論 文

確率密度函數의 逐次母數推定方式에 관한 研究

正會員 韓 榮 烈* 正會員 朴 鎭 秀**

A Study on the Recursive Parameter Estimation Algorithm of the Probability Density Function

Young Yeul HAN* and Jin Soo PARK**, *Regular Members*

要 約 本論文에서는 平均値가 確率密度函數의 퍼래미터의 函數일 때 確率 1 과 mean square로 수렴하는 새로운 母數推定알고리즘을 提案한다. 提案된 逐次알고리즘은 推定하려는 퍼래미터가 다수의 퍼래미터일 경우라도 적용시킬 수 있으며 컴퓨터 시뮬레이션에 의해 결과의 타당성을 입증하였다.

ABSTRACT We propose a new parameter estimation algorithm that converges with probability one and in mean square, if the mean is the function of parameter of the probability density function. This recursive algorithm is applicable also even though the parameters we estimate are multiparameter case. And the results are shown by the computer simulation.

1. 序 論

어떤 $x_1, x_2, \cdots x_n$의 確率密度函數는 알려져 있으나 母數 θ는 알려지지 않은 確率密度函數의 標示値이고 θ가 實數의 벡터 $\theta = \theta_1, \theta_2, \cdots \theta_k$일 때 確率密度函數는 k個의 母數를 가지고 있다고 말할 수 있다.

여기서 $x_1, x_2, \cdots x_n$이 推定하려는 母數를 포함한 確率密度函數 $f(x : \theta)$의 標本値일 때 θ의 推定値 θ는 $x_1, x_2 \cdots x_n$의 函數로서 다음과 같이 표시할 수 있다.

$$\theta = g(x_1, x_2, \cdots x_n)$$

* 漢陽大學校工科大學電子通信工學科
Dept. of Electronic Communication Engineering, Han Yang University Seoul, 133 Korea
** 淸州大學校理工大學電子工學科
Dept. of Electronic Engineering, Cheongju University Cheongju, 310 Korea
論文番號 : 84－22 (接受 1984. 10. 12)

이러한 母數를 推定하는 방법으로서 가장 많이 사용되고 있는 방법은 最尤度推定 (maximum likelihood estimate) 方法과 모우먼트推定 (moment estimate) 方法이 있다.

最尤度推定方法은 標本値 $x_1, x_2, \cdots x_n$이 어떤 母數의 값을 갖는 確率密度函數로부터 標本한 것인가를 알아내는 방법으로 다음과 같이 定義되는 尤度函數 (Likelihood Function)를 사용하고 있다.

$$L(\theta : x_1, x_2, \cdots x_n) = \prod_{i=1}^{n} f(x_i : \theta) \qquad (1)$$

여기서 n은 標本의 수이고 θ는 k個의 母數를 표시하고 있다.

대부분의 確率密度函數는 1개 혹은 2개의 母數를 포함하고 있으며 그 이상 포함하고 있는 경우는 극히 드물다. 예로서 正規確率密度函數 (normal probability density function)는 平均値 (mean) μ와 分散 (variance) σ^2의 2 개의 母數를 갖는다.

따라서 最尤度推定은 다음 k개의 方程式을 만족시키는 $\theta_1, \theta_2, \cdots \theta_k$의 式이다.

$$\frac{\partial L(\theta : x_1, x_2, \cdots x_n)}{\partial \theta_1} = 0$$
$$\frac{\partial L(\theta : x_1, x_2, \cdots x_n)}{\partial \theta_2} = 0$$
$$\vdots$$
$$\frac{\partial L(\theta : x_1, x_2, \cdots x_n)}{\partial \theta_k} = 0 \qquad (2)$$

$L(\theta : x_1, x_2, \cdots x_n)$과 $\log L(\theta : x_1, x_2, \cdots x_n)$은 θ의 同一値의 最大値를 가지므로 尤度函數의 代數를 취하여 풀어도 같은 식을 얻는다. 대부분의 確率密度函數의 母數는 最尤度推定方法으로 母數의 推定値를 구할 수 있다.

