Journal of Statistical Theory & Methods
1997, Vol. 8, No. 1, pp. 9~20

퓨리에 급수기법에 의한 밀도함수추정의 최적화 고찰

김 종 태, 이 성 호, 김 경 무[1]

요약 밀도함수를 추정하는 방법에 있어서 퓨리에(Fourier) 급수기법과 핵(kernel) 기법, 스플라인(spline)평활기법들이 많은 통계학자들의 관심의 대상이 되어 왔다. 이 연구는 확률밀도함수의 추정에 있어서 전통적으로 각각 독립적으로 사용하여 왔던 정지규칙(stopping rule)과 승수규칙(selection multiplier)을 조합하여 퓨리에 급수기법을 이용한 새로운 추정기법을 연구하였다. 모의 실험을 통해 제시된 추정 기법이 기존의 연구기법들보다 다소 우월 하다는 결론을 얻었다.

주제어 : 밀도함수 추정, 푸리에 시리즈 급수

1. 서 론

일반적인 퓨리에 급수를 이용한 밀도함수 추정에 있어서 밀도함수 $f(x)$는 다음과 같이 표현되어진다.

$$f(x) = \sum_{k=-\infty}^{\infty} B_k \Phi_k(x).$$

이때 $\Phi_k(x)$는 정규직교의 조건을 만족하는 급수이고, $B_k = \int \Phi_k(x) f(x) dx$로 일반화된 퓨리에 계수들이다.

밀도함수의 추정에 있어서 크게 두가지 전략적인 방법이 통계학자들의 관심의 대상이 되어왔다. 첫번째 방법은 정지규칙(stopping rule), 밀도함수를 최적화시키는 평활모수(smoothing parameter)를 찾는 방법으로 다음과 같은 추정함수로서 표현되어진다.

$$\hat{f}(x) = \sum_{k=-m}^{m} \hat{B}_k \Phi_k(x).$$

이때 m은 평활모수이고, $\hat{B}_k$는 표본으로부터 추정된 계수로서 $\hat{B}_k = n^{-1}\sum_i \Phi_k(x_i)$ 이고 B_k의 불편추정량이 된다. Tarter 와 Kronmal(1976), Hart(1985), Diggle 과 Hall(1986) 등이 정지규칙에 대한 연구를 하여왔다.

두 번째 방법은 밀도함수에 승수(multiplier), b_k라고 하는 가중모수를 곱하여 밀도함수를 최적화 시키는 b_k를 찾는 전략이다. 즉 승수 b_k는 k가 증가함에 따라 '0'에 수렴함으로서 밀도함수를 추정할 수 있다. 추정된 밀도함수의 표현식은 다음과 같다.

---

[1] (712-714) 경북 경산시 진량면 내리동 15, 대구대학교 통계학과

$$f^{*}(x)=\sum_{k=-\infty}^{\infty} b_k \hat{B}_k \Phi_k(x), \quad 0<b_k \le 1.$$

여기에 대한 연구는 Waston (1969), Fellner (1974), Wahba (1981), 그리고 Lock (1991) 등에 의해 연구되어져 왔다. 승수규칙의 전략은 정지규칙의 절단전략보다 최적화과정에서 잠재적으로 더 큰 제어를 제공한다. 이러한 추정치들의 정확도를 평가하기 위한 판단기준으로 평균누적제곱오차, $R=E\int\{\hat{f}(x)-f(x)\}^2dx$ 가 보편적으로 많이 사용되어져 왔다.

이 연구의 주된 목적은 정지규칙과 승수규칙의 두 기법을 조합하는 새로운 추정방법을 보이고, 기존의 각각의 추정방법들과 비교 분석하였다. 그 결과 두 기법을 조합한 새로운 기법이 기존의 두 전략에서 각각 밀도함수를 추정하는 것보다 더 최적화 된 기법임을 모의실험을 이용하여 보였다.

2. 정지규칙에 의한 밀도함수의 추정

2.1 개 요

밀도함수를 코사인 (cosine) 퓨리에 급수를 이용하여 다음과 같이 정의하자.

$$f(x)=\sum_{k=0}^{\infty} a_k \Phi_k(x), \quad 0<x<1. \tag{2.1}$$

이때 $a_k=\int_0^1\Phi_k(x)f(x)dx$ 이고, $\Phi_0(x)=1,\ \Phi_k(x)=\sqrt{2}\cos(k\pi x),\ k=1,2,3,\cdots$ 라고 할때 $\{\Phi_k(x)\}_{k=0}^{\infty}$ 는 $L^2(0,1)$ 에서 정규직교기저(orthonormal basis)이다. 이러한 코사인 수열의 사용은 계산을 간편하게 할 뿐 아니라 곡선의 변동을 잘 관찰할수 있는 이점이 있다.

