한국수학사학회지 제17권 제2호(2004년 5월), 15-20

비모수 통계학에서 밀도 추정의 평활에 관한 역사적 고찰

서경대학교 소프트웨어학과 이승우
swlee@skuniv.ac.kr

본 논문에서는 밀도 추정에 관한 통계량으로서 불편성과 일치성에 관하여 제시하고 밀도함수에 관한 평활 방법으로서 히스토그램과 커널 밀도 추정 및 극소적응평활(local adaptive smoothing)에 관하여 보이고자 한다. 그리고 과거에서 현재까지 비모수 밀도 추정에 관한 연구에 관하여 조사하고 논하고자 한다.

주제어 : 빈너비, 띠너비, 밀도평활, 극소적응평활

0. 서론

비모수 통계학의 연구 분야는 통계 분석에서 새로운 대안으로서 최근에 널리 연구되고 있는 최신 학문이며 비모수 밀도 추정(nonparametric density estimation)은 통계 연구 분야에서 가장 활발하게 연구되고 있는 주제 중에 하나이다. $X_1, X_2, \cdots, X_n$ 이 연속확률밀도함수 f로부터 d-차원의 표본에서 생성되고 다음을 만족시킨다고 하자.

$$f(x) \geq 0, \quad \int_{R^d} f(x)\,dx = 1$$

그러면 특정한 모형에 대한 가정 없이 확률밀도함수 f를 추정할 수 있다.

대부분 통계적 절차에서 모집단이나 확률변수가 특정한 종류의 분포를 따른다는 가정 하에서 논의되었으나, 비모수 통계학은 특정한 모형에 대한 가정 없이 추정할 수 있기 때문에 f는 특정한 모형에 포함되어 있는 모수, 즉 평균 및 분산과 같은 중요한 변수들만 정의된 모형에만 적용되는 것이 아니라 관측값들이 어떠한 확률표본을 표현하는지 또는 관측값의 순서가 어떠한 경향이 있는지를 알아낼 수 있다. 평활도(smoothness)의 조건으로서 확률밀도함수 f는 연속이며 미분 가능할 때 가능하다.

일변량 밀도 함수 f의 비모수 추정량으로서 히스토그램이 최초로 사용되었다. 커널 방법(kernel method), 직교열 방법(orthogonal series method), 최근접 이웃 방법(nearest neighbor method)는 비모수 판별 분석에서 응용되어졌으며 점차 발전되어졌다. 그 후 벌점 가능도(penalized likelihood), 다항 스플라인 모형(polynomial spline),

가변 커널(variable kernel), 사영 추적(projection pursuit) 등과 같은 방법들이 비모수 통계학의 다른 분야에 접목되어 소개되어졌다. 오늘날 컴퓨터의 급속한 발전으로 인하여 다양한 통계 패키지와 통계 소프트웨어의 발달 및 고화질의 그래픽으로 인하여 비모수 밀도 추정을 보다 더 정확하게 신뢰할 수 있는 기회를 제공하고 있다.

1. 밀도 추정량의 통계적 성질

유한 표본에서 비모수 밀도 추정량의 성질은 특별한 모형에 대한 가정을 배제하기 때문에 특이한 상황 하에서도 추정이 가능하다. 일반적으로 비모수 통계학의 연구는 대표본 하에서 확률밀도함수 f를 추정하였고 계속 연구 발전되어졌다.

1.1. 불편성(unbiasedness)

확률밀도함수 f의 추정량 $\hat{f}$가 모든 $x \in R^d$ 에 대하여 $E_f[\hat{f}(x)] = f(x)$ 라면 f에 대한 불편 추정량(unbiased estimator)이라고 부른다. 비모수 밀도 추정량에 대한 수열 $\{\hat{f}_n\}$이 모든 $x \in R^d$에 대하여 $E_f[\hat{f}(x)] \rightarrow f(x)$, $n \rightarrow \infty$라면 f에 대한 점근적 불편 추정량(asymptotically unbiased estimator)이라고 부른다.

1.2. 일치성(consistency)

만약 확률밀도함수 f가 $\hat{f}$의 불편 추정량이라면 f의 평균제곱오차(mean squared error)는 모든 $x \in R$에 대하여 다음과 같이 주어진다.

$$MSE(x) = E_f[\hat{f}(x) - f(x)]^2 = var[\hat{f}(x)] + \{bias[\hat{f}(x)]\}^2 \tag{1.1}$$

단, $var[\hat{f}(x)] = E_f\{\hat{f}(x) - E_f[\hat{f}(x)]\}^2$이고 $bias[\hat{f}(x)] = E_f[\hat{f}(x)] - f(x)$이다.