모우먼트推定方法은 $f(x : \theta)$가 k개의 母數를 갖는 確率密度函數일 때 $f(x : \theta)$의 r次모우먼트는 다음 式으로 定義된다.

$$U_r = E[X^r] \qquad (3)$$

$E[X^r]$는 期待函數이다.

$x_1, x_2, \cdots x_n$이 確率密度函數 $f(x : \theta)$의 n개 標本일 때 j次標本모우먼트(sample moment)는

$$M_j = \frac{1}{n} \sum_{i=1}^{n} X_i^j \qquad (4)$$

이다. 그러므로 다음 k개의 方程式으로부터 $\theta_1, \theta_2, \cdots \theta_k$의 推定値를 구할 수 있다.

$$M_j = U_j \ (j = 1, 2, \cdots k) \qquad (5)$$

이와 같이 最尤度推定方法이나 모우먼트推定方法에는 θ_j $(j=1, 2, \cdots k)$에 대하여 각각 推定式이 주어져 있으나 本論文에서는 母數가 平均値의 函數일 때 stochastic approximation方法을 사용하여 母數를 逐次的으로 推定하는 방법을 제안하고 平均値가 2개 이상의 母數의 函數일 때 동시에 2개의 母數를 推定하는 방법을 실례를 들어 타당성을 입증하고자 한다.

2. 逐次母數推定 알고리즘

$Y_n (n=1, 2, \cdots)$이 數値의 系列(sequency)일 때

$$Y_n = F_n(\theta_1) + Z_n \ (n = 1, 2, \cdots) \qquad (6)$$

여기서 $F_n(\theta_1)$은 θ_1의 函數이고 θ_1은 推定하려는 母數이다.

그리고 Z_n은 랜덤變數(random variable)로 $F_n(\theta_1)$에 雜音(noise)으로 가해진다. 따라서 Z_n의 平均値가 0이고 有限한 分散의 값을 가지고 있는 獨立的이고 동일한 分布를 갖는 確率密度函數로부터의 標本値일 때 Y_n의 平均値는 다음과 같다.

$$E[Y_n] = E[F_n(\theta_1)] + E[Z_n] = F_n(\theta_1) \qquad (7)$$

式(7)은 Z_n의 平均値가 0이라는 것만 알고 있으면 Z_n의 確率密度函數는 定義되지 않아도 성립한다.

여기서 推定하려는 母數 θ_1이 하나이고 $\{\theta_1, n\}$이 다음 逐次關係式으로 定義되는 推定値의 系列일 때 $(n+1)$번의 θ_1의 推定値 $\theta_{1,n+1}$은 다음과 같이 주어진다.

$$\hat{\theta}_{1,n+1} = \theta_1, n + [a_n(\hat{\theta}_{1,1}, \hat{\theta}_{1,2}, \cdots \hat{\theta}_{1,n})]$$
$$\cdot [Y_n - F_n(\hat{\theta}_{1,n})]$$
$$(\because a_n(\hat{\theta}_{1,1}, \hat{\theta}_{1,2}, \cdots \hat{\theta}_{1,n}) = \frac{\dot{F}_n(\hat{\theta}_{1,0})}{\sum_{k=1}^{n} \dot{F}_2^k(\hat{\theta}_{1,0})} \qquad (8)$$

$F_n(\hat{\theta}_{1,n})$은 θ_1에 대한 $F_n(\theta_1)$의 導函數로 n番推定値를 의미한다. 만일 θ_1의 참값에 대한 사전지식을 가지고 있고 이 때의 값을 $\theta_{1,0}$라 하면 a_n으로 다음 式을 사용할 수 있다.

$$a_n(\hat{\theta}_{1,1}, \hat{\theta}_{1,2} \cdots \hat{\theta}_{1,n}) = \frac{\dot{F}_n(\hat{\theta}_{1,0})}{\sum_{k=1}^{n} \dot{F}_k^2(\hat{\theta}_{1,0})} \qquad (9)$$

Albert는 $\hat{\theta}_{1,n}$의 참값 θ_1에 대하여 確率1(probability one)과 mean square로 수렴함을 증명하였다.

$$P_r[\lim_{n \to \infty} (\hat{\theta}_{1,n-\theta_1}) = 0] = 1 \qquad (10)$$

$$\lim_{n \to \infty} E[(\hat{\theta}_{1,n-\theta_1})^2] = 0 \qquad (11)$$

이러한 방법을 stochastic approximation이라 하고 이 式은 앞에서 論義한 바와 같이 Z_n의 確率密度函數가 定義되지 않아도 성립한다.

