한국통계학회 논문집 제1권1호
1994년 12월 pp. 141-148

입소전 범죄 횟수를 고려한 재소자 모형

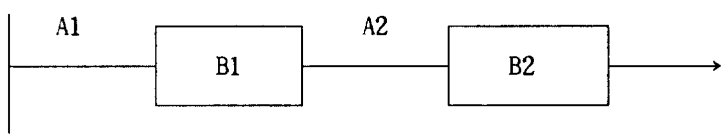
김철웅[1)]

요약

범죄행태에 관한 연구를 하기 위하여 재소자들에 관한 자료를 이용하는 경우가 많다. 그러나 이들 자료는 범죄자 전체의 모집단을 대표하는 표본이 될 수없으므로 추정치에 편기가 존재하여 그대로 사용할 수없다. 단순 교대재생과정에 기초를 두고 재소자들의 특성을 고려하여 모형을 세우면, 편기가 제거된 범죄횟수, 재소기간등에 관한 추정치를 얻을 수있다.

1.머리말

단순 교대재생과정(alternating renewal process)에서는 시간변수 A1, B1, A2, B2, A3, B3,...에 대하여 다음과 같은 가정을 한다. (1) 변수 A와 B는 독립이다. (2) A변수 끼리 서로 독립이고, B변수 끼리 서로 독립이다. (3) A변수들는 서로 같은 분포를 따르고, 또한 B변수들도 서로 같은 분포를 따른다.



<그림 1> 단순 교대재생과정

재소자 모형은 교대재생과정에 근거를 두고 만들어 볼 수있다. 재소자가 출소를 하게되면 다시 범죄를 저지르고 감옥에 오는 것을 반복하게 된다. 감옥 밖에 있는 기간(자유시간)과 안에 있는 기간(재소기간)이 계속 연이어 지게된다. 물론 재소자들의 행태가 단순 교대재생과정를 따르는 것은 아니다. 재소자들의 행태를 살펴서 모형이 현실에 부합되도록 교대재생과정를 수정해 보려는 것이 목적이다. 범죄의 종류에 따라 여러 특성으로 나누어 지는 데, 절도나 강도 범죄자에 한하여 모형을 국한하려 한다.

1) (120-749) 서울 서대문구 신촌동 134, 연세대학교 응용통계학과 조교수.

2. 재소자 모형

절도나 강도 범죄자들의 자료에서 찾아 볼 수있는 특성은 다음과 같은 것들이 있다(Kim, 1989).

(1) 자료는 경제적, 시간적인 제약 때문에 재소자에게서만 얻고 있다. 잡히지 않고 범죄활동을 벌이고 있는 범죄자나 출소한 사람의 자료는 얻기 힘들다 (심영희, 1991).

(2) 첫번째 재소기간은 두번째 이후의 재소기간보다 짧은 경향이 있다.

(3) 재소자 나이 분포가 20대 초 중반이 대부분이고 그 이후는 급격히 줄어 든다.

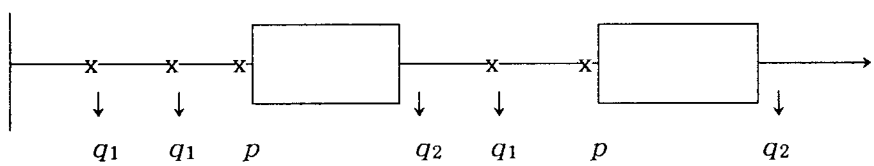
위의 특성들로 인하여 모형설정에 다음 사항들을 고려하여야 한다.

(1) 길이 편기(length bias)의 문제가 있으며 이를 수정하기 위하여는 자료 취득시에 감옥에 있을 확률을 구하여 조건부 밀도함수(conditional likelihood)를 세워야 한다.

(2) 첫번째 재소기간의 분포를 두번째 이후의 재소기간 분포와 다르게 모형화 한다.

(3) 출소하고 난 후에 다시 감옥으로 돌아오지 않는 사람들이 많다. 즉 언젠가는 재생과정(renewal process)이 끝이 나게 되어, 이를 반영할 수 있는 장치를 모형에 추가하여야 한다.

