활 적합도 검정들에 있어서 Rice의 분산 추정량 보다 뛰어난 성질을 가지는 GSJS 추정량을 사용함으로 검정 통계량들에 대한 검정력에 미치는 영향을 조사하는데 그 목적을 둔다. 또한 분산의 값들의 변화가 진동수와 진폭에 따른 선형 모형에서의 검정력들에 미치는 영향을 관찰하였다.

주제어: 적합도 검정, 오차 분산 추정, 퓨리에 급수, 비모수 선형모형.

1. 서론

비모수 회귀모형이나 선형모형에 있어서의 평활 적합도 검정에 대하여 최근에 많은 연구들이 활발히 진행되어 왔다. 그러나 이러한 모형들에 대하여 평활 적합도 검정 문제들을 다루는 경우에 필연적으로 발생하는 오차분산의 추정 문제에 대해서는 매우 소홀히 다루어지고 있음을 발견한다. 예를 들어 Buckley(1991), Eubank 와 Hart(1992), Eubank 와 Hart(1993) 등이 발표한 연구들은 Rice(1984)의 오차분산 추정량을 사용하여 평활 검정통계량을 제시하였다. 그러나 김종태(1998)의 차분에 기저한 분산추정의 연구에서 Rice가 제시한 추정량은 Gasser, Sroka 와 Jennen-Steinmetz(1984)가 제시한 GSJS추정량과 비교하여 볼때 Rice의 추정량이 GSJS추정량보다 분산을 추정함에 있어서 좋지 않은 결과를 가지고 있음을 발견했다.

이러한 사실을 기초로 하여 Buckeley의 검정법, von Neumann의 검정법과 Eubank와 Hart의 검정법에 있어서 그들이 제시한 검정통계량들에 대한 분산추정량을 다르게 적용함으로써 적합도 검정통계량의 검정력에 대한 분산추정의 영향을 파악하는 것이 본 연구의 목적이다. 또한 분산의 값들에 대한 변화에 따라서 검정력들이 얼마나 변화 되는가를 관찰할 것이다. 본 연구에서의 관찰은 평활 적합도 검정에 대한 이론들이나 실제적인 문제들에 대한 적용에 있어서 많은 아이디어와 문제점들을 제시할 것으로 믿는다.

¹ 이 논문은 1998년도 대구대학교 학술 지원비에 의한 논문임

² 경북 경산시 진량면 대구대학교 통계학과, 712-714, 교수

³ 경북 경산시 진량면 대구대학교 통계학과, 712-714, 조교수

⁴ 경북 경산시 점촌동 경산대학교 통계학과, 712-240, 조교수

예측변수 t 에 대한 설계점들 $t_i = (i - 1/2)/n$ 에서 관찰된 반응변수 y 의 값들인 y_1, \dots, y_n 을 가지는 다음의 모형을 생각하자.

$$y_i = f(t_i) + \epsilon_i, \quad i = 1, \dots, n. \quad (1)$$

여기서 f 는 미지의 함수이고, ϵ_i 는 독립이며 평균 0과 공통분산 σ^2 을 가지는 정규확률변수들이다. 이 모형에 대하여 본 논문에서 다루는 문제는 예측변수 t 가 반응변수 y 에 대하여 “효과가 없음(no effect)”을 가지는가를 검정하는데 목적이 있다. 즉 이검정은 $H_0: f = 0$ 인가를 검정하는 것과 동일하다. 이러한 “no effect” 모형에 대한 검정은 고전적 적합도 검정법이지만 다른 모형들을 검정함에 있어서 기초가 되는 중요한 검정법이다. 다음절에서는 위의 가설에 대한 Buckley 검정과 von Neumann의 검정과 Eubank와 Hart의 검정법들을 소개하고, 3절에서는 Gasser, Sroka와 Jennen-Steinmetz의 GSJS 오차분산추정량을 다루고, 4절에서는 GSJS추정량과 Rice추정량을 2절에서 소개한 검정통계량에 각각 적용하여 모의 실험결과를 제시하고 비교분석 할 것이다.

