누적합 통계량을 이용한 축차검정에 관한 연구+

박 창순* 최 기철**

요 약

이 논문에서는 통계량의 누적합을 이용하여 누적합 검정을 정의하고 그 특성에 대해 연구하였으며, 축차확률비 검정과의 상대효율을 정의하였다. 누적합 검정의 검사특성함수와 평균표본수는 Wald의 근사공식과 Wiener과정 근사에 의해 표현될 수 있음을 보였다. 또한 누적합 검정에서 축차확률비 검정과 점근적으로 동일한 효율성을 나타내는 통계량을 선정할 수 있음을 보였다. 관측값이 지수 분포를 할 때 누적합 검정과 축차확률비 검정의 효율을 예를 들어 비교해 보았다.

1. 서 론

통계적 가설검정은 표본의 크기가 관측값이 얻어지기 전에 결정되느냐 또는 관측값에 따라 결정되느냐에 따라 고정표본 검정과 축차 검정으로 구분되며, 축차 검정의 대표적인 예로서 Wald(1947)의 축차확률비 검정(sequential probability ratio test)을 들 수 있다.

축차검정에서 가설에 대한 결정을 내릴 수 있는 가장 작은 표본의 수를 표본수(sample number)라 하며, 검정의 특성은 이 표본수에 의해 결정되어 진다. 축차검정의 검사특성함수(operating characteristic function)와 평균표본수(average sample number)는 일반적으로 그 정확한 표현이 구해지지 않고 있으나 Wald의 근사공식 또는 Wiener과정을 이용한 근사공식으로 표현할 수 있다.

이 논문에서는 축차확률비 검정과 유사한 형태로서 누적합 검정을 정의하고 그 성질에 대한 연구를 하는 것을 주목적으로 하고 있다. 누적합 검정의 절차는 모수의 변화에 민감한 통계량을 선정하여 이 통계량의 누적합이 주어진 경계점을 통과하면 가설에 대한 결정을 내리게 되고, 그렇지 않은 경우는 다음 관측값을 얻어 같은 과정을 반복하게 되는 절차를 나타낸다.

따라서 이 논문에서는 먼저 누적합 검정의 절차에 대한 검사특성함수와 평균표본수의 표현을 구하고 그 다음 누적합 검정의 축차확률비 검정에 대한 상대효율을 정의하여 검정의 효율을 최대로 할 수 있는 통계량의 선정에 대해서 연구하였다.

+ 이 논문은 1988년 한국과학재단 연구비 지원을 받아 수행되었음.

*중앙대학교 정경대학 응용통계학과, 서울 동작구 흑석동 221

**부산 외국어대학 통계학과, 부산 남구 우암동 55-1

2. 누적합검정

누적합 검정은 통계량의 누적합을 이용하여 축차 검정을 하는 것으로서, 이것은 누적합 관리도(cumulative sum control chart)가 통계량의 누적합을 사용하여 공정의 변화를 알아내려는 것과 유사한 방법이다.

축차적으로 얻어지는 관측값 $\{X_i,\ i=1, 2, \cdots\}$가 서로 독립이며 동일한 분포를 따를때 가설

$$H_0: \theta=\theta_0, \quad H_1: \theta=\theta_1 \tag{2. 1}$$

에 대한 누적합 검정은 다음과 같이 정의한다.

정의 1: 가설 (2. 1)에 대한 누적합 검정은 통계량 $T(X_i)$를 이용하여 상수 a_c, $b_c(b_c<a_c)$와 k에 대하여 n번째 표본에서

(i) $\sum_{i=1}^{n}\{T(X_i)-k\}\le b_c$이면, H_o를 채택

(ii) $\sum_{i=1}^{n}\{T(X_i)-k\}\ge a_c$이면, H_1을 채택

(iii) $b_c\le \sum_{i=1}^{n}\{T(X_i)-k\}\le a_c$이면, 다음 관측값 X_{n+1}을 얻는다.

누적합 검정에서 k의 선정은 Wiener과정을 이용하여 근사적으로 최적값인

$$k=\frac{E\{T(X)\ ;\ \theta_0\}+E\{(T)\ ;\ \theta_1\}}{2} \tag{2. 2}$$

으로 설정한다(박창순, 1987). 통계량 T(X)는 누적합 검정의 특성을 결정하는 중요한 요인이므로 모수 θ의 변화에 민감한 통계량(일반적으로 충분통계량)을 선정하여 평균표본수를 작게 하는 것이 바람직하다.

누적합 검정의 검사특성함수와 평균표본수는 관측값의 분포가 알려진 경우에는 Wald근사에 의해, 통계량 T(X)의 평균과 분산만이 알려진 경우에는 Wiener과정 근사에 의해 구할 수 있다. 누적합 통계량의 분포에 Wiener과정 근사를 이용한 예로서 Anderson(1960)과 Reynolds(1975)를 들 수 있다.

