⊠ 연구논문

불량갯수에 대한 축차 샘플링검사

이재헌
광주대학교 응용통계학과
박창순
중앙대학교 응용통계학과
박종태
평택대학교 전산통계학과

Sequential Sampling Inspection Plans for Defectives

Jae-Heon Lee
Dept. of Applied Statistics, Kwangju University
Chang-Soon Park
Dept. of Applied Statistics, Chung-Ang University
Jong-Tae Park
Dept. of Computer Science and Statistics, Pyongtaek University

Abstract

The sequential sampling inspection method is an extension of the double-sampling and multiple-sampling methods and its theory is based on the sequential probability ratio test(SPRT). In this paper, the characteristics of SPRT for testing the propotion of defectives are approximated by using the estimated excess over the boundaries. The use of the estimated excess shows good performances in estimating the operating characteristic function and the average sample number of SPRT compared to the method by neglecting the excess. It also makes it possible to determine the boundary values which satisfy the desired error probabilities.

1. 서론

소비자에게 만족스러운 품질의 제품을 공급하기 위해서는 생산하는 여러 단계에서 품질검사를 실시하여야 한다. 제품의 품질을 검사하는 방법으로는 전수검사(100% inspection)와 샘플링검사(sampling inspection)로 나누어 생각할 수 있다. 샘플링검사는 전수검사가 바람직하지 못한 경우, 즉 파괴검사이거나, 검사량이 너무 많거나, 검사비용이 많이 들 때 유용하게 사용하는 검사방법이다.

축차 샘플링검사(sequential sampling inspection plans)는 다회 샘플링검사의 개념을 확장시켜 만들어진 것으로 동일한 검사특성함수(operating characteristic function : OC function)를 갖는 여러 형식의 샘플링검사 중 평균표본수(average sample number : ASN)를 가장 작게 하는 검사방식이다. 축차 샘플링검사는 Wald(1947)의 축차확률비검정(sequential probability ratio test : SPRT)에 그 이론적 근거를 두고 있다.

Wald(1947)는 두 단순가설 $H_0 : \theta = \theta_0$, $H_1 : \theta = \theta_1$에 대해 다음과 같은 축차확률비검정의 절차를 제안하였다. 확률밀도함수 $f(x\, ;\, \theta)$를 갖고 축차적으로 얻어지는 표본 X_1, X_2, …에서 $Z_i = \log f(X_i\, ;\, \theta_1)/f(X_i\, ;\, \theta_0)$라 하면, 상수 a, $b(-\infty < b < a < \infty)$에 대해 n번째 표본에서

(1) $\sum_{i=1}^{n} Z_i \le b$ 이면 H_0를 채택하고,

(2) $\sum_{i=1}^{n} Z_i \ge a$ 이면 H_0를 기각하고,

(3) $b < \sum_{i=1}^{n} Z_i < a$ 이면 표본추출을 계속한다.

축차확률비검정의 특성을 나타내는 대표적 특성치로는 귀무가설을 채택할 확률을 나타내는 검사특성함수와 검정에 필요한 최소의 표본수를 나타내는 결정표본수(decisive sample number)가 있다. 축차확률비검정의 결정표본수를 N, $S_n = \sum_{i=1}^{n} Z_i$라 하면 결정표본수, 검사특성함수($OC(\theta)$)와 평균표본수(ASN)는 다음을 나타낸다.

$$N = \min\{n\, ;\, S_n \notin (b, a)\}\ ,\quad OC(\theta) = P(S_N \le b\, ;\, \theta)\ ,\quad ASN = E(N)\ .$$

검사특성함수와 평균표본수는 Wald(1947)의 이론(fundamental identity)에 의해 다음과 같이 표현된다. 여기서 $d(\theta)$는 $E(e^{d(\theta)\cdot Z}) = 1$을 만족하는 0이 아닌 유일근이다.

