〈研究報告〉

Journal of the Korean Institute of Industrial Engineers
Vol. 13, No. 2, Dec. 1987

FBI (Fortran Based Interpreter)를 이용한 확률변수 변환의 시뮬레이션 모델에 관한 연구

A Study On the Simulation Model of the Transformation of Random Variables Using FBI (Fortran Based Interpreter)

金 原 經*

Abstract

Although there are many theoretical methods for the transformation of random variables, it is difficult to find probability density functions for the new random variables because of the complexity in mathematics. The author developed a simulation model solving the above difficulties using FBI (Fortran Based Interpreter) routines. The FBI is a kind of language interpreter analyzing the arithmetic statement in character data forms. In this paper, the FBI routines will be explained and the structure and applications of simulation model will be also demonstrated. Polynomial curve fitting method is applied to define the probability density function which can not be defined by well-known pdf. This program can also be used for instructing mathematical statistics and identifying distribution of the simulated data.

I. 서 론

확률변수의 변환(Transformation of Random Variable)에 관한 이론적 방법은 여러가지가 있다. 그러나 변환을 통한 새로운 확률변수의 밀도함수(pdf)를 구하고자 할 때 수학적인 계산절차상의 어려움 때문에 실제 문제의 해결에는 한계에 부딪히는 경우가 많이 있다. 특히 서로 다른 종류의 확률변수가 복수개일 경우에는 이론적으로 계산하는 것은 매우 어렵다. 이를 해결하는 하나의 방법으로써 시뮬레이션으로 접근하는 방법을 모색하여 볼 수 있다. 즉 새로운 확률변수의 pdf를 구하는 방법으로써 독립변수인 확률변수의 샘플데이타의 난수(Random Number)들을 생성시켜서 원하는 변환을 거친 후 그 결과로써 나온 샘플데이타들의 이론적 분포함수의 pdf를 여러가지 방법으로 적합(fitting) 시켜서 구해보는 방법이다. 그러나 시뮬레이션의 과

*경남대학교 공과대학 산업공학과

정중에 변환되는 함수식을 일일이 새로 고쳐가며 실행하기에는 너무나 번거로움이 많다. 또한 필요한 난수의 생성도 기존의 알려진 방법으로는 생성할수가 없는 경우도 있으므로 이때에는 그 확률변수의 pdf만 실행시에 입력시켜서 난수생성을 할 수 있다면 시뮬레이션을 더욱 편리하게 할 수 있을 것이다.

본 논문은 이러한 변환을 필요로 하는 변수의 생성과 변환식등을 사용자가 원하는대로 입력시켜서 변환된 결과의 분포함수를 손쉽게 구할수 있도록 하는 시뮬레이션 모델과 팩케이지 프로그램에 대한 연구이다. 제Ⅱ장에서 원하는 pdf 또는 변환식등을 해석, 계산하여 주는 FBI(Fortran Based Interpreter) 루틴의 개요에 대하여 알아보며, 제Ⅲ장에서는 이를 이용한 시뮬레이션 모델에 관한 프로그램의 구성과 내용에 대한 논의를 한 후, 제Ⅳ장에서 이 모델의 응용예를 몇가지 들고자 한다.

Ⅱ. FBI 프로그램

1. 목적과 기능

이 프로그램의 목적은 Fortran 프로그램내에서 연산식을 계산하고자 할 때 이 연산식을 Source 프로그램내에서 정의하지않고 사용자가 컴퓨터를 실행시킬때에 연산식을 입력시켜주면 정의된 연산식대로 계산토록 하는 일종의 interpreter 구실을 하는 프로그램이다. FBI를사용하므로써 Source 프로그램의 내용을 바꾸고자 할 때 일일이 Editing을 새로 하지 않아도 되므로 여러가지 다른 연산식을 편리하게 바꾸어 볼 수 있다. 단순한 연산 또는 산술식 뿐만아니라 사용자가 특정분야에서 자주 필요로 하는 루틴들, 예를 들어서 수치계산에서 자주 쓰이는 적분, 합, 미분방정식 등을 미리 작성해 놓았다가 필요하면 Fortran의 Library 함수처럼 사용할 수가 있다. 이 루틴에서 다룰 수 있는 독립변수의 갯수는 프로그램의 DIMENSION문의 크기만 조절하면 몇개라도 가능하며 연산식과 변수 또는 함수등의 사용규칙은 Fortran의 문법규칙과 동일하다.