만일 $F_n(\theta_1)$이 常數이고 平均値가 0인 Z_n의 確率密度函數를 알고 있다고 가정하면 Y_n은 平均値가 $F_n(\theta_1)$인 確率分布를 갖는 標本値이다. 즉 미지의 母數가 θ_1일 때 平均値가 θ_1의 函數이면 stochastic approximation方法을 사용하여 θ_1을 推定할 수 있다.

$F_n(\theta_1)$이 θ_1의 線型函數일 때 初期推定値 $\hat{\theta}_{1,1}=0$으로 놓으면

$$F_n(\theta_1) = C\theta_1$$
$$A_n = \frac{1}{nc} \qquad (12)$$

이고 여기서 C는 常數이다. 따라서

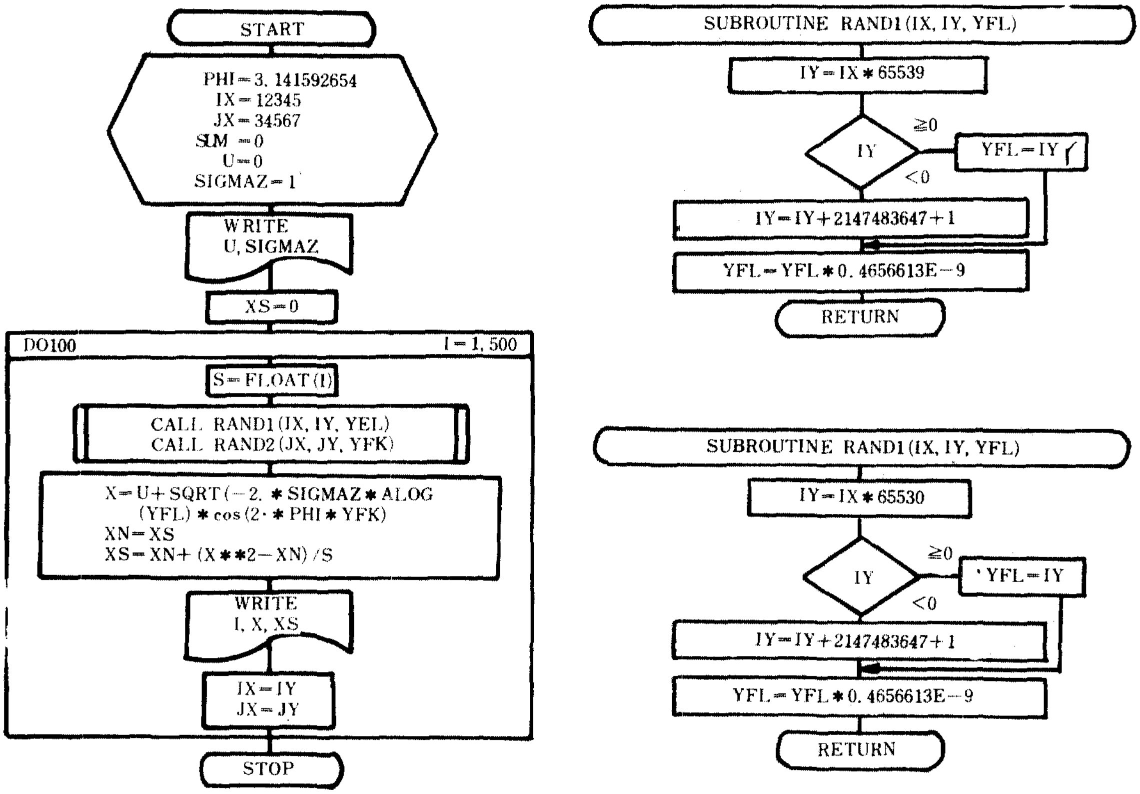


그림 1 正規分布函數의 平均値推定program의 flow chart
The flow chart of mean estimation program of normal distribution function.

$$\hat{\theta}_{1,2}=\frac{Y_1}{C}$$

$$\hat{\theta}_{1,3}=\frac{(Y_1+Y_2)}{2C}$$

$$\vdots$$

$$\hat{\theta}_{1,nH}=\frac{\sum_{k=1}^{n} Y_k}{\sum_{k=1}^{n} C} \tag{13}$$

와 같이 된다.