2. 비모수 선형모형에 대한 적합도 검정들

Buckley(1991)는 베이지안 관점으로부터 H_0 에 대한 검정을 유도했다. 그는 모수 b 에 대해 $f = bg$ 을 이용하여 귀무가설 H_0 과 동일한 가설인 $b = 0$ 에 대한 검정을 생각했다. 이것을 바탕으로 $(g(t_1), \dots, g(t_n))$ 에 대한 특별한 사전분포를 가지고 $b = 0$ 에 대하여 부분적으로 최강력 검정을 유도했다. Buckley가 제시한 검정통계량은 다음과 같다.

$$S_B = n^{-1} \sum_{k=1}^n \left(\sum_{j=1}^k (y_j - \bar{y}) \right)^2 / \hat{\sigma}^2 \quad (2)$$

이때 $\bar{y} = n^{-1} \sum_{i=1}^n y_i$ 이고,

$$\hat{\sigma}^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (y_{i+1} - y_i)^2. \quad (3)$$

여기서 제시한 $\hat{\sigma}^2$ 은 σ^2 의 일치추정량으로서 Rice(1984)에 제시된 Rice 분산 추정량이다. Eubank와 Hart(1993)는 표본 푸리에 코사인 급수의 계수, \tilde{a}_{kn} 을 가지고 Nair(1986)의 결과에 의해 S_B 가 다음과 같음을 보였다.

$$S_B = \left(\sum_{k=1}^{n-1} \frac{\tilde{a}_{kn}^2}{\gamma_k} \right) / \hat{\sigma}^2. \quad (4)$$

여기서

$$\tilde{a}_{kn} = \frac{\sqrt{2}}{\sqrt{n}} \sum_{j=1}^n y_j \cos(k\pi t_j), \quad k = 1, \dots, n-1, \quad (5)$$

이고 $\gamma_k = [2n \sin(\frac{k\pi}{2n})]^2$.

von Neumann(1941)은 자료들의 첫번째 차분들에 대한 제곱합에 기초한 검정통계량을 다음과 같이 제시하였다.

$$S_N = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\hat{\sigma}^2}. \quad (6)$$

그리고 Munson과 Jernigan (1989)은 von Neuman의 검정통계량을 변형 개발하였고, 그들의 통계량은 회귀함수가 상수일때 Darbin과 Waston(1971)에 의한 검정통계량과 같아진다. Eubank와 Hart(1993)는 푸리에 코사인 급수의 계수를 이용하여 von Neumann 검정통계량이 다음의 평활 검정통계량과 점근적으로 같아짐을 보였다.

$$S_N = \frac{1}{\hat{\sigma}^2} \sum_{j=1}^{n-1} \tilde{a}_{jn}^2. \quad (7)$$

이때 $\hat{\sigma}^2$ 는 Rice의 분산 추정량을 사용하였다.

Eubank와 Hart(1993)는 f 에 대한 추정량으로서 푸리에 코사인 급수 추정량

$$f_m(t) = \sqrt{2} \sum_{j=1}^m \tilde{a}_{jn} \cos(j\pi t) / \sqrt{n}, \quad 1 \leq m \leq n, \quad (8)$$

을 사용하여 다음과 같은 추정량을 제시하였다.

$$S_E = \frac{\sum_{k=1}^m \tilde{a}_{jn}^2}{\hat{\sigma}^2} \quad (9)$$

이때 만약 $m = n - 1$ 인 경우 Eubank와 Hart의 추정량은 위에서 말한 von Neuman의 추정량인 S_N 과 같아진다. 그러나 Eubank와 Hart는 $m \ll n$ 인 보다 일반적인 경우를 강조하였고, f 가 제곱 적분 도함수를 가지는 연속함수라면 표준 평활실행은 m 이 $n^{1/3}$ 의 비율보다 빠르지 않게 커진다는 것을 보였다. 그러나 Eubank와 Hart(1993)는 m 을 찾는 방법을 소개하지 않았지만, Eubank와 Hart(1992)의 연구에서

$$\sum_{j=1}^m \tilde{a}_{jn}^2 - 2\hat{\sigma}^2 m/n \quad (10)$$

을 최대로 하는 m 을 선택하는 방법을 제시하였다. 다음절에서는 Rice의 분산추정량 보다 뛰어난 성질을 가지는 GSJS 추정량을 소개할 것이다.