모든 $x \in R$에 대하여 $n \rightarrow \infty$일 때 $MSE(x) \rightarrow 0$이라면, $\hat{f}$를 이차평균(quadratic mean)에서 f의 점일치 추정량(pointwise consistent estimator)이라고 부른다.

적합도(goodness of fit)의 척도로서 평균제곱오차 (1.1)를 적분함으로서 누적평균제곱오차(integrated mean squared error)를 다음과 같이 정의한다.

$$IMSE = \int_{-\infty}^{\infty} E_f[\hat{f}(x) - f(x)]^2 dx \tag{1.2}$$

또 다른 척도로서 적분제곱오차(integrated squared error)를 다음과 같이 정의한다.

$$ISE = \int_{-\infty}^{\infty} [\hat{f}(x) - f(x)]^2 dx \tag{1.3}$$

식 (1.3)에서 f에 대하여 기대치를 취하면 평균적분제곱오차(mean integrated squared error)는 $MISE = E_f(ISE)$으로 주어진다. 그리고 $MISE = IMSE$이다.

주어진 관측값으로 생성된 ISE는 추정량 $\hat{f}$이 확률밀도함수 f와 어느 정도 차이가 있는지, 혹은 어느 정도로 근접해 있는지를 판단하는 기준으로 ISE가 $MISE$보다 추정량의 일치성을 판단하기에 보다 더 적합할 수 있다. 반면 $MISE$는 주어진 모든 관측값들의 평균을 사용한다. 비교적 약한 제한 조건 하에서, ISE는 $MISE$에 근사할 수 있다. 그러나 주어진 관측값으로서 통계 추정을 할 때, $MISE$는 ISE보다 더 유용한 판단 기준일 수 있다.

2. 밀도평활(density smoothing)

2.1. 히스토그램(histogram)

히스토그램은 확률밀도함수 f의 일반적인 형태에 관한 시각적인 정보를 제공하는데 사용된다. 1차원 확률밀도함수의 추정에 대한 히스토그램의 일반적인 계산은 다음과 같은 절차에 의하여 특성화된다.

- 실선을 여러 개의 빈(bin)들로 나눈다.

$$B_j = [x_0 + (j-1)h,\ x_0 + jh]\quad j \in Z$$

여기서 $h > 0$을 빈너비(binwidth)라 하고 x_0를 히스토그램의 원점이라고 하자.

- 각각의 빈(bin) 구간에 놓여있는 관측값의 개수를 계산한다.

일반화된 히스토그램의 함수 추정식은 다음과 같이 표현된다.

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n}\sum_{j} I(X_i \in B_j) I(x \in B_j) \tag{2.1}$$

h는 히스토그램에서 중요한 모수이다. h가 변화할 때마다 $\hat{f}_h(x)$의 여러 가지 다른 형태의 히스토그램을 얻을 수 있다. 만약 h를 증가시키면, 더 많은 관측값이 빈너비(binwidth)에 포함되기 때문에 히스토그램의 형태는 보다 더 매끈하게(smooth) 된다. 만약 h를 감소시키면, 즉 $h \to 0$이면 각 빈너비(binwidth)에 소량의 관측값이 포함되어 히스토그램의 추정은 매우 난해한 형태를 가진다. 만약 $h \to \infty$이면 히스토그램의 추정은 상자 모양(box-shape, overly smooth)의 형태를 가진다. 그러므로 히스토그램의 비모수 밀도 추정 하에서 h를 선택하는 방법이 가장 중요한 문제이다.

$H(\Omega)$를 Ω상에서 정의된 실변수 함수의 집합이라고 가정하자. 최대우도(Maximum likelihood, ML)의 문제는 우도함수 $L(f) = \prod_{i=1}^{n} f(X_i)$을 최대화는 확률밀도함수 f를 찾는 것이다. 여기서 $f \in H(\Omega)$, $\int_{\Omega} f(t)\,dt = 1$, 그리고 $f(t) \ge 0$ $(\forall t \in \Omega)$이다.

$H(\Omega)$이 유한집합이고 이 문제에 관한 해가 존재한다면 f의 ML 추정량이라고 부른다. 히스토그램은 확률표본 $X_1, \cdots, X_n$에서 생성된 유일한 ML 추정량이다.

