韓國數學敎育學會誌 시리즈 A <數學敎育>
1996. 6. 제 35권, 제 1호, 57-59.

Journal of the Korea Society of Mathematical Education
Series A: Jun. 1996, Vol. 35, No. 1, 57-59.

표준정규분포에 있어서 누적분포함수의 근사식

이 장 택 (단국대학교)
오 희 정 (단국대학교 박사과정)

I. 서 론

일반적으로 정규분포에서의 확률계산은 정규확률변수를 표준정규확률변수로 바꾸어서 구한다. 하지만 표준정규분포에 있어서 다음 (식 1)과 같이 주어지는 누적분포함수 $F(z)$의 값은 적분을 필요로 하며 이 적분은 구하기가 매우 어렵다. 따라서 주어진 z에 대한 누적분포함수 값을 계산하기 위해서는 교재에 첨부되는 표준정규분포표를 이용하는 것이 보통이다.

$$F(z)=\int_{-\infty}^{z}(2\pi)^{-\frac{1}{2}}e^{-\frac{x^2}{2}}dx \qquad (식\ 1)$$

하지만 표준정규분포표가 없는 경우에는 정규분포에 대한 확률계산이 거의 불가능하여지므로 (식 1)을 z에 대한 함수식으로 표현하기 위한 노력이 진행되었는데, Schader와 Schmid (1989)는 이항분포의 정규근사의 절대오차의 최대값을 줄이는데 역점을 두었으며, Norton (1989)은 Shah (1985)의 선택함수를 발전시켜 표준정규분포의 근사값을 구하였다. 이외에도 Page(1977), Lim(1989) 등에 의해 여러 가지 근사식이 제시되었다. 본 연구에서도 이와 같은 목적으로 기존의 연구들에서 제시된 방법들보다 좀 더 정확한 값을 제공하는 근사식을 만드는데 초점을 맞춘다.

II. 누적분포함수의 근사식

(식 1)에서 z와 누적분포함수 $F(z)$의 관계식을 함수로 표현하기 위하여 다음 (식 2)와 같은 변수변환을 고려하여 보자.

$$t=\ln\left(\frac{p^{\lambda}}{1-p^{\lambda}}\right),\quad \lambda>0 \qquad (식\ 2)$$

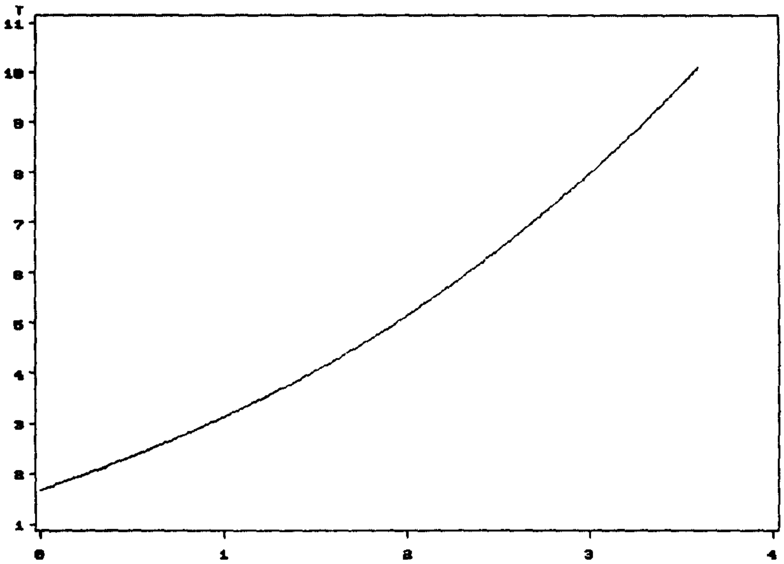
단, $p=F(z)$

변수변환된 t와 z의 관계그래프를 그림으로 표시하면 <그림 1>과 같이 되는데 이는 t와 z의 관계식이 2차식으로 표현될 수 있음을 의미한다.

t와 z의 관계를 2차회귀방정식으로 표현하는데 정확한 값과 근사값의 오차를 줄이기 위하여 여러번의 시행을 거쳐서 (식2)에서의 λ값을 $\lambda=1/4$로 택하였으며, z의 구간을 $0\le z<1$, $1\le z<2$, $z\ge2$와 같이 세부분으로 나누어 각 구간별로 회귀모형을 고려하였다. 각 구간에 대한 회귀모형의 회귀계수 값을 구하기 위하여 z의 값을 0부터 3.59까지 0.01간격으로 사용하였으며 통계팩키지 SAS를 이용하여 구한 회귀방정식은 다음 (식 3)과 같이 주어진다. 이렇게 구하여진 모든 회귀모형은 결정계수의 값이 1이 됨을 확인할 수 있다.

$$t=\begin{cases}1.666939+1.228807z+0.222419z^2, & 0\le z<1\\ 1.799733+0.968713z+0.353971z^2, & 1\le z<2\\ 2.198462+0.575436z+0.452146z^2, & z\ge2\end{cases} \qquad (식\ 3)$$

(식 2)를 이용하면 표준정규분포의 누적분포함수 $F(z)$의 근사식 t은 다음 (식 4)와 같이 나타내어진다. 여기서 의 값은 (식 3)으로 계산되어진다.



