Journal of the KSQC
Vol. 14, No. 2, 1986

마이크로 컴퓨터를 利用한 正規確率群 分布 函數의 近似計算에 관한 硏究

Approximations for the Normal Family Distribution Function Using Micro Computer

閔 聖 基*
孫 惠 淑**
尹 德 均***

ABSTRACT

A different kind of approximation has been developed in connection with calculation of the normal family distributions in digital computer. These approximations usually employ polynomial expressions. They give quite high accuracy, sometimes only within definite limits on the values of the variable. Outside these limits they may give quite poor approximations. In this paper we compare these approximations by criteria of C.P.U. time and accuracy using micro computer. Approximation formulas given by Zelen and Severo (1984) are proven to give the most accurate results within allowable C.P.U. time.

Ⅰ. 서 론

1. 연구목적

마이크로 컴퓨터의 보급이 대중화 되어 감에 따라 생산현장에서는 마이크로 컴퓨터를 통한 업무의 전산화를 실시할 수 있게 되었다. 특히 품질관리에 있어서는 업무를 단순, 용이하게 할 수 있었을 뿐 아니라 작업에 필요한 정보를 효율적으로 관리할 수 있게 되었다.

그러나 아직까지 이러한 상황에서 전산화된 통계적 품질관리를 실시함에 있어 기본이 되는 정규확률군에 대한 본격적인 연구가 실시되지 못하고 있는 형편이다.

현재 마이크로 컴퓨터에 사용되고 있는 정규확률군의 수치들은 단순히 어떤 표에 나와있는 수치를 기억시켜 사용하는 형편이다. 그렇기 때문에 생산현장에서 원하는 다양한 데이타를 사용할 수 없는 형편이다.

그러므로 본 논문에서는 정규확률군에 대한 근사식 및 전개도들을 프로그램화하여 마이크로

*漢陽大學校 産業工學科
**漢陽大學校 電子計算學科
***漢陽大學校 産業工學科 敎授

컴퓨터에서 가장 효율적으로 사용될 수 있는 근사식 및 전개식을 찾아 내어, 마이크로 컴퓨터에 선택된 근사식 및 전개식에 대한 알고리즘을 기억시킴으로써 원하는 데이타를 신속, 정확하게 얻을 수 있도록 하고자 한다.

2. 연구범위 및 방법

(1) 정규분포

프로그램화된 정규분포의 누적분포함수의 근사식들을 여러가지 측면에서 평가할 수 있겠지만 여기에서는 마이크로 컴퓨터에 의한 C.P.U. 시간과 정확도를 평가기준으로 검토하였다. 정확도는 Max Error와 Mean Square Error를 사용하였다.

(2) χ^2 분포

부분적분을 이용하여 구한 전개식의 C.P.U. 시간과 정확도의 관계를 검토하였다.

(3) F분포

부분적분을 이용하여 구한 전개식들의 C.P.U. 시간을 비교하여 최적의 식을 검토하였다.

Ⅱ. 수학적 배경

정규확률군에 대한 누적분포함수의 근사식 및 전개식을 통하여 본 논문의 수학적 배경을 제시하고자 한다.

1. 정규분포

정규분포는 확률론에 있어서 중심적인 위치를 차지하고 있다. 어떤 임의의 확률변수가 어떤 확률분포를 하는지 알수 없는 경우 중심극한정리에 의하여 확률변수에 대한 확률들을 근사적으로 계산할 수 있기 때문이다. 그렇기 때문에 많은 통계학자들에 의하여 정규분포의 누적분포 함수의 근사식에 대한 많은 연구가 있었다.

먼저 Hart는 $Pr(X \leq x)$의 근사식으로 $\Phi(x)$ 를 다음과 같이 제시하였다.

$$\Phi(x) \fallingdotseq 1-(x\sqrt{2\pi})^{-1} [\exp(-x^2/2)] \times \left[1 - \frac{(1+bx^2)^{\frac{1}{2}} (1+ax^2)^{-1}}{x\sqrt{\pi/2} + [\frac{1}{2}\pi x^2 + (1+bx^2)^{\frac{1}{2}} \cdot (1+ax^2)^{-1} \exp(-x^2/2)]^{\frac{1}{2}}} \right] \quad \cdots\cdots (1)$$

$$a = \frac{1}{2\pi} [1 + (1+6\pi - 2\pi^2)^{\frac{1}{2}}]$$

$$b = \frac{1}{2\pi} [1 + (1+6\pi - 2\pi^2)^{\frac{1}{2}}]^2$$

이 식은 $5 \leq x \leq 10$에 대하여 0.5×10^{-8}의 평균오차를 나타내며 $x = 5$ 일때는 0.39×10^{-6}, $x = 10$일때는 0.42×10^{-7}의 오차를 갖는다.

