Journal of the Korean Data & Information Science Society
2011, **22**(1), 89–97

한국데이터정보과학회지

임의중단자료에서의 조건부 평균잔여수명함수 추정

이원기[1] · 송명언[2] · 정성화[3]

[1]경북대학교 의학전문대학원 · [2]경북대학교 통계학과 · [3]대구한의대학교 보건학부

접수 2010년 12월 30일, 수정 2011년 01월 17일, 게재확정 2011년 01월 20일

요 약

본 연구에서는 Buckley와 James의 방법을 이용하여 중도절단된 자료를 보완한 조건부생존함수 추정량으로부터 조건부평균잔여수명함수를 추정하는 방법을 제안하고,모의실험을 통하여 제안된 방법의 효율성을 평가하였다. 모의실험 결과 비례위험모형이 아닌 경우 제안된 방법으로 추정한 조건부평균잔여수명함수의 평균제곱오차가 Cox모형이나 Beran의 비모수적 방법을 이용하여 구한 추정치의 평균제곱오차보다 작게 나타났으며, 비례위험모형인 경우에는 제안된 방법으로 추정한 결과들이 Cox 모형을 이용하여 얻은 결과들과 비슷하게 나타났다. 또한 K대학교병원 외과에서 위암 수술을 받은 1,192명의 환자 자료를 이용하여 제안된 방법의 임상적 적용의 적절성을 평가하였다.

주요용어: 임의중단, 조건부생존함수, 조건부평균잔여수명함수.

1. 서론

생존자료분석은 생존함수와 위험함수의 추정이 주요관심의 대상이지만 또 하나의 관심대상은 평균잔여수명함수이다. 특히, 암 수술 등 침습적 시술이 필요한 분야에서는 치료 후 얼마간의 시간경과가 매우 중요하다. 왜냐하면 치료 후 일정시간을 경과한 환자는 처치 직후에 비해 생존 가능성이 높다. 따라서 이를 반영한 평균잔여수명의 추정은 중요하다.

평균잔여수명함수 (mean residual life function)의 추정은 모수적 및 비모수적 방법으로 많은 학자들에 의해 연구되었다. 중도절단이 없는 완전한 자료에 대하여 모수적 방법을 이용한 접근으로 Morrison (1978)은 평균잔여수명함수의 최대우도 추정량을 연구하였으며, Kotz와 Shanbhang (1980)은 평균잔여수명함수의 극한에 대한 평균잔여수명함수의 임의순차적 수렴성 (convergence of an arbitrary sequence)의 확실한 결과를 세웠으며, Hall과 Wellner (1981)는 평균잔여수명함수의 특성을 요약하였다. 비모수적 접근방법으로 Yang (1978)은 생존함수의 경험적 추정량을 이용하여 평균잔여수명함수를 추정하였고, 추정량의 일양적 강일치성과 가우시안 과정으로 분포수렴함을 보였다.

임의중도절단 자료에서 평균잔여수명함수에 대한 비모수적 연구는 Yang (1977)이 생존함수의 Nelson Aalen 추정량 (Nelson, 1972; Aalen, 1978)을 이용한 평균잔여수명함수의 추정량을 제안하고, 유계구간에서 가우시안과정으로 분포수렴함을 보였다. Kumazawa (1987)는 생존함수의 Kaplan-Meier추정량 (Kaplan과 Meier, 1958)을 이용한 평균잔여수명함수의 추정량을 제안하고, 유계구간에서 점근적 성질을 밝혔다. Jeong 등 (1996)은 Ghorai와 Rejtö (1987)가 제안한 생존함수의 커넬 평활추정량을 이

[1] (700-422) 대구광역시 중구 동인동2가 101, 경북대학교 의학전문대학원, 조교.
[2] (702-701) 대구광역시 북구 산격동 1370, 경북대학교 통계학과, 박사.
[3] 교신저자 : (712-715) 경상북도 경산시 유곡동 290, 대구한의대학교 보건학부, 전임강사,
E-mail: jeongsh@dhu.ac.kr

용하여 추정량을 제안하고, 일양강일치성과 정규과정으로의 분포수렴성을 밝혔다. 뿐만 아니라 평균잔여수명함수의 추정에 있어 잔여생존함수 (residual survival function)를 이용하거나 또는 부분적률근사 (partial moment approximation)를 적용한 연구들도 있었다 (Jeong 등, 1997; Cha, 2004).