3. 오차분산 추정

Gasser와 Sroka, Jennen-Steinmetz(1986)의 연구에서, 그들이 제시한 2차 차분 (second order differences)을 이용한 오차분산의 추정량은 편의와 분산을 줄이는 작용을 하였다. Gasser와 Sroka, Jennen-Steinmetz(1986)이 제시한 추정량을 GSJS 추정량이라고 이름을 짓

고 $\hat{\sigma}_{GSJS}^2$ 이라고 표기하겠다. GSJS의 추정량을 연구하기 앞서, 먼저 차분에 기저한 분산 추정에 있어서 의사오차(pseudo-residuals), $\tilde{\epsilon}_i$ 는 t_{i-1}, t_i, t_{i+1} 의 계획된 점들(design points)의 연속적인 삼중관계(continuous triples)에 의해 얻어지는데 일직선상의 두 개의 외곽 관찰값(outer observations)들을 결합하고 난 다음 이 일직선과 그 중앙의 관찰값 $y(t_i)$ 사이의 차분을 계산한다. 수식적으로 표현하면 다음과 같다.

$$\tilde{\epsilon}_i = \frac{t_{i+1} - t_i}{t_{i+1} - t_{i-1}} y(t_{i-1}) + \frac{t_i - t_{i-1}}{t_{i+1} - t_{i-1}} y(t_{i+1}) - y(t_i) \quad (11)$$

$$= a_i y(t_{i-1}) + b_i y(t_{i+1}) - y(t_i), i = 2, \dots, n-1. \quad (12)$$

GSJS 추정량의 기본적인 개념은 위의 오차들이 $f = 0$ 일때 $\tilde{\epsilon}_i^2$ 이 σ^2 에 대한 불편성을 갖게 하기 위하여 정규화(normalization)를 시키고 이러한 정규화 된 오차의 평균을 σ^2 의 추정량으로서 사용한다는 것이다. 다음은 의사 오차를 행렬로 표현하여 오차의 분산에 대한 GSJS 추정량을 구한것이다. 식(12)을 표현하기 위해 $(n-2) \times n$ 행렬 A를 다음과 같이 두자.

$$A = \begin{pmatrix} a_2 & -1 & b_2 & 0 & \dots & 0 \\ 0 & a_3 & -1 & b_3 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & a_{n-1} & -1 & b_{n-1} \end{pmatrix} \quad (13)$$

또한 B는 연속적인 삼중관계의 가중치를 나타내는 $(n-2) \times (n-2)$ 의 대각 행렬(diagonal matrix)로서 다음과 같이 표현된다.

$$B = \text{diag}\left[\frac{1}{\sqrt{a_2^2 + b_2^2 + 1}}, \frac{1}{\sqrt{a_3^2 + b_3^2 + 1}}, \dots, \frac{1}{\sqrt{a_{n-1}^2 + b_{n-1}^2 + 1}}\right] \quad (14)$$

$$= \text{diag}[c_2, c_3, \dots, c_{n-1}]. \quad (15)$$

위의 두 행렬을 이용하여 $D = BA$ 는 $(n-2) \times n$ 행렬로서 다음과 같이 계산이 된다.

$$D = \begin{pmatrix} c_2 a_2 & -c_2 & c_2 b_2 & 0 & \dots & 0 \\ 0 & c_3 a_3 & -c_3 & c_3 b_3 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & c_{n-1} a_{n-1} & -c_{n-1} & c_{n-1} b_{n-1} \end{pmatrix} \quad (16)$$

그러므로 식 (12)의 의사오차 $\tilde{\epsilon} = Dy = Df + D\epsilon$ 는 다음과 같은 행렬로서 표현되어진다.

$$\tilde{\epsilon} = Dy = \begin{pmatrix} c_2 a_2 y_1 & -c_2 y_2 & +c_2 b_2 y_3 \\ c_3 a_3 y_2 & -c_3 y_3 & +c_3 b_3 y_4 \\ \dots & \dots & \dots \\ c_{n-1} a_{n-1} y_{n-2} & -c_{n-1} y_{n-1} & +c_{n-1} b_{n-1} y_n \end{pmatrix} \quad (17)$$

그러면 의사오차 $\tilde{\epsilon}$ 의 오차분산의 추정량 $\hat{\sigma}_{GSJS}^2$ 은 다음과 같이 제시되어진다.

$$\hat{\sigma}_{GSJS}^2 = \frac{\tilde{\epsilon}^T \tilde{\epsilon}}{tr(D^T D)} = \frac{y^T D^T D y}{tr(D^T D)} = \frac{1}{(n-2)} \sum_{i=2}^{n-1} \tilde{\epsilon}_i^2. \quad (18)$$