2. 1 누적합 검정의 Wald근사

서로 독립이며 동일한 분포를 하는 확률변수 $\{X_i,\ i=1, 2, \cdots\}$의 분포가 일모수 지수족에 속할 때 누적합 검정에서의 검사특성함수와 평균표본수를 구해보자. 먼저 관측값의 확률밀도함수는 다음과 같다.

$$f(x\ ;\ \theta)=\exp\{Q(\theta)T(x)+D(\theta)+S(x)\}, \quad {}^{\forall}x\in R \tag{2 3}$$

이 때 가설 (2. 1)에 대한 누적합 검정의 임계부등식은

$$b_c < \sum_{i=1}^{n} \left\{ T(X_i) - \frac{E\{T(X_i)\,;\,\theta_0\} + E\{T(X_i)\,;\,\theta_1\}}{2} \right\} < a_c \tag{2. 4}$$

가 된다. 같은 확률변수 $\{X_i,\ i=1,\ 2,\ \cdots\}$를 이용하여 또 다른 가설

$$H_0^* : \theta = \theta_2, \quad H_1^* : \theta = \theta_3 \tag{2. 5}$$

에 대한 축차확률비 검정의 절차를 알아보자. 이때 축차확률비 검정의 임계부등식은

$$b_s < \sum_{i=1}^{n} [\{Q(\theta_3) - Q(\theta_2)\}T(X_i) + \{D(\theta_3) - D(\theta_2)\}] < a_s \tag{2. 6}$$

로 표현되며 θ_2와 θ_3를

$$\frac{D(\theta_3) - D(\theta_2)}{Q(\theta_3) - Q(\theta_2)} = \frac{E\{T(X_i)\,;\,\theta_0\} + E\{T(X_i)\,;\,\theta_1\}}{2}$$

을 만족하는 값이라 하면 가설(2. 1)에 대한 누적합검정은 가설(2. 5)에 대한 축차확률비 검정과 동치(equivalent)임을 알 수 있다. 위 내용으로부터 다음정리를 얻는다.

정리 2. 1 : 확률변수 $\{X_i,\ i=1,\ 2,\ \cdots\}$가 (2. 3)의 분포를 할 때, 가설

$$H_0 : \theta = \theta_0, \quad H_1 : \theta = \theta_1$$

에 대한 누적합 검정과 동치인 또 다른 가설

$$H_o^* : \theta = \theta_2, \ H_1^* : \theta = \theta_3$$

에 대한 축차확률비 검정이 존재한다.

따라서 가설 (2. 1)에 대한 누적합 검정의 검사특성함수와 평균표본수는 가설 (2. 5)에 대한 축차확률비 검정의 검사특성함수와 평균표본수의 표현과 동일하며, 이 표현은 Wald의 근사공식을 적용시켜 얻을 수 있다.

가설 (2. 5)에 대한 축차확률비 검정의 표본수를 N_s라 할 때 검사특성함수와 평균표본수는 Wald의 근사공식을 사용하면

$$OC_s(\theta) = \frac{\exp\{a_s h(\theta)\} - 1}{\exp\{a_s h(\theta)\} - \exp\{b_s h(\theta)\}} \tag{2. 7}$$

$$E(N_s\,;\,\theta) = \frac{OC_s(\theta) b_s + \{1 - OC_s(\theta)\} a_s}{E(Z\,;\,\theta)} \tag{2. 8}$$

단, $Z=\log\{f(X;\theta_1)/f(X;\theta_0)\}=\{Q(\theta_3)-Q(\theta_2)\}T(X)+D(\theta_3)-D(\theta_2)$이고 $h(\theta)$는 $E[\exp\{h(\theta)Z\}]=1$이 되는 0이 아닌 유일한 해이다. 따라서 가설 (2. 1)에 대한 누적합 검정의 표본수를 N_c라 할 때 검사특성함수와 평균표본수는 식 (2. 7), (2. 8)로 부터

$$OC_c(\theta)=\frac{\exp\{a_c h(\theta)\}-1}{\exp\{a_c h(\theta)\}-\exp\{b_c h(\theta)\}} \tag{2. 9}$$

$$E(N_c;\theta)=\frac{OC_c(\theta)b_c+\{1-OC_c(\theta)\}a_c}{E(Z;\theta)} \tag{2. 10}$$

가 되며, 여기서 누적합 검정의 정지한계선 a_c와 b_c는 주어진 오류의 확률 α와 β에 대해

$$OC_c(\theta)=1-\alpha,\quad OC_c(\theta)=\beta \tag{2. 11}$$

의 관계식을 만족하는 값으로 설정된다.

2. 2 누적합 검정의 Wiener과정 근사

누적합검정은 Wiener과정에 의해서도 근사될 수 있다. 가설 $H_0:\theta=\theta_0$, $H_1:\theta=\theta_1$에 대한 누적합검정의 임계부등식은

$$b_c<\sum_{i=1}^{n}\{T(X_i)-k\}<a_c$$

이고 표본수 N_c는 다음과 같이 표현된다.

$$N_c=\min[n;\ \sum_{i=1}^{n}\{T(X_i)-k\}\notin(b_c,\ a_c)] \tag{2. 12}$$

통계량 $T(X)$의 평균과 분산을 각각 $\mu(\theta)$와 $v(\theta)$라 할 때 $v(\theta)<\infty$이면 $\{T(X_i)-k\}$는 $n\to\infty$일 때 평균 $\mu(\theta)t$, 분산 $\nu(\theta)t$인 Wiener과정에 따른다. 이 Wiener과정의 표본시간은

$$T^*=\min\{t;X(t)\notin(b_c,\ a_c),\quad t\geq 0\} \tag{2. 13}$$

로 표시되며, T^*의 분포는 N_c의 분포의 근사에 적용될 수 있다. 다음 두 정리는 Wiener과정을 이용하여 누적합 검정을 근사시켰을 때 최적 모수값 k와 정지한계선 a_c, b_c를 선정하는 방법을 설명하고 있다(박창순, 1987).