한편 평균잔여수명은 처치 후 경과시간 뿐만 아니라 처치 전 생존시간에 영향이 미칠 만한 여러 가지 공변량에 영향을 받는다. 예를 들면, 암 수술을 받은 환자가 수술시의 암의 크기, 암의 진행정도, 다른 장기로의 전이여부 등의 공변량은 t시간까지 생존했을 때 앞으로의 평균잔여수명에 큰 영향을 끼치므로 이들 공변량을 고려하여 조건부 평균잔여수명을 추정하는 것이 보다 정확한 추정치를 얻을 수 있을 것이다.

따라서 본 연구에서는 Lee와 Song (2004)이 제안한 세 가지 조건부 생존함수 추정량을 이용하여 조건부 평균잔여수명함수의 추정량을 제안하고, 모의실험을 통하여 이들 추정량의 효율성을 비교하였다. 또한 위암환자 자료에 제안된 방법의 임상적 적용성을 평가하였다.

2. 조건부 평균잔여수명함수의 추정

생존시간 T_1, T_2, $\cdots$, T_n은 서로 독립이면서 동일한 분포 F를 따르고, 중도절단시간 C_1, C_2, $\cdots$, C_n는 마찬가지로 서로 독립이면서 동일한 분포 G를 따른다고 하자. 이들 자료에 대한 공변량은 Z_1, Z_2, $\cdots$, Z_n이며, 여기서 $Z_i = (z_{i1}, z_{i2}, \cdots, z_{ip})^t$이라 하자. 그러면 $\{(X_i, \delta_i, Z_i), i = 1, 2, \cdots, n\}$을 관측하게 되며 여기서 관측시간은 $X_i = T_i \wedge C_i$으로 생존시간과 중도절단시간 중 적은값이며, δ_i는 지시함수로 $\delta_i = I(T_i <= C_i)$이다. 또한 $\{ Z_i \}_{i=1}^n$는 유계이고 주어진 Z_i에서 T_i 와 C_i는 조건부 독립이라고 가정한다.

또한 시간 t에서의 평균잔여수명함수 $e(t)$는 생존시간 확률변수 T가 연속분포함수 F를 따른다고 할 때

$$e(t) = E[T - t | T > t]$$

로 정의한다. 여기서, $S(t) = 1 - F(t)$는 확률변수 T의 생존함수이며 $S(t) = 0$이면 $e(t) = 0$으로 정의한다. 또한 평균잔여수명함수 $e(t)$는 $S(t) > 0$에 대하여

$$e(t) = \frac{1}{S(t)} \int_t^\infty S(u) du$$

로 표현되어 질 수 있으므로 $t = 0$이면 평균수명이 된다. 또한 T가 연속확률변수일 경우 생존함수 $S(t)$와 위험함수 $\lambda(t)$와는 일대일 관계로서

$$S(t) = \frac{e(0)}{e(t)} \exp\left[-\int_0^t \frac{1}{e(u)} du \right]$$

$$\lambda(t) = \frac{\left(\frac{d}{dt} e(t) + 1 \right)}{e(t)}$$

$$e(t) = \frac{\int_t^\infty (u - t) f(u) du}{S(t)}$$

들을 얻을 수 있다.

또한 공변량 $Z = (z_1,\ z_2, \cdots,\ z_p)$가 주어졌을 때 조건부 평균잔여수명함수는

$$e(t|z) = E[T - t | T > t, Z = z]$$

로 정의되며 $S(t|z) > 0$에 대하여 다음과 같이 표현되어 질 수 있다.

$$e(t|z) = \frac{1}{S(t|z)} \int_t^{\infty} S(u|z) du \tag{2.1}$$

여기서 $t = 0$이면 공변량 $Z = z$가 주어진 조건부 평균수명이 되며 적분이 존재하기 위해

$$\sup\{t : S(t|z) > 0\} \leq \sup\{t : 1 - G(t|z) > 0\}$$

을 만족한다고 가정한다.