식(2.1)에서 a_k 의 불편추정량 $\hat{a}_{kn}=(1/n)\sum_{i=1}^{n}\sqrt{2}\cos(k\pi x_i)$ 를 대입하여 밀도함수 추정량 $\hat{f}_m$ 을 다음과 같이 얻을수 있다.

$$\hat{f}_m(x)=\sum_{k=0}^{m}\hat{a}_{kn}\sqrt{2}\cos(k\pi x). \tag{2.2}$$

여기서 m 은 절단점이다. 정지규칙에 있어서의 최적화 된 밀도함수를 찾는 방법은 밀도함수 f 와 추정량 $\hat{f}_m$ 에 대한 차이를 최소화 하는 m 을 선택하는 것이다.

2.2 Diggle 과 Hall 의 정지규칙

Diggle 과 Hall(1986)은 평균누적제곱오차를 최소화하는 절단점 m 을 발견하기 위해 다음과 같은 연구를 하였다. 밀도함수 $f(x)$ 와 추정량 $\hat{f}_m(x)$ 의 차이에 대한 측도로서의 평균누적제곱오차 $R(m)$ 은 다음과 같다.

$$R(m)=\mathrm{E}\int_0^1\left(\hat{f}_m-f(x)\right)^2dx$$

$$= \sum_{k=1}^{m} \sigma_k^2 + \sum_{k=m+1}^{\infty} a_k^2. \qquad (2.3)$$

이때 $\sigma_k^2 = Var(\hat{a}_{kn}) = 1/n\{1+(1/\sqrt{2})a_{2k} - a_k^2\}$이다. 평균누적제곱오차 $R(m)$에 대한 추정량 $\hat{R}(m)$은 σ_k^2의 불편추정량 $\hat{\sigma}_{kn}^2 = \{1+(1/\sqrt{2})\hat{a}_{2kn} - \hat{a}_{kn}^2\}/(n-1)$과 a_k^2의 불편추정량을 대입함으로 쉽게 얻을 수 있다. 따라서

$$\hat{R}(m) = \sum_{k=1}^{m} \hat{\sigma}_{kn}^2 + \sum_{k=m+1}^{\infty} \left(\hat{a}_{kn}^2 - \hat{\sigma}_{kn}^2\right). \qquad (2.4)$$

으로 표현된다. 식 (2.4)의 우편 첫 번째 항은 $\sum_{k=1}^{m} \hat{\sigma}_{kn}^2$은 n이 커짐에 따라 mn^{-1}로 수렴하는 것은 Hall(1982)에 의해 증명되었다. 두 번째 항은 만약 m이 무한대로 감에 따라 $\lambda(m)$의 값이 무한대로 가는 λ의 함수로 정의하면 다음과 같이 표현되어진다.

$$\sum_{k=m+1}^{\infty} \left(\hat{a}_{kn}^2 - \hat{\sigma}_{kn}^2\right) \sim \sum_{k=m+1}^{\lambda(m)m} \left(\hat{a}_{kn}^2 - \hat{\sigma}_{kn}^2\right).$$

따라서, 첫번째 항과 두 번째 항에 이 사실을 결합시키면 다음과 같은 $R(m)$의 추정값 $R^*(m)$을 구할 수 있다.

$$R^*(m) = \frac{m}{n} + \sum_{k=m+1}^{\lambda(m)m} \left(\hat{a}_{kn}^2 - \hat{\sigma}_{kn}^2\right).$$

Diggle과 Hall은 모의실험을 통해 $\lambda(m) = 4m^{1/2}$을 선택하고 $R^*(m)$을 최소화하는 m을 선택함으로 밀도함수를 최적화 하였다.