여기서 C가 1일 때는 θ_1의 推定値는 標本平均値(sample mean)와 같다.

만일 確率密度函數의 平均値가 θ_1과 θ_2의 函數일 때 처음 $\hat{\theta}_{1,1}$을 계산하고 다음 $\hat{\theta}_{2,1}$을 계산할 때 $\hat{\theta}_{1,1}$의 값을 사용하여 推定하면 동시에 2개의 母數를 推定하는 즉 probability one과 mean square로 수렴하는 推定式을 얻을 수 있다.

k개의 母數를 $\theta_i(i=1, 2, \cdots k)$라 하고 θ_i를 逐次的으로 推定하는 推定値를 $\hat{\theta}_{i,j}$ $(j=1, 2, \cdots n)$라 할 때 k개의 母數를 推定하는 알고리즘은 다음과 같다.

(1) 推定하려는 母數의 사전 근사치를 알고 있으면 初期値를 근사치로 놓고 그렇지 않으면 $\hat{\theta}_{i,j}=0$으로 놓는다.

(2) $\hat{\theta}_{1,2}$를 다음 式을 이용하여 구한다.

$$\hat{\theta}_{1,2}=\hat{\theta}_{1,1}+[a_1(\hat{\theta}_{1,1},\ \hat{\theta}_{2,1}\cdots\hat{\theta}_{k,1})]$$
$$\cdot[Y_1-F_1(\hat{\theta}_{1,1},\hat{\theta}_{2,1},\cdots\hat{\theta}_{k,1})]$$

$$\left(\because\ a_1(\hat{\theta}_{1,1},\hat{\theta}_{2,1},\cdots\hat{\theta}_{k,1})=\frac{\dot{F}_1(\hat{\theta}_{1,1},\hat{\theta}_{2,1}\cdots\hat{\theta}_{k,1}}{\dot{F}_2(\hat{\theta}_{1,1},\hat{\theta}_{2,1}\cdots\hat{\theta}_{k,1}}\right)$$

마찬가지로 $\hat{\theta}_{2,2}, \hat{\theta}_{3,2}, \cdots\hat{\theta}_{k,2}$를 이미 구한 $\hat{\theta}_{i,1}$의 값을 이용하여 구한다.

(3) $\hat{\theta}_{i,n}$의 값을 구한다.

$$\hat{\theta}_{i,n}=\hat{\theta}_{i,n+1}+[a_n(\hat{\theta}_{1,n},\hat{\theta}_{2,n}\cdots\hat{\theta}_{k,n})]$$
$$+[Y_n-F_n(\hat{\theta}_{i,n})]$$

$$\left(\because\ a_n(\hat{\theta}_{1,n},\hat{\theta}_{2,n},\cdots\hat{\theta}_{k,n})=\frac{F_n(\hat{\theta}_{i,n})}{\sum_{k=1}^{n}F_n(\hat{\theta}_{i,n})}\right.$$