Cox (1972)의 조건부 평균잔여수명함수 추정량은 Lee와 Song (2004)이 제안한 Cox 비례위험모형 하에서의 조건부 생존함수 추정량 $\widehat{S}^{Cox}(t|z)$를 식 (2.1)에 대입하여 다음과 같이 제안할 수 있다.

$$\widehat{e}^{Cox}(t|z) = \frac{1}{\widehat{S}^{Cox}(t|z)} \int_t^{\infty} \widehat{S}^{Cox}(u|z) du \tag{2.2}$$

또한 조건부 생존함수 추정에 Beran (1981)의 비모수적 방법을 적용한 $\widehat{S}^{Beran}(t|z)$와 Buckley와 James (1979)의 회귀보완법을 적용하여 얻은 조건부 생존함수의 추정량 $\widehat{S}^{BJ}(t|z)$를 식 (2.1)에 각각 대입하여 다음과 같은 조건부 평균잔여수명함수의 추정량을 제안한다.

$$\widehat{e}^{Beran}(t|z) = \frac{1}{\widehat{S}^{Beran}(t|z)} \int_t^{\infty} \widehat{S}^{Beran}(u|z) du, \tag{2.3}$$

$$\widehat{e}^{BJ}(t|z) = \frac{1}{\widehat{S}^{BJ}(t|z)} \int_t^{Y} \widehat{S}^{BJ}(u|z) du \tag{2.4}$$

한편 Cox의 조건부 생존함수의 추정량 $\widehat{S}^{Cox}(t|z)$와 회귀보완법을 이용한 조건부 생존함수의 추정량 $\widehat{S}^{BJ}(t|z)$에 대해서는 아직까지 점근적 성질이 밝혀져 있지 않으므로 조건부 평균잔여수명함수의 추정량 $\widehat{e}^{Cox}(t|z)$와 $\widehat{e}^{BJ}(t|z)$에 대한 점근적 성질은 규명할 수 없지만 $\widehat{e}^{Beran}(t|z)$는 Beran(1981)과 Dabrowska(1987)에 의해 밝혀진 $\widehat{S}^{Beran}(t|z)$의 강일치성과 분포수렴성을 이용하여 다음과 같이 증명된다.

공변량 $Z = z$가 주어졌을 때 $\widehat{S}^{Beran}(t|z)$을 추정하기 위해

$$H_1(t|z) = \Pr(X > t, \delta = 1|Z = z),$$
$$H_2(t|z) = \Pr(X > t|Z = z)$$

로 정의하면, 앞에서 정의한 자료 $\{(X_i,\ \delta_i,\ Z_i\ ),\ i = 1,\ 2, \cdots, n\}$에서 경험적 조건부 누적생존함수는 각각

$$H_{1n}(t|z) = \sum_{i=1}^{n} I(X_i > t, \delta_i = 1) W_i(z),$$
$$H_{2n}(t|z) = \sum_{i=1}^{n} I(X_i > t) W_i(z)$$

와 같이 경험적으로 추정될 수 있다. 여기서 $W_i(z)$는 단지 공변량 z에만 의존하는 음수가 아닌 가중치로서 커넬과 이웃근접 (nearest neighborhood) 가중치이다. 또한 $H_i(t|z)$는 연속이며, $\phi(z)$는 z에만 의

존하는 함수라 가정하고, $i = 1, 2$에 대하여

$$\begin{aligned} W_{in}(t|z) =& \sqrt{n}\left\{H_{in}(t|z) - H_i(t|z)\right\}, \\ W_n(t|z) =& \left\{W_{1n}(t|z), W_{2n}(t|z)\right\}, \\ L_n(t|z) =& \sqrt{n}\left\{\widehat{S}^{Beran}(t|z) - S(t|z)\right\} \end{aligned}$$

라 하자.

보조정리 2.1 (Beran, 1981; Dabrowska, 1987)
(1) $\tau(z) < \sup\{s \mid H_2(s|z) > 0\}$라 하고, $i = 1, 2$에 대하여 $n \to \infty$일 때

$$\begin{aligned} &\sup\nolimits_{0 \le t \le \tau(z)} |H_{in}(t_-|z) - H_i(t_-|z)| \to 0 \quad a.s. \\ &\sup\nolimits_{0 \le t \le \tau(z)} |H_{in}(t|z) - H_i(t|z)| \to 0 \quad a.s. \end{aligned}$$

이면

$$\sup\nolimits_{0 \le t \le \tau(z)} |\widehat{S}^{Beran}(t|z) - S(t|z)| \to 0 \quad a.s.$$

