

로버스트 축차 확률비 검정의 설계 및 구현

최인훈¹⁾ · 박노진²⁾

본 논문에서는 정보 전달 능력이 향상된 변형된 축차확률비검정을 소개한다. 새로운 검정은 기존의 검정과 수학적 면에서 공통된 점들을 갖고 있으나, 이상치의 영향을 덜 받고 더욱이 그 존재에 대하여 시각적으로 보여주며, 또한 자료의 변화에 보다 민감하게 반응하는 특성이 있다. 새롭게 제안된 로버스트 축차확률비검정 시스템을 Microsoft사의 Visual Basic 6.0 언어로 구현하여 본 연구에서 제안한 모델과 기존의 검정모델을 비교한 결과 제안된 검정 모델의 우수성과 실용가능성을 실증적으로 확인하였다.

주요용어: 축차 확률비 검정, 비주얼 베이직

1. 서론

단순귀무가설에 대한 단순대립가설을 검정하는 경우 우도비 검정법을 통하여 바람직한 검정을 도출할 수 있다. 만일 표본의 크기가 여러 가지 상황에 의해 미리 고정되지 않는 경우 우도비검정을 확장한 축차확률비검정을 통하여 믿을 만한 검정을 도출할 수 있다. 통계학에 있어서 축차분석론은 고정된 샘플수하에서 다루는 기존의 통계적 개념과 달리 자료에 따라 샘플수를 확률변수로 간주하는 여러 가지 통계적 이론과 방법을 다루는 것이다.

기존의 축차확률비 검정의 경우 사용되는 검정통계량이 이상치의 영향을 직접적으로 받아 검정에 오류가 발생할 수 있기 때문에 이를 방지하기 위해 로버스트 M-추정법을 적용하여 이상치의 영향을 덜어 로버스트 하면서도 정보의 손실을 막을 수 있는 축차확률비 검정을 박노진(1996)이 제안하였다. 본 논문의 마지막 부분의 예에서 볼 수 있듯이 기존의 검정은 이상치가 존재하는 경우 그 능력을 상실할 수 있으나, 제안된 로버스트 검정은 이상치의 영향을 덜 받는 것을 볼 수 있다. 그러나, 여러 가지 수학 계산적 복잡성 때문에 실제 구현을 하기 어려움이 있었다. 본 논문은 제안된 로버스트 축차확률비검정을 누구나 쉽게 접근할 수 있도록 알고리즘을 개발하고 그에 따른 프로그램을 Microsoft사의 Visual Basic 6.0 언어를 이용하여 개발함으로써 다양한 데이터에 의한 결과를 기존의 검정방법과 쉽게 비교, 분석할 수 있도록 하였다.

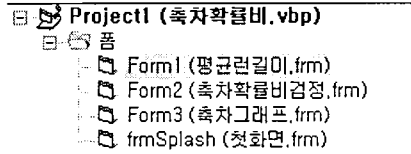
2. 알고리즘과 시스템의 구현

제안된 검정의 구현을 위한 알고리즘은 <표 1>에 정리하였다. 본 장에서는 2장에 제시했던 로버스트 축차확률비검정을 수행하는 방법을 Microsoft사의 Visual Basic 6.0 언어(이하 VB)를

1) 단국대학교 전산통계학과 대학원, (주) 육선

2) 단국대학교 전산통계학과 부교수

기반으로 프로그램을 구현하였다. 프로그램을 정리하면 <그림 1>과 같다. 프로그램은 네 개의 Form으로 구성되어 있다.



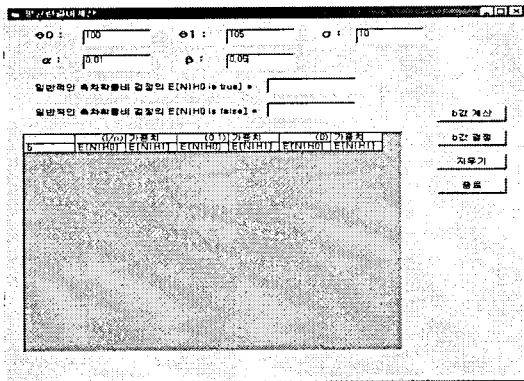
<그림 1> 로버스트 측차확률비 시스템 폼 구조도

프로그램의 작동 방법을 설명하기 위해 측차검정의 대표적인 예제인 Mood의 2인(1974, p471)의 예제27과 같은 가설을 사용하였다. 표본들이 독립적으로 동일하게 $N(\theta, 100)$ 에서 추출되었을 때, $H_0: \theta = \theta_0 = 100, H_1: \theta = \theta_1 = 105$ 인 가설검정을 생각했다.