Streak는

$$\Phi(x) \fallingdotseq \begin{cases} \frac{1}{2}[1 + \mathrm{erf}(x/\sqrt{2})] & \text{if } x > 0 \\ \frac{1}{2}[1 - \mathrm{erf}(x/\sqrt{2})] & \text{if } x < 0 \end{cases} \quad \cdots\cdots (2)$$

$$\mathrm{erf}(x) \fallingdotseq \frac{2}{\pi} \left[\frac{x}{5} + \sum_{n=1}^{37} \exp\left\{-\left(\frac{n}{5}\right)^2\right\} \cdot \sin(2nx/5)/n \right] \quad \text{for } |x| \leq \frac{5}{2}\pi$$

를 정규분포의 누적분포함수의 근사식으로 제시하였다.

Ladwell은

$$\Phi(x) \fallingdotseq \frac{1}{2} \left[1 + \{ 1 - \exp(-2\pi^{-1} x^2 + \frac{2}{3}\pi^{-1}(\pi-3)x^4) \}^{\frac{1}{2}} \right] \quad \cdots\cdots (3)$$

을 근사식으로 제시하였는데 이식의 최대오차는 $x=2.5$일 때 0.0007이다. 그러나 이 식에

$$-0.0005x^6 + 0.00002x^8 \quad \cdots\cdots (4)$$

을 첨가 시킴으로써 최대오차를 0.00005로 줄일수 있었다.

또한 Hasting은

$$\Phi(x) \fallingdotseq 1-\phi_t(C_1 t + C_2 t^2 + C_3 t^3 + C_4 t^4 + C_5 t^5) \quad \cdots\cdots (5)$$

$$t = 1/(1+px)$$

$$\phi_t = \frac{1}{\sqrt{2}} \exp\{-x^2/2\}$$

$P = 0.2316419$
$C_1 = 0.31938153$
$C_2 = -0.356563782$
$C_3 = 1.78147937$
$C_4 = -1.821255978$
$C_5 = 1.330274429$

을 근사식으로 제시하였다.

Pólya는

$$\Phi(x) \fallingdotseq \frac{1}{2}\left[1+\{1-\exp(-2x^2/\pi)\}^{\frac{1}{2}}\right] \quad \cdots\cdots (6)$$

을 근사식으로 제시하였는데 $x=1.6$일때 0.003의 최대오차를 갖는다.

Burr의 누적분포함수의 근사식

$$\Phi(x) \fallingdotseq \frac{1}{2}\left[G(x)+1-G(-x)\right] \quad \cdots\cdots (7)$$

$$G(x) \fallingdotseq 1-\left[1+(\alpha+\beta x)^C\right]^{-K}$$

$\alpha = 0.644693$
$\beta = 0.161984$
$c = 4.874$
$K = -6.158$

은 $x \fallingdotseq \pm 0.6$일 때, 0.00046의 최대오차를 갖는다.

Zelen과 Severo는 2개의 정규분포의 누적분포함수의 근사식을 제시하였다.
첫번째 근사식은

$$\Phi(x) \fallingdotseq 1-(a_1 t + a_2 t^2 + a_3 t^3)\,\xi(x) \quad \cdots\cdots (8)$$

$$\xi(x) \fallingdotseq (a_4 + a_5 x^2 + a_6 x^4 + a_7 x^6)^{-1}$$

$$t = (1+0.33267x)^{-1}$$

$a_1 = 0.4361836$
$a_2 = -0.1201676$
$a_3 = 0.937298$
$a_4 = 2.490895$
$a_5 = 1.466003$
$a_6 = -0.024393$
$a_7 = 0.178257$

인데 $x \geq 0$에 대하여 오차는 1×10^{-5}이하 이다.
두번째 근사식은

$$\Phi(x) \fallingdotseq 1-\frac{1}{2}(1+a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4)^{-4} \quad \cdots\cdots (9)$$

$a_1 = 0.196854$
$a_2 = 0.115194$
$a_3 = 0.000344$
$a_4 = 0.019527$

인데, $x \geq$에 대하여 오차는 2.5×10^{-4}이하 이다.