이다.
(2) 과정 $W_n(t|z)$ 가 $D[0, \tau(z)] \times D[0, \tau(z)]$에서 이차원 평균 0인 가우스과정

$$W(t|z) \;=\; \{W_1(t|z), W_2(t|z)\}$$

로 분포수렴하고, 공분산 행렬이 $s \le t$에 대하여

$$Cov\left\{W_i(s|z), W_j(t|z)\right\} \;=\; \left\{H_{i \wedge j}(t|z) - H_i(s|z)H_j(t|z)\right\}\phi(z)$$

이라 하자. 그러면 과정 $L_n(t|z)$는 $D[0, \tau(z)]$에서 평균 0인 가우시안 과정 $L(t|z)$로 분포수렴하고 공분산행렬은 다음과 같다.

$$Cov\left\{L(s|z), L(t|z)\right\} \;=\; S(s|z)S(t|z)\int_0^{s \wedge t} \frac{\Lambda(du|z)}{H_2(u|z)}\phi(z).$$

보조정리 2.1을 이용하면 제안한 조건부 평균잔여수명함수 추정량 $\widehat{e}^{Beran}(t|z)$은 다음과 같은 점근적 성질을 가진다.

정리 2.1 보조정리 2.1 (1)항의 가정하에서 $n \to \infty$일 때

$$\sup_{0 \le t \le t\tau(z)} \left|\widehat{e}^{Beran}(t|z) - e(t|z)\right| \;\xrightarrow{p}\; 0$$

이다.

증명. 고정된 $t \in [0, \tau(z)]$에 대하여

$$\begin{aligned}
&|\widehat{e}^{Beran}(t|z) - e(t|z)| \\
=&|\widehat{S}^{Beran}(t|z)S(t|z)|^{-1} \times |S(t|z) \int_t^\infty [\widehat{S}^{Beran}(u|z) - S(u|z)]du \\
&- [\widehat{S}^{Beran}(t|z) - S(t|z)] \int_t^\infty S(u|z)du| \\
\le&|\widehat{S}^{Beran}(t|z)|^{-1} \times \sup_{0 \le t \le t\tau(z)} |\widehat{S}^{Beran}(u|z) - S(u|z)| \times (\tau(z) - t) \\
&+ |\widehat{S}^{Beran}(t|z)S(t|z)|^{-1} \times |\widehat{S}^{Beran}(t|z) - S(t|z)| \int_t^\infty S(u|z)du
\end{aligned}$$

이므로 보조정리 2.1 (1)항에 의해 증명된다.

정리 2.2 보조정리 2.1 (1)과 (2)항의 가정이 만족될 때 과정 $\sqrt{n}\left[\widehat{e}^{Beran}(t|z) - e(t|z)\right]$는 $D[0, \tau(z)]$에서 평균 0인 가우스과정 $B(t|z)$로 분포수렴하며,

$$B(t|z) = [S(t|z)]^{-1} \left[\int_t^{\tau(z)} L(u|z)du - e(t|z)L(u|z) \right]$$

이고, $L(t|z)$은 보조정리 2.1 (2)항에 있다.

증명. 고정된 $t \in [0, \tau(z)]$에 대하여 과정 $\sqrt{n}\left[\widehat{e}^{Beran}(t|z) - e(t|z)\right]$는

$$\begin{aligned}
&\sqrt{n}\left[\widehat{e}^{Beran}(t|z) - e(t|z)\right] \\
&= S(t|z)^{-1} \int_t^{\tau(z)} \sqrt{n}[\widehat{S}^{Beran}(u|z) - S(u|z)]du \\
&- (\widehat{S}^{Beran}(t|z)S(t|z))^{-1}\sqrt{n}[\widehat{S}^{Beran}(t|z) - S(t|z)] \int_t^\infty S(u|z)du
\end{aligned}$$

으로 표현되며, 보조정리 2.1 (1)과 (2)항 그리고 $e(t|z)$의 정의에 의해 증명된다.

3. 모의실험

조건부 평균잔여수명함수 추정량들의 효율성을 비교하기 위하여 비례위험모형인 경우와 비례위험모형이 아닌 경우로 나누어 모의실험을 하였으며, 제안된 추정량들의 평균제곱오차 (mean square error)를 계산하였다.