2.3 Hart에 의한 정지규칙

Hart(1985)는 평균누적제곱오차 $R(m)$을 다음과 같이 계산하였다.

$$R(m) = \sum_{k=1}^{m} \left(\sigma_k^2 - a_k^2\right) + \sum_{k=1}^{\infty} a_k^2$$

$$= -M(m) + \sum_{k=1}^{\infty} a_k^2. \qquad (2.5)$$

식 (2.5) 두 번째 항은 m에 의존하지 않으므로 평균누적오차 $R(m)$을 최소로 m의 값은 $M(m)$을 최대로 하는 m의 값과 일치한다. 여기서 σ_j^2과 a_j^2의 불편추정량을 대입하면 $M(m)$의 불편추정량을 다음과 같이 구할수 있다.

$$\hat{M}(m) = \sum_{k=1}^{m} \left(\hat{a}_{kn}^2 - 2\hat{\sigma}_{kn}^2\right). \qquad (2.6)$$

식 (2.6)을 최대로 하는 m은 $R(m)$을 최소로 하는 값과 동일함으로 이러한 평활모수 m을 이용하여 최적화된 확률밀도함수를 추정하는 것이 Hart에 의한 정지규칙 방법이다.

3. 승수규칙에 의한 밀도함수 추정방법

3.1 개 요

밀도함수의 추정에 있어서 승수 b_j를 곱하는 방법으로 추정된 밀도함수의 추정식을 다음과 같이 정의하자.

$$f^*(x) = \sum_{k=0}^{\infty} b_k \hat{a}_{kn} \sqrt{2}\cos(j\pi x), \quad 0 < x < 1. \tag{3.1}$$

이때 b_k는 $0 \le b_k \le 1$로서, $k \to \infty$에 따라 $b_k \to 0$이 되므로 어떤상수 m이 존재하여 $k \le m$일 경우 b_k는 b_k그 자체의 값을 가지고, 만약 $k > m$인 경우 $b_k = 0$이 된다고 가정하자. 그러면 식 (3.1)은 다음과 같이 된다.

$$f^*(x) = \sum_{k=0}^{m} b_k \hat{a}_{kn} \sqrt{2}\cos(j\pi x).$$

f^*와 f의 평균누적 제곱오차 $R(b_k)$은 다음과 같다.

$$R(b_k) = \sum_{k=1}^{m}\left(b_k^2\left(a_k^2 + \sigma_k^2\right) - 2b_k a_k^2\right) + \sum_{k=1}^{\infty} a_k^2. \tag{3.2}$$

이 때 σ_k^2 앞 장에서 정의한 것과 같다.

승수에 기저한 밀도함수 추정은 주어진 m에 대해 평균누적제곱오차 $R(b_k)$의 값을 가장 최소화하는 b_k를 찾는 방법이다.

3.2 Watson 의 승수 기법

Watson(1965)의 승수수열은 식 (3.2)의 $R(b_k)$를 최소화하기 위해 $R(b_k)$를 b_k에 대해 미분하여 구하는데 다음과 같다.

$$b_k^{\omega} = \frac{a_k^2}{\sigma_k^2 + a_k^2}, \quad k = 1,\ 2, \cdots \tag{3.3}$$

이때 구간 $[0,1]$ 안에서 a_k^2이 ‘0’으로 접근함에 따라 b_k^{ω} 도 ‘0’으로 간다. 따라서 k가 무한대로 접근함에 따라 퓨리에 급수 추정량의 계수들 a_k가 점차적으로 적어져서 ‘0’으로 접근하는 성질을 가진다. 그리고 식 (3.3)에 추정량을 대입하면 추정된 승수수열을 얻게 되는데 이 방법은 다음 장에서 중요한 역할을 하게 된다.

3.3 Fellner 의 승수 기법

Fellner(1974)의 승수기법은 평균누적제곱오차의 불편추정량 $\hat{R}(b_k)$를 최소화하는 승수를 찾는 방법이다. 식 (3.2)에 있는 $R(b_k)$의 불편추정량 $\hat{R}(b_k)$는 다음과 같이 얻어진다.

$$\hat{R}(b_k) = \sum_{k=1}^{m}\left(b_k^2 \hat{a}_{kn}^2 - 2b_k\left(\hat{a}_{kn}^2 - \hat{\sigma}_{kn}^2\right)\right) + \sum_{k=1}^{\infty}\left(\hat{a}_{kn}^2 - \hat{n}\sigma_{kn}^2\right). \tag{3.4}$$

Fellner 의 승수수열은 식 (3.4)를 최소화하는 b_k 를 얻기 위해 $\hat{R}(b_k)$에 대해 b_k를 미분하여 다음과 같은 b_k의 추정치 $\hat{b}_k^F = (\hat{a}_{kn}^2 - \hat{\sigma}_{kn}^2)/\hat{a}_{kn}^2$, $k = 1, 2, \cdots$ 을 얻었다.