여기서

$$\tilde{\epsilon}_i = c_i(a_i y_{i-1} + b_i y_{i+1} - y_i). \quad (19)$$

행렬 W 를 $W = D^T D$ 로 두면, σ^2 에 대한 추정량들의 일반적인 형태는 다음과 같은 자료의 이차형태(quadratic form)로 표현되어진다.

$$\hat{\sigma}_{GSJS}^2 = y^T W y / tr(W). \quad (20)$$

여기서 W 는 대칭이며 양정치(positive-definite) 행렬이다. 분모에있는 $tr(W)$ 은 추정량이 $f = 0$ 일때 σ^2 에 대한 불편성 추정량임을 보장하여 준다.

4. 모의 실험에 의한 검정력 비교

이절에서는 모의실험을 이용하여 오차분산 추정량인 GSJS 추정량을 이용하여 Rice 분산 추정량을 적용한 기존의 검정통계량인 von Neuman 검정통계량, Buckley 검정통계량, Eubank와 Hart의 검정통계량에 대해 검정력 비교를 할 것이다. 편의상 Rice 오차 분산 추정량을 사용한 검정통계량을 S_{BR} , S_{NR} , S_{ER} 로 두고 GSJS 오차분산 추정량을 사용한 검정통계량을 S_{BG} , S_{NG} , S_{EG} 로 두자. 그리고 모형(1)에 있는 $t_i = (i - 1/2)/n$, $i = 1, \dots, n$ 로 두고 함수 f 에 대해서는 다음의 6개의 모형들을 조사하였다.

$$\text{모형 A: } f_1(t) = t + .5 \exp(-50(t - .5)^2). \quad (21)$$

$$\text{모형 B: } f_2(t) = 5. \exp(-5t). \quad (22)$$

$$\text{모형 C: } f_3(t) = \exp(-t/2) \cos(t). \quad (23)$$

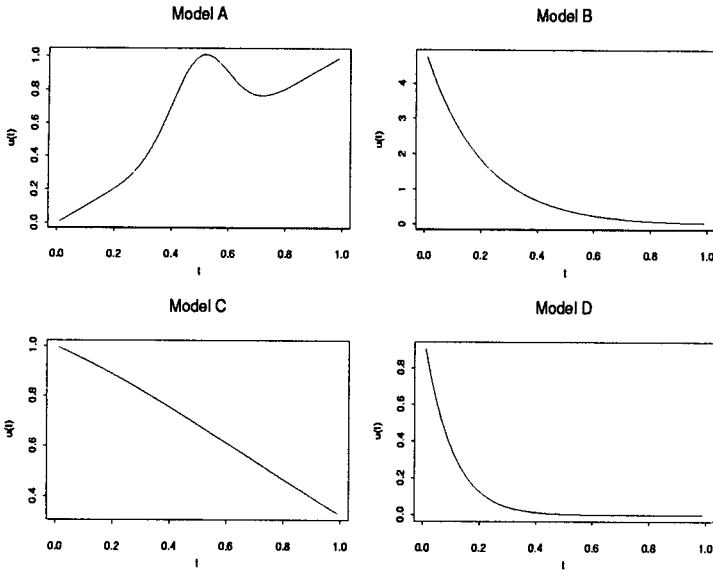
$$\text{모형 D: } f_4(t) = \exp(-t/0.1) \cos(t/0.01). \quad (24)$$

$$\text{모형 E: } f_5(t) = b(e^{4t} - \frac{e^4 - 1}{4}) (\frac{e^8 - 1}{8} - \frac{(e^4 - 1)^2}{4})^{-1/2}. \quad (25)$$

$$\text{모형 F: } f_6(t) = \lambda \cos(j\pi t). \quad (26)$$

모형 E에서 $b = i/4$, $i = 1, 2, 3, 4$ 인 경우에 대하여 조사하였고, 모형 F인 경우에는 $\lambda = 1$ 에서의 $j = 1, 3, 5, 7$ 인 진동의 변화에 따른 검정력을 조사하였다. 그림 1은 모형 A에서 모형 D까지의 4개의 함수 f 들에 대한 분포를 조사한 것이고, 그림 2는 모형 E에서 $b = i/4$, $i = 1, 2, 3, 4$ 인 경우의 분포의 폭의 변화에 대하여 조사하였고, 그림 3은 모형 F에서 $\lambda = 1$ 에서의 $j = 1, 3, 5, 7$ 인 진동의 변화에 대한 분포를 조사한 것이다.