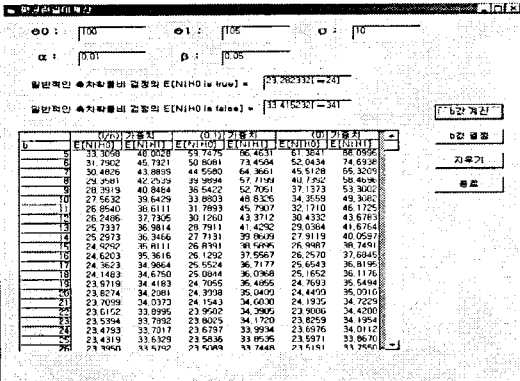
단계1: 평균런길이 계산의 실행

$\alpha = 0.01, \beta = 0.05$ 로 하고 로버스트 측차확률비검정 프로그램에 검정에 사용될 $\theta_0, \theta_1, \sigma, \alpha, \beta$ 값을 입력하여 [b값계산] 명령버튼을 클릭하면 <그림 2>, 기존의 검정의 평균런길이가와 각 가중법에 따른 평균런길이가 계산된다 <그림 3>.

일반적인 측차확률비 검정의 경우, $E[N | H_0 \text{ is true}] = 23.2823[\ll 24]$ 그리고 $E[N | H_1 \text{ is true}] = 33.4152[\ll 34]$ 로 계산된다.



<그림 2> 평균런길이가 계산의 입력 화면



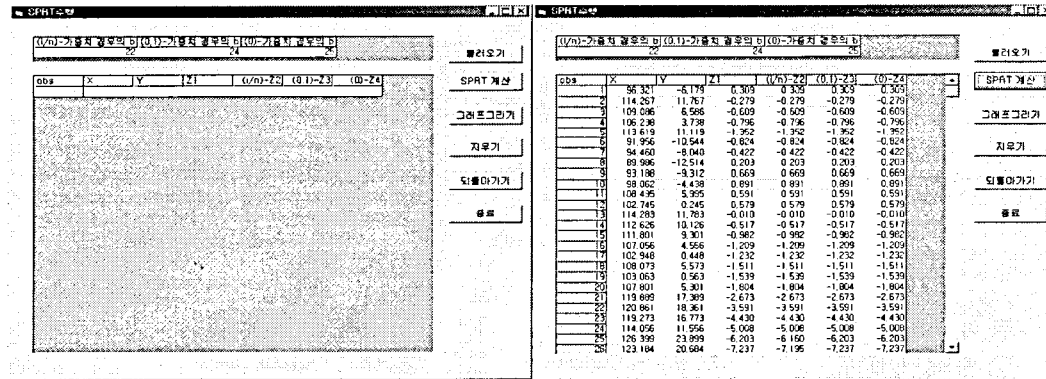
<그림 3> 각 가중치별 평균런길이가 계산 화면

단계 2: 가중치에 따른 Huber-함수의 b값 결정

각각의 가중법들에 대해 Huber의 b값이 주어질 때의 주요 평균런길이들이 <표 2>과 같이 구해진다. 각 가중치에 대하여 b가 커질수록 평균런길이는 기존에 함수를 사용하는 경우에 접근하고, b가 지나치게 작아질수록 기존에 사용되는 측차 확률비 검정 보다 더 많은 표본을 요구하게 된다. 즉, 부적절한 Huber-함수를 사용할 경우 새로운 검정의 효과를 상실한다고 할 수 있다. 그래서 새로운 검정의 평균런길이가 기존 검정의 평균런길이를 넘어서지 않게 하는 b값 중에서 가장 큰 값을 사용할 것을 제안했다(박노진, 1996). <표 2>를 보면 가중법①의 경우 제

안된 검정의 평균런길이가 기존 검정의 평균런길이가 넘어서지 않으면 가장 근사하는 b 값 중에서 가장 큰 값이 22가 된다. 다른 가중법에 대하여도 같은 방법으로 적당한 b 을 구할 수 있게 된다. 프로그램에서는 [b 값결정] 명령버튼을 클릭하면 자동적으로 구해지도록 했다.