2. χ^2분포

자유도가 ϕ인 χ^2분포의 확률밀도함수는

$$f(x,\phi) = \frac{1}{\Gamma(\frac{\phi}{2}) \cdot 2^{\frac{\phi}{2}}} \cdot x^{\frac{\phi}{2}-1} e^{-\frac{x}{2}}$$

이다. 누적분포함수

$$P(x,\phi) = \int_0^x \frac{1}{\Gamma(\frac{\phi}{2}) \cdot 2^{\frac{\phi}{2}}} \cdot t^{\frac{\phi}{2}-1} e^{-\frac{t}{2}}\,dt$$

$$= 1-\frac{1}{\Gamma(\frac{\phi}{2})\, 2^{\frac{\phi}{2}}} \int_x^\infty t^{\frac{\phi}{2}-1} e^{-\frac{t}{2}}\,dt$$

라 할때

$$Q(x,\phi) = \frac{1}{\Gamma(\frac{\phi}{2})\, 2^{\frac{\phi}{2}}} \int_x^\infty t^{\frac{\phi}{2}-1} e^{-\frac{t}{2}}\,dt$$

라 하면 이때 다음과 같은 전개식을 얻을 수 있다.

$$Q(x,\phi) = \frac{(\frac{x}{2})^{\frac{\phi}{2}-1} e^{-\frac{x}{2}}}{\Gamma(\frac{\phi}{2})} + Q(x, \phi-2) \quad \cdots\cdots (10)$$

$Q(x, 2) = e^{-\frac{x}{2}}$

$Q(x, 1) = 2\Phi(-\sqrt{x})$

$\Phi(x)$는 정규분포의 누적확률함수 이다.

3. F분포

자유도가 ϕ_1, ϕ_2인 F분포의 확률밀도함수 $f(F; \phi_1, \phi_2)$는

$$f(F; \phi_1, \phi_2) = \frac{1}{B(\frac{\phi_1}{2}, \frac{\phi_2}{2})} \phi_1^{\frac{\phi_1}{2}} \phi_2^{\frac{\phi_2}{2}} \cdot F^{\frac{1}{2}(\phi_1-2)} (\phi_2+\phi_1 F)^{-\frac{1}{2}(\phi_1+\phi_2)}$$

이다. 이때 누적분포함수 $P(F; \phi_1, \phi_2)$는

$$P(F; \phi_1, \phi_2) = \int_0^F f(t; \phi_1, \phi_2)\,dt = 1-Q$$

$$Q(F; \phi_1, \phi_2) = \int_F^\infty \frac{1}{B(\frac{\phi_1}{2}, \frac{\phi_2}{2})} \phi_1^{\frac{\phi_1}{2}} \phi_2^{\frac{\phi_2}{2}} t^{\frac{1}{2}(\phi_1-2)} (\phi_2+\phi_1 t)^{-\frac{1}{2}(\phi_1+\phi_2)}\,dt$$

$B(\frac{\phi_1}{2}, \frac{\phi_2}{2})$ 는 Beta 함수

이다. $Q(F; \phi_1, \phi_2)$와 $P(F; \phi_1, \phi_2)$에 대한 전개식을 구해보면 다음과 같다.

가. ϕ_1이 우수일때

$x = \frac{\phi_2}{\phi_2+\phi_1 F}$라 하면 부분적분에 의하여

$$Q(F; \phi_1, \phi_2) = x^{\frac{\phi_2}{2}} \{ 1+\frac{\phi_2}{2}(1-x) + \frac{\phi_2(\phi_2+2)}{2 \cdot 4}(1-x)^2 + \cdots\cdots + \frac{\phi_2(\phi_2+2)(\phi_2+4)\cdots(\phi_1+\phi_2-4)}{2 \cdot 4 \cdot 6 \cdots\cdots (\phi_1-2)} (1-x)^{\frac{1}{2}(\phi_1-2)} \} \quad \cdots\cdots (11)$$