그림 1 : 모형 A-D에 대한 대립가설 함수 f 의 분포



각 모형들의 검정력에 대한 비교 분석을 위하여 오차는 평균 0이고 분산의 값은 $\sigma^2 = 0.005, 0.01, 0.05, 0.1, 0.5, 1.0$ 인 정규 분포를 따르게 하였다. 그리고 관찰값들에 대한 표본의 크기는 임의로 10인 경우로 제한하였다. 검정력을 구하기 위한 모의 실험의 반복 횟수는 1000으로 하였고, 계산시 검정력의 유의수준은 $\alpha = 0.05$ 로 설정하였다. 귀무가설이 참인 경우 기각 영역의 값은 $f = 0$ 인 경우에서의 모의 실험을 이용하여 $\sigma^2 = 0.005, 0.01, 0.05, 0.1, 0.5, 1.0$, 반복횟수 1000번을 사용하여 구하였다.

우리는 다음의 표 1 - 표 12에 있는 모의 실험 결과들에서 매우 흥미로운 사실들을 발견할 수 있다. 표 1 - 표 9에 있는 모형 A - 모형 E와 모형 F (1) 까지의 결과들과 표 10 - 표 12에 있는 모형 F (2)에서 모형 F (4) 까지의 결과들이 매우 상반된 검정력을 가진다는 것을 알 수 있다. 이는 대립 분포들이 그 자신이 가지고 있는 진동수에 따라 검정의 결과에 큰 영향을 미치고 있음을 알게 한다.

먼저 표 1에서 표 8에 있는 모형 A 에서 모형 E 까지의 검정력의 결과들을 조사해 보자.

오차분산의 값들의 변화에 따른 검정력의 결과에 있어서 모형 A 에서 모형 F (1)에 대한 표 1에서 표 9까지의 검정력에 대한 비교 결과들에서 분산의 값이 적을 수록 강한 검정력을 가지고 분산의 값이 증가 할 수록 검정력의 값이 떨어짐을 알 수 있다. 이는 분산의 값이 증가 할 때에는 대립 분포가 주는 정보량이 매우 감소하여서 생기는 현상이다. 표 1과 표 3 - 표 6에 있는 모형 A와 모형 C - 모형 E (2) 들에 대한 결과들에서 분산의 값이 클 때 검정력의 값이 매우 감소됨이 관찰된다. 또한 한가지의 사실을 더 알 수 있는 것은 표 2와 표 6 - 표 7에 있는 진폭이 큰 모형 B, 모형 E (3), E (4)와 표 1과 표 3 - 표 6에 있는 진폭의 적은 모형 A, 모형 C - 모형 E (2)를 비교해 볼 때 분포의 진폭이 클 수록 검정력의 값이 증