2.2. 커널 밀도 추정(kernel density estimation)

로젠블래트(Rosenblatt, 1956)가 커널 추정량에 관한 개념을 최초로 소개했다. 커널 추정의 장점은 수학적으로 취급하기 쉽고 어떠한 모델이라고 유용하게 적용할 수 있다는 것이다. 히스토그램의 경우처럼, 커널 평활기(kernel smoother)는 매끄러움(smooth)의 정도를 조절하는 평활모수(smoothing parameter)이다.

확률밀도함수 f의 다변량 커널 밀도 추정량은 다음과 같이 표현된다.

$$\hat{f}_h(x) = (nh^d)^{-1} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right), \quad x \in R^d \tag{2.2}$$

여기서 커널함수 K과 윈도우 너비(window width) $h = h_n > 0$의 선택은 확률밀도함수 f의 추정량으로서 $\hat{f}_h$의 형태를 결정한다. 그리고 K는 가중 함수로 정의한다. 커널이 밀도함수이기 때문에 커널 추정은 밀도함수이며 다음을 만족한다.

$$K(x) \geq 0, \quad \int_{R^d} K(x)dx = 1 \Rightarrow \int_{R^d} \hat{f}_h(x)dx = 1$$

커널의 매끄러움의 성질은 $\hat{f}_h(x)$에 의하여 영향받는다. 만약 K가 n번 미분 가능하다면 $\hat{f}_h(x)$도 역시 n번 미분 가능하다. 커널 추정은 히스토그램과 같이 원점의 선택에 영향을 받지 않는다. 만약 주어진 관측값 집합에서 띠너비(bandwidth) h와 커널 밀도 K가 임의로 주어진다면, 커널 밀도 함수는 주어진 관측값 집합에서 유일하다.

임의로 주어진 띠너비 $h = h_0$으로 균일 커널(uniform kernel), 삼각 커널(triangle kernel)을 이용하여 추정을 한다면, 약간의 차이 있을 정도의 각각 다른 밀도 추정을 얻을 수 있으나, 시각적으로 커널을 변경하여 얻은 커널 추정에는 별 차이가 없다.

히스토그램의 빈너비(binwidth)를 증가함으로서 평활의 정도가 증가하는 것과 마찬가지로 커널에서의 띠너비 h는 중요한 모수이다. 만약 h를 감소시키면, 즉 $h \to 0$이면 각 띠너비에 소량의 관측값이 포함되어 커널의 추정은 매우 난해한 형태를 가진다. 만약 h를 증가시키면, 더 많은 관측값이 각 띠너비에 포함되기 때문에 커널의 형태는 보다 더 매끄럽게 된다. 만약 $h \to \infty$이면 커널 추정은 매우 완만하고 평평한 형태를 가진다. 그러므로 실제적인 비모수 커널 밀도 추정 상황 하에서 커널의 선택보다는 띠너비 h를 선택하는 방법이 추정에서 중요한 문제이다.

2.3. 국소적응평활(local adaptive smoothing)

커널 밀도 추정은 밀도 함수 추정에서 커널을 유용한 도구로 사용하고 있다. 주어진 구간 내에서 커널을 이용한 밀도추정은 가능하지만, 주어진 구간의 경계에서의 커널을 이용한 밀도추정은 평활 문제가 발생한다. 그러므로 이러한 단점을 극복할 수 있는 또 다른 밀도 추정량으로서 극소적응평활(local adaptive smoothing)을 사용한다. 극소적응평활은 주어진 다른 구간 상에서 커널 밀도 추정을 할 경우, 극소 띠너비

(local bandwidth)를 사용하는 방법이다. 즉, 주어진 구간의 경계근방에서는 경계의 효과를 제거 또는 감소시키기 위하여 주어진 구간에서 자유롭게 띠너비를 조절함으로서 추정할 수 있는 방법이다. 이 방법을 적응커널밀도추정(adaptive kernel density estimation)이라고 부르며 오늘날 활발한 연구가 진행되고 있는 연구 주제이다. 제시된 적응커널(adaptive kernel)의 기법은 밀도추정에서 뾰족한 끝, 첨단에서는 과소추정(underestimate)하고 움푹 패어 들어간 곳, 골짜기에서는 과대추정(overestimate)하는 경향을 보임으로써 커널 밀도 추정의 또 다른 단점을 제거 또는 감소하는데 적절하다. 그리고 주어진 구간에서 고정된 전체적 띠너비(globally bandwidth)가 아닌 주어진 구간에서 자유롭게 띠너비를 조절할 수 있는 극소 띠너비(local bandwidth)를 사용함으로서 추정의 편의를 제거 또는 감소시킬 수 있는 장점이 있다.