와

$$Q(F; \phi_1, \phi_2) = x^{\frac{1}{2}(\phi_1+\phi_2-2)} \{ 1+\frac{\phi_1+\phi_2-2}{2} \cdot (\frac{1-x}{x}) + \frac{(\phi_1+\phi_2-2)(\phi_1+\phi_2-4)}{2 \cdot 4} \cdot (\frac{1-x}{x})^2 + \cdots + \frac{(\phi_1+\phi_2-2)(\phi_1+\phi_2-4)\cdots(\phi_2+2)}{2 \cdot 4 \cdots\cdots (\phi_1-2)} \cdot (\frac{1-x}{x})^{\frac{1}{2}(\phi_1-2)} \} \quad \cdots\cdots (12)$$

나. ϕ_2가 우수일때

$x = \frac{\phi_2}{\phi_2+\phi_1 F}$라 하면 부분적분에 의하여

$$Q(F; \phi_1, \phi_2) = 1-(1-x)^{\frac{\phi_1}{2}} \{ 1+\frac{\phi_1}{2}x + \frac{\phi_1(\phi_1+2)}{2 \cdot 4}x^2 + \cdots\cdots + \frac{\phi_1(\phi_1+2)(\phi_1+4)\cdots\cdots(\phi_1+\phi_2-4)}{2 \cdot 4 \cdot 6 \cdots\cdots (\phi_2-2)} x^{\frac{1}{2}(\phi_2-2)} \} \quad \cdots\cdots (13)$$

와

$$Q(F; \phi_1, \phi_2) \fallingdotseq 1-(1-x)^{\frac{1}{2}(\phi_1+\phi_2-2)} \{ 1+\frac{(\phi_1+\phi_2-2)}{2} \cdot (\frac{x}{1-x}) + \frac{(\phi_1+\phi_2-2)(\phi_1+\phi_2-4)}{2 \cdot 4}(\frac{x}{1-x})^2 + \cdots\cdots + \frac{(\phi_1+\phi_2-2)(\phi_1+\phi_2-4)\cdots(\phi_1+2)}{2 \cdot 4 \cdots\cdots (\phi_2-2)} (\frac{x}{1-x})^{\frac{1}{2}(\phi_2-2)} \} \quad \cdots\cdots (14)$$

를 얻을 수 있다.

다. ϕ_1과 ϕ_2가 기수일때

$\theta = \tan^{-1}\sqrt{\frac{\phi_1}{\phi_2}F}$ 라 할때

$$Q(F; \phi_1, \phi_2) = \beta(\phi_1, \phi_2) + 2Q(\theta, \phi_2) \quad \cdots\cdots$$

$$\beta(\phi_1, \phi_2) = \begin{cases} \dfrac{2}{B(\frac{1}{2}, \frac{\phi_2}{2})} \sin\theta \cos^{\phi_2}\theta \\ \{ 1 + \dfrac{\phi_2+1}{3} \sin^2\theta + \cdots\cdots\cdots\cdots \\ + \dfrac{(\phi_2+1)(\phi_2+3)\cdots(\phi_1+\phi_2+4)}{3\cdot 5 \cdots\cdots\cdots (\phi_1-2)} \\ \sin^{\phi_1-3}\theta \} \quad (\phi_1 > 1) \\ 0 \quad (\phi_1 = 1) \end{cases}$$

$$Q(\theta, \phi_2) = \begin{cases} \dfrac{1}{2} - \dfrac{1}{\pi} [\theta + \sin\theta \{ \cos\theta \\ + \dfrac{2}{3}\cos^2\theta + \cdots\cdots\cdots\cdots \\ + \dfrac{2\cdot 4 \cdots\cdots (\phi_2-3)}{1\cdot 3 \cdots\cdots (\phi_2-2)} \cos^{\phi_2-2}\theta \}] \\ (\phi_2 > 1) \\ \dfrac{1}{2} - \dfrac{\theta}{\pi} \quad (\phi_2 = 1) \end{cases}$$

을 얻었다.

라. Dorrer의 전개식

$$P(F; \phi_1, \phi_2) = P((\frac{\phi_2 - 2}{\phi_2}) F ; \phi_1, \phi_2 - 2)$$

$$+ g(\phi_2, \phi_1, Z)$$

$$= P((\frac{\phi_1}{\phi_1 - 2}) F ; \phi_1 - 2, \phi_2)$$

$$+ g(\phi_1, \phi_2, w) \cdots\cdots\cdots\cdots (16)$$

$$w = F(\phi_1 / \phi_2)$$

$$Z = (\phi_2 / \phi_1) F$$

$$g(\phi_1, \phi_2, Y) = \frac{2\Gamma((\phi_1+\phi_2)/2) \cdot Y^{(\phi_1-2)/2}}{\Gamma(\frac{\phi_1}{2}) \cdot \Gamma(\frac{\phi_2}{2}) \cdot (Y+1)^{(\phi_1+\phi_2-2)/2} \cdot (\phi_1+\phi_2-2)}$$