4. 제안된 조합방법에 의한 밀도함수의 추정

4.1 제안된 밀도함수의 추정

이 연구의 주된 목적은 밀도함수 추정을 위한 앞의 두 장에서 설명한 두 전략을 조합하여 새로운 방법을 제시하는 것으로서, 정지규칙은 추정량에서 포함하는 항의 총 수인 m을 결정하는데 사용되고, 승수수열은 추정량의 처음 m항들에 적용되어진다. 앞에서 승수수열을 이용한 방법에서의 밀도함수 추정량은 평활모수 m을 선택하는 방법에 상관없이 이용될 수 있음을 알 수 있다. 따라서 먼저 정지규칙에 의해 최적화된 평활(smoothness)을 결정하는 평활모수 $\hat{m}$을 결정하고 다음으로 승수수열을 정지규칙에 의해 추정된 함수에 대입하여 새로운 밀도함수 추정식을 제시하는 전략이다.

식 (3.3)에서 Watson 의 승수수열에 대한 추정량을 σ_k^2에 대한 불편추정량 $\hat{\sigma}_{kn}^2$과 a_k에 대한 불편추정량 $\hat{a}_{kn}$를 대입하면 다음과 같은 새로운 추정량을 얻을 수 있다.

$$\hat{b}_k^{\omega} = \frac{\hat{a}_{kn}^2}{\hat{\sigma}_{kn}^2 + \hat{a}_{kn}^2} \tag{4.1}$$

그러므로 제안된 조합 방법에 의한 밀도함수의 추정식은 다음과 같다.

$$\hat{f}_m^*(x) = \sum_{k=0}^{m} \hat{b}_k^{\omega} \hat{a}_{kn} \sqrt{2}\cos(k\pi x).$$

다음 절에서 $\hat{b}_k^{\omega}$에 대한 통계학적 성질을 연구할 것이다.

4.2 제안된 승수 수열

3.3 절에서 Fellner 승수수열은 추정된 평균누적제곱합수를 최소화하는 b_k의 실수값들을 고려하였다. 승수수열을 선택하는 또다른 접근방법은 승수수열 그 자체의 추정량들로서 승수들을 취급하도록 하는 방법이다. 이 연구에서 제안된 승수수열은 Watson 의 수열 추정량을 새로운 승수수열으로서 이용한 것이다.

승수수열에 대한 특성은 다음과 같다. $b_k = 1$이면 原 퓨리에급수 추정량의 승수수열이 되며, n이 커짐에 따라 결과적으로 모든 승수수열들의 점근적 불편추정량이 되어진다. 즉, 이는 n이 무한대로 감에 따라 b_k^{ω}를 포함한 모든 승수수열은 '1'로 수렴하기 때문이다. 그러므로 작은 표본의 크기를 갖는 집단에서의 승수수열 b_k^{w}와 이에 대한 점근적 불편추정량 1과의 편차는 다음과 같이 나타낼 수 있다.

$$\mathrm{E}\left(1 - b_k^{\omega}\right) = \frac{\sigma_k^2}{\sigma_k^2 + a_k^2} .$$

이때 $\sigma_k^2 = (1/n)\{1 + (1/\sqrt{2})a_{2k} - a_k^2\}$, $a_k = (1/n)\{\sum_{i=1}^{n} \sqrt{2}\cos(k\pi x)\}$ 으로 $b_k = 1$에 대한 편차는 단

지 n^{-1}로서 0으로 감소한다.

그러나 이 연구에서 제안된 $\hat{b}_k^{\omega} = \hat{a}_{kn}^2 \big/ \left(\hat{\sigma}_{kn}^2 + \hat{a}_{kn}^2\right)$의 편차는 n^{-1}로서 '0'으로 감소함으로 보다 효율적이고 좋은 승수수열을 얻을수 있다.

이러한 주장을 증명하기 위해 다음의 보조정리들을 필요로 한다.

보조정리 4.1 1) $\mathrm{E}\left(\hat{a}_{kn}^4\right) = a_k^4 + \mathrm{O}\left(n^{-1}\right)$.

2) $\mathrm{E}\left(\hat{a}_{kn}^2\right) = a_k^2 + \mathrm{O}\left(n^{-1}\right)$.

증 명. Ross(1976)의 multinomial 정리를 이용하면 계산과정은 매우 길지만 쉽게 얻을 수 있다.