위의 사항들을 고려하여 재소자 모형을 <그림 2>와 같이 생각해 볼 수 있다. x로 표시한 것은 범죄를 저지른 시점을, 네모는 재소기간을, ↓표는 재소자과정을 벗어나는 것을 표시한다. 각 범죄 시점사이의 시간분포는 1/θ를 평균으로 갖는 지수분포, 재소기간은 1/δ를 평균으로 갖는 지수분포로 가정하고, 범죄직후 중단할 확률을 q_1, 범죄직후 입소할 확률을 p, 출소직후 중단할 확률을 q_2로 가정한다.



<그림 2> 재소자 모형

2.1. 재소확률

재소확률을 구하기 위하여는 무한수열의 합을 구하여야하는데, 라플라스 변환(Laplace transformation)을 이용하면 편리하게 구할 수 있다. 함수 $f(t)$의 라플라스 변환 $f^*(\xi)$은

$$f^*(\xi) = \int_0^\infty f(t)\, e^{-\xi t}\, dt$$

로 정의된다.

t시점에서 재소상태일 확률을 $P(t)$, t시점에서 k번째 재소상태일 확률을 $P_k(t)$로 표시하면,

$$P(t) = \sum_{k=1}^{\infty} P_k(t)$$

의 관계식을 이용하여 계산하게 된다.

[정리 1] $P(t)$의 라플라스 변환 $P^*(\xi)$은

$$P^*(\xi) = \frac{P_1^*(\xi)}{1-(p+q_1)f_{G+S}^*(\xi)}$$

이다.

[증명] t시점에서 k+1번째 재소상태일 확률은 k번째 재소상태일 확률과

$$\begin{aligned} P_{k+1}(t) &= \Pr\{G_1+S_1+\cdots+G_{k+1}<t<G_1+S_1+\cdots+G_{k+1}+S_{k+1}\} \\ &= \int_0^t \Pr\{G_1+S_1+\cdots+G_k<t-x<G_1+S_1+\cdots+G_k+S_k\} \\ &\qquad \times \Pr\{\text{출소후 중단않음}\}\ dF_{G+S}(x) \\ &= \int_0^t P_k(t-x)(1-q_2)\ dF_{G+S}(x) \end{aligned}$$

의 관계가 있으므로, 라플라스 변환은

$$\begin{aligned} P_{k+1}^*(\xi) &= \int_0^\infty e^{-\xi t} P_{k+1}(t)\ dt \\ &= \int_0^\infty e^{-\xi t} \int_0^t P_k(t-x)(1-q_2)\ dF_{G+S}(x)\ dt \\ &= \int_0^\infty \int_0^t e^{-\xi t} P_k(t-x)\ dt\ (1-q_2)\ dF_{G+S}(x) \\ &= \int_0^\infty \int_0^\infty e^{-\xi(u+x)} P_k(u)\ du\ (1-q_2)\ dF_{G+S}(x) \\ &= \int_0^\infty e^{-\xi u} P_k(u)\ du\ (1-q_2) \int_0^\infty e^{-\xi x}\ dF_{G+S}(x) \\ &= P_k^*(\xi)\ (1-q_2)\ f_{G+S}^*(\xi) \end{aligned}$$

으로 표시되고, 이 관계를 반복하여 적용하면

$$P_{k+1}^*(\xi) = P_1^*(\xi)\left[(1-q_2)\ f_{G+S}^*(\xi)\right]^k$$

으로 되므로

$$\begin{aligned} P^*(\xi) &= \sum_{k=1}^{\infty} P_k^*(\xi) \\ &= \frac{P_1^*(\xi)}{1-(1-q_2)\ f_{G+S}^*(\xi)} \end{aligned}$$

의 결과를 얻는다.

위의 정리에서 주의할 것은 첫번째 자유시간과 재소시간의 분포형태가 두번째 이후의 것들과 달라도 계속 성립한다는 점이다.