$$OC(\theta)=\begin{cases}\dfrac{E(e^{d(\theta)\cdot S_N}\mid S_N\geq a)-1}{E(e^{d(\theta)\cdot S_N}\mid S_N\geq a)-E(e^{d(\theta)\cdot S_N}\mid S_N\leq b)} & ,\ E(Z)\neq 0\text{인 경우}\\ \dfrac{E(S_N\mid S_N\geq a)}{E(S_N\mid S_N\geq a)-E(S_N\mid S_N\leq b)} & ,\ E(Z)=0\text{인 경우}\end{cases} \tag{1.1}$$

$$E(N)=\begin{cases}\dfrac{E(S_N\mid S_N\geq a)\cdot(1-OC(\theta))+E(S_N\mid S_N\leq b)\cdot OC(\theta)}{E(Z)}, & E(Z)\neq 0\text{인 경우}\\ \dfrac{E(S_N^2\mid S_N\geq a)\cdot(1-OC(\theta))+E(S_N^2\mid S_N\leq b)\cdot OC(\theta)}{E(Z^2)}, & E(Z)=0\text{인 경우}\end{cases} \tag{1.2}$$

그러나 (1.1)과 (1.2)식과 같이 정확한 표현식이 있음에도 불구하고, 대부분의 관찰값 분포에서 위 식에 포함되어 있는 조건부($S_N\geq a$ 또는 $S_N\leq b$) 기대값을 계산할 수 없기 때문에 검사특성함수와 평균표본수의 정확한 값을 산출하는데 어려운 점이 있다.

Wald는 $S_N\geq a$ 또는 $S_N\leq b$ 일 때에 통계량 S_N 이 경계선을 초과하는 양을 무시하고, 위의 식 (1.1)과 (1.2)에서 단순히 $S_N=a$ 또는 $S_N=b$ 를 대입하여 다음과 같은 근사식을 제안하였다.

$$OC(\theta)=\begin{cases}\dfrac{e^{a\cdot d(\theta)}-1}{e^{a\cdot d(\theta)}-e^{b\cdot d(\theta)}} & ,\ E(Z)\neq 0\text{인 경우}\\ \dfrac{a}{a-b} & ,\ E(Z)=0\text{인 경우}\end{cases} \tag{1.3}$$

$$E(N)=\begin{cases}\dfrac{a\cdot(1-OC(\theta))+b\cdot OC(\theta)}{E(Z)} & ,\ E(Z)\neq 0\text{인 경우}\\ \dfrac{a^2\cdot(1-OC(\theta))+b^2\cdot OC(\theta)}{E(Z^2)} & ,\ E(Z)=0\text{인 경우}\end{cases} \tag{1.4}$$

그러나 이 Wald의 근사식은 경계선의 초과를 무시하고 유도된 식이기 때문에 실제로 적용시킬 경우 제1종 또는 제2종의 오류의 확률이 작은 경우에는 정확성이 떨어지게 된다. 이러한 단점을 보완하기 위해 여러 방법들이 연구되어 지고 있으며 그 주된 방법은 경계선의 초과량을 추정하는 것이다. (Lorden(1970), Siegmund(1979), Khan (1978), 박창순(1992), 이재헌 · 박창순 · 김병천(1994))

이재헌 · 박창순 · 김병천(1994)에서는 관측값의 분포가 연속형인 경우 경계선의 초과를 추정하여 축차확률비검정의 특성치를 계산하는 방법을 제안하였다. 본 논문에서는 관측값이 불량품 또는 양호품을 나타내는 이산형분포인 경우 축차확률비검정의 특성치(검사특성함수, 평균표본수)를 간편하면서도 정확하게 계산하는 방법과 주어진 제1종의 오류와 제2종의 오류를 좀 더 정확하게 만족시키는 경계선을 설정하는 방법을 제안하고 있다. 또한 모의실험을 이용하여 제안된 방법의 정확성을 비교하고 있다.

2. 축차확률비검정의 특성치 계산

기호의 단순화를 위해 S_N^a, Z_N^a를 $S_N \geq a$를 조건부로 하는 S_N, Z_N이라 하고, $S_N \leq b$를 조건부로 하는 S_N, Z_N을 S_N^b, Z_N^b라 하자.