정리 2. 2 : 가설 $H_0:\theta=\theta_0$, $H_1:\theta=\theta_1$에 대한 누적합검정에서 $E(N_c;\theta_0)$과 $E(N_c;\theta_1)$을 동시에 최소로 하는 최적값 k는

$$k=\{\mu(\theta_0)+\mu(\theta_1)\}/2$$

정리 2. 3 : 가설 $H_0:\theta=\theta_0$, $H_1:\theta=\theta_1$에 대한 누적합 검정에서

$k = \frac{\mu(\theta_0)+\mu(\theta_1)}{2}$ 일 때 제1종, 2종 오류 α, β를 만족하는 모수 a_c, b_c값은

$$a_c = \frac{v(\theta_0)}{\mu(\theta_1)-\mu(\theta_0)} \log \frac{1-\beta}{\alpha}$$

$$b_c = \frac{v(\theta_0)}{\mu(\theta_1)-\mu(\theta_0)} \log \frac{\beta}{1-\alpha}$$

위 두 정리에서 k의 최적값은 α, β에 관계없이 결정되고, a_c, b_c는 주어진 α, β를 만족시키도록 설정할 수 있음을 보였다. 정리 2. 2와 정리 2. 3의 결과를 이용하면 Wiener과정 근사를 통하여 얻은 누적합 검사특성함수와 평균표본수는 Darling과 Siegert(1953)에 의해 다음과 같게 된다.

$$OC_c(\theta) = \frac{\exp[-2a_c\{\mu(\theta)-k\}/v(\theta)]-1}{\exp[-2a_c\{\mu(\theta)-k\}/v(\theta)]-\exp[-2b_c\{\mu(\theta)-k\}/v(\theta)]} \tag{2. 14}$$

$$E(N_c\,;\theta) = \frac{OC_c(\theta)b_c+\{1-OC_c(\theta)\}a_c}{\mu(\theta)-k} \tag{2. 15}$$

단, $k = \{\mu(\theta_0)+\mu(\theta_1)\}/2$

$$a_c = \frac{v(\theta_0)}{\mu(\theta_1)-\mu(\theta_0)} \log \frac{1-\beta}{\alpha}$$

$$b_c = \frac{v(\theta_0)}{\mu(\theta_1)-\mu(\theta_0)} \log \frac{\beta}{1-\alpha}$$

3. 상대효율

Wald와 Wolfowitz(1948)는 가설 (2. 1)에 대한 축차확률비 검정은 모수 θ의 참값이 θ_0와 θ_1일 경우에 평균표본수가 가장 작다는 점에서 최적(optimum)검정임을 보였다. 그러나 θ의 참값이 θ_0와 θ_1이 아닌 경우에는 최적 성질이 성립되지 않고 평균표본수가 커지는 경향이 있다. 여기서는 이점을 고려하여 축차확률비 검정에 대한 누적합 검정의 상대효율을 최대화시키는 통계량 $T(X_i)$의 선정에 대해 알아보자.

먼저 서로 독립이며 동일분포를 따르는 확률변수 $\{X_i,\ i=1, 2, \cdots\}$가 확률밀도함수 $f(x\,;\theta)$를 가진다고 하자. 축차확륙비 검정을 S, 누적합검정을 C라 하고, N_s는 S의 표본수, N_c는 C의 표본수라 하면 가설 $H_0: \theta=\theta_0$, $H_1: \theta=\theta_1$에 대한 누적합 검정과 축차확률비 검정의 상대효율을 다음과 같이 표현할 수 있다.

$$RE(C\mid S) = \frac{\int_{\theta_0}^{\theta_1} E(N_s\,;\theta)d\theta}{\int_{\theta_0}^{\theta_1} E(N_c\,;\theta)d\theta} \tag{3. 1}$$

그러나 이러한 상대효율은 오류의 확률 α, β와 θ_0, θ_1 그리고 관측값의 분포에 의존하게 되어 일반적인 결론을 얻는데 있어서 어려움이 있으므로 점근상대효율을 다음과 같이 정의한다.

$$ARE(C \mid S) = \lim_{\theta_1 \to \theta_0} \frac{\int_{\theta_0}^{\theta_1} E(N_s\,;\,\theta)d\theta}{\int_{\theta_0}^{\theta_1} E(N_c\,;\,\theta)d\theta} \tag{3. 2}$$

$\triangle = \theta_1 - \theta_0$, $r = \dfrac{\theta - \theta_0}{\triangle}$ 라 하면 식 (3. 2)의 점근상대효율은

$$ARE(C \mid S) = \lim_{\triangle \to 0} \frac{\int_0^1 E(N_s\,;\,r)dr}{\int_0^1 E(N_c\,;\,r)dr} \tag{3. 3}$$

로 표현된다.

오류의 확률 α, β가 $\alpha > 0$, $\beta > 0$이며, $\alpha + \beta < 1$인 조건을 만족할 때, $\theta_1 \to \theta_0$일 경우는 N_c와 N_s는 무한히 커진다. $\theta_1 \to \theta_0$이고, $\theta \in (\theta_0, \theta_1)$인 경우에 축차확률비 검정에 대한 Wald의 근사식과 누적합검정에 대한 Wiener과정 근사식은 그 표현이 정확하다는 것이 알려져 있다. ARE(C | S)를 구하기 위해 먼저 $\theta_1 \to \theta_0$일 경우에 축차확률비 검정과 누적합 검정의 평균표본수의 표현에 대해 알아보자.

3. 1 축차확률비 검정의 점근 평균표본수

축차확률비 검정의 점근 평균표본수는 식 (2. 7)와 (2. 8)의 표현을 이용하여 구할 수 있다. 먼저 관측값의 확률밀도함수 $f(x\,;\,\theta)$에 관하여 다음 조건을 가정한다.