그림 2 : 모형 E에 대한 대립가설 함수 f 의 분포

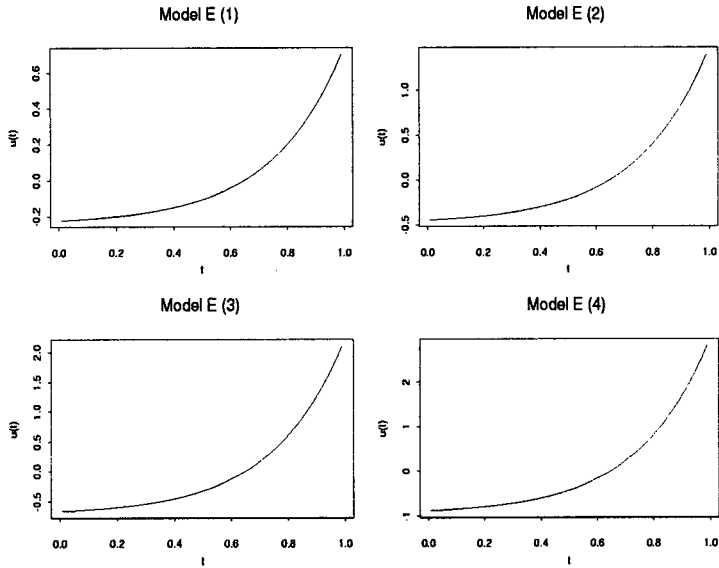


그림 3 : 모형 F에 대한 대립가설 함수 f 의 분포

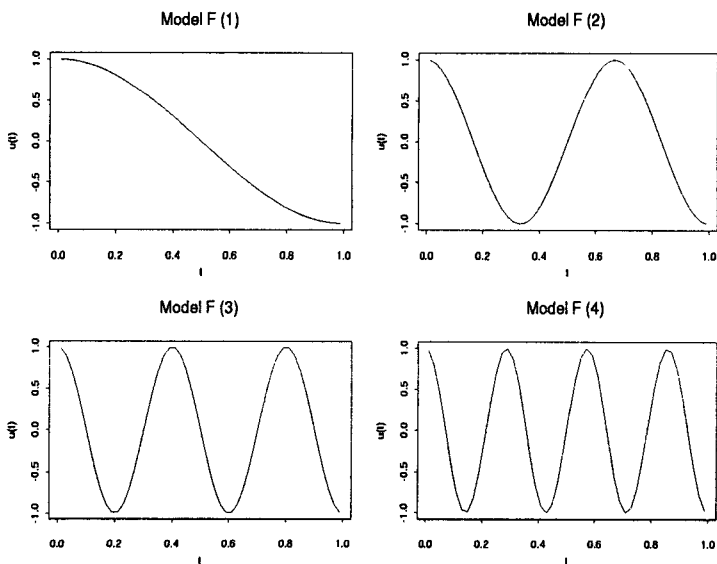


표 1: 모형 A의 분산 추정에 대한 검정력 비교

	σ^2	0.005	0.01	0.05	0.10	0.50	1.00
S_{BR}		1.000	1.000	0.927	0.673	0.183	0.128
S_{BG}		1.000	1.000	0.935	0.710	0.226	0.168
S_{NR}		1.000	1.000	0.849	0.529	0.171	0.125
S_{NG}		1.000	1.000	0.876	0.609	0.266	0.236
S_{ER}		1.000	1.000	0.852	0.536	0.172	0.124
S_{EG}		1.000	1.000	0.879	0.611	0.266	0.236

가 함을 알 수 있다. 특히 표 5 - 표 8의 모형 E (1), (2), (3), (4)에서 진폭의 변화에 따른 검정력의 값을 비교해 봄으로서 우리는 이 사실을 명백히 알 수 있다.

표 1에서 표 9까지의 결과들에서 Berckley의 검정 통계량이 거의 대부분의 경우에 있어서 von Neumann 이나 Eubank와 Hart의 검정 통계량들 보다 더 강력한 검정력을 가지고 있음이 관찰되나 표 10에서 표 12까지의 결과들에서 진동의 수가 증가할 수록 Berckley의 검정 통계량의 검정력이 매우 떨어짐을 알 수 있다. 이것은 Eubank 와 Hart(1993)의 결과에서도 알려진 사실이다.

Rice의 분산 추정 기법에 의한 통계량의 검정력 보다 GSJS 분산 기법에 의한 검정 통계량이 보다 좋은 검정 통계량을 가지고 있음을 알 수 있다.

마지막으로 모형 표 9 - 표 12에 있는 E (1), (2), (3), (4)에서의 검정의 결과는 매우 흥미로운 사실을 가지고 있다. 이는 진동수가 증가함에 따라 Berckley의 검정 통계량이나 von Neumann, Eubank와 Hart의 검정 통계량들은 검정 통계량으로서의 자신의 역할을 충분히 하지 않음을 알 수 있다. 즉 진동수가 일정한 구간안에 증가 할 수록 진동수에 따른 변화를 진단하는데 매우 둔함을 알 수 있다. 이 문제에 대한 해결 방안은 Kim (1995)에서 찾을 수 있을 것이다.

표 2: 모형 B의 분산 추정에 대한 검정력 비교

σ^2	0.005	0.01	0.05	0.10	0.50	1.00
S_{BR}	1.000	1.000	1.000	1.000	0.934	0.730
S_{BG}	1.000	1.000	1.000	1.000	0.957	0.775
S_{NR}	1.000	1.000	1.000	1.000	0.907	0.683
S_{NG}	1.000	1.000	1.000	1.000	0.924	0.762
S_{ER}	1.000	1.000	1.000	1.000	0.907	0.681
S_{EG}	1.000	1.000	1.000	1.000	0.924	0.762