관측시점 N에서 통계량 S_N의 경계선 a와 b의 초과에 대한 기대값을 각각 u와 l이라 할 때, u와 l은 다음과 같이 나타낼 수 있다.

$$u = E(S_N^a) - a \ , \quad l = E(S_N^b) - b \ . \tag{2.1}$$

본 논문에서는 이 u와 l을 추정한 다음 기존의 경계선 (b, a)을 새로운 경계선 $(b+l, a+u)$로 대체하여 Wald의 근사식의 부정확성을 보완하고자 한다. 이 때에 경계선의 초과에 대한 기대값 u와 l은 경계선 a와 b에 대한 수정항의 역활을 하게 되는 것이다. 새로운 경계선 $(b+l, a+u)$를 사용하여 수정한 검사특성함수와 평균표본수의 근사식은 다음과 같다.

$$OC(\theta) = \begin{cases} \dfrac{e^{(a+u)\cdot d(\theta)} - 1}{e^{(a+u)\cdot d(\theta)} - e^{(b+l)\cdot d(\theta)}} & , \ E(Z) \neq 0\text{인 경우} \\ \dfrac{a+u}{a+u-(b+l)} & , \ E(Z) = 0\text{인 경우} \end{cases} \tag{2.2}$$

$$E(N) = \begin{cases} \dfrac{(a+u)\cdot(1-OC(\theta)) + (b+l)\cdot OC(\theta)}{E(Z)} & , \ E(Z) \neq 0\text{인 경우} \\ \dfrac{(a+u)^2\cdot(1-OC(\theta)) + (b+l)^2\cdot OC(\theta)}{E(Z^2)} & , \ E(Z) = 0\text{인 경우} \end{cases} \tag{2.3}$$

Wald(1947, pp. 196)는 등식

$$\frac{1-\beta}{\alpha} = E_{\theta_0}(e^{S_N^a}) \ , \quad \frac{1-\alpha}{\beta} = E_{\theta_1}(e^{-S_N^b})$$

을 제안하고, $S_N^a = a$와 $S_N^b = b$ 를 대입하여 제1종의 오류 α와 제2종의 오류 β가 주어졌을 경우 경계선 (b, a)를 다음과 같이 설정하였다.

$$a \approx \log\frac{1-\beta}{\alpha}\ , \quad b \approx \log\frac{\beta}{1-\alpha}\ . \tag{2.4}$$

위의 근사식은 검사특성함수와 평균표본수와 마찬가지로 경계선의 초과를 무시하고 얻어진 식이기 때문에, S_N^a 와 S_N^b를 $a+u$와 $b+l$ 로 수정하여 대입하면

$$a \approx \log\frac{1-\beta}{\alpha} - u\ , \quad b \approx \log\frac{\beta}{1-\alpha} - l \tag{2.5}$$

와 같이 수정된 경계선을 얻을 수 있다.

3. 불량갯수에 대한 축차 샘플링검사

불량갯수에 대한 축차 샘플링검사는 로트로 부터 1개씩 표본을 추출하여 검사해 가면서 누적 불량갯수를 합격 또는 불합격 판정갯수와 비교함으로써, 로트의 합격여부를 결정하는 검사방법이다. 이것을 축차확률비검정의 절차로 바꾸어 나타내 보자.

$\{X_1, X_2, \cdots\}$가 불량률이 p 인 $i.i.d.$ 베르누이 확률변수라면 $H_0 : p = p_0$, $H_1 : p = p_1$ $(0 \le p_0 < p_1 \le 1)$에 대한 축차확률비검정의 절차는 다음과 같다.

(1) $S_n \le b$ 이면 H_0를 채택(로트 합격)하고,

(2) $S_n \ge a$ 이면 H_0를 기각(로트 불합격)하고,

(3) $b < S_n < a$ 이면 표본추출을 계속한다.

베르누이분포에서의 확률비

$$Z_i = \log\frac{1-p_1}{1-p_0} + X_i \cdot \log\frac{p_1(1-p_0)}{p_0(1-p_1)}$$

로 부터

$$P\{Z_i = \log\frac{p_1}{p_0}\} = p\ , \quad P\{Z_i = \log\frac{1-p_1}{1-p_0}\} = 1-p$$

가 되어, $\log \frac{p_1}{p_0} > 0$, $\log \frac{1-p_1}{1-p_0} < 0$임을 이용하면

$$Z_N^a = \log \frac{p_1}{p_0} \quad , \quad Z_N^b = \log \frac{1-p_1}{1-p_0} \tag{3.1}$$

이 성립한다. 그리고,

$$P\{a \le S_N^a < a + Z_N^a\} = 1 \quad , \quad P\{b + Z_N^b < S_N^b \le b\} = 1$$

에 (3.1)식을 대입하면

$$P\{a \le S_N^a < a + \log \frac{p_1}{p_0}\} = 1 \quad , \quad P\{b + \log \frac{1-p_1}{1-p_0} < S_N^b \le b\} = 1 \tag{3.2}$$

을 얻을 수 있다.