보조정리 4.2 $Var\left(\hat{a}_{kn}^2\right) = \mathrm{O}\left(n^{-1}\right)$.

증 명.
$$\begin{aligned} Var\left(\hat{a}_{kn}^2\right) &= \mathrm{E}\left(\hat{a}_{kn}^4\right) - \left(\mathrm{E}\left(\hat{a}_k^2\right)\right)^2 \\ &= a_k^4 + \mathrm{O}\left(n^{-1}\right) - \left(a_k^2 + \mathrm{O}\left(n^{-1}\right)\right)^2 \\ &= \mathrm{O}\left(n^{-1}\right). \end{aligned}$$

보조정리 4.3 $MSE\left(\hat{a}_{kn}^2, a_k^2\right) = \mathrm{E}\left(\hat{a}_{kn}^2 - a_k^2\right)^2 = \mathrm{O}\left(n^{-1}\right)$.

증 명.
$$\begin{aligned} \mathrm{E}\left(\hat{a}_{kn}^2 - a_k^2\right)^2 &= Var\left(\hat{a}_{kn}^2\right) + \left(\mathrm{E}\left(\hat{a}_{kn}^2\right) - a_k^2\right)^2 \\ &= \mathrm{O}\left(n^{-1}\right). \end{aligned}$$

위의 보조정리들을 이용하여 이 절의 중요한 정리 4.1을 증명할 수 있다.

정리 4.1 b_k^{ω}와 $\hat{b}_k^{\omega}$은 각각 식 (3.2)와 식 (4.1)에 의해 정의되고 $\hat{a}_{kn}$과 a_k가 둘다 '0' 이 아닐때, $\mathrm{E}\left(\hat{b}_k^{\omega}\right) = b_k^{\omega} + \mathrm{O}(n^{-2})$이다.

증 명. $\hat{a}_{kn}^2$과 a_k는 [0,1]안에 영역을 가지는 실수이다. 그리고 $\hat{\sigma}_{kn}^2$은 σ_k^2의 불편추정량으로 $n\hat{\sigma}_{kn}^2$은 '1'로 수렴 한다는 것이 Hall(1982)에 의해 증명되었고, 이를 이용하여 b_k^{ω}와 $\hat{b}_k^{\omega}$을 다음과 같이 쓸 수 있다.

$$b_k^{\omega} \cong \frac{na_k^2}{n\left(a_{kn}\right)^2 + 1} .$$

$$\hat{b}_k^{\omega} \cong \frac{\left(\hat{a}_{kn}\right)^2}{\left(\hat{a}_{kn}\right)^2 + \frac{1}{n}} = \frac{n\left(\hat{a}_{kn}\right)^2}{n\left(\hat{a}_{kn}\right)^2 + 1} .$$

한편 다음의 연속실수함수를 생각하여 보자.

$$h(x) = \frac{nx}{nx + 1} .$$

이때 $h\left(a_k^2\right) = b_k^{\omega}$이다. 그러면 주어진 $h(x)$는 [0,1] 에서 두번 연속 미분이 가능하므로 a_k^2에서 $h\left(\hat{a}_{kn}^2\right)$에 대한 근사식을 Taylor 정리를 이용하여 다음과 같이 구할 수 있다.

$$\hat{b}_k^{\omega} = h\left(a_k^2\right) + h'\left(a_k^2\right)\left(\hat{a}_{kn}^2 - a_k^2\right) + R_n$$

$$= b_k^{\omega} + \frac{n}{\left[na_k^2 + 1\right]^2}\left(\hat{a}_{kn}^2 - a_k^2\right) + R_n \ .$$

이때 $R_n = (1/2)h''(\xi)\left(\hat{a}_{kn}^2 - a_k^2\right)^2 = \{-n^2/[n\xi + 1]^2\}(\hat{a}_{kn}^2 - a_k^2)^2$ 이고 ξ 는 $\hat{a}_{kn}^2$ 과 a_k^2 사이의 실수이고, $\left|R(n)\right| = O\left(n^{-1}\right)\left(\hat{a}_{kn}^2 - a_k^2\right)$ 이다.