2. 프로그램의 절차와 구성

다음의 그림 2.1에에서 Step 1 은 연산식의정의와 컴퓨터 실행을 위한 데이타 입력루틴이며 Step 2 부터 Step 4 까지는 일종의 컴파일러와 링커의 기능을 가진 루틴들로써 만일 연산식과 변수, 상수등에 오류가 없다면 Step 5 로 넘어가며 오류발생시에는 Step 1 부터 재입력을요구한다. Step 5 는 실행 루틴으로써 매번의 실제계산시 독립변수의 값이 주어지면 반복되어 불리우는 루틴으로써 정의된 연산식을 계산하여 준다.

註) 본 팩케이지 프로그램의 list는 저자에게 요구하면 받아볼 수가 있으며 그래픽터미날 또는 플롯타가 필요하다.

3. 연산함수의 종류

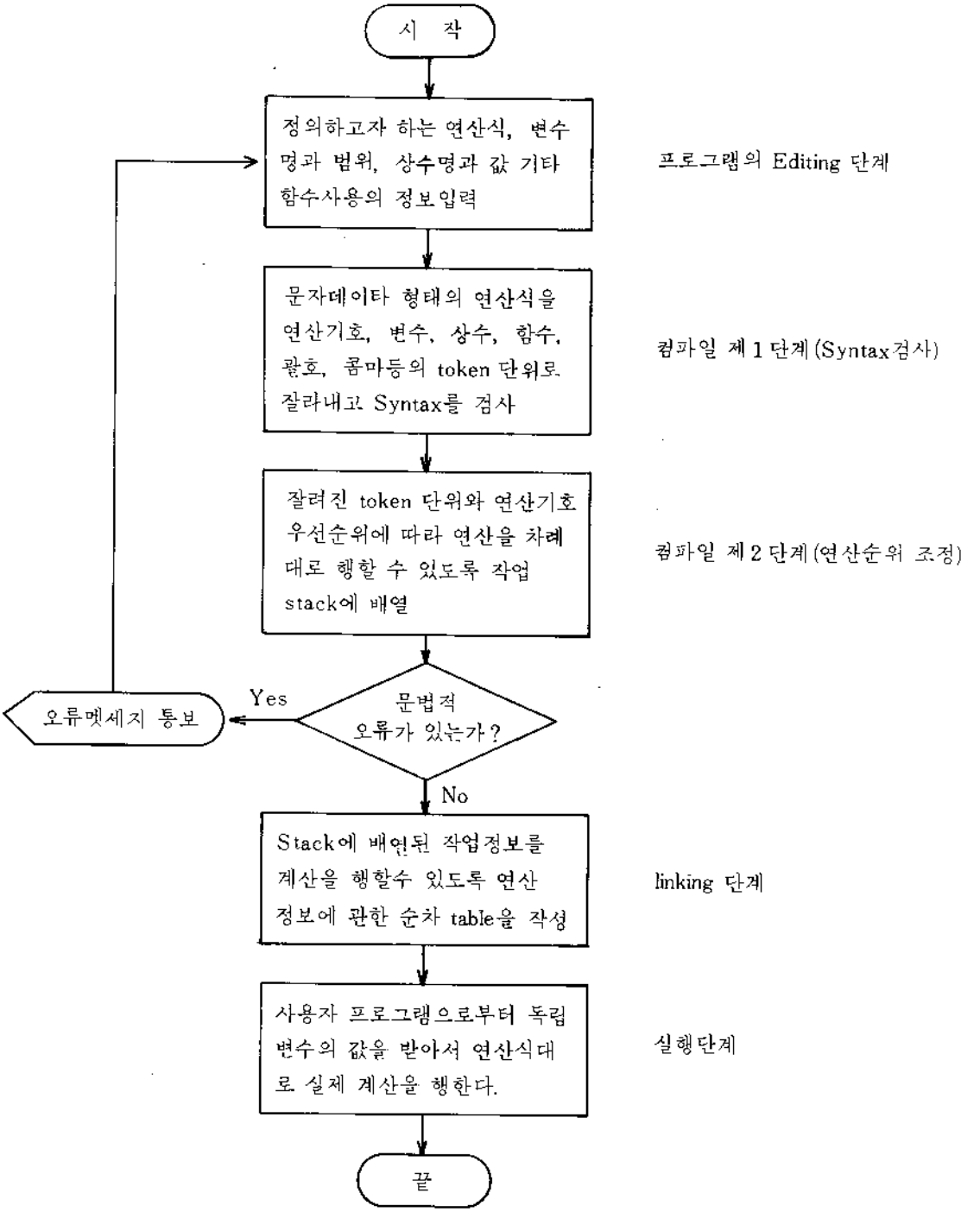


그림2.1 FBI 루틴의 작업절차 및 내용

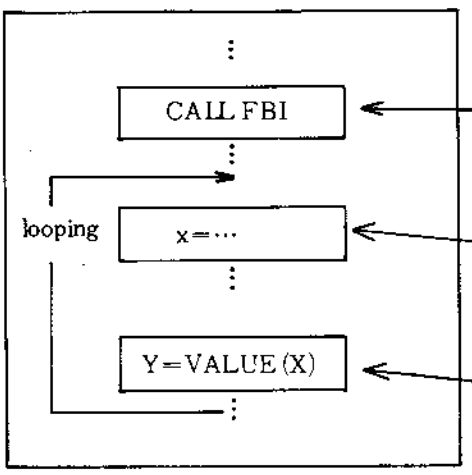
기존의 Fortran에서 제공되는 내장함수는 그대로 불러 사용할 수가 있으며 사용자가 특별히 요구하는 임의의 연산함수를 미리 정의하여 넣어줄 수가 있다. 본 프로그램에서 쓰이는 함수는 다음의 표2.1와 같으며 이외의 다른 함수가 필요하면 언제든지 추가시킬 수 있다.

표2.1 FBI 루틴에서 사용되는 연산 함수

기 호	명 칭	내 용	수 식 예	사용형태예