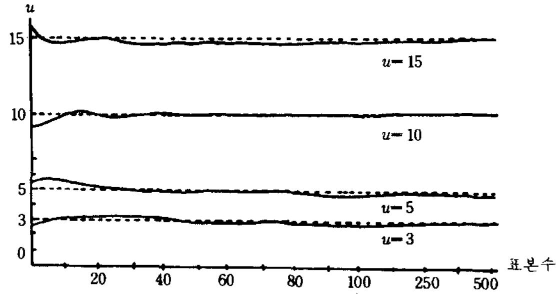
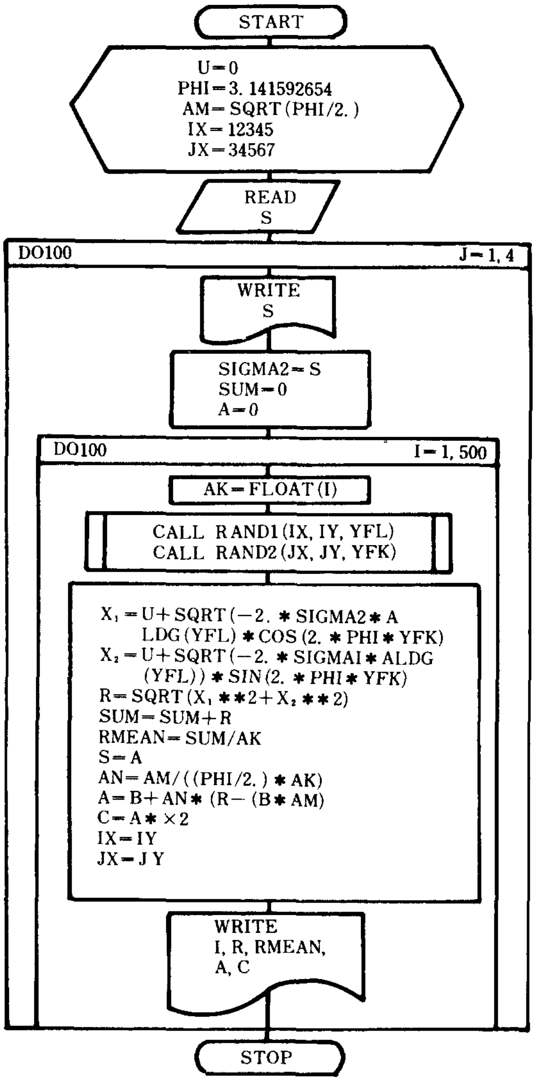


그림 2 正規分布函數의 平均値推定
The mean estimation of normal distribution function.



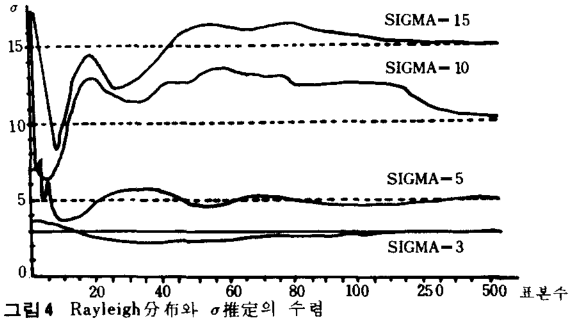


그림 4 Rayleigh分布와 σ推定의 수렴
The convergence of rayleigh's distribution and a estimation.

(4) $\hat{\theta}_i$, $(i=1, 2, \cdots k)$가 수렴할 때까지 逐次的으로 計算한다.

3. 컴퓨터 시뮬레이션

確率密度函數의 母數를 推定하는 방법을 실례를 들어 타당성을 입증하기 위하여 먼저 正規分布와 그의 平均値를 구해 보기로 한다.

$$f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-(x-\mu)^2/2\sigma^2} \quad (-\infty<x<\infty)$$

$$E(x)=\mu \tag{14}$$

正規랜덤數發生器(normal random number generator)로는

$$x_j=\mu+(-2\sigma^2\log U_j)^{1/2}\cos 2\pi U_{j+1}$$
$$x_{j+1}=\mu+(-2\sigma^2\log U_{j+1})^{1/2}\cos 2\pi U_{j+1}$$
$$(j=1, 3, 5, \cdots) \tag{15}$$

를 使用하였다.