그러므로 보조정리 4.1 과 보조정리 4.2 를 이용하여 다음을 구할 수 있다.

$$E\left(\hat{b}_k^{\omega}\right) = b_k^{\omega} + O\left(n^{-1}\right)E\left(\hat{a}_{kn}^2 - a_k^2\right) + O\left(n^{-1}\right)E\left(\hat{a}_{kn}^2 - a_k^2\right)^2$$

$$= b_k^{\omega} + O\left(n^{-2}\right) \ .$$

5. 모의실험

이 절에서 소개되는 모의실험은 Fellner 승수와 제시된 승수 절차들을 비교하기 위한 것 뿐만 아니라 이 주장의 타당성을 경험적으로 제공해 주기 위해 계획되었다. 그리고 결과들은 앞에서 설명된 정지규칙들의 실행을 비교하는데 사용될 수도 있다. 모의실험의 기본계획은 다음과 같다 :

(1) 알고 있는 분포 $f(x)$로 부터 확률표본이 생성되면 이 자료에서 표본의 개수 만큼의 표본 Fourier 계수 $\tilde{a}_{jn}$, $j = 1, 2, \cdots$ 이 계산된다.

(2) 2 개의 평활모수 m_{DH}, m_H 을 각각 Diggle-Hall 과 Hart 의 정지규칙에 의해 결정 한다.

(3) (2)단계에서 결정된 평활모수들을 이용해서 3 개의 밀도함수추정치가 계산된다 :

(a) m 항에서 절단된 原 Fourier 급수 추정치,

(b) 추정치에 승수수열 $b_j = \hat{b}_j^F$, $j = 1, 2, \cdots, m$, $b_j = 0$, $j > m$ 에 의해 결정된 Fellner 승수를 곱한 값,

(c) 추정치에 승수수열 $b_j = \hat{b}_j^{\omega}$, $j = 1, 2, \cdots, m$, $b_j = 0$, $j > m$ 에 의해 결정된 Watson 승수를 곱한 값 : 그래서, 각 표본마다 세개의 밀도함수 추정치가 구해진다.

(4) 각 세개 추정치의 평균누적제곱오차가 계산된다.

(5) (4)단계를 1000 번 반복해서 구한 1000 개 평균누적제곱오차의 평균을 계산한다. 각각 표본에서 평균누적제곱오차의 표준편차도 계산한다.

두 개의 다른 정지규칙에 두개의 승수를 결합시키기 위해서, 모의실험은 2 개의 분포와 몇 개의 표본크기에서 반복된다. 여기에서 사용되는 분포는 표준정규분포, 베

타분포이며, 표본의 크기는 각각 25, 50, 100, 200 으로 실행하였다.

(1) Diggle Hall 정지규칙에 의한 추정량들의 비교

다음의 표에서 비교되고 있는 3 개의 밀도함수 추정량에 포함된 항의 개수 m_{DH}은 Diggle-Hall 정지규칙에 의해 결정되었다. 추정량들은 승수수열에 의해 결정된다 :

原 추정량 : $b_j = 1$, $|j| \le m_{DH}$, $b_j = 0$, $|j| > m_{DH}$;

Fellner 추정량 : $b_j = \hat{b}_j^F$, $|j| \le m_{DH}$, $b_j = 0$, $|j| > m_{DH}$;

제시된 추정량 : $b_j = \hat{b}_j^{\omega}$, $|j| \le m_{DH}$, $b_j = 0$, $|j| > m_{DH}$.

첫 번째 열은 밀도함수의 표본의 크기를 설명하고 있고, 각 분포와 표본크기에 대해 1000 개의 표본이 생성되었다. Diggle-Hall 정지규칙에 의해 선택된 항 개수의 평균이 두 번째 열에서 위의 숫자이고, 아래 숫자는 그것의 표준편차이다. 나머지 오른쪽 3 개의 열에서 위의 숫자는 原, Fellner, Watson 추정량에 대한 평균누적제곱오차이고, 아래 숫자는 그것의 표준편차이다. 평균누적제곱오차의 값이 작을수록 좋은 추정치임을 의미한다.

표 5.1 표준정규분포

표본 크기	항의 수(m)	原 추정량	Fellner 추정량	제시된 추정량
25	2790 (0.525)	0.13632 (0.10740)	0.12220 (0.11001)	0.11880 (0.11069)
50	1.8920 (0.592)	0.11310 (0.07317)	0.10780 (0.07395)	0.10657 (0.07421)
100	2.2400 (0.554)	0.07020 (0.04406)	0.06853 (0.04432)	0.06811 (0.04445)
200	2.5090 (0.715)	0.04338 (0.02655)	0.04262 (0.02663)	0.04246 (0.02667)