SUM	합	수학식의 $\sum$ 기호에 해당되는 연산	$\sum_{i=1}^{n} f(i)$	SUM (I, 1, N, f (i))
INT	적 분	정적분 $\int_a^b f(x)dx$의 값	$\int_a^b f(x)dx$	INT (x, a, b, dx, f (x))
DEQ	미분방정식	n차 상미분방정식의 초기치 문제의 해의 K차 도함수식	$y''=f(x, y, y')$	DEQ (2, x, dx, k, f (x, y, y')
GAM	감마함수	$\Gamma(\alpha)=\int_0^\infty e^{-x} x^{\alpha-1}dx$	$\int_0^\infty e^{-x} x^{\alpha-1} dx$	GAM (α)
FAC	계 승	factorial (!) 기호의 계산	n !	FAC (N)
INV	역 적 분	$\alpha=\int_a^x f(t)dt$의 연산에서 α값 주어질때 범위 x를 구한다.	$\alpha=\int_a^x f(t)dt$	INV (t, a, b, dt, α, f (t))
기 타	내장함수	Fortran에서 사용하는 library 함수들	sin x ln x ⋮	SIN (X) ALOG (X) ⋮

4. FBI 프로그램의 사용방법

Fortran Source 프로그램내에서 어떤 연산문장의 내용을 컴퓨터 실행시마다 바꾸어야할 경우가 있다. 이때에 연산문장의 위치에 FBI 루틴을 대신 넣어준다. 실제로는 다음과 같은 요령으로 Source 프로그램을 작성한다.