모든 $\theta \in (\theta_0, \theta_1)$에 대해

$$\frac{\partial^i}{\partial \theta^i} \int \log\ f(x\,;\,\theta)d\theta = \int \frac{\partial^i}{\partial \theta^i} \log\ f(x\,;\,\theta)d\theta, \quad i = 1,\ 2 \tag{3. 4}$$

$\theta_1 \to \theta_0$일 때 식 $E[\exp\{h(\theta)Z\}] = 1$의 0이 아닌 유일한 해 $h(\theta)$는 다음 보조정리에 의해 얻어진다.

보조정리 3. 1 : 확률변수 X가 가정 (3. 4)를 만족하고 $Z = \log \dfrac{f(X\,;\,\theta_1)}{f(X\,;\,\theta_0)}$ 라 하자.

$E[\exp\{h(\theta)Z\}] = 1$의 0이 아닌 유일한 해를 $h(\theta)$라 하면

$$\lim_{\theta_1 \to \theta_0} h(\theta) = 1 - 2r$$

단, $r = \dfrac{\theta - \theta_0}{\theta_1 - \theta_0}$

증명 : (부록 참조)

축차확률비 검정의 임계부등식에서 두 경계선 a_s와 b_s는 Wald근사에 의해 다음과 같이 설정된다.

$$a_s = \log\frac{1-\beta}{\alpha}, \quad b_s = \log\frac{\beta}{1-\alpha} \tag{3. 5}$$

식 (2. 7)과 (2. 8)에 식 (3. 5)를 대입하고 보조정리 3. 1과 부록 (A. 3)을 이용하면 축차확률비 검정의 점근 평균표본수는 다음과 같다.

$$\lim_{\triangle\to 0} E(N_s\,;\, r) = \lim_{\triangle\to 0}\frac{\log\frac{1-\beta}{\alpha}\{1-OC_s{*}(r)\} + \log\frac{\beta}{1-\alpha}OC_s{*}(r)}{I(\theta_0)\triangle^2(r-1/2)+o(\triangle^2)} \tag{3. 6}$$

단, $\triangle = \theta_1 - \theta_0$

$$r = \frac{\theta-\theta_0}{\triangle}$$

$$I(\theta) = E\left[\frac{\partial \log f(x\,;\,\theta)}{\partial\theta}\right]^2$$

$$OC_s{*}(r) = \lim_{\triangle\to 0} OC_s(r) = \frac{\left(\frac{1-\beta}{\alpha}\right)^{1-2r} - 1}{\left(\frac{1-\beta}{\alpha}\right)^{1-2r} - \left(\frac{\beta}{1-\alpha}\right)^{1-2r}}$$

3. 2 누적합검정의 점근 평균표본수

가설 $H_0 : \theta=\theta_0$, $H_1 : \theta=\theta_1$에 대한 누적합검정의 임계부등식은 Wiener과정 근사에 의하여 주어진 α, β를 만족시키기 위해 다음과 같이 설정한다.

$$\frac{v(\theta_0)}{\mu(\theta_1)-\mu(\theta_0)}\log\frac{\beta}{1-\alpha} < \sum_{i=1}^{n}\left\{T(X_i) - \frac{\mu(\theta_0)+\mu(\theta_1)}{2}\right\} < \frac{v(\theta_0)}{\mu(\theta_1)-\mu(\theta_0)}\log\frac{1-\beta}{\alpha} \tag{3. 7}$$

이 경우 누적합검정의 검사특성함수는 식 (2. 14)로 부터 아래와 같이 얻어진다.

$$OC_c(\theta) = \frac{\exp[P(\theta)]-1}{\exp[P(\theta)]-\exp[Q(\theta)]} \tag{3. 8}$$

단, $P(\theta) = \left\{-2\frac{v(\theta_0)}{\mu(\theta_1)-\mu(\theta_0)}\left(\log\frac{1-\beta}{\alpha}\right)\left(\frac{\mu(\theta)-k}{v(\theta)}\right)\right\}$

$$Q(\theta) = \left\{-2\frac{v(\theta_0)}{\mu(\theta_1)-\mu(\theta_0)}\left(\log\frac{\beta}{1-\alpha}\right)\left(\frac{\mu(\theta)-k}{v(\theta)}\right)\right\}$$

$$k = \frac{\mu(\theta_0)+\mu(\theta_1)}{2}$$

Taylor정리에 의하여 $\mu(\theta)=\mu(\theta_0)+\mu'(\theta)(\theta-\theta_0)+o(\triangle^2)$ 이므로
$\theta\in(\theta_0,\ \theta_1)$에 대하여

$$\begin{aligned}&\lim_{\triangle\to 0}\frac{\mu(\theta)-k}{\mu(\theta_1)-\mu(\theta_0)}\,\frac{v(\theta_0)}{v(\theta)}\\&=\lim_{\triangle\to 0}\frac{2\{\mu(\theta_0)-\mu'(\theta_0)r\triangle+o(\triangle)\}-\{\mu(\theta_0)+\mu'(\theta_0)\triangle+o(\triangle)\}-\mu(\theta_0)}{2\{\mu(\theta_0)+\mu'(\theta_0)\triangle+\mu(\theta_0)+o(\triangle)\}}\\&\qquad\times\lim_{\triangle\to 0}\frac{v(\theta_0)}{v(\theta)}\\&=r-1/2\end{aligned}\tag{3. 9}$$