첫번째 자유시간의 밀도함수는

$$
\begin{aligned}
f_{G_1}(t) &= \sum_{n=1}^{\infty} \Pr\{n-1\text{회 중단하지 않음}\} \\
&\quad \times [t\text{ 시점에서의 } n\text{회 범죄시간의 밀도함수}] \times \Pr\{\text{투옥 또는 중단}\} \\
&= \sum_{n=1}^{\infty} (1-p-q_1)^{n-1}\ Gam(n,\theta;\ t)\ (p+q_1) \\
&= (p+q_1)\theta\ e^{-(p+q_1)\theta t} \\
&= Exp((p+q_1)\theta;\ t)
\end{aligned}
$$

으로 구하여진다. 위 식에서 $Gam(n,\theta;t)$는 n/θ를 평균으로 갖는 t시점에서의 감마분포밀도함수, $\mathrm{Exp}((p+q_1)\theta;\ t)$는 $1/[(p+q_1)\theta]$를 평균으로 갖는 t시점에서의 지수분포밀도함수를 나타낸다. t 시점에서 첫번째 재소상태일 확률은

$$
\begin{aligned}
P_1(t) &= \Pr\{G_1 < t < G_1 + S_1\} \\
&= \int_0^t \Pr\{\text{입소/자유시간 종결}\}\Pr\{t-x < S_1\}\ dF_{G_1}(x) \\
&= \int_0^t \frac{p}{p+q_1} e^{-\delta(t-x)}(p+q_1)\theta e^{-(p+q_1)\theta x} dx \\
&= \frac{p\theta}{(p+q_1)\theta-\delta}[e^{-\delta t} - e^{-(p+q_1)\theta t}]
\end{aligned}
$$

으로 구하여지고, 라플라스 변환은

$$
\begin{aligned}
P_1^*(\xi) &= \frac{p\theta}{(p+q_1)\theta-\delta}\left[\frac{1}{\xi+\delta} - \frac{1}{\xi+(p+q_1)\theta}\right] \\
&= \frac{p\theta}{(\xi+\delta)[\xi+(p+q_1)\theta]}
\end{aligned}
$$

이 된다. 자유시간과 재소시간의 분포및 라플라스 변환은

$$
f_G(t) = (p+q_1)\theta e^{-(p+q_1)\theta t}, \qquad f_G^*(\xi) = \frac{(p+q_1)\theta}{\xi+(p+q_1)\theta}
$$

$$
f_S(t) = \delta e^{-\delta t}, \qquad f_S^*(\xi) = \frac{\delta}{\xi+\delta}
$$

이므로, 두 분포의 독립성 가정에 의하여 두 분포의 합성(convolution)의 라프라스 변환은

$$
f_{G+S}^*(\xi) = f_G^*(\xi) f_S^*(\xi) = \frac{(p+q_1)\theta}{\xi+(p+q_1)\theta}\ \frac{\delta}{\xi+\delta}
$$

으로 얻어진다. 따라서 정리 1에 의하여

$$
\begin{aligned}
P^*(\xi) &= \sum_{k=1}^{\infty} P_k^*(\xi) \\
&= \frac{P_1^*(\xi)}{1-(1-q_2)\, f_{G+S}^*(\xi)} \\
&= \frac{\dfrac{p\theta}{(\xi+\delta)[\xi+(p+q_1)\theta]}}{1-(1-q_2)\dfrac{(p+q_1)\theta}{\xi+(p+q_1)\theta}\,\dfrac{\delta}{\xi+\delta}} \\
&= \frac{p\theta}{\xi^2+[(p+q_1)\theta+\delta]\xi+(p+q_1)\theta\delta-(1-q_2)p\theta\delta} \\
&= \frac{p\theta}{(\xi+A)(\xi+B)}\,.
\end{aligned} \tag{1}
$$

단, 위 식에서

$$
\begin{aligned}
A &= \frac{1}{2}\left\{\, [(p+q_1)\theta+\delta]+\sqrt{[(p+q_1)\theta+\delta]^2-4(q_1+pq_2)\theta\delta}\,\right\} \\
B &= \frac{1}{2}\left\{\, [(p+q_1)\theta+\delta]-\sqrt{[(p+q_1)\theta+\delta]^2-4(q_1+pq_2)\theta\delta}\,\right\}
\end{aligned}
$$

와 같이 계산되므로 재소확률은

$$
P(t)=\frac{p\theta}{A-B}(e^{-At}-e^{-Bt}) \tag{2}
$$

으로 구하여진다.