연산식을 미리 정의하는 것으로써, 이 위치에서는 실제 계산을 하는 것이 아니고 연산식의 syntax 검사 및 실행준비를 한다

독립변수의 값을 정의하며, 복수개의 변수이면 X를 배열(array)로 선언한다.

실제 연산을 하는 위치이며, VALUE는 FBI 루틴중의 하나인 함수로 변수X값에 따른 연산 결과치를 변수 Y에 넘겨준다.

이제 하나의 예로써 위 프로그램에서 Y=VALUE(X)라는 문장의 위치에서 다음과 같은 표준정규분포의 누적확률값이 필요하다고 하자.

$$y = \int_0^x 1/\sqrt{2\pi} \cdot \exp\left(-\frac{t^2}{b}\right) dt,\ b=2.0 < x < 5$$
$$\pi = 3.141592$$

실제 컴퓨터의 실행시에는 컴퓨터가 필요로 하는 일련의 질문을 다음과 같이 '=' 기호의 오른편에 입력시킨다.

① NVAR, NCEF = 1, 2
② FR, TO, VAR = 0,5, X
③ COEF, CNAME = 3.141592, PI
COEF, CNAME = 2, B
④ FUNCTION = INT(T, 0, X, 0.01, EXP(−T*T/B)/SQRT(2*PI))

위와 같은 일련의 질문과 답은 다음과 같은 의미이다.
① 독립변수의 갯수는 1개, 상수는 2개
② 독립변수의 범위 및 이름
③ 두개의 상수에 대한 값과 이름
④ 실제 계산하고자 하는 적분식의 정의

Ⅲ. 시뮬레이션 모델

1. 목적

이 모델은 확률변수의 수학적인 변환을 시뮬레이션 기법으로 처리하기 위한 것으로써, 원하는 분포들의 난수발생과 변환결과로써 나온 분포의 샘플데이타를 수집하여 검정을 위한 제반 통계처리를 한다. 순서통계량, 분포함수의 중첩, 통계량의 계산, 모수의 추정, 이론 분포와의 유의차 검정, 다항식(polynomial) 분포로의 적합, 출력을 위한 그래픽 처리등이 주요 내용이다. 작업중 필요하면 FBI 루틴을 호출시켜 그 편의성을 더하여 준다.

출력결과를 단순히 수치적인 데이타만을 제시한다면 분포함수의 추정이 어려우므로 그래픽 터미날 또는 플롯타를 통하여 실제 결과치와 이론적 결과치를 히스토그램으로 보여주므로써 시각적인 판단을 용이하게 하여준다.

2. 프로그램의 절차와 구성

프로그램의 절차는 7단계로써 각 단계별 기능과 내용에 대한 흐름도가 그림3.1에 나타나 있다.

〈Step 1〉 데이타 입력

시뮬레이션 프로그램을 실행시키기 위한 데이타를 입력시킨다. 변수변환을 함수에 의해서 행하고자 할 경우에는 변환 함수식을 FBI 루틴에서 입력시킨다. 예를 들면, 확률변수 X가 Normal(μ, α^2)을 따르고, 확률변수 Y는 Gamma(α, β)를 따른다고 하자. 이때 $Z=X^2+Y$라는 새로운 변수변환을 했을때 Z의 분포함수를 다음과 같이 컴퓨터 실행시에 입력시켜 구할 수 있다.

```
NVAR        = 2
VAR         = X
VAR         = Y
FUNCTION = X**2 + Y
```

```
DST, PA, PB = NOR, μ, α²
DST, PA, PB = GAM, α, β
```