따라서

$$\lim_{\theta_1\to\theta_0}P(\theta)=\left\{\log\frac{\beta}{1-\alpha}\right\}^{1-2r}\tag{3. 10}$$

$$\lim_{\theta_1\to\theta_0}Q(\theta)=\left\{\log\frac{1-\beta}{\alpha}\right\}^{1-2r}\tag{3. 11}$$

이 되므로 $\theta_1\to\theta_0$일 때 누적합 검정의 검사특성함수는

$$OC_c^*(r)=\lim_{\triangle\to 0}OC_c(r)=\frac{\left(\frac{1-\beta}{\alpha}\right)^{1-2r}-1}{\left(\frac{1-\beta}{\alpha}\right)^{1-2r}-\left(\frac{\beta}{1-\alpha}\right)^{1-2r}}\tag{3. 12}$$

이 되어 축차확률비 검정의 점근 검사 특성함수 $OC_s^*(r)$과 같다. 이 때 $\theta=\theta_0$이면 $r=0$, $\theta=\theta_1$이면 $r=1$이 되어 식 (3. 12)로부터

$$OC_c^*(r)=\begin{cases}1-\alpha, & \theta=\theta_0\text{일 때}\\ \beta, & \theta=\theta_1\text{일 때}\end{cases}\tag{3. 13}$$

임을 쉽게 알 수 있다.

Taylor정리에 의하여

$$\mu(\theta)-k=\mu(\theta_0)+\mu'(\theta_0)r\triangle+o(\triangle)-\{\mu(\theta_0)+\mu(\theta_1)\}/2$$

이고, 또

$$\mu(\theta_1)=\mu(\theta_0)+\mu'(\theta_0)\triangle+o(\triangle)$$

이므로

$$\mu(\theta)-k=\triangle\mu'(\theta_0)(r-1/2)+o(\triangle)$$

이 되어

$$\begin{aligned}
&\frac{\{\mu(\theta)-k\}\{\mu(\theta_1)-\mu(\theta_0)\}}{v(\theta_0)} \\
&= \frac{\{\triangle\mu'(\theta)(r-1/2)+o(\triangle)\}\{\mu'(\theta)\triangle+o(\triangle)\}}{v(\theta_0)} \\
&= \{\triangle^2(r-1/2)+o(\triangle^2)\}\frac{\{\mu'(\theta_0)\}^2}{v(\theta_0)} \qquad (3.\ 14)
\end{aligned}$$

따라서 누적합 검정의 평균표본수는 식 (2. 15)로부터 다음과 같이 표현되어진다.

$$\begin{aligned}
\lim_{\triangle\to 0} E(N_c\ ;\ r) &= \lim_{\triangle\to 0}\frac{\dfrac{v(\theta_0)}{\mu(\theta_1)-\mu(\theta_0)}\log\dfrac{1-\beta}{\alpha}[1-OC_c^*(r)]}{\mu(\theta)-k} \\
&\quad+\frac{\dfrac{v(\theta_0)}{\mu(\theta_1)-k(\theta_0)}\log\dfrac{\beta}{1-\alpha}OC_c^*(r)}{\mu(\theta)-k} \\
&= \lim_{\triangle\to 0}\frac{\log\dfrac{1-\beta}{\alpha}\{1-OC_c^*(r)\}+\log\dfrac{\beta}{1-\alpha}OC_c^*(r)}{\{\triangle^2(r-1/2)+o(\triangle^2)\}\dfrac{\{\mu'(\theta_0)\}^2}{v(\theta_0)}} \qquad (3.\ 15)
\end{aligned}$$

3. 3 점근상대효율

3. 1과 3. 2절에서 축차확률비 검정과 누적합 검정의 점근 평균표본수의 표현에 대해 알아보았다. 식 (3. 6)과 (3. 15)를 식 (3. 3)에 대입하면 점근상대효율은 다음과 같다.

$$\begin{aligned}
ARE(C\mid S) &= \lim_{\triangle\to 0}\frac{\displaystyle\int_0^1\frac{\log\dfrac{1-\beta}{\alpha}\{1-OC_s^*(r)\}+\log\dfrac{\beta}{1-\alpha}OC_s^*(r)}{I(\theta_0)\triangle^2(r-1/2)+0(\triangle_2)}dr}{\displaystyle\int_0^1\frac{\log\dfrac{1-\beta}{\alpha}\{1-OC(r)\}+\log\dfrac{\beta}{1-\alpha}OC(r)}{[\{\mu'(\theta)\}^2/v(\theta)]\{\triangle^2(r-1/2)+o(\triangle^2)\}}dr} \\
&= \frac{\{\mu'(\theta_0)\}^2}{I(\theta_0)v(\theta_0)} \qquad (3.\ 17)
\end{aligned}$$

한편, 통계량 T(X)는 $\mu(\theta)$의 불편추정량이므로 Cramer-Rao의 부등식에 의해

$$v(\theta)\geq\frac{\{\mu'(\theta)\}^2}{I(\theta)}$$

이고, 따라서 모든 θ에 대해

$$ARE(C\mid S) = \frac{\{\mu'(\theta_0)\}^2}{v(\theta_0)I(\theta_0)}\leq 1 \qquad (3.\ 18)$$

임을 알 수 있다. Cramer-Rao부등식에서 등호가 성립하는 경우는 T(X)가 $\frac{\partial}{\partial\theta}\log f(X;\theta)$의 선형함수일 때이므로 누적합 검정에서 사용되는 통계량 T(X)를 $\frac{\partial}{\partial\theta}\log f(X;\theta)$의 선형함수로 선정하면 점근상대효율은 1이 되어 최대값이 된다.