여기서 S_N^a와 S_N^b의 분포에 대해 알기가 어렵기 때문에, S_N을 브라운 운동과정(Brownian motion process)으로 근사시키고 식 (3.2)의 성질을 이용하여 다음과 같이 각 변수의 기대값을 추정하였다. (부록 참조)

$$E(S_N^a) \approx a + \frac{1}{2} \log \frac{p_1}{p_0} \quad , \quad E(S_N^b) \approx b + \frac{1}{2} \log \frac{1-p_1}{1-p_0} \tag{3.3}$$

따라서 a의 초과에 대한 기대값 u는

$$u = E(S_N^a) - a \approx \frac{1}{2} \log \frac{p_1}{p_0} \tag{3.4}$$

이 되고, 유사한 방법으로 b의 초과에 대한 기대값 l은

$$l = E(S_N^b) - b \approx \frac{1}{2} \log \frac{1-p_1}{1-p_0} \tag{3.5}$$

과 같이 추정할 수 있다.

식 (3.4)과 (3.5)를 (2.2)과 (2.3)에 대입하면 검사특성함수와 평균표본수를 계산할 수 있다. 여기서 $d(p)$는 $p(p_1/p_0)^{d(p)} + (1-p)[(1-p_1)/(1-p_0)]^{d(p)} = 1$을 만족하는 0이

아닌 유일근이다. 또한 (2.5)에 대입하면 주어진 제1종의 오류(생산자 위험)과 제2종의 오류(소비자 위험)에 대한 경계선을 설정할 수 있다.

4. 모의실험

이 장에서는 모의실험을 통하여 2장과 3장에서 제안한 방법의 정확성을 알아보고자 한다. 불량률에 대한 축차확률비검정의 검사특성함수와 평균표본수를 계산하는 수치해석적인 방법이 있으나(Ghosh(1970), pp. 148-153), 이 방법은 계산절차가 너무 복잡하여 실제 공정에서 사용하기에 매우 어렵다는 단점이 있다. 여기서는 모의실험한 결과를 참값으로 간주하여 Wald의 근사식(WALD)과 새로 제안된 방법(NEW)의 계산결과를 <표 1>에서 비교하였다. 모의실험은 축차확률비검정을 독립적으로 10000번 반복하여 얻은 수치이다. <표 1>을 살펴보면, p_1/p_0이 커질수록 Wald의 방법은 정확성이 크게 떨어짐을 알 수 있으며, 새로 제안된 방법도 정확성이 조금 떨어지나 전반적으로 정확한 결과를 주는 것을 알 수 있다. (여러가지 다른 p_0와 p_1에 대해서도 유사한 결과를 얻었음)

또한 <표 2>에는 제1종과 제2종의 오류 α와 β가 주어졌을 경우 Wald의 (2.4)식과 본 논문에서 제안한 (2.5)식을 이용하여 계산된 경계선 (b, a) 값이 나타나 있다. <표 2>의 결과를 볼 때, Wald의 경계선은 동일한 α와 β에 대하여 p_0와 p_1에 상관없이 동일하지만, (2.5)식을 이용한 경계선은 p_0와 p_1에 따라 달라짐을 알 수 있다. 또한 모의실험을 통하여 정확성을 확인한 결과 본 논문의 방법으로 추정된 경계선이 Wald의 방법에 비해 주어진 제1종과 제2종의 오류를 훨씬 잘 만족하는 것으로 나타났다.

5. 결론

샘플링검사는 품질관리활동에서 매우 중요한 역할을 담당하고 있으며 수입검사, 공정검사, 최종검사, 출하검사 등에 많이 사용된다. 본 논문에서는 불량갯수에 대한 축차샘플링검사에서 검사특성함수와 평균표본수를 추정하는 방법과 제1종의 오류(생산자 위험)과 제2종의 오류(소비자 위험)이 주어졌을 경우 경계선(로트의 합격과 불합격을판단하는 판정선)을 설정하는 방법을 제안하였다.