$$P(F; 2,2) = w / (w+1)$$

$$P(F; 2,1) = 1 - 1\sqrt{w+1}$$

$$P(F; 1,2) = \sqrt{w / w+1}$$

$$P(F; 1,1) = \frac{2}{\pi} \tan^{-1}(\sqrt{w})$$

Ⅲ. 전산결과 분석

(1) 정규분포

Ⅱ에서 제시한 9개의 정규분포 누적분포함수의 근사식을 프로그램화한 결과는 표 1과 같다. 표1.에 나타난 결과에 의하면 C.P.U. 시간과 정확도 측면에서 절대우위에 있는 근사식은 찾을 수 없었다. 그러나 이 근사식들의 사용목적에 따라 Hasting의 근사식과 Zelen과 Severo의 두번째 근사식(9)을 정규분포 누적분포함수에 대한 최적의 근사식으로 제시하고자 한다.

(2) χ^2분포

χ^2분포의 누적분포함수에 대한 근사는 전개식 (10)을 사용함에 있어 어떠한 정규분포의 근사식을 사용하는가에 달려있다.

그리하여 본 논문에서 앞서 정규분포에 대한 최적 근사식으로 제시한 Hasting의 근사식을 사용한 전개식 (10)에 의한 χ^2분포의 누적분포함수의 근사에 대한 정확도는 거의 오차가 없었으며 자유도변화에 따른 C.P.U. 시간의 변화는 그림 1과 같다.

(3) F분포

F분포에 대한 전개식들을 비교 검토한 결과, 정확도는 거의 오차가 없었으며 C.P.U. 시간측면에서 비교해 본 결과는 자유도ϕ_1이 우수인 경우는 전개식 (11)이, 자유도ϕ_2가 우수인 경우는 전개식 (13)이, 자유도ϕ_1과 ϕ_2가 모두 기수인 경우는 전개식 (15)가 F분포에 대한 최적의 전개식으로 채택되었다.

표 1. **Comparision of Normal C.D.F Approximations**

근사식번호 \ 구분	C.P.U. Time (Sec.)	Max Error	Mean Square Error
1	0.42	1×10^{-4}	5.6×10^{-10}
2	3	1×10^{-4}	2.6×10^{-10}
3	0.19	7×10^{-4}	1.55×10^{-7}
4	0.32	1×10^{-4}	1.94×10^{-9}
5	0.31	1×10^{-4}	2.8×10^{-10}
6	0.16	3.2×10^{-3}	3.2061×10^{-8}
7	0.39	5×10^{-4}	2.25×10^{-8}
8	0.26	3.2×10^{-3}	1.32×10^{-6}
9	0.15	2×10^{-4}	2.55×10^{-8}