표 5.2 베타분포

모수	표본크기	항의 수(m)	原 추정량	Fellner 추정량	제시된 추정량
p=2.0 q=3.0	25	1.5970 (0.644)	0.18243 (0.11428)	0.17846 (0.11443)	0.17787 (0.11442)
	50	1.9680 (0.711)	0.12644 (0.06997)	0.12560 (0.06997)	0.12569 (0.06998)
	100	2.3110 (0.659)	0.08507 (0.04383)	0.08485 (0.04380)	0.08489 (0.04380)
	200	2.6840 (0.804)	0.05653 (0.02672)	0.05638 (0.02670)	0.05640 (0.02670)

p=2.0 q=2.0	25	1.5750 (0.589)	0.17637 (0.11052)	0.15781 (0.11175)	0.15336 (0.11240)
	50	1.8780 (0.526)	0.10981 (0.06151)	0.10034 (0.06179)	0.09827 (0.06186)
	100	2.0450 (0.446)	0.06619 (0.03742)	0.06163 (0.03734)	0.06051 (0.03732)
	200	2.1550 (0.507)	0.03896 (0.02069)	0.03662 (0.02066)	0.03603 (0.02065)

(2) Hart의 정지규칙에 의한 추정량들의 비교

다음의 표에서 비교되고 있는 3개의 밀도함수 추정량에 포함된 항의 개수 m_H은 Hart의 정지규칙에 의해 결정되었다. 추정량들은 승수수열에 의해 결정된다 :

原 추정량 : $b_j = 1$, $|j| \le m_H$, $b_j = 0$, $|j| > m_H$;

Fellner 추정량 : $b_j = \hat{b}_j^F$, $|j| \le m_H$, $|j| > m_H$;

제시된 추정량 : $b_j = \hat{b}_j$, $|j| \le m_H$, $b_j = 0$, $|j| > m_H$.

첫번째 열은 밀도함수의 표본의 크기를 설명하고 있고, 각 분포와 표본크기에 대해 1000개의 표본이 생성되었다. Hart의 정지규칙에 의해 선택된 항 개수의 평균이 두 번째 열에서 위의 숫자이고, 아래 숫자는 그것의 표준편차이다. 나머지 오른쪽 3개의 열에서 위의 숫자는 原, Fellner, 제시된 추정량에 대한 Hart의 $M(m)$값이고, 아래 숫자는 그것의 표준편차이다. $M(m)$의 값이 클수록 평균누적제곱오차는 작아진다.

표 5.3 표준정규분포

표본 크기	항의 수(m)	원 추정량	Fellner 추정량	제시된 추정량
25	1.45700 (1.718)	0.08905 (0.12598)	0.21839 (0.15278)	0.22201 (0.15427)
50	2.69300 (1.567)	0.23601 (0.17450)	0.28020 (0.17073)	0.28174 (0.17085)
100	3.10400 (1.452)	0.39137 (0.17247)	0.40479 (0.17365)	0.40549 (0.17376)
200	3.56400 (1.506)	0.53741 (0.16569)	0.54424 (0.16624)	0.54460 (0.16628)

표 5.4 베타분포

모수	표본크기	항의 수(m)	원 추정량	Fellner 추정량	제시된 추정량
p=2.0 q=3.0	25	2.60100 (1.555)	0.25999 (0.17769)	0.31417 (0.18786)	0.31750 (0.18848)
	50	3.32500 (1.639)	0.30735 (0.12686)	0.32891 (0.13068)	0.33010 (0.13090)
	100	3.76100 (1.603)	0.33052 (0.08789)	0.34003 (0.08837)	0.34057 (0.08839)
	200	4.41300 (1.560)	0.34810 (0.06312)	0.35290 (0.06278)	0.35319 (0.06274)
p=2.0 q=2.0	25	2.00500 (1.812)	0.11766 (0.14110)	0.24990 (0.15666)	0.25359 (0.06274)
	50	2.61600 (1.512)	0.14350 (0.10058)	0.19340 (0.09951)	0.19500 (0.09976)
	100	3.13100 (1.611)	0.17401 (0.07523)	0.19699 (0.07567)	0.19791 (0.07571)
	200	3.44900 (1.667)	0.18250 (0.04914)	0.19474 (0.04839)	0.19524 (0.04842)
p=3.0 q=2.0	25	2.65400 (1.614)	0.25760 (0.17887)	0.31262 (0.18663)	0.31604 (0.18729)
	50	3.24500 (1.624)	0.31161 (0.12561)	0.33153 (0.13017)	0.33275 (0.13049)
	100	3.66700 (1.582)	0.32842 (0.08854)	0.33742 (0.08939)	0.33793 (0.08945)
	200	4.30500 (1.503)	0.34724 (0.06446)	0.35177 (0.06412)	0.35024 (0.06409)

6. 결 론

표 5.1, 5.4 에서 Diggle-Hall 과 Hart 의 정지규칙에 Fellner 승수와 제시된 승수수열을 결합시킨 새로운 기법에서의 추정량과 原 추정량을 비교하였다. 모의실험을 통한 결과를 요약하면 다음과 같다.