〈Step 2〉 난수 생성

독립변수인 확률변수에 대한 정보를 입력시키면 그에 따른 난수를 생성시킨다. 기존의 방법을 사용할 수도 있으며, 만일 pdf 또는 cdf는 알고 있으나 기존의 방법으로 생성이 어려운 경우라면 FBI 루틴에서 그 함수식을 직접 넣어주어 난수생성을 한다. 이 경우의 기법은 역변환 방법을 사용하였다. 즉 발생되는 난수의 예상되는 최소 및 최대치를 주고 이 범위내에서 수치적분을 행하여 pdf로 부터 cdf를 구한다. 매적분구간마다의 누적되는 cdf값을 테이블에 기억시킨 후, 역변환시에 대응되는 일양난수에 해당

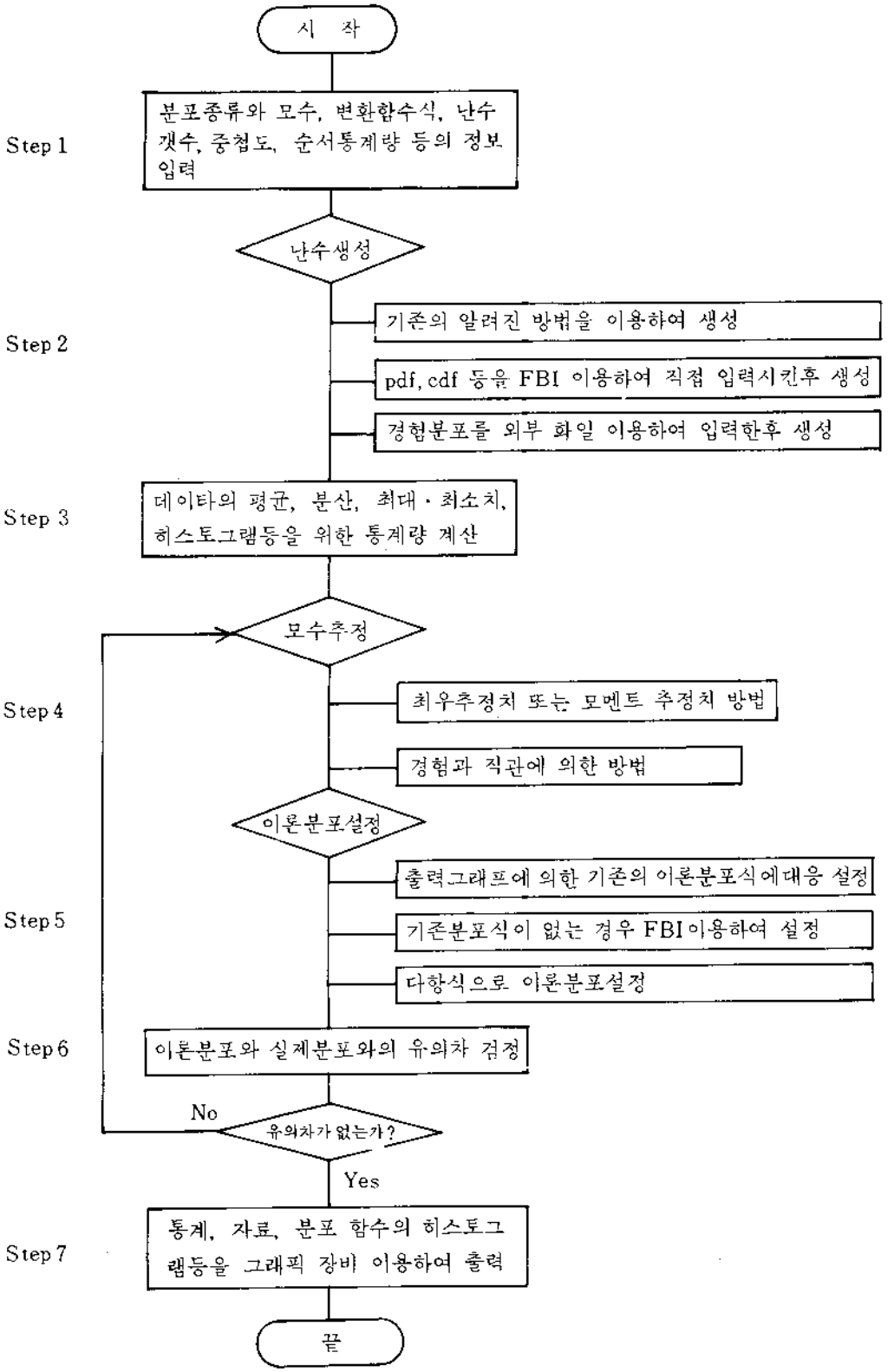


그림3.1 시뮬레이션 프로그램의 절차흐름도

되는 값을 읽어내면 원하는 pdf의 난수가 된다. 만일 cdf를 안다면 수치적분과정이 생략되고, 역변환 함수를 알고 있다면 테이블에서 대응 값을 읽어내는 과정이 생략된다. 경험분포의경우에는 외부 화일에서 그 정보를 입력시키면 위의 pdf를 아는 경우와 동일한 과정으로 난수생성을 하여 준다.

예를 하나 들어서, 확률변수 X의 pdf가 $f(x)=\sin x$, $0 \leq x \leq \pi/2$ 라고 하면 실제 컴퓨터 실행시에 다음과 같이 입력시킨다.

```
DST, PA, PB=PDF
NVAR          =1
FR, TO, VAR=0, 1.570796, X
FUNCTION      =SIN(X)
```

〈Step 3〉 통계량 계산

모수 추정과 출력설계, 유의수준 검정등을 위한 통계량을 계산한다.

〈Step 4〉 모수 추정