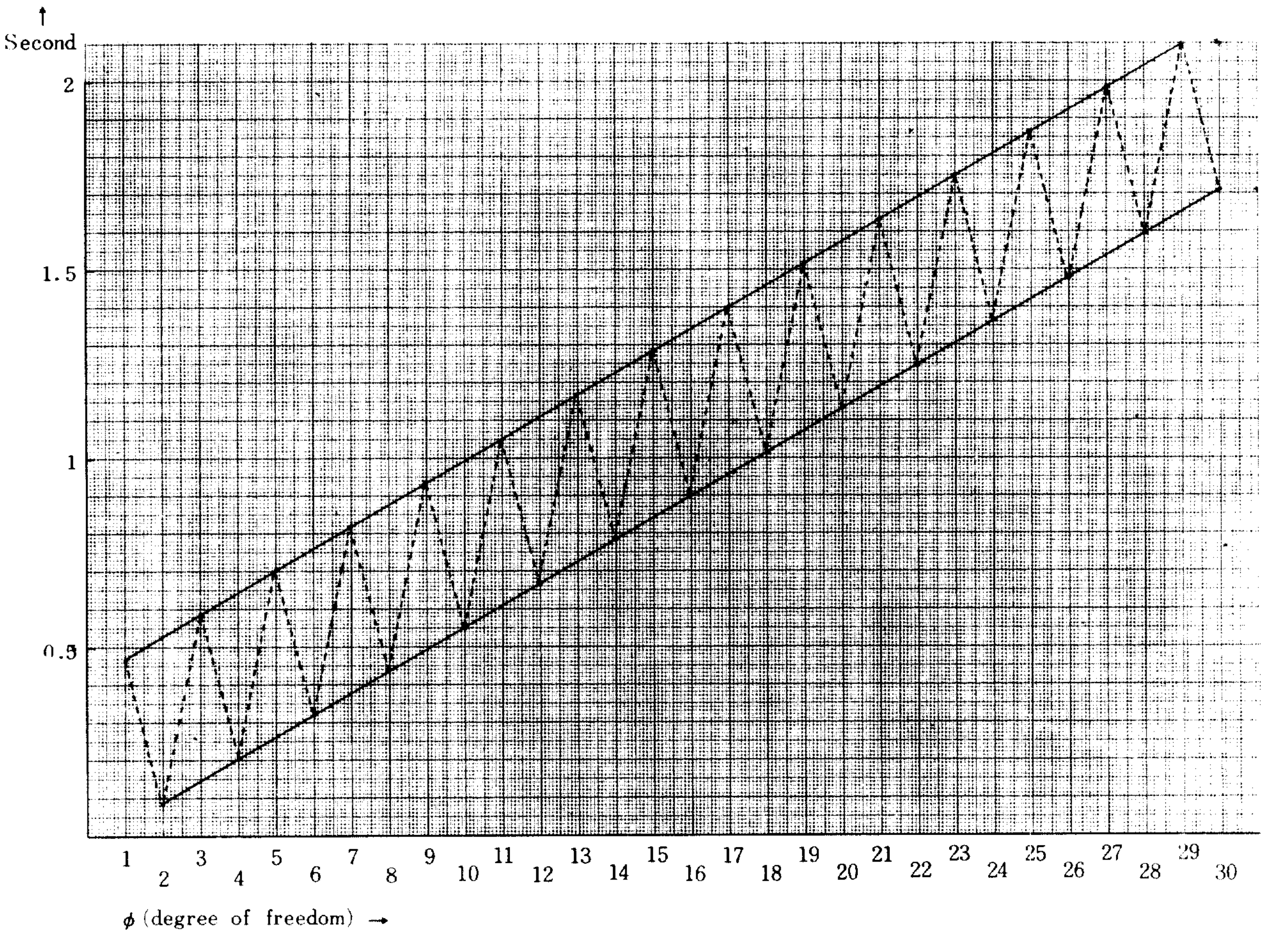

그림 1. **C.P.U. Time of Equation (10)**

Ⅳ. 결　　론

본 연구에서는 기초확률분포의 누적분포함수에 대한 근사식과 전개식을 마이크로 컴퓨터를 이용하여 C.P.U.시간과 정확도를 평가기준으로 하여 비교한 결과 정규분포에서는 Hasting의 식과 Zelen과 Severo의 두번째 근사식을, χ^2분포에서는 (10)식을, F분포에서는 각 경우에 대하여 (11)식, (13)식, (15)식을 마이크로 컴퓨터에서 사용할 수 있는 최적의 근사식과 전개식으로 제시한다.

본 연구는 정규분포 확률변수에 대한 누적확률을 구하는 것이다. 그러나 수치해석적 기법을 사용하여, 이와 반대되는 경우, 즉 누적확률에 대한 확률변수값을 구할 수 있다. 이 확률변수값을 구하는 흐름도는 그림 2와 같다.