U_j, U_{j+1}은 IBM/360에 使用하는 uniform random number인 RANDU를 使用하였으며 이 때 computer

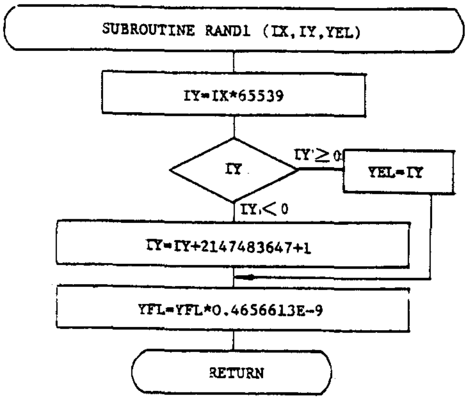
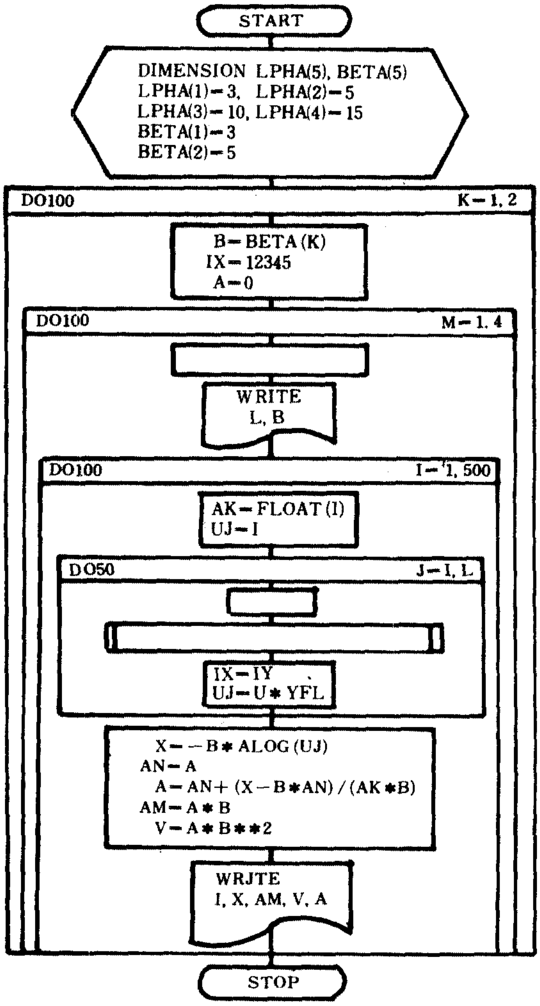


그림 3 Rayleigh分布函數의 σ推定program flow chart.
The flow chart of a estimation program of rayleigh's distribution function.



simulation flow chart는 그림 1 과 같다.

여기서 平均値 $\mu=3, 5, 10, 15$일 때 시뮬레이션結果는 그림 2 와 같다.

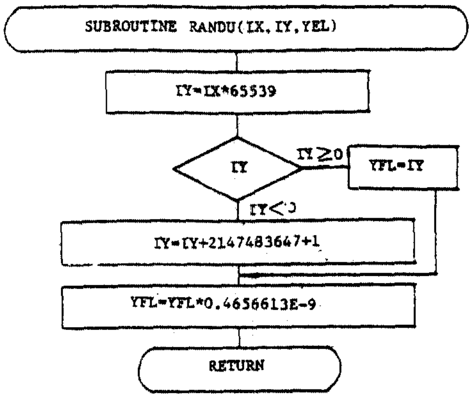


그림 5 Gamma分布函數의 α 및 β推定program flow chart.
The flow chart of α and β estimation program of gamma distribution function.

平均値 $E[x]=\mu$이므로 제시한 平均値 μ를 推定하는 알고리즘은 最尤度推定法과 동일하다.

그림 2 에서 보는 바와 같이 標本値가 많을수록 본래의 母數에 수렴하고 있음을 보여주고 있다.

다음은 x_1과 x_2가 正規分布를 가지고 있을 때 $=(x_1+x_2)^{1/2}$의 確率密度函數 즉 Rayleigh確率密度函數와 그의 平均値는 다음과 같다.

$$f(x)=\frac{x}{\sigma^2}e^{-x^2/2\sigma^2} \qquad (x\geq 0)$$

$$E[x]=\sqrt{\frac{\pi}{2}}\,\sigma \tag{16}$$

다음은 $\sigma^2=3, 5, 10, 15$일 때 시뮬레이션結果를 그림 4 에 나타내었다.

그림 4 에서 보는 바와 같이 Rayleigh確率密度函數의 母數 σ^2을 推定하는 방법은 本論文에서 제시한 방법의 有用性을 잘 나타내고 있다.