결론적으로, 본 논문에서 제안된 방법은 경계선의 초과를 추정하여 Wald의 근사식을 보완한 것으로 실제공정에서 쉽게 이용할 수 있으며, 전반적으로 정확한 결과를 얻을 수 있다. 또한 이 방법은 이산형자료에 대한 누적합관리도의 평균런의 길이를 계산할 때에도 유용하게 사용할 수 있을 것이다.

< 표 1 > 검사특성함수와 평균표본수

(1) $p_0=0.1$, $p_1=0.2$

		검사특성함수			평균표본수		
(b, a)	p	SIM.	WALD	NEW	SIM.	WALD	NEW
a=2 b=-2	0.10	0.9032	0.8808	0.9155	44.68	41.52	45.97
	0.12	0.7751	0.7424	0.7877	52.63	47.36	54.88
	0.16	0.3730	0.3610	0.3790	54.24	46.46	56.57
	0.18	0.2138	0.2135	0.2170	48.00	40.65	49.34
	0.20	0.1180	0.1192	0.1168	39.35	34.30	41.26
a=4 b=-4	0.10	0.9856	0.9820	0.9873	108.14	105.10	107.71
	0.12	0.9049	0.8925	0.9104	161.52	153.40	161.49
	0.16	0.2454	0.2420	0.2454	185.02	172.52	190.87
	0.18	0.0666	0.0687	0.0671	130.71	122.43	134.22
	0.20	0.0179	0.0180	0.0170	93.13	86.84	94.66
a=3 b=-2	0.10	0.9686	0.9567	0.9691	51.13	48.60	51.57
	0.12	0.8735	0.8563	0.8802	67.17	62.61	68.95
	0.16	0.4382	0.4277	0.4348	80.46	72.01	83.28
	0.18	0.2425	0.2424	0.2389	70.54	63.43	72.93
	0.20	0.1239	0.1295	0.1237	57.74	52.98	60.31

(2) $p_0=0.01$, $p_1=0.03$

		검사특성함수			평균표본수		
(b, a)	p	SIM.	WALD	NEW	SIM.	WALD	NEW
a=2 b=-2	0.010	0.9219	0.8808	0.9316	179.64	165.23	184.24
	0.014	0.7602	0.7156	0.7876	207.86	181.81	219.67
	0.020	0.4499	0.4155	0.4630	207.74	171.44	222.40
	0.024	0.2836	0.2588	0.2821	183.06	149.61	195.87
	0.030	0.1208	0.1192	0.1248	142.34	115.73	150.46
a=4 b=-4	0.010	0.9861	0.9820	0.9896	421.02	418.29	425.37
	0.014	0.8906	0.8636	0.8946	634.80	613.26	655.38
	0.020	0.3824	0.3357	0.3547	711.95	666.72	767.71
	0.024	0.1180	0.1087	0.1115	533.09	485.49	557.59
	0.030	0.0185	0.0180	0.0179	323.07	292.98	333.98
a=3 b=-2	0.010	0.9697	0.9567	0.9750	199.41	193.44	202.99
	0.014	0.8583	0.8323	0.8726	264.98	244.91	274.56
	0.020	0.5226	0.4962	0.5261	298.26	263.17	316.71
	0.024	0.3061	0.2987	0.3103	269.86	233.65	282.89
	0.030	0.1281	0.1295	0.1306	203.28	178.75	214.50

(3) $p_0 = 0.01$, $p_1 = 0.05$

		검사특성함수			평균표본수		
(b, a)	p	SIM.	WALD	NEW	SIM.	WALD	NEW
a=2 b=-2	0.010	0.9382	0.8808	0.9471	69.02	61.58	71.36
	0.022	0.6402	0.5768	0.6693	79.58	62.31	86.22
	0.030	0.4283	0.3846	0.4426	74.46	55.77	80.82
	0.042	0.2016	0.1924	0.2100	59.21	43.80	63.79
	0.050	0.1207	0.1192	0.1256	49.77	36.89	53.25
a=4 b=-4	0.010	0.9894	0.9820	0.9920	157.47	155.89	159.67
	0.022	0.6821	0.6500	0.7043	272.75	243.50	286.33
	0.030	0.2974	0.2808	0.3008	248.98	211.81	259.74
	0.042	0.0572	0.0537	0.0543	147.93	127.11	154.02
	0.050	0.0176	0.0180	0.0178	105.35	93.39	112.56
a=3 b=-2	0.010	0.9748	0.9567	0.9806	75.82	72.09	77.13
	0.022	0.7269	0.6894	0.7491	107.02	90.71	113.47
	0.030	0.4773	0.4574	0.4930	105.43	86.13	112.72
	0.042	0.2221	0.2166	0.2229	83.31	68.26	89.24
	0.050	0.1305	0.1295	0.1300	70.43	56.98	73.80