이론분포를 설정하기 위한 모수의 추정을 행한다. 분포함수의 종류에 따라 최우추정치 또는 모멘트 추정치를 이용한 모수를 해석적인방법으로 구하고 만일 해석적으로 어려우면 수치해석적 방법으로 구해낸다. 경험적으로 pdf를 설정하거나 모수의 추정이 어려운 경우라면 Step 5에서 FBI 루틴을 이용하여 pdf를 직접 입력시킨다.

〈Step 5〉 이론분포의 설정

출력된 샘플데이타들의 이론적인 분포함수식을 유도하여 내고, 이 함수식의 계급 구간치에 따른 확률값을 계산한다. 만일 Step 4 에서 모수 추정을 못했거나 특수한 pdf라고 간주되면 FBI 루틴을 불러서 함수식을 입력시킨다. 위의 방법으로도 이론분포의 설정이 어려운 경우에는 다항식을 이용하여 이론분포식을 유도한다.

〈Step 6〉 유의수준 검정

관측된 샘플데이타가 설정된 이론분포에 잘 적합하는가를 검정하기 위한 카이 자승 검정통계량을 계산한다. 계급구간수와 이론 돗수에 따른 자유도가 계산되면 카이 자승 밀도함수식을 0부터 계산된 검정통계량 값까지 수치적분을 하여 확률값을 구한다. 구한 확률값이 95% 또는 99%등 원하는 신뢰수준의 확률값 이내에 들어오는지를 보고 이론분포의 채택 여부를 결정하여 준다.

계급구간의 폭을 여러가지로 변경해서 자유도 크기에 따른 검정을 비교해 볼수도 있고, 아니면 다른 난수 샘플을 통해 나온 데이타를 보고 검정을 여러번 쉽게 되풀이 할 수도 있도록 하였다.

〈Step 7〉 결과의 출력

출력된 샘플데이타들의 결과를 단순히 수치적인 데이타나 통계량만 갖고는 분포함수를 예측하기가 어렵다. 따라서 그래픽 터미날 또는 플롯타등으로 데이타의 히스토그램을 그려주면 사용자도 이 그래프를 보고 분포함수의 추정을 시각적으로 용이하게 할 수 있다. 이때는 분포의 추정을 여러가지 방법으로 시행착오의 과정을 거쳐서 행할 수 있도록 실제 분포는 히스토그램으로 표시하고 이론분포는 함수식의 모양을 그려준다.