다음은 2 개의 母數를 갖는 Gamma確率密度函數의 母數를 推定하여 본다.

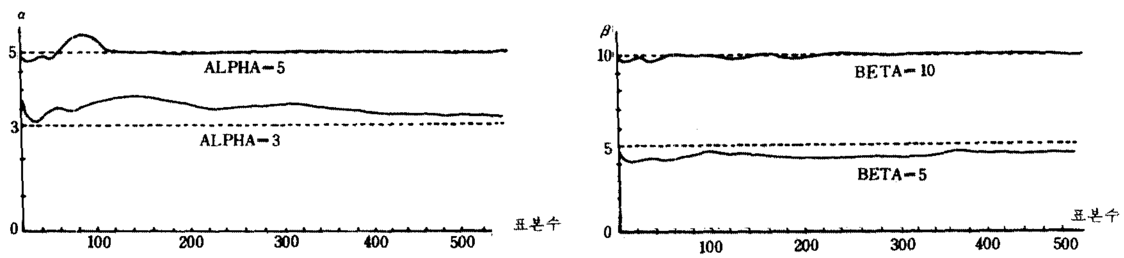
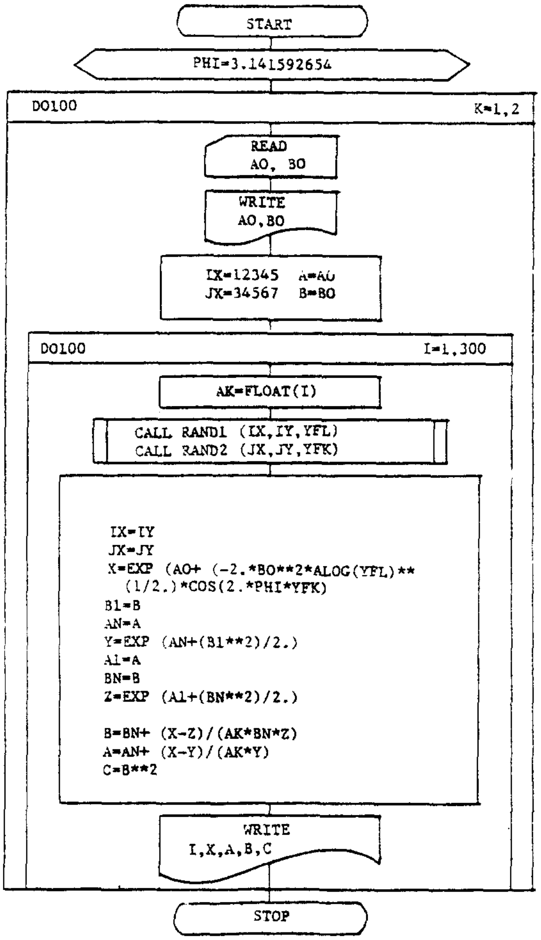


그림 6 Gamma分布의 α, β의 수렴
The convergence of α and β of gamma distribution.



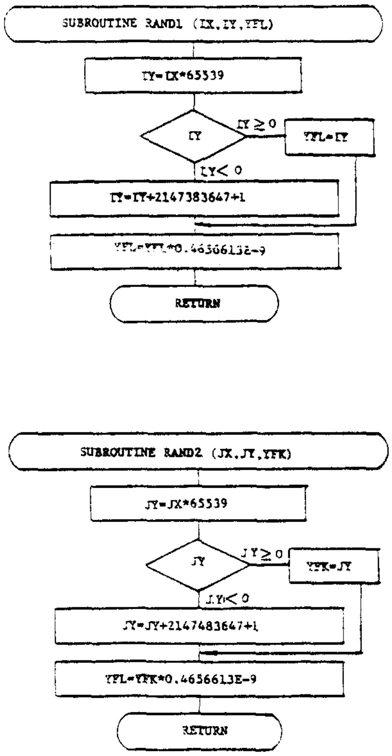


그림 7 Log Normal分布函數의 variance 및 mean의 推定 program flow chart.
The flow chart of estimation program of variance and mean for log normal distribution function.