모의실험에서 생존시간 T는 와이블분포, 중도절단시간 C는 지수분포, 그리고 공변량 Z는 $U(0,1)$를 가정하였다. 공변량의 개수는 1개, 주어진 공변량의 값은 0.55, 중도절단률은 30%, 그리고 표본수는 100으로 가정하였다. 비례위험모형인 경우에는

$$\lambda(t|z) = e^z,$$

비례위험모형이 아닌 경우에는

$$\lambda(t|z) = e^z I(z \le 0.5) + 2te^z I(z > 0.5)$$

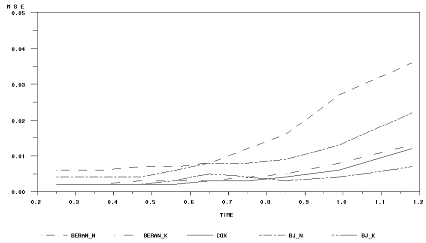

그림 3.1 비례모형하에서의 평균제곱오차 ($F \sim Weib(2.0, e^z), Z \sim U(0,1)$, _N: 이웃근접가중치, _K: 커넬가중치)

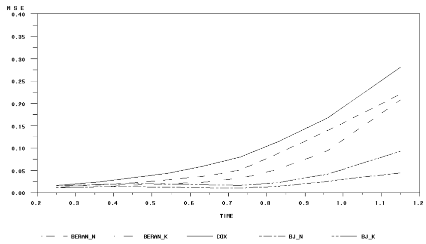

그림 3.2 비례모형이 아닌 경우 추정량 별 평균제곱오차
($F \sim Weib(1.0, e^z)I(z \leq 0.5) + Weib(2.0, e^z)I(z > 0.5), Z \sim U(0,1)$, _N: 이웃근접가중치, _K: 커넬가중치)

을 고려하였다.

모의실험결과 비례위험가정이 타당한 경우에는 Cox의 조건부 평균잔여수명함수의 추정량의 평균제곱오차가 다른 추정량들에 비해 조금 적게 나타났다. 또한 Beran과 Buckley-James 방법에서는 이웃근접가중치를 사용하는 경우가 커넬가중치를 사용하는 경우보다 평균제곱오차가 더 크게 나타났다 (그림 3.1). 반면 비례위험모형이 아닌 경우에는 Buckley-James 방법을 적용한 추정량의 평균제곱오차가 가장 작았고, Cox 방법을 적용한 추정량의 평균제곱오차가 가장 크게 나타났다. 이는 Buckley-James 방법은 회귀보완법을 이용하여 중도절단된 자료를 개선함으로써 중도절단 자료가 없기 때문에 평균오차제곱이 작게 나타나는 것으로 해석된다 (그림 3.2).

4. 실제 자료의 적용 예

예제 자료는 K대학교병원에서 10여년간 위암으로 인하여 위절제수술을 받았던 환자 1,282명 중 교통사고 등 다른 요인으로 사망한 90명을 제외한 1,192명의 생존시간을 관찰한 자료이며 이 중 713명(59.8%)이 중도절단되어 관찰되었다.

수술 후 생존시간에 유의한 영향을 주리라고 생각되어지는 변수로 Depth, Distance, Size, Borrmann을 고려하였으며, 이들 변수들은 수술 전 각종 검사를 통하여 알 수 있는 변수이다. 여기서 Depth (1, 2, 3, 4)는 종양이 위벽으로부터 얼마나 깊이 위치하고 있는지를 나타내며, Distance (0: no, 1: yes)는 종양이 주변 장기나 임파선 등으로 전이되었는지의 여부를 나타내고, Size는 종양의 장축지름을 측정한 것이며, Borrmann (0, 1, 2, 3, 4)은 종양의 육안적 형태를 나타낸 것이다. 또한 수술 후에는 수술시 제거한 림프절 중 전이된 림프절의 백분율로 나타나는 NF 변수를 추가로 관찰할 수 있다. 이들 공변량을 이용하여 제안한 조건부 평균잔여수명함수의 추정량을 이 자료에 적용하였다.