Ⅳ. 시뮬레이션 모델의 응용예

1. 다항식으로의 분포함수 설정

지금까지 살펴본 모델을 응용하는 예로써 다음과 같은 경우를 생각해 보자. 확률변수 X가 모수 $\alpha=6.2$, $\beta=2.8$인 감마분포를 따른다고 하고, 이 분포로 부터 크기가 5인 확률변수 표본의 순서통계량을 $X_1, X_2, \cdots, X_5$라 하자. 만일 다음과 같이 $Z=\sin(X_5-X_1)*\sqrt{X_1+X_5}/4$ 의 pdf를 찾고자 할 경우에 이론적인 방법으로는 계산상의 난점이 많으므로 다음과 같이 데이타를 주어서 시뮬레이션을 행하여 본다.

```
NVAR          =2
VAR           =X1
VAR           =X5
```

FUNCTION =SIN(X5-X1)*SQRT(X5+X1)/4
DST, PA, PB, NA, NO=GAM, 6.2, 2.8, 5, 1
DST, PA, PB, NA, NO=GAM, 6.2, 2.8, 5, 5

위에서 변수 NA와 NO는 각각 순서통계량 X_1과 X_5의 샘플크기와 순위를 나타낸다. 시뮬레이션의 결과로써 나온 Z의 히스토그램이 그림 4.1에 나타나 있다. 분포함수의 모양이 특이한 형상으로 기존의 잘 알려진 분포로는 이론분포함수를 추정하기가 용이하지 않을 것이므로 여기서는 다항식을 이용하는 방법을 살펴보자. 즉 Z의 이론분포식을 $f(Z)=a_0+a_1Z+a_2Z^2+\cdots+a_nZ^n$으로 설정한다. 다음 n을 적당한 차수부터 차츰 늘려가면서 카이자승 검정 통계량값에 따른 확률값이 최소인 차수 n을 구하고 이때의 계수들을 구한다. n의 값에 따라 실제 계산된 카이자승 검정통계량과 자유도 그리고 확률값 등이 표4.1에 나타나 있다.

표4.1 다항식으로 설정한 경우의 카이자승 검정

차수n	검정통계량	자유도	확 률 값
14	161.66	60	100.00%
15	149.15	59	100.00%
16	55.89	59	40.36%
17	55.31	58	41.84%
18	55.68	57	46.92%
19	58.49	56	61.02%
20	62.36	54	79.50%
21	53.51	54	50.03%
22	49.72	52	42.99%
23	65.41	52	90.13%
24	86.83	50	100.00%