(4) $p_0 = 0.001$, $p_1 = 0.01$

		검사특성함수			평균표본수		
(b, a)	p	SIM.	WALD	NEW	SIM.	WALD	NEW
a=2 b=-2	0.001	0.9394	0.8808	0.9628	256.29	226.05	268.99
	0.003	0.7282	0.6348	0.7656	280.17	209.19	308.85
	0.006	0.4023	0.3478	0.4173	240.08	166.12	272.82
	0.008	0.1978	0.1831	0.2066	186.13	127.97	210.61
	0.010	0.1246	0.1192	0.1297	156.98	108.28	176.49
a=4 b=-4	0.001	0.9923	0.9820	0.9943	586.28	572.28	586.57
	0.003	0.8100	0.7513	0.8249	849.57	780.09	931.65
	0.006	0.2663	0.2215	0.2414	721.68	608.15	802.77
	0.008	0.0577	0.0478	0.0491	444.20	365.18	474.70
	0.010	0.0183	0.0180	0.0181	340.10	274.13	354.41
a=3 b=-2	0.001	0.9789	0.9567	0.9863	281.85	264.65	285.01
	0.003	0.8089	0.7526	0.8348	356.46	295.96	383.27
	0.006	0.4426	0.4110	0.4537	341.02	257.94	370.80
	0.008	0.2183	0.2052	0.2155	261.89	199.26	285.15
	0.010	0.1348	0.1295	0.1329	218.34	167.25	236.97

< 표 2 > 주어진 α와 β에 대한 경계선의 설정 및 참오류의 확률

(1) $p_0 = 0.1$, $p_1 = 0.2$

		WALD			NEW		
α	β	(b, a)	α^*	β^*	(b, a)	α^*	β^*
0.01	0.01	(-4.60 , 4.60)	0.0076	0.0087	(-4.54 , 4.25)	0.0107	0.0111
0.01	0.05	(-2.99 , 4.55)	0.0083	0.0487	(-2.93 , 4.21)	0.0100	0.0492
0.01	0.10	(-2.29 , 4.50)	0.0067	0.0941	(-2.23 , 4.15)	0.0119	0.0961
0.05	0.01	(-4.55 , 2.99)	0.0377	0.0104	(-4.49 , 2.64)	0.0578	0.0099
0.05	0.05	(-2.94 , 2.94)	0.0369	0.0482	(-2.89 , 2.60)	0.0554	0.0514
0.05	0.10	(-2.25 , 2.89)	0.0397	0.0956	(-2.19 , 2.54)	0.0587	0.0952
0.10	0.01	(-4.50 , 2.29)	0.0816	0.0089	(-4.44 , 1.95)	0.1087	0.0090
0.10	0.05	(-2.89 , 2.25)	0.0799	0.0484	(-2.83 , 1.90)	0.1106	0.0493
0.10	0.10	(-2.20 , 2.20)	0.0778	0.1002	(-2.14 , 1.85)	0.1070	0.1011

(2) $p_0 = 0.01$, $p_1 = 0.03$

		WALD			NEW		
α	β	(b, a)	α^*	β^*	(b, a)	α^*	β^*
0.01	0.01	(-4.60 , 4.60)	0.0078	0.0090	(-4.58 , 4.05)	0.0114	0.0103
0.01	0.05	(-2.99 , 4.55)	0.0067	0.0467	(-2.98 , 4.00)	0.0120	0.0515
0.01	0.10	(-2.29 , 4.50)	0.0072	0.1003	(-2.28 , 3.95)	0.0141	0.0975
0.05	0.01	(-4.55 , 2.99)	0.0356	0.0093	(-4.54 , 2.44)	0.0607	0.0104
0.05	0.05	(-2.94 , 2.94)	0.0353	0.0518	(-2.93 , 2.40)	0.0601	0.0522
0.05	0.10	(-2.25 , 2.89)	0.0348	0.1036	(-2.24 , 2.34)	0.0577	0.0991
0.10	0.01	(-4.50 , 2.29)	0.0698	0.0113	(-4.49 , 1.74)	0.1246	0.0111
0.10	0.05	(-2.89 , 2.25)	0.0679	0.0487	(-2.88 , 1.70)	0.1233	0.0468
0.10	0.10	(-2.20 , 2.20)	0.0697	0.1009	(-2.19 , 1.65)	0.1244	0.0965