위 결과로 부터 다음정리를 얻는다.

정리 3. 1 : 확률변수 $\{X_i, i=1, 2, \cdots\}$가 서로 독립이고 동일한 일모수 지수족의 분포를 따르며 그 확률밀도함수가

$$f(x;\theta) = \exp\{Q(\theta)T(x)+D(\theta)+S(x)\} \quad {}^{\forall}x\in R$$

일 때, 축차확률비 검정에 대한 누적합 검정의 점근상대효율은

$$ARE(C \mid S) = 1$$

4. 예

서로 독립이며 동일한 분포를 따르는 확률변수 $\{X_i, i=1, 2, \cdots\}$가 모수 λ를 갖는 지수분포를 따를 때 누적합 검정의 특성과 상대효율을 Wald근사에 의해 알아보자.

$\{X_i\}$가 지수분포를 하는 경우 확률밀도함수는

$$f(x;\lambda) = \lambda e^{-x},\ \lambda>0,\ x>0 \tag{4. 1}$$

가설

$$H_0;\lambda=\lambda_0,\quad H_1;\lambda = \lambda_1(\lambda_0<\lambda_1) \tag{4. 2}$$

에 대한 누적합 검정의 임계부등식은 식 (2. 4)에 의해

$$b_c < \sum_{i=1}^{n}\left(-X_i + \frac{1/\lambda_0 + 1/\lambda_1}{2}\right) < a_c \tag{4. 3}$$

로 나타내어진다. 만약 λ_2, λ_3가

$$(1/\lambda_0+1/\lambda_1)/2 = \left(\log\frac{\lambda_3}{\lambda_2}\right)/(\lambda_3-\lambda_2) \tag{4. 4}$$

을 만족하는 상수라면, (4. 3)식은

$$(\lambda_3-\lambda_2)b_c < \sum_{i=1}^{n}\{-(\lambda_3-\lambda_2)X_i+\log(\lambda_3/\lambda_2)\} < (\lambda_3-\lambda_2)a_c \tag{4. 5}$$

가 된다.

또 다른 가설

$$H_0^* : \lambda = \lambda_2,\ H_1^* : \lambda = \lambda_3,\ (\lambda_2 < \lambda_3) \tag{4. 6}$$

에 대한 축차확률비 검정은

$$Z_i = \log \frac{f(X_i\,;\,\lambda_3)}{f(X_i\,;\,\lambda_2)} = -(\lambda_3-\lambda_2)X_i + \log(\lambda_3/\lambda_2)$$

이므로, 그 임계부등식은

$$b_s < \sum_{i=1}^{n} \{-(\lambda_3-\lambda_2)X_i + \log(\lambda_3/\lambda_2)\} < a_s \tag{4. 7}$$

이 된다.

따라서 가설 (4. 1)에 대한 누적합 검정과 가설 (4. 6)에 대한 축차확률비 검정은 동치임을 알 수 있다. 누적합 검정의 검사특성함수와 평균표본수는 (4. 7)식을 임계부등식으로 사용한 축차확률비 검정에서 Wald의 근사식에 의해

$$OC_c(\lambda) = \frac{\exp\{(\lambda_3-\lambda_2)a_c h(\lambda)\}-1}{\exp\{(\lambda_3-\lambda_2)a_c h(\lambda)\}-\exp\{(\lambda_3-\lambda_2)b_c h(\lambda)\}} \tag{4. 8}$$

$$ASN_c(\lambda) = \frac{OC_c(\lambda)\,(\lambda_3-\lambda_2)b_c + \{1-OC_c(\lambda)\}\,(\lambda_3-\lambda_2)a_c}{-(\lambda_3-\lambda_2)/\lambda+\log(\lambda_3/\lambda_2)} \tag{4. 9}$$

가 되며, $h(\lambda)$는 $E(e^{hZ}) = \dfrac{\lambda(\lambda_3/\lambda_2)^h}{\lambda+(\lambda_3-\lambda_2)^h} = 1$의 0이 아닌 해이다.

검사특성함수와 평균표본수에 대한 계산은 a_c, b_c가 먼저 결정되어야 하며, 이것은 오류의 확률이 α, β로 주어질 때 식 (4. 8)에서

$$OC_c(\lambda_0) = 1-\alpha$$
$$OC_c(\lambda_1) = \beta$$

를 만족하는 값으로서 Newton방법을 사용하여 구할 수 있다. 이때 a_c, b_c의 초기값으로서 $\log\{(1-\beta)/\alpha\}$와 $\log(\beta/(1-\alpha))$가 사용되었다.

상대효율의 계산을 위해 가설 (4. 2)에서 $\lambda_0=1$로 고정시킨다. $\lambda_0\neq1$인 경우는 단순한 변수변환으로 이를 수정할 수 있다. 이때, $\lambda_1=1.1$, 1.3, 1.5, 2.0에 대한 누적합 검정과 축차확률비 검정의 평균표본수를 $1\leq\lambda\leq\lambda_1$인 λ에 대해 식(4. 9)을 이용하여 구하였다. 축차확률비 검정에 대한 누적합 검정의 상대효율은 식 (3. 1)에 의해 $1\leq\lambda\leq\lambda_1$인 λ에 대하여

$$RE(C\mid S) = \frac{\int_1^{\lambda_1} E(N_s\,;\,\lambda)d\lambda}{\int_1^{\lambda_1} E(N_c\,;\,\lambda)d\lambda} \tag{4. 10}$$

로 정의하고, α, β는 각각 0.001, 0.01, 0.05, 0.1인 경우에 식 (4. 10)의 상대효율을 〈표 1〉에 나타내었다.