조건부 평균잔여수명함수 추정에서 공변량 값은 Depth=3, Distance=0, Size=50, Borrmann=2를 사용하였고, 수술 후 추가되는 공변량 NF가 조건부 평균잔여수명함수에 미치는 영향을 살펴보기 위하여 다른 조건은 동일하게 설정하고 NF 값만 변화시켜 추정량들을 비교하였다.

수술 전 정보를 이용한 모형에서는 수술 후 약 4년까지 평균잔여수명이 일정하게 나타나는데 이는 시간이 흘렀는데도 불구하고 평균잔여수명이 일정하게 줄지 않고 비슷하게 추정되었기 때문으로 실제적으로는 수술직후의 평균잔여수명 예측치보다 늘어난 것으로 볼 수 있다. 이런 이유는 수술 후 일정기간 동안 생존하였다면 사망할 가능성이 줄어든다는 것을 잘 반영해 주는 것으로 보인다 (그림 4.1).

수술 후 전이된 림프절의 수가 0% (NF=0)인 경우 전체적인 곡선의 형태는 수술 전 정보를 사용한 모형에 비교하여 전반적으로 유사하였지만 수술 직후 평균잔여수명 추정치는 좀 더 길 것으로 예측되었다. 이는 상식적인 상황이 모형에 잘 반영된 것으로 생각된다. 제안한 추정량간의 비교에서는 회귀보완법을 이용한 BJ_N과 BJ_K의 추정치는 일정하게 줄어듦을 보이고 있는 반면 Cox 방법을 이용한 추정치는 수술 전 모형과 비교하여 큰 변화가 없는 것으로 나타났다 (그림 4.2).또한 NF의 값이 40으로 림프절의 전이 비율이 상당히 높은 경우는 Cox 방법을 이용한 추정치는 평균적으로 약 2년 정도 짧아지는 것으로 나타났으나, 다른 추정치들은 그보다 더 짧게 나타났다. 특히 커넬가중치를 이용한 비모수적 방법은 1년 이상 생존하지 못하는 것으로 나타났다. 이는 주어진 공변량의 조건에 근접하는 자료가 많지 않아 나타난 현상으로 해석된다 (그림 4.3).

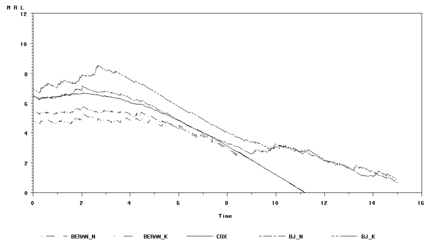

그림 4.1 수술 전 공변량을 이용한 조건부 평균잔여수명
(Depth=3, Distance=0, Size=50, Borrmann=2, _N: 이웃근접가중치, _K: 커넬가중치)

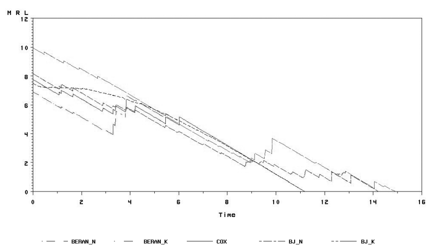

그림 4.2 수술 후 공변량을 이용한 조건부 평균잔여수명
(Depth=3, Distance=0, Size=50, Borrmann=2, NF=0, _N: 이웃근접가중치, _K: 커넬가중치)

위 결과를 종합해 볼 때 평균잔여수명을 추정하는데 있어 어떤 방법이 효율적인지판단하기 어렵지만 이들 방법에 의해 평균적인 잔여수명을 개략적으로 예측할 수 있을 것으로 판단된다. 그러나 가중치를 사용하는 방법은 자료의 수가 적은 경우 추정량의 정확도가 떨어지는 단점이 있으므로, 자료의 수에 따

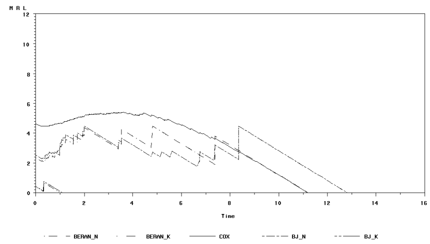

그림 **4.3** 수술 후 공변량자료를 이용한 조건부평균잔여수명함수
(Depth=3, Distance=0, Size=50, Borrmann=2, NF=40, _N: 이웃근접가중치, _K: 커넬가중치)

라 또한 공변량의 비례위험가정 만족여부에 따라 모형을 선택적으로 사용할 필요가 있겠다. 또한 이 자료에 대한 추적자료를 확보하여 각 추정량의 효율성을 비교하는 연구도 필요하겠다.