[정리 2] 첫번째 재소시간의 모수를 두번째 이후의 것들과 달리하기 위하여 δ_0로 표시하면,

$$
P(t)=\frac{p\theta}{A-B}(e^{-At}-e^{-Bt})-p\theta(\delta-\delta_0)\frac{(A-B)e^{-\delta_0 t}+(B-\delta_0)e^{-At}+(\delta_0-A)e^{-Bt}}{(\delta_0-A)(A-B)(B-\delta_0)}. \tag{3}
$$

[증명] 식 (1)에서

$$
\begin{aligned}
P^*(\xi) &= \frac{\dfrac{p\theta}{(\xi+\delta_0)[\xi+(p+q_1)\theta]}}{1-(1-q_2)\dfrac{(p+q_1)\theta}{\xi+(p+q_1)\theta}\,\dfrac{\delta}{\xi+\delta}} \\
&= \frac{p\theta(\xi+\delta)}{(\xi+\delta_0)(\xi+A)(\xi+B)} \\
&= \frac{p\theta}{(\xi+A)(\xi+B)}+\frac{p\theta(\delta-\delta_0)}{(\xi+\delta_0)(\xi+A)(\xi+B)}
\end{aligned}
$$

으로 바뀌어 지므로, 라플라스 변환을 풀면 된다.

2.2. 평균 범죄수와 평균 재소수

각 입소전 범죄횟수는 $1/(p+q_1)$를 평균으로 갖는 기하분포를 따른다. M_k를 k번째 자유시간 동안의 평균 범죄횟수라고 하면 중단전까지의 평균 총 범죄횟수는

$$\begin{aligned}
\text{평균총범죄수} &= \sum_{k=1}^{\infty}[\Pr\{\text{입소}\}\Pr\{\text{출소직후 중단않음}\}]^{k-1}M_k \\
&= \sum_{k=1}^{\infty}\left[\frac{p}{p+q_1}(1-q_2)\right]^{k-1}\frac{1}{(p+q_1)} \\
&= \frac{\dfrac{1}{(p+q_1)}}{1-\dfrac{p}{p+q_1}(1-q_2)} \\
&= \frac{1}{q_1+pq_2},
\end{aligned}$$

중단전까지의 평균 총 재소수는

$$\begin{aligned}
\text{평균재소수} &= \sum_{k=1}^{\infty}\Pr\{k\text{번 입소}\}\Pr\{k\text{번째 출소직후 중단}\}\ k \\
&= \sum_{k=1}^{\infty}[\Pr\{\text{입소}\}]^{k}[\Pr\{\text{출소직후 중단않음}\}]^{k-1}\Pr\{\text{출소직후 중단}\}\ k \\
&= \frac{\dfrac{p}{p+q_1}q_2}{\left[1-\dfrac{p}{p+q_1}(1-q_2)\right]^2} \\
&= \frac{pq_2(p+q_1)}{(q_1+pq_2)^2}
\end{aligned}$$

로 계산되므로 자료에서 얻은 추정치를 대입하면 범죄자들이 범죄활동을 중단하기 전까지 평균범죄수와 평균재소수를 추정할 수있다.

2.3. 모수의 추정

k번째 입소해 있는 재소자 i의 j번째 자유시간을 g_{ij}, 재소시간을 s_{ij}, j번째 자유시간에 저지른 범죄수를 k_{ij}로 표시하면, 표본조사시점 $t=\sum_{j=1}^{k_{ij}}(g_{ij}+s_{ij})$이 되므로, i번째 재소자에 대한 우도함수는

$$L_i(\theta,\delta,p,q_1,q_2)=\left[\prod_{j=1}^{k_i}(1-p+q_1)^{k_{ij}-1}\theta e^{-(p+q_1)g_{ij}}p\right]\left[\prod_{j=1}^{k_i-1}\delta e^{-s_{ij}}(1-q_2)\right]e^{-s_{ik_i}}\Big/P(\textstyle\sum_{j=1}^{k_{ij}}(g_{ij}+s_{ij}))$$