Diggle-Hall 의 정지규칙에서 표본크기, 분포에 상관없이 일정한 m 에 대해서는 승수수열을 결합시킨 새로운 기법의 평균주적제곱오차 값이 原 추정량보다 작다. 또한 Hart 의 정지규칙에서도 표본크기, 분포에 상관없이 승수수열을 결합시킨 새로운 기법의 $M(m)$ 값이 原 추정량보다 크게 나타난다. 이는 $M(m)$ 의 값이 클수록 평균누적제곱오차가 작음을 의미한다.

따라서, 기존의 기법보다 새로운 기법이 더 최적화된 밀도함수를 추정할 수 있다는 근거를 제시해 준다.

정지규칙에 조합하는 Fellner 와 제시된 승수수열을 비교해 보면 매우 근소한 차이로 대부분 Fellner 추정량의 평균누적제곱오차 값이 작지만, 표준편차는 제시된 추정량보다도 크므로 Fellner 의 승수수열을 적용하는 것이 좋다고 단정하기는 어렵다.

정지규칙의 비교에서는 Hart 의 정지규칙이 Diggle-Hall 정지규칙보다 항의 수 m 도 많고, 표본크기가 증가할수록 항의 수 m 의 증가폭도 크다. 또한 표준편차도 Hart 의 정지규칙에서 높게 나타난다.

위의 결과들에서 살펴보면 기존의 최적 밀도함수 추정방법인 정지규칙이나 승수수열 방법은 각각 적용하는 것보다 두 방법을 조합한 새로운 방법을 사용하면 정지규칙, 분포, 표본크기에 상관없이 더 최적화된 밀도함수를 추정할 수 있음을 알 수 있다.

참 고 문 헌

1. Diggle, P. J. and Hall, P. (1986). The Selection of Terms in an Orthogonal Series Density Estimator, *Journal of American Statistical Association*, 81, 230-233.
2. Fellner, W. H. (1974). Heuristic Estimation of Probability Densities, *Biometrika*, 61, 485-492.
3. Hall, P. (1982). Comparison of Two Orthogonal Series Methods of Estimating a Density and its Derivatives on an Interval, *Journal of Multivariate Analysis*, 12, 432-449.
4. Hart, J. D. (1985). On the Choice of Truncation Point in Fourier Series Density Estimation, *Journal of Statistical Computation and Simulation*, 21, 95-116.
5. Lock, M. D. (1990). *Optimizing Density Estimates Based on Unweighted and Weighted Mean Integrated Square Error*, Unpublished doctoral thesis, University of California, Berkeley.
6. Ross, S. (1976). *A First Course in Probability*, Macmilian Publishing Company: New York.
7. Tarter, M. E. and Kronmal, R. A. (1976). An Introduction to the Implementation and Theory of Nonparametric Density Estimation, *American Statistician*, 30, 105-112.
8. Wahba, G. (1981). Data-based Optimal Smoothing of Orthogonal Series Density Estimates, *The Annals of Statistics*, 9, 146-156.
9. Watson, G. S. (1969). Density Estimation by Orthogonal Series, *The Annals of Mathematical Statistics*, 40, 1496-1498.

A study on Optimizing Fourier Series Density estimates

Jong-Tae Kim , Sung-Ho Lee , Kyung-Moo Kim[2]

Abstract

Several methods are proposed for optimizing Fourier series estimators with respect to Mean Integrated Square Error metrics. Traditionally, such method have followed. one of two basic strategies; A stopping rules or the rules of determine multipliers. A central hypothesis of this study is that better estimates can be obtained by combining the two strategies. A new multiplier sequence is proposed, which used in conjunction with any of the stopping rules, is shown to improve the performance of estimator which relies solely on a stopping rule.

[2] Department of Statistics, Taegu University, Taegu, 712-714, Korea