기존의 축차 확률비 검정과의 비교를 위해 각 가중치의 평균런길이가 24개와 34개가 되면서 새로운 검정의 효과를 얻도록 하려면 <그림 3>의 상태에서 [b 값결정] 명령 버튼을 클릭하면 Huber-함수의 b 값을 Form2에 출력하게 된다. 즉, (i/n) - 가중법의 경우 $b = 22$, $(0, 1)$ - 가중법의 경우 $b = 24$, (0) - 가중법의 경우 $b = 25$ 로 각각 정해진다.



<그림 4> Huber-함수의 b 값 출력 화면

<그림 5> 기존의 축차확률비검정과 새로 제안된 축차확률비검정의 값 출력 화면

단계 3: 로버스트 축차확률비 검정의 결과

새로운 검정과 기존 검정의 수행능력을 비교하기 위해 처음 30개의 데이터들 $N(110, 100)$ 에서, 나머지 후반부 15개의 데이터는 $N(100, 100)$ 에서 추출하였다. 따라서 데이터가 입력되기 시작하면 $H_0: \theta = 100$ 를 기각하기 위해 검정량이 아래 방향으로 움직일 것이고 후반부에는 $H_1: \theta_1 = 105$ 를 기각하기 위해 위쪽으로 값들이 움직일 것이다. 관측치(X)들을 입력하고 [SPRT계산] 버튼을 클릭하면 각각의 관측치에 대한 $x_i - (\theta_0 + \theta_1)/2$ 값들이 Y열에, 기존의 검정에 의한 z_i 값들이 Z1열에, 가중합 $[(\theta_0 - \theta_1)/\sigma^2][\phi(y_i) + w_i(y_i - \phi(y_i))]$ 값들이 가중법에 따라 Z2열, Z3열, Z4열에 계산되어 각각 출력된다 <그림 5>.

단계 4: 그래프 출력하기

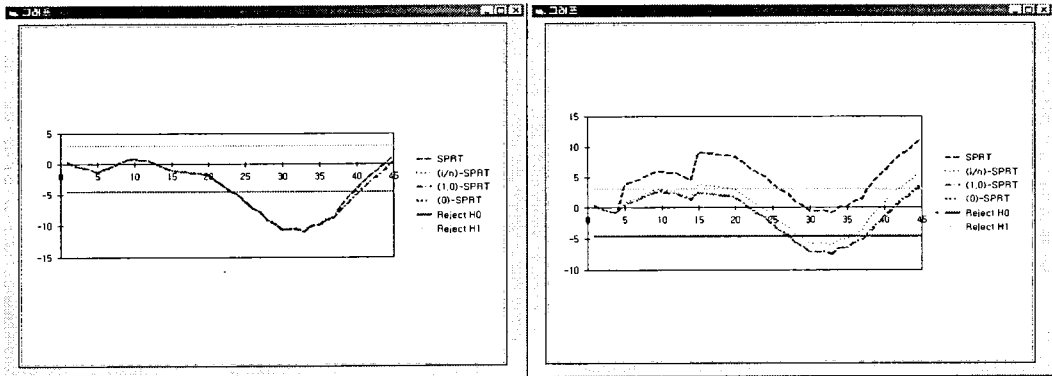
이상치가 없는 경우, 기존의 검정과 새로운 검정들이 같은 방향으로 움직이는 것을 볼 수 있다. <그림 6>에서 보듯이 23번째 자료가 입력될 때 H_1 을 기각하게 되고 30번째 자료가 입력된 후부터 H_0 를 기각하도록 그래프가 위쪽으로 움직이는 것을 볼 수 있다.

한편, 임의로 5번째 관측치 인 113.619는 11.36으로 그리고 15번째 관측치 인 111.801은 11.18로 이상치의 영향을 관측하기 위해 대체했다.