<그림 1> 변수변환된 t와 z의 관계그래프

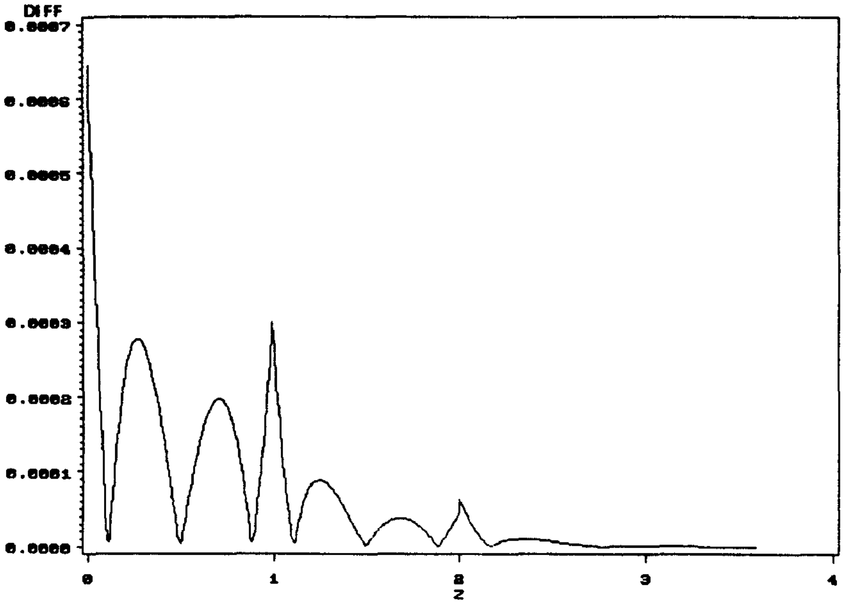
$$F(z) = P(Z < z) \approx \left[\frac{e^t}{1+e^t} \right] \quad \text{(식 4)}$$

<그림 2>는 표준정규분포의 누적분포함수의 값과 본 논문에서 제시된 근사식으로부터 얻어진 누적분포함수의 근사값간의 오차를 그래프로 도식화한 것이다. 이 때 사용된 z의 값은 0부터 3.59까지 0.001간격으로 선택되어졌으며, 각 z값에 따른 누적분포함수의 정확한 값과 (식3)과 (식4)를 이용한 근사값의 차이를 구하여 SAS/GRAPH의 PROC GPLOT과 spline 옵션을 이용하여 작성되었다.

<그림 2>을 보면 정확한 값과 근사적인 값간의 최대오차가 0.0007 미만임을 알 수 있는데, 이 결과는 기존의 연구 결과에서 제시된 방법들의 오차가 대부분 0.01미만인 것에 비하면 본 논문에서 제시된 방법은 그 오차를 크게 줄인다고 할 수 있다.

III. 결 론

표준정규분포에 있어서 누적분포함수의 계산값들은 많은 확률계산에 사용하는 것이 일반적이나 확률밀도함수의 적분식을 계산하는 것이 결코 쉬운 일이 아니며 교재가 없을 때에는 표준정규분포표를 이용하는 것이 거의 불가능하다. 따라서 본 연구에서 제시된 근사식은 기존에 제시된 방법들보다도 표준정규분포의 누적분포함수의 계산값과 근사식으로 계산된 값간의 오차를 0.0007미만으로 줄여주며, 간단한 계산기만 있으면 표준정규분포의 누적분포함수값을 쉽게 계산할 수 있다.



<그림 2> 표준정규분포의 누적분포함수의 정확한 값과 근사값의 오차

참 고 문 헌

Page, E. (1977). Approximations to the cumulative normal function and its inverse for use on a pocket calculator. *Applied Statistics*, Vol. 26. 75-76.

Lim, Jinn-Tyan (1989). Approximating the normal tail probability and its inverse for use on a pocket calculator. *Royal Statistical Society*, Vol. 38. 69-70.

Norton, R. M. (1989). Pocket-calculator approximation for areas under the standard normal curve. *The Americican Statistician*, Vol. 43. 24-26.

Schader & Schmid (1989). Two Rules of thumb for the approximation of the binomial distribution by the normal distribution. *The American Statistician*, Vol. 43. 23-24.

Shch, A. K. (1985). A Simpler approximation for arears under the standard normal curve, *The American Statistician*, Vol. 39. 80.