Gamma確率密度函數와 그 平均値는 다음과 같다.

$$f(x)=\frac{1}{\Gamma(\alpha)}\left(\frac{1}{\beta}\right)^{\alpha}x^{\alpha-1}e^{-x/\beta}$$

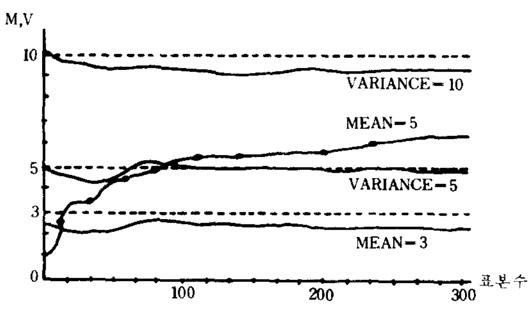


그림 8 Log Normal分布函數의 variance 및 mean의 수렴
The convergence of variance and mean of log normal distribution function.

$$(0\leq x,\ \alpha>0,\ \beta>0)$$

$$E(x)=\alpha\beta \tag{17}$$

式(17)에 의해 α 및 β를 推定하는 program flow chart는 그림 5와 같다.

그리고 Alpha= 3, 5 이고 Beta= 5, 10일 때 시뮬레이션結果는 그림 6과 같다.

끝으로 Log Normal確率分布函數의 母數인 平均値(mean)와 分散(variance)을 推定한다.

Log Normal確率密度函數와 平均値는 다음과 같다.

$$f(x)=\exp\{\mu+(-2\sigma^2\log YFL)^{1/2}\cos\cdot(2\pi YFK)\}$$

$$E(x)=\frac{\Sigma xi}{N} \tag{18}$$

式(18)에 의한 平均値 및 分散의 推定program

flow chart는 그림 7 과 같다.

그리고 mean= 3 , 5 이고 variance= 5 , 10일 때 시뮬레이션結果는 그림 8 과 같다.

4. 結　論

本論文에서는 推定하려는 母數가 平均値의 函數일 때 母數를 推定하는 새로운 逐次알고리즘을 제시하였다.

이와 같은 방법은 stochastic approximation方法을 이용하고 있으며 推定하려는 母數가 2 개 이상일 때에도 逐次的으로 동시에 2 개의 母數를 推定할 수 있음을 실례를 들어 타당성을 입증하였다.

參 考 文 獻

(1) A. M. Mood, F. A. Graybill and D. C. Boes, "Introduction to the theory of statistics," New York; McGraw-Hill, 1974.

(2) A. E. Albert and L. A. Gardner Jr, "Stochastic approximation and nonlinear regression," Cambridge, Mass. M. I. T. Press, 1967.

(3) G. S. Fishman, "Concepts and methods in discrete event digital simulation," John Wiley and Sons, 1973.

(4) J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," Ann. Moth. Stat. pp. 462-466, 1952.

(5) B. A. Einstein, "A self-learning estimator for tracking," IEEE Trans on sys. Man and Cybernetics, pp. 281-284, April, 1972.

韓 榮 烈(Young Yeul HAN)　正會員

1938年 6 月10日生

1960年 2 月：서울工大電子工學科卒業

1976年 8 月：Missouri州立大學大學院 通信專攻(工學碩士)

1979年 8 月：Missouri州立大學大學院 通信專攻(工學博士)

1961年 8 月～1964年 8 月：西獨Ziemens會社(株)에서 電子分野研修

1964年 8 月～1969年11月：한영공업(株)勤務

1969年11月～1970年10月：韓國科學技術研究所勤務

1980年 8 月～現在：漢陽大學校電子通信工學科副教授

美國Sigma Xi 및 IEEE正會員

本學會理事

朴 鎭 秀(Jin Soo PARK)　正會員

1948年 8 月30日生

1975年 2 月：漢陽大學校工科大學電子工學科卒業

1977年 2 月：漢陽大學校大學院電子通信工學科卒業(工學碩士)

1980年 9 月～現在：漢陽大學校大學院電子通信工學科 (博士過程)

1978年 3 月～現在：清州大學校電子工學科助教授