<그림 7>에서 보듯이 기존의 축차 검정은 이상치에 현격하게 영향을 받아 전혀 검정으로서 역할을 감당하지 못하고 있다. 하지만 새로 제안된 검정들은 이상치의 영향에서 거의 벗어나거나 기존의 축차 검정보다 상대적으로 영향을 덜 받고 있음을 알 수 있다. 새로운 검정에 의하

로버스트 축차 확률비 검정의 설계 및 구현

면 26 또는 27번째 자료가 입력되면 H_0 를 기각하게 되고, 43 또는 45번째 자료가 입력되면 H_1 을 기각하게 된다. 이는 이상치가 없을 경우 기존의 검정을 사용하는 경우와 거의 유사한 결과이나, 만일 위의 예가 실제 상황이라면 기존의 검정을 사용하는 경우 5번째 자료가 입력된 뒤 H_1 을 기각하고 검정을 종결했을 가능성이 크다고 할 수 있다.



<그림 6> 이상치가 없는 경우의 그래프 출력 화면 <그림 7> 이상치가 있는 경우의 그래프 출력 화면

3. 맺음말

본 논문에서 소개한 로버스트 축차 검정법은 가능한 많은 방법 중에 하나에 불과하지만 기존의 검정과 함께 사용하면 불필요한 실수를 줄일 수 있는 근거를 제공하리라 믿는다. 프로그램을 수행하기 위해 필요한 많은 내용들이 지면상 생략되었다. 보다 자세한 내용과 프로그램을 원하면 저자에게 연락하면 구할 수 있다.

단 계	내 용
단계1	검정에 사용될 $\alpha, \beta, \mu_0, \mu_1, \sigma$ 의 값을 입력받아 각 가중치에 따른 평균런 길이를 계산한다.
단계2	기존의 축차확률비검정과 비교를 위해 각 가중치의 Huber의 b 값을 결정한다.
단계3	실제 데이터를 입력받아 기존의 축차확률비 검정의 Z_i 값과 각 가중치에 따른 로우버스트 축차확률비 검정의 Z_i 의 값을 계산한다.
단계4	각각의 그래프를 출력한다.

<표 1> 로버스트 축차확률비 시스템 단계별 설계과정

b	(i/n) - 가중치		(0,1) - 가중치		(0) - 가중치	
	$E[N H_0]$	$E[N H_1]$	$E[N H_0]$	$E[N H_1]$	$E[N H_0]$	$E[N H_1]$
5	33.3098	48.0028	59.7475	86.4631	31.3841	88.0996
15	24.9292	35.8111	26.8391	38.5895	26.9987	38.7491
20	23.8374	34.2081	24.3998	35.0409	24.4499	35.0910
21	23.7099	34.0373	24.1543	34.6838	24.1935	34.7229
22	<u>23.6152</u>	<u>33.8995</u>	23.9582	34.3985	23.9886	34.4288
23	23.5394	33.7892	23.8025	34.1720	23.8259	34.1954
24	23.4793	33.7017	<u>23.6797</u>	<u>33.9934</u>	23.6976	34.0112
25	23.4319	33.6329	23.5836	33.8535	<u>23.5971</u>	<u>33.8670</u>
35	23.2882	33.4238	23.2941	33.5121	23.2947	33.5151
50	23.2823	33.4152	23.2823	33.4152	23.2823	33.4152
60	23.2823	33.4152	23.2823	33.4152	23.2823	33.4152

<표 2> Huber-함수를 사용할 때 세 가지 가중 방법에 따른 평균런길이 비교

참고문헌

- [1] 박노진(1996), "정보전달이 보다 효과적인 변형된 축차확률비검정", 「응용통계연구」, 9권 2호, pp.109-117.
- [2] 송문섭(1995), 「로버스트통계」, 자유아카데미.
- [3] 이형배(1998), 「이형배의 비주얼베이직」, 사이버출판사.
- [4] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. J.(1986), *Robust Statistics, the Approach Based on Influence Function*, New York: Wiley, 1986.
- [5] Huber, P. J.(1981), *Robust Statistics*, New York: Wiley.
- [6] Mood, A. M., Graybill, F. A., Boes, D. C. (1974), *Introduction to the Theory of Statistics*, New York: McGraw Hill, 1974.
- [7] Richard, L. B., J. Douglas, F.(1999), *Numerical Analysis*, International Thomson Publishing Asia.
- [8] Staudte, R. G. and Sheather, S. J.(1990), *Robust Estimation and Testing*, New York: Wiley.