식 (2.1)은 고정된 분할(fixed partition)인 경우로 제한되어져있기 때문에 히스토그램은 f의 추정량으로서 보다 더 많은 관측값들에 영향을 받도록 연구됐으며 웨그만(Wegman)에 의하여 제시되어진 가변분할 히스토그램(variable partition histogram)으로 정의되었다. 가변분할 히스토그램은 고정된 분할(fixed partition)로 표현된 히스토그램과 유사한 방식으로 고안되어졌으나 가변분할 히스토그램은 분할(partition)이 순서통계량 $X_{(1)}, \cdots, X_{(n)}$간의 간격에 의존한다.

픽스(Fix)와 호저스(Hodges, 1951)는 비모수 판별분석에서 최근접이웃 추정량(nearest neighbor estimator, k-NN)을 제안했다. 임의의 점 x와 임의의 정수 k에서 $D_k(x)$를 $X_1, \cdots, X_n$중에서 x와 k번째 최근접이웃간의 유클리드 거리라고 하자. 그리고 $Vol_k(x) = c_d[D_k(x)]^d$를 반지름 $D_k(x)$의 d-차원 구(sphere)의 부피, c_d를 d-차원 구(sphere)의 부피라고 하자. 이때 k-NN밀도추정은 다음과 같이 정의된다.

$$\hat{f}(x) = \frac{k/n}{vol_k(x)} = \frac{1}{n[D_k(x)]^d}\sum_{j=1}^{n} K\left(\frac{x-X_j}{D_k(x)}\right) \tag{2.3}$$

k-NN 추정량과 관련된 문제점을 보완하기 위한 가변 커널 추정량(variable kernel estimator)는 다음과 같이 정의된다[3, p. 143].

$$\hat{f}(x) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{H_{jk}^{d}} K\left(\frac{x-X_j}{H_{jk}}\right) \tag{2.4}$$

여기서 가변 윈도우 너비(variable window width) $H_{jk} = hD_k(X_j)$은 (2.3)처럼 x에 의존하지 않으며 h는 평활모수(smoothing parameter)이며 k는 H_{jk}의 극소적 형태(local behavior)를 통제한다. 그리고 가변 커널 추정량(2.4)이 적응 커널 추정량(adaptive kernel estimator)를 유도한다. 확률밀도함수 f를 추정하는 데 관심을 가지고 있는 아브람슨(Abramson)은 자료에 적응 가능한 윈도우 너비(data-adaptive window width)를 계산할 수 있는 적응 커널 추정량(adaptive kernel estimator)를 다음과 같이 제시했다.

$$\hat{f}_h(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{h_j^d} K\left(\frac{x - X_j}{h_j} \right) \tag{2.5}$$

그리고 또 다른 밀도 추정으로서, 세노브(Cenov, 1962)에 의하여 소개되어진 직교열 밀도 추정량(Orthogonal series density estimators) 등이 있으며 패턴 인식, 판별 분석, 군집 분석 등 여러 가지 다른 분야에서 응용되고 있다[2, p. 50].

참고 문헌

1. Eubank, R. L., *Spline Smoothing and Nonparametric Regression,* Marcel Dekker, New York, 1988
2. Härdle, W., *Applied Nonparametric Regression,* Cambridge University Press, Cambridge, 1990
3. Härdle, W., *Smoothing Techniques with Implementation in S,* Springer-Verlag, New York, 1991.
4. Silverman, B. W., *Density Estimation for Statistics and Data Analysis,* Chapman and Hall, New York, 1986.
5. Wand, M. P. and Jones, M. C., *Kernel Smoothing,* Chapman and Hall, 1995.

Historical Study on Density Smoothing in Nonparametric Statistics

Department of Software, Seokyeong University **Seung-Woo Lee**

We investigate the unbiasedness and consistency as the statistical properties of density estimators. We show histogram, kernel density estimation, and local adaptive smoothing as density smoothing in this paper. Also, the early and recent research on nonparametric density estimation is described and discussed.

Key words: binwidth, bandwidth, density smoothing, local adaptive smoothing