α^*, β^* : 추정된 경계선을 사용했을 경우의 simulation 결과 (참오류의 확률)

(3) $p_0 = 0.01$, $p_1 = 0.05$

		WALD			NEW		
α	β	(b, a)	α^*	β^*	(b, a)	α^*	β^*
0.01	0.01	(-4.60 , 4.60)	0.0053	0.0089	(-4.57 , 3.79)	0.0133	0.0111
0.01	0.05	(-2.99 , 4.55)	0.0053	0.0475	(-2.97 , 3.75)	0.0114	0.0470
0.01	0.10	(-2.29 , 4.50)	0.0062	0.0957	(-2.27 , 3.70)	0.0131	0.0980
0.05	0.01	(-4.59 , 2.99)	0.0287	0.0113	(-4.53 , 2.18)	0.0698	0.0083
0.05	0.05	(-2.94 , 2.94)	0.0302	0.0527	(-2.92 , 2.14)	0.0636	0.0519
0.05	0.10	(-2.25 , 2.89)	0.0284	0.1030	(-2.23 , 2.09)	0.0670	0.0986
0.10	0.01	(-4.50 , 2.29)	0.0622	0.0107	(-4.48 , 1.49)	0.1234	0.0100
0.10	0.05	(-2.89 , 2.25)	0.0609	0.0534	(-2.87 , 1.45)	0.1364	0.0486
0.10	0.10	(-2.20 , 2.20)	0.0590	0.1058	(-2.18 , 1.39)	0.1371	0.1026

(4) $p_0 = 0.001$, $p_1 = 0.01$

		WALD			NEW		
α	β	(b, a)	α^*	β^*	(b, a)	α^*	β^*
0.01	0.01	(-4.60 , 4.60)	0.0050	0.0113	(-4.59 , 3.44)	0.0151	0.0105
0.01	0.05	(-2.99 , 4.55)	0.0040	0.0528	(-2.98 , 3.40)	0.0165	0.0531
0.01	0.10	(-2.29 , 4.50)	0.0038	0.1029	(-2.29 , 3.35)	0.0155	0.1020
0.05	0.01	(-4.59 , 2.99)	0.0264	0.0113	(-4.55 , 1.83)	0.0882	0.0098
0.05	0.05	(-2.94 , 2.94)	0.0242	0.0505	(-2.94 , 1.79)	0.0915	0.0516
0.05	0.10	(-2.25 , 2.89)	0.0223	0.1027	(-2.25 , 1.74)	0.0900	0.0933
0.10	0.01	(-4.50 , 2.29)	0.0421	0.0107	(-4.50 , 1.14)	0.1531	0.0099
0.10	0.05	(-2.89 , 2.25)	0.0399	0.0500	(-2.89 , 1.10)	0.1494	0.0486
0.10	0.10	(-2.20 , 2.20)	0.0439	0.1009	(-2.19 , 1.05)	0.1506	0.0916

α^*, β^* : 추정된 경계선을 사용했을 경우의 simulation 결과 (참오류의 확률)

부록

독립인 확률변수들의 합을 브라운 운동과정(Brownian motion process)으로 근사시키는 것은 축차분석에서 많이 사용하는 방법중에 하나이다. 이 방법은 Reynolds(1975)와 Siegmund(1979, 1986) 등에서 적용된 예를 찾을 수 있다.

$X(t)$가 평균이 $E(Z)\cdot t$, 분산이 $Var(Z)\cdot t$ 인 브라운 운동과정을 따른다고 하고, stopping time인 τ를 $\tau=\inf\{t: X(t)\notin(b,a)\}$라 정의하자. 그러면 결정표본수 N은 브라운 운동과정에서의 τ로 추정될 수 있으며, 또한 S_N은 N이 어느 정도 큰 경우 $X(t)$로 근사시킬 수 있다. 따라서 S_N^a와 S_N^b의 기대값을 $X(t)$를 이용하여 추정하여 보자.