〈표 1〉 지수분포에서 축차확률비 검정에 대한 누적합 검정의 상대효율

α	β	$\lambda_1=1.1$	$\lambda_1=1.3$	$\lambda_1=1.5$	$\lambda_1=2.0$
0.001	0.001	0.998	0.993	0.987	0.964
	0.01	0.995	0.981	0.968	0.934
	0.05	0.991	0.972	0.954	0.912
	0.1	0.990	0.968	0.949	0.903
0.01	0.001	1.004	1.005	1.005	0.992
	0.01	0.999	0.993	0.986	0.960
	0.05	0.996	0.984	0.972	0.940
	0.1	0.997	0.981	0.968	0.933
0.05	0.001	1.007	1.015	1.02	1.017
	0.01	1.003	1.003	1.0008	0.985
	0.05	0.999	0.994	0.988	0.967
	0.1	0.998	0.991	0.984	0.962
0.1	0.001	1.007	1.015	1.02	1.017
	0.01	1.003	1.007	1.007	0.996
	0.05	1.0006	0.998	0.994	0.979
	0.1	0.999	0.995	0.991	0.975

〈표 1〉의 결과를 보면 α, β가 고정된 경우에 λ_1이 2.0, 1.5, 1.3, 1.1과 같이 1에 가까이 갈수록 상대효율은 증가하여 점점 1에 가까와지는 것을 알 수 있고, 또한 α가 β보다 상대적으로 큰 경우에는 누적합 검정이 축차확률비 검정보다 좀 더 효율적임을 보여주고 있다. 그러나 상대효율의 값이 1보다 크거나 또한 작은 경우에도 그 값이 거의 1에 가까움을 역시 알 수 있다.

5. 결 론

이 논문에서는 누적합검정에 대한 일반적인 정의와 함께 축차확률비 검정과의 상대효율을 연구하였다. 관측값이 일모수 지수족의 분포를 따를 때 누적합 검정의 특성을 결정짓는 검사특성함수와 평균표본수는 이 검정과 동치인 축차확률비 검정을 찾아내어 Wald의 근사공식에 의해 표현할 수 있고 또한 Wiener과정 근사를 이용하여 표현할 수 있음을 보였다. 일반적으로 축차확률비 검정은 그 검정의 최적성질 때문에 이보다 더 효율적인 축차 검정을 찾아내기는 어려운 점이 있으나, 누적합 검정은 축차확률비 검정과 점근적으로 동일한 효율성을 나타내는 통계량을 선정할 수 있음을 보였다. 축차확률비 검정의 적용을 위해서는 관

측값의 분포의 형태를 정확히 알아야 하지만 누적합 검정은 통계량 T(X)의 평균과 분산만 알려져 있으면 Wiener과정을 이용하여 근사적으로 실행될 수 있는 장점을 가지고 있다. 그 예로서 누적합 검정에서 비모수적 통계량을 사용하는 경우(박창순, 1987)를 들 수 있다. 관측값이 지수분포를 할 때에는 누적합검정이 축차확률비 검정보다 더 효율적인 경우도 있음을 예를 통해 알 수 있었다.

부록 : 보조정리 3. 1의 증명

$z = \log \frac{f(x\,;\,\theta_1)}{f(x\,;\,\theta_0)}$ 라 하면 z는 Taylor정리에 의해

$$z = \triangle \frac{\partial \log f(x\,;\,\theta_0)}{\partial\theta_0} + \frac{1}{2}\triangle^2 \frac{\partial \log f(x\,;\,\theta_0)}{\partial\theta_0^2} + o(\triangle^2) \qquad \text{(A. 1)}$$

또한 $f(x\,;\,\theta)$의 θ에 관한 Taylor정리에 의해

$$\begin{aligned} f(x\,;\,\theta) &= f(x\,;\,\theta_0) + (\theta-\theta_0)\frac{\partial f(x\,;\,\theta_0)}{\partial\theta} + \frac{(\theta-\theta_0)^2}{2}\frac{\partial^2 f(x\,;\,\theta)}{\partial\theta^2} + o(\triangle^2) \\ &= f(x\,;\,\theta_0) + r\triangle\frac{\partial f(x\,;\,\theta_0)}{\partial\theta} + \frac{r^2\triangle^2}{2}\frac{\partial^2 f(x\,;\,\theta_0)}{\partial\theta^2} + o(\triangle^2) \qquad \text{(A. 2)} \end{aligned}$$

조건 (3. 4)로 부터

$$E\left[\frac{\partial \log f(x\,;\,\theta)}{\partial\theta}\right] = 0$$

$$E\left[\frac{\partial \log f(x\,;\,\theta)}{\partial\theta}\right] = -E\left[\frac{\partial^2 \log f(x\,;\,\theta)}{\partial\theta^2}\right] = I(\theta)$$