참고문헌

Aalen, O. O. (1978). Nonparametric inference for a family of counting processes. *Annals of Statistics*, **6**, 701-726.

Beran, R. J. (1981). *Nonparametric regression with randomly censored data*, Technical report, University of California, Berkeley.

Buckley, J. J. and James, I. R. (1979). Linear regression with censored data. *Biometrika*, **66**, 429-436.

Cha, Y. J. (2004). Nonparametric estimation of mean residual life by partial moment approximation under proportional hazard model. *Journal of Korean Data & Information Science Society*, **15**, 965-971.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Ser. B*, **34**, 187-220.

Dabrowska, D. M. (1987). Non-paramertic regression with censored survival time data. *Scandinavian Journal of Statistics*, **14**, 181-198.

Ghoriai, J. K. and Rejtö L. (1987). Estimation of mean residual life with censored data under proportional hazard model. *Communications in Statistics - Theory and Methods*, **16**, 2097-2114.

Hall, W. J. and Wellner, J. A. (1981). Mean residual life. In *Statistics and related topics*, Eds. Csörgö, M., Dawson, D. A., Rao, J. N. K. and Saleh, M. E., North-Holland, New York.

Jeong, D. M., Song, M. U. and Song, J. K. (1996). Smoothing mean residual life with censored data. *The Korean Communications in Statistics*, **3**, 129-138.

Jeong, D. M., Song, M. U. and Song, J. K. (1997). A simple estimator of mean residual life function under random censoring. *Journal of Korean Data & Information Science Society*, **8**, 225-230.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of American Statistical Association*, **53**, 457-481.

Kotz, S. and Shanbhag, D. N. (1980). Some new approaches to probability distributions. *Advanced in Applied Probability*, **12**, 903-921.

Kumazawa, Y. (1987). A note an estimator of life expectancy with random censoring. *Biometrika*, **74**, 655-658.

Lee, W. K. and Song, M. U. (2004). A study on the conditional survival function with random censored data. *Journal of Korean Data & Information Science Society*, **15**, 405-411.

Morrison, D. G. (1978). On linearly increasing mean residual lifetimes. *Journal of Applied Probability*, **15**, 617-620.

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometics*, **14**, 945-965.

Yang, G. L. (1977). Life expectancy under random censorship. *Stochastic Processes and their Applications*, **6**, 33-39.

Yang, G. L. (1978). Estimation of biometric function. *Annals of Statistics*, **6**, 112-116.

Journal of the Korean Data & Information Science Society
2011, **22**(1), 89–97

한국데이터정보과학회지

Estimation of conditional mean residual life function with random censored data

Won Kee Lee[1] · Myung Unn Song[2] · Seong Hwa Jeong[3]

[1]School of Medicine, Kyungpook National University
[2]Department of Statistics, Kyungpook National University
[3]Faculty of Health Science, Daegu Haany University

Received 30 December 2010, revised 17 January 2011, accepted 20 January 2011

Abstract

The aims of this study were to propose a method of estimation for mean residual life function (MRLF) from conditional survival function using the Buckley and James's (1979) pseudo random variables, and then to assess the performance of the proposed method through the simulation studies. The mean squared error (MSE) of proposed method were less than those of the Cox's proportional hazard model (PHM) and Beran's nonparametric method for non-PHM case. Futhermore in the case of PHM, the MSE's of proposed method were similar to those of Cox's PHM. Finally, to evaluate the appropriateness of practical use, we applied the proposed method to the gastric cancer data. The data set consist of the 1,192 patients with gastric cancer underwent surgery at the Department of Surgery, K-University Hospital.

Keywords: Conditional mean residual life function, conditional survival function, random censoring.

[1] Research Assistant, School of Medicine, Kyungpook National University, 700-422 Daegu, Korea.
[2] Ph.D., Department of Statistics, Kyungpook National University, Daegu 702-701, Korea.
[3] Corresponding Author: Full-time lecturer, Faculty of Health Science, Daegu Haany University, Gyeongsan 712-715, Korea. E-mail: jeongsh@dhu.ac.kr