$\mu=E(Z)$, $\sigma^2=Var(Z)$, $a^*=a+\log\frac{p_1}{p_0}$ 이고, $\phi(\cdot)$와 $\Phi(\cdot)$는 각각 표준정규분포의 확률밀도함수와 누적분포함수를 나타낼 때,

$$\begin{aligned} E(S_N^a) &= E(S_N \mid a\le S_N\le a^*) \\ &\approx E(X(t)\mid a\le X(t)\le a^*) \\ &= \frac{\int_{(a-\mu t)/\sigma\sqrt{t}}^{(a^*-\mu t)/\sigma\sqrt{t}} y\cdot\phi(y)\,dy}{\int_{(a-\mu t)/\sigma\sqrt{t}}^{(a^*-\mu t)/\sigma\sqrt{t}} \phi(y)\,dy} \\ &= \mu t - \sigma\sqrt{t}\cdot\frac{\phi(\frac{a^*-\mu t}{\sigma\sqrt{t}})-\phi(\frac{a-\mu t}{\sigma\sqrt{t}})}{\Phi(\frac{a^*-\mu t}{\sigma\sqrt{t}})-\Phi(\frac{a-\mu t}{\sigma\sqrt{t}})} \end{aligned} \qquad (A.1)$$

와 같이 근사된다. 여기서 일반화 평균치정리(the generalized law of the mean)를 이용하면

$$\frac{\phi(\frac{a^*-\mu t}{\sigma\sqrt{t}})-\phi(\frac{a-\mu t}{\sigma\sqrt{t}})}{\Phi(\frac{a^*-\mu t}{\sigma\sqrt{t}})-\Phi(\frac{a-\mu t}{\sigma\sqrt{t}})} = \frac{\phi(\xi)\cdot(-\xi)}{\phi(\xi)} = -\xi \qquad (A.2)$$

를 만족하는 ξ가 $(a-\mu t)/\sigma\sqrt{t}$와 $(a^*-\mu t)/\sigma\sqrt{t}$ 사이에 존재하며, ξ를 ξ가 취할 수 있는 구간의 중앙값으로 추정하면

$$\xi \approx \frac{a+a^*-2\mu t}{2\sigma\sqrt{t}} \qquad (A.3)$$

가 된다. 따라서 식 (A.2)와 (A.3)를 (A.1)에 대입하면

$$E(S_N^a) \approx \frac{a+a^*}{2} = a+\frac{1}{2}\log\frac{p_1}{p_0}$$

이 되며, 유사한 방법으로

$$E(S_N^b) \approx b+\frac{1}{2}\log\frac{1-p_1}{1-p_0}$$

을 얻을 수 있다.

참고문헌

[1] Ghosh, B.K.(1970), *Sequential Tests of Statistical Hypotheses*, Addison-Wesley Publishing Company.

[2] Khan, R.A.(1978), "Wald's Approximations to the Average Run Length in CUSUM Procedures," *Journal of Statistical Planning and Inference*, Vol. 2, pp. 63-77.

[3] Lee, J., Park, C., and Kim, B.(1994), "An Estimation Method for the Excess over the Boundaries in the SPRT and its Applications," *Sequential Analysis*, Vol 13, No. 2, pp. 127-143.

[4] Lorden, G.(1970), "On Excess over the Boundaries," *Annals of Mathematical Statistics*, Vol. 41, No. 2, pp. 520-527.

[5] Park, C.(1992), "An Approximation Method for the Characteristics of the Sequential Probability Ratio Test," *Sequential Analysis*, Vol. 11, No. 1, pp. 55-72.

[6] Reynolds, M.R., Jr.(1975), "Approximations to the Average Run Length in Cumulative Sum Control Charts," *Technometrics*, Vol. 17, No. 1, pp. 65-71.

[7] Siegmund, D.(1979), "Corrected Diffusion Approximations in Certain Random Walk Problems," *Advances in Applied Probabilities*, Vol. 11, pp. 710-719.

[8] Siegmund, D.(1986), "Boundary Crossing Probabilities and Statistical Applications," *Annals of Statistics*, Vol. 14, No. 2, pp. 361-404.

[9] Wald, A.(1947), *Sequential Analysis*, Wiley, New York.