표 3: 모형 C의 분산 추정에 대한 검정력 비교

σ^2	0.005	0.01	0.05	0.10	0.50	1.00
S_{BR}	1.000	0.997	0.657	0.406	0.120	0.079
S_{BG}	1.000	0.997	0.676	0.437	0.137	0.105
S_{NR}	0.999	0.974	0.480	0.276	0.101	0.078
S_{NG}	0.999	0.964	0.539	0.361	0.194	0.174
S_{ER}	0.999	0.979	0.488	0.277	0.102	0.078
S_{EG}	0.999	0.964	0.543	0.364	0.195	0.174

표 4: 모형 D의 분산 추정에 대한 검정력 비교

σ^2	0.005	0.01	0.05	0.10	0.50	1.00
S_{BR}	0.950	0.759	0.292	0.207	0.068	0.058
S_{BG}	0.992	0.912	0.382	0.267	0.108	0.090
S_{NR}	0.962	0.729	0.278	0.159	0.076	0.074
S_{NG}	0.996	0.931	0.475	0.306	0.198	0.179
S_{ER}	0.961	0.743	0.281	0.162	0.078	0.070
S_{EG}	0.996	0.937	0.479	0.309	0.200	0.178

표 5: 모형 E (1)의 분산 추정에 대한 검정력 비교

	σ^2	0.005	0.01	0.05	0.10	0.50	1.00
S_{BR}	1.000	1.000	0.761	0.468	0.131	0.103	
S_{BG}	1.000	1.000	0.797	0.493	0.164	0.140	
S_{NR}	1.000	0.996	0.632	0.332	0.120	0.104	
S_{NG}	1.000	0.995	0.702	0.436	0.222	0.202	
S_{ER}	1.000	0.997	0.634	0.334	0.119	0.104	
S_{EG}	1.000	0.995	0.704	0.442	0.220	0.201	

표 6: 모형 E (2)의 분산 추정에 대한 검정력 비교

	σ^2	0.005	0.01	0.05	0.10	0.50	1.00
S_{BR}	1.000	1.000	1.000	0.967	0.363	0.231	
S_{BG}	1.000	1.000	1.000	0.975	0.408	0.269	
S_{NR}	1.000	1.000	0.992	0.893	0.281	0.202	
S_{NG}	1.000	1.000	0.994	0.908	0.388	0.300	
S_{ER}	1.000	1.000	0.992	0.896	0.282	0.200	
S_{EG}	1.000	1.000	0.994	0.910	0.389	0.298	

표 7: 모형 E (3)의 분산 추정에 대한 검정력 비교

	σ^2	0.005	0.01	0.05	0.10	0.50	1.00
S_{BR}	1.000	1.000	1.000	0.999	0.676	0.430	
S_{BG}	1.000	1.000	1.000	0.999	0.713	0.482	
S_{NR}	1.000	1.000	1.000	0.997	0.557	0.359	
S_{NG}	1.000	1.000	1.000	0.995	0.631	0.464	
S_{ER}	1.000	1.000	1.000	0.997	0.556	0.358	
S_{EG}	1.000	1.000	1.000	0.995	0.635	0.463	

표 8: 모형 E (4)의 분산 추정에 대한 검정력 비교

σ^2	0.005	0.01	0.05	0.10	0.50	1.00
S_{BR}	1.000	1.000	1.000	1.000	0.885	0.643
S_{BG}	1.000	1.000	1.000	1.000	0.902	0.675
S_{NR}	1.000	1.000	1.000	1.000	0.806	0.557
S_{NG}	1.000	1.000	1.000	1.000	0.848	0.631
S_{ER}	1.000	1.000	1.000	1.000	0.808	0.555
S_{EG}	1.000	1.000	1.000	1.000	0.848	0.631

표 9: 모형 F (1)의 분산 추정에 대한 검정력 비교

σ^2	0.005	0.01	0.05	0.10	0.50	1.00
S_{BR}	1.000	1.000	1.000	1.000	0.721	0.472
S_{BG}	1.000	1.000	1.000	1.000	0.729	0.487
S_{NR}	1.000	1.000	1.000	0.997	0.578	0.343
S_{NG}	1.000	1.000	1.000	0.991	0.605	0.434
S_{ER}	1.000	1.000	1.000	0.997	0.581	0.341
S_{EG}	1.000	1.000	1.000	0.993	0.608	0.433