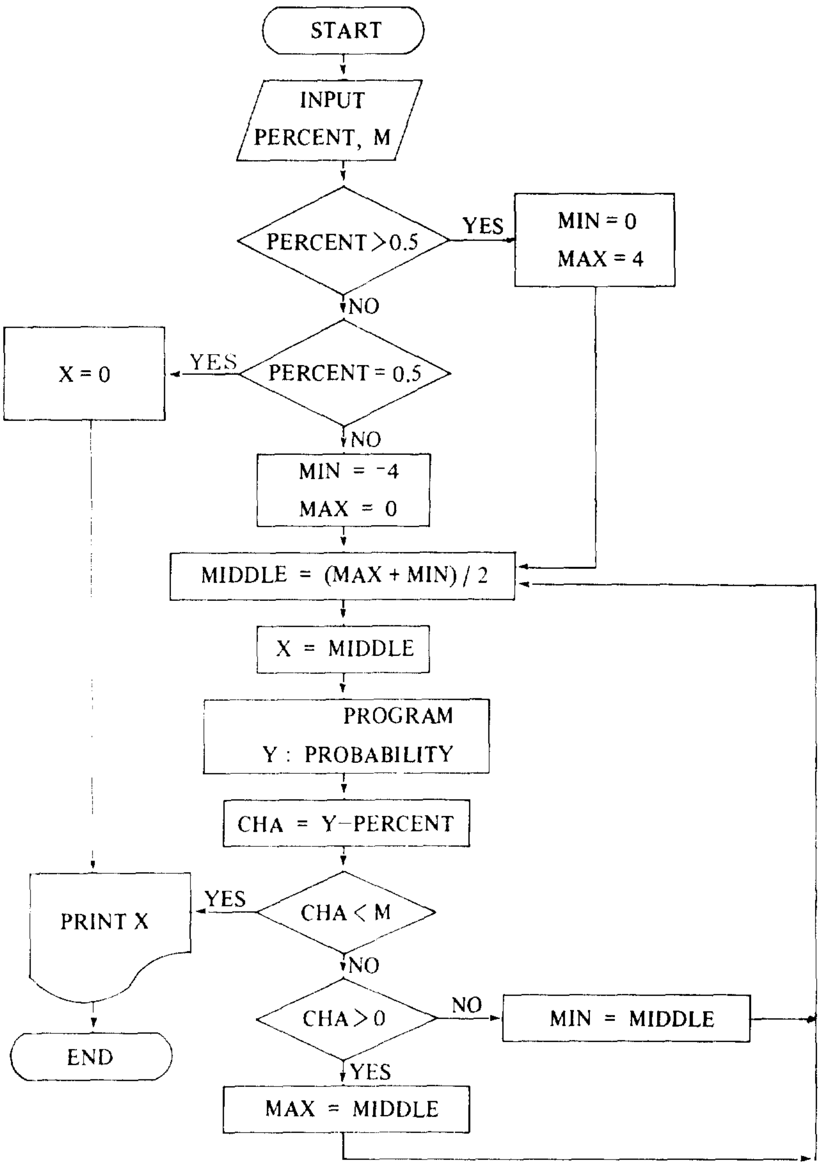

그림 2. **Flow-chart of Percent Point.**

본 연구에서는 t분포에 대하여 논하지 않았으나 t분포와 F분포의 관계는

$$t^2(\phi, \alpha) = F(1, \phi; \alpha)$$

$$P(t;\phi) = \begin{cases} \dfrac{1+P(t^2;1,\phi)}{2} & \text{if} \quad t \geqq 0 \\ \dfrac{1-P(t^2;1,\phi)}{2} & \text{if} \quad t \leqq 0 \end{cases}$$

$P(F;\phi_1,\phi_2)$, F분포의 누적분포 함수

이다. 그러므로 본 연구에서 구한 F분포의 전개식을 사용하여 t분포의 누적분포함수를 구할수 있다.

앞으로 본 연구가 마이크로 컴퓨터에 의한 정규확률군의 연구에 기초가 되기를 바라며 앞으로도 이에 관한 많은 연구가 있기를 바란다.

參 考 文 獻

1. 김영휘, (1972), 공업통계학, 동양사
2. Burr, I.W., (1967), *"A Useful Approximation to the Distribution with Application to Simulation,"* Technometrics, Vol. 19, pp. 647-651, 1967.
3. Craig, R.J., (1984), *"Normal Family Distribution Functions; FORTRAN and BASIC Programs,"* Quality Technology, Vol. 16, No. 4, 1984.
4. Hart, R.G., (1966) *"A Close Approximation Related to the Error Function,"* Mathematics of Computation, Vol. 20, pp. 600-602, 1966.
5. Hastings, C., (1955) Approximations for Digital Computers, Princeton University Press.
6. Johnson, N.L., Kotz, S., (1970) Continuous Univariate Distribution, Houghton Mifflin Company.
7. Zelen. and Severo. N.C., (1984) *"Probability Functions"* in Handbook of Mathematical Function, Ed. N. Abramowitz. and I.A., Stegun., US Dept. of Commerce, Applied Mathematics Series, No. 55, pp. 925-995.