로 표시된다. 여기서 $P(\sum_{j=1}^{k_{ij}}(g_{ij}+s_{ij}))$는 식 (2)의 공식을 이용한다. n명의 재소자 자료에서 모수를 추정하는 우도함수는

$$L_i(\theta,\delta,p,q_1,q_2)=\prod_{i=1}^{n} L_i(\theta,\delta,p,q_1,q_2)$$

의 형태로 구하여진다. 만일 첫번째 재소기간의 모수를 δ_0로 바꾸려면

$$L_i(\theta,\delta_0,\delta,p,q_1,q_2)=\left[\prod_{j=1}^{k_i}(1-p+q_1)^{k_{ij}-1}\theta e^{-(p+q_1)g_{ij}}p\right]$$
$$\times\delta_0 e^{-s_{i1}}(1-q_2)\left[\prod_{j=2}^{k_i-1}\delta e^{-s_{ij}}(1-q_2)\right]e^{-s_{ik_i}}\Big/P\left(\sum_{j=1}^{k_{ij}}(g_{ij}+s_{ij})\right)$$

을 쓰면된다. 이 때 분모에 있는 확률은 δ_0를 고려한 식 (3)의 확률이다.

범죄시점사이의 분포형태에 대한 연구는 Holden(1983), Rolph, Chaiken, and Houchen(1981)등에 의해 연구되었는 바, 지수분포가정을 몇가지 범죄유형에만 적용할 수 있다는 결론을 내렸다. Lehoczky(1984)는 지수분포가정을 단순하고 다루기 쉬운 이유로 채택하였다. 표본자료에서 범죄사이의 시간이 지수분포에 적합하지 않게 나오는 것은 범죄자들이 동일(homogeneous)하지 않기 때문이라 생각한다. 이를 해결하기위하여 우리는 확률모수(random parameter) 모형을 생각할 수 있다. 즉, 각 범죄자들의 서로 다른 특성을 반영하기 위하여 각 범죄자들 별로 범죄시간의 모수는 θ_i로 구분지어 생각하고, 이들 각각 θ_i은 범죄집단 전체를 대표하는 어떤 상위의 분포, 예를 들어 감마분포를 따르는 확률변수로 생각할 수있다. 이렇게 하면 자료에서 보이는 이질성(heterogeneity) 문제를 해결할 수있다. 또 다른 방법으로는 각 범죄자들의 q_1, p, q_2를 확률모수로 생각하여 상위분포를 베타분포로 가정하는 것이다. 어떤 방법이 적합할지는 자료에 따라 다르리라고 생각된다.

θ_i가 $Gam(\alpha,\beta)$를 따를 때의 우도함수는

$$L(\alpha,\beta,\delta,p,q_1,q_2)=\prod_{i=1}^{n}\int L_i(\theta_i,\delta,p,q_1,q_2)\frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta_i^{\alpha-1}e^{-\beta\theta_i}d\theta_i$$

로 구하여진다. q_1, p, q_2를 확률모수로 가정하는 경우도 같은 방법으로 우도함수를 세울 수있다.

3. 맺는말

재소자로부터 얻은 자료를 이용하여 재생과정에 변형을 시켜 모형을 만들어 보았다. 단순재생과정에 중단확률, 입소확률, 모수의 확률변수화등을 도입하여 융통성이 높은 모형을 세우므로써 현실 자료에 부합할 수있게 하였다. 주어진 자료에서 첫번째 재소시간 모수 δ_0를 따로 고려하여할지, 자유시간 모수 θ를 확률변수로 하여야할 지, 또는 입소확률 p를 확률변수로 하여야할 지등은 자료에서 추정치를 구하고 우도함수의 최대값을 서로 비교하여 결정할 수있다. 즉, 우도함수의 비율을 χ^2분포에 접근시켜 검정하는 방법을 사용할 수있다.

참고문헌

[1] 심영희(1991). "강도범죄의 실태에 관한 연구," 제 4회 형사정책세미나, 한국형사정책연구원.

[2] Holden, R.T. (1983). "Failure Time Models for Criminal Recidivism," Technical Report, Yale University, Department of Sociology, New Haven, Connecticut, USA.

[3] Kim, Chul E. (1989). *Stochastic Modeling of Incarceration Processes*, Ph.D thesis, Carnegie-Mellon University, Pittsburgh, Pennsylvania, USA.

[4] Lehoczky, John P. (1984). "Random Parameter Stochastic Process Models of Criminal Careers," Technical Report No. 343, Carnegie-Mellon University, Department of Statistics, Pittsburgh, Pennsylvania, USA.

[5] Rolph, J.E., Chaiken, J.M. and Houchen, R.E. (1981). *Methods for Estimating Crime Rates of Individuals*, Report R-2730-NIJ, Rand Corporation, Santa Monica, California, USA.