표 10: 모형 F (2)의 분산 추정에 대한 검정력 비교

σ^2	0.005	0.01	0.05	0.10	0.50	1.00
S_{BR}	0.000	0.000	0.000	0.000	0.017	0.022
S_{BG}	1.000	0.996	0.694	0.412	0.116	0.105
S_{NR}	1.000	1.000	0.973	0.736	0.188	0.135
S_{NG}	1.000	1.000	1.000	0.983	0.584	0.424
S_{ER}	1.000	1.000	0.975	0.744	0.191	0.130
S_{EG}	1.000	1.000	1.000	0.983	0.586	0.421

표 11: 모형 F (3)의 분산 추정에 대한 검정력 비교

σ^2	0.005	0.01	0.05	0.10	0.50	1.00
S_{BR}	0.000	0.000	0.000	0.000	0.008	0.014
S_{BG}	0.000	0.000	0.000	0.000	0.016	0.033
S_{NR}	0.000	0.000	0.000	0.000	0.006	0.014
S_{NG}	0.000	0.000	0.024	0.048	0.115	0.149
S_{ER}	0.000	0.000	0.000	0.000	0.006	0.013
S_{EG}	0.000	0.000	0.023	0.042	0.114	0.147

표 12: 모형 F (4)의 분산 추정에 대한 검정력 비교

σ^2	0.005	0.01	0.05	0.10	0.50	1.00
S_{BR}	0.000	0.000	0.000	0.000	0.005	0.010
S_{BG}	0.000	0.000	0.000	0.000	0.004	0.014
S_{NR}	0.000	0.000	0.000	0.000	0.002	0.008
S_{NG}	0.000	0.000	0.000	0.000	0.003	0.032
S_{ER}	0.000	0.000	0.000	0.000	0.002	0.008
S_{EG}	0.000	0.000	0.000	0.000	0.003	0.032

참고문헌

1. Buckley, M.J. (1991). Detecting a Smoothing Signal: Optimizing of Cusum Based Procedures, *Biometrika*, 78, 253-262.
2. Durbin, J. and Watson, G.S. (1971). Testing for Serial Correlation in least squares regression III, *Biometrika*. 58, 1-19.
3. Eubank, R.L. and Hart, J.D. (1992). Testing Goodness-of Fit in Regression via Order Selection Criteria, *The Annals of Statistics*, 20, 1412-1425.
4. Eubank, R.L. and Hart, J.D. (1993). Commonality of Cusum, von Neumann and Smoothing-Based Goodness-of-Fit Tests, *Biometrika*, 80, 89-98.
5. Gasser, T., Sroka, L., Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression, *Biometrika* 73, 625-33.
6. Kim, J.T. (1995). Goodness of Fit Test Based on Smoothing Parameter Selection Criteria, *The Korean Communications in Statistics*, 2, 122-136.
7. Munson, P.J. and Jernigan, R.W. (1989). A Cubic Spline Extension of The Durbin-Watson Test, *Biometrika*, 76, 39-47.
8. Nair, V.N. (1986). On Testing Against Ordered Alternatives in Analysis of Variance Model, *Biometrika*, 73, 493-499.
9. Rice, J. (1984). Bandwidth choice for nonparametric kernel regression, *Annals of Statistics*, 12, 1215-30.
10. von Neumann, J. (1941). Distribution of the Ratio of the Mean Squared Successive Difference to the Variance, *The Annals of Mathematical Statistics*. 12, 367-395.
11. 김 종태. (1998). 비모수 회귀모형의 차분에 기저한 분산의 추정에 대한 고찰, *The Korean Communications in Statistics*, 5, 121-131.

A Study On Variance Estimation in Smoothing Goodness-of-Fit Tests

Yonghwa Yoon⁵ · Jongtae Kim⁶ · Woo-Dong Lee⁷

Abstract

The goal of this paper is to study on variance estimation - Rice variance estimation, Gasser, Sroka and Jennen-Steinmetz's variance estimation - in smoothing goodness-of-fit tests. The comparisons of powers on test statistics are conducted by the change of variance, the number of oscillations, the amplitude of the alternative sample distribution.

Key Words and Phrases : Goodness-of-fit test, Residual variance Estimation, Fourier series, Nonparametric linear model.

⁵Professor, Dept. of Statistics, Taegu University, Kyungbuk 712-714

⁶Assistant Professor, Dept. of Statistics, Taegu University, Kyungbuk 712-714

⁷Assistant Professor, Dept. of Statistics, Kyungsan University, Kyungbuk 712-240