를 알 수 있고 따라서 식 (A. 1)와 (A. 2)로 부터

$$\begin{aligned} E(Z\,;\,\theta) &= \int \log\frac{f(x\,;\,\theta_1)}{f(x\,;\,\theta_0)} f(x\,;\,\theta)dx \\ &= \int\left[\triangle\frac{\partial \log f(x\,;\,\theta_0)}{\partial\theta_0} + \frac{\triangle^2}{2}\frac{\partial^2 \log f(x\,;\,\theta_0)}{\theta_0^2} + o(\triangle^2)\right] \\ &\qquad \left[f(x\,;\,\theta_0) + r\triangle\frac{\partial f(x\,;\,\theta_0)}{\theta_0} + \frac{r^2\triangle^2}{2}\frac{\partial^2 f(x\,;\,\theta_0)}{\theta_0} + o(\triangle^2)\right]dx \end{aligned}$$

$$= \int \left[\triangle \frac{\partial \log f(x\,;\theta_0)}{\partial\theta_0} + f(x\,;\theta_0)\frac{\triangle^2}{2}\frac{\partial^2 \log f(x\,;\theta_0)}{\partial\theta_0} f(x\,;\theta_0) \right.$$

$$\left. + r\triangle^2 \frac{\partial \log f(x\,;\theta_0)}{\theta_0} \frac{\partial f(x\,;\theta_0)}{\theta^0} + o(\triangle^2) \right] dx$$

$$= \frac{\triangle^2}{2} E\left[\frac{\partial \log f(x\,;\theta_0)}{\theta_0}\;;\theta_0\right] + r\triangle E\left[\frac{\partial \log f(x\,;\theta_0)}{\theta_0}\;;\theta_0\right]^2 + o(\triangle^2)$$

$$= I(\theta_0)\triangle^2(r-1/2)+o(\triangle^2) \tag{A. 3}$$

이제 N을 축차확률비 검정의 표본수라 하고 $h(\theta)$를 $\psi(h)=E\{\exp(hZ)\}=1$의 0이 아닌 유일한 해라면, $S_N = \sum_{i=1}^{N} Z_i$라 할 경우 Wald의 기본항등식(fundamental identity)

$$E\{\exp[hS_N(\psi(h))^{-N}]\}=1 \tag{A. 4}$$

$$E\{\exp[h(\theta)S_N]\} = 1 \tag{A. 5}$$

을 얻는다.

또한 모든 $n>N$에 대해

$$E\{\exp[h(\theta)S_n]\} = 1 \tag{A. 6}$$

따라서 $\theta_1\to\theta_0(\triangle\to 0)$일 경우 $N\to\infty$이므로 모든 양의 정수 $n>N$에 대해 $n\to\infty$이다.

식 (A. 6)와 중심극한 정리에 의해

$$1 = \lim_{\substack{\triangle\to 0\\(n\to\infty)}} E\{\exp[h(\theta)Sn]\}$$

$$= \lim_{\substack{\triangle\to 0\\(n\to\infty)}} \exp\left[E(Z)h(\theta) + \frac{Var(Z)h(\theta)^2}{2}\right]^n \tag{A. 7}$$

임을 알 수 있고, 따라서

$$\lim_{\triangle\to 0}\left\{E(Z)h(\theta) + \frac{Var(Z)h(\theta)^2}{2}\right\} = 0 \tag{A. 8}$$

이다.

$h(\theta)\neq 0$인 값이므로

$$\lim_{\triangle\to 0} h(\theta) = \lim_{\triangle\to 0}\frac{-2E(Z\,;\theta)}{Var(Z\,;\theta)} \tag{A. 9}$$

Ghosh(1970, P139)에 의해

$$\lim_{\triangle\to 0} Var(Z\,;\theta) = \triangle^2 I(\theta_0) + o(\triangle^2) \tag{A. 10}$$

식 (A. 3)과 (A. 10)을 식 (A. 9)에 대입하면

$$\lim_{\triangle \to 0} h(\theta) = 1 - 2r \tag{A. 11}$$

참 고 문 헌

1) Anderson, T.W. (1960), A modification of the sequential probability ratio test to reduce the sample size, The Annals of Mathematical Statistics vol. 31, 165-197.
2) Darling, D.A. & Siegert, A.J.F. (1953), The first passage problem for a continuous Markov process, The Annals of Mathematical Statistics vol. 24, 624-639.
3) Ghosh, B.K. (1976), Sequential Tests of Statistical Hypothesis, Addison-Wesley Publishing Company, Reading, Massachusetts.
4) Park, C.S. (1987), A nonparametric sequential test based on observations in groups, The Korean Journal of Applied Statistics, vol. 1, No. 2, 66-81.
5) Reynolds, M.R. (1975), Approximation to the average run legnth in cummulative sum control charts, Technometrics, vol. 17, 1, 65-71.
6) Wald, A. (1947), Sequential Analysis, Wiley, New York.
7) Wald, A & Wolfowitz, J. (1948), Optimum character of the sequential probability ratio test, The Annals of Mathematical Statistics, vol. 19, 326-329.

A Study on Sequential Test Based on Cumulative Sum of Statistics

Changsoon Park*, Ki Chul Choi**

Abstract

In this paper, a sequential test procedure is defined by using cumulative sum (CUSUM) of statistics. The properties as well as the efficiency of the CUSUM test are studied in comparison with the sequential probability ratio test (SPRT). It was shown that, the operating characteristic function and the average sample number can be derived by Wald and Wiener process approximations. Also it was shown that the statistics used in the CUSUM test is determined to provide asymtotically equivalent efficiency compared to the SPRT. The efficiency of the CUSUM test and the SPRT are cpmpared by an example for some limited number of cases in the exponential distribution.

* Dept. of Applied Statistics, Chung-ang Univ., Dongiak-gu Huksuk-dong 221, Seoul.
** Dept. of Statistics, Pusan Univ. of Foreign Studies, Uam-dong 55-1, Nam-gu, Pusan.