

A Nonparametric Additive Risk Model Based On Splines

Cheolyong Park¹⁾

Abstract

We consider a nonparametric additive risk model that are based on splines. This model consists of both purely and smoothly nonparametric components. As an estimation method of this model, we use the weighted least square estimation by Huffer and McKeague (1991). We provide an illustrative example as well as a simulation study that compares the performance of our method with the ordinary least square method.

Keywords : nonparametric additive risk model, weighted least square

1. 머리말

McKeague와 Sasieni (1994)에 의해 제안된 부분모수적 가법위험모형(partly parametric additive risk model)은 Lin과 Yang (1994)에 의해 고려된 Cox (1972)의 비례위험모형의 가법모형을 포함하는 준모수 모형(semiparametric model)으로 t 시점의 위험함수(hazard function)가

$$\lambda(t|x, z) = \alpha(t)^T x + \beta^T z$$

형태로 주어진다. 단 여기서 x, z 는 q, p 차원의 공변량, $\alpha(\cdot)$ 은 미지의 비모수함수이며 β 는 미지의 모수이다. Park (2006)은 위의 모형을 확장하여 미지의 모수 부분이 $\beta(t) = \beta \cdot f(t)$ 형태로서 $f(\cdot)$ 는 알려진 함수로 주어지는 일반화 부분모수적 가법위험모형을 고려하였다.

이 논문에서는 위의 모형을 더욱 확장하여 $\beta(t)$ 가 부드러운(smooth) 비모수 함수로 주어지며 따라서 t 시점의 위험함수(hazard function)가

$$\lambda(t|x, z) = \alpha(t)^T x + \beta(t)^T z \quad (1-1)$$

로 주어지는 모형을 고려한다. 따라서 이 모형은 Aalen (1980)의 순수 비모수 부분인 $\alpha(t)^T x$ 와 부드러운 비모수 부분인 $\beta(t)^T z$ 의 가법모형인 것이다. 이 논문에서 제안하고 있는 것은 부드러운 비모수 함수인 $\beta(t)$ 를 3차 스플라인 함수(cubic spline function)로 근사화시켜 일반화 부분모수적 가법위험모형에서 적용했던 기법들을 그

1) Associate Professor, Department of Statistics, Keimyung University, Taegu 704-701
E-mail : cypark1@kmu.ac.kr

대로 적용하는 것이다. 구체적으로 이 논문에서 고려하는 형태는 $\beta(t) = Bb$ 이다. 여기서 $B = (\beta_{ij})$ 는 $p \times m$ 인 모수 행렬이고 $b = (B_1(t), B_2(t), \dots, B_m(t))^T$ 로서 $B_1(t), B_2(t), \dots, B_m(t)$ 는 3차 B-스플라인의 기저(B-spline basis)이다. 이 논문에서는 기저의 숫자 선택에 대한 문제는 고려하지 않고 있으며 자의적으로 너무 커지도 않고 너무 작지도 않다고 판단하는 $m = 8$ 을 주로 사용하였다.

앞에서 고려된 모형에서 추정의 초점은 모수인 B 및 (순수) 비모수 부분의 누적위험함수(cumulative hazard function)

$$A(t) = \int_0^t a(s) ds$$

이다. 이 논문에서는 Huffer와 McKeague (1991)에서 그룹화 자료(grouped data)의 위험함수의 추정에 사용하였던 조각별 고정위험(piecewise constant risk)을 사용하여 모수 B와 위험함수의 추정에 적용하려고 한다. 구체적으로 주어진 추적기간(follow-up period)을 작은 구간으로 분할하고 각 구간 안에서는 위험이 동일하다는 가정 하에서 가중최소제곱법(weighted least squares method)에 의해 모수 B와 구간별 위험함수를 추정하는 방법이다. 이 방법은 실제로 적용하기 아주 쉬운 방법이며 또한 계산이 용이한 보통최소제곱법(ordinary least square method)에 비해 효율성이 좋은 것으로 Huffer와 McKeague (1991)와 Park (2006)의 연구에서 밝혀졌기 때문에 사용되었다.

참고문헌

1. Aalen, O. O. (1980). A model for nonparametric regression analysis of counting processes. In *Mathematical Statistics and Probability Theory, Lecture Notes in Statistics, 2*, Ed. W. Klonecki, A. Kozek and J. Rosinski, pp. 1-25. Springer-Verlag, New York.
2. Huffer, F. W, and McKeague, I. W. (1991). Weighted least squares estimation for Aalen's additive risk model. *Journal of American Statistical Association*, 86, 38-53.
3. Lin, D. Y. , and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika*, 81, 61-71.
4. McKeague, I. W., and Sasieni, P. D. (1994). A partly parametric additive risk model. *Biometrika*, 81, 501-514.
5. Park, C. (2006). A generalized partly-parametric additive risk model. *Journal of the Korean Data & Information Science Society*, 17, 401-409.