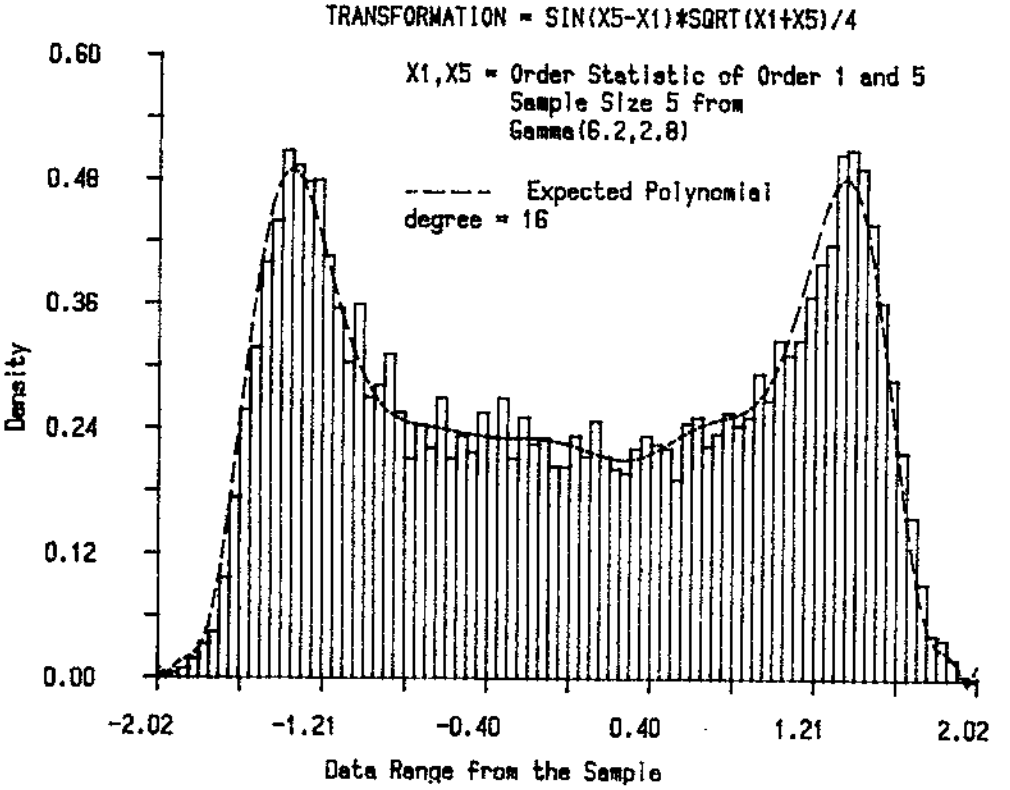


그림 4.1 예제의 변환함수식에 의한 출력

표4.1을 살펴보면 검정의 유의수준을 5%로 볼 때 책택되는 차수는 16차 부터 23차 까지임을 알 수 있고, 16차인 경우에 확률값이 40.36%로써 가장 최소이다. 즉 이 분포는 16차의 다항식으로 이론분포를 삼을 때 실제의 데이타와 가장 잘 일치한다고 볼 수 있다. 이때의 각 계수들은 표4.2에 나타나 있다.

표4.2 차수 n=16인 경우의 다항식의 계수

i	계 수 a_i
0	0.2253913039343529
1	−6.0217864427878669E−02
2	−0.1445719292614004
3	0.3429715487201813
4	1.322873789960357
5	−0.6987110412336043
6	−3.5234400120638296
7	0.6956539022707532
8	4.397847405187427
9	−0.3791410542314638
10	−2.739182747564850
11	0.1150668943699722
12	0.8855603470188948
13	−1.8193505948353435E−02
14	−0.1429128458408241
15	1.155498937082578 1E−03
16	9.1246213519941125E−03

2. 감마분포를 따르는 두 변수의 중첩

감마분포는 두개의 모수(α, β)를 갖고 있는데 두번째 모수 β값이 같을 경우에만 감마분포를 따르는 두 변수의 중첩시킨 결과의 분포함수를 이론적으로 알 수 있다. 그러나 서로 다른 임의의 두 모수를 갖는 감마분포의 중첩시킨 결과도 또한 감마분포를 따른다는 것을 시뮬레이션 방법으로 알 수가 있다. 예를 들어서 변수 X가 모수 (α, β)=(2.5, 1.7)인 감마분포를 따르고 변수 Y가 모수(α, β)=(6.3, 5.8)인 감마분포를 따르는 경우에 두 변수의 중첩 Z=X+Y의 결과를 표4.3에서 보여주고 있다. 모두 10번의 반복에서 Z의 모수 (α, β)가 대체로 큰 변화없이 비슷한 값을 갖고 있으며 유의수준 5%에서 모두 채택되고 있음을 알 수가 있다.

표4.3에서 자유도는 구간의 갯수에서 추정된 모수의 갯수를 빼고 이론돗수가 5미만인 구간의 갯수를 뺀 것으로써 매번의 실험마다 약간의 차이가 있다.

이 자유도에 따른 카이자승 함수를 0부터 카이자승 검정통계량값까지 적분하면 확률값을 얻고, 이 확률값이 95% 이상이면 유의수준 5%에서 기각을 한다. 다른 모수의 경우도 모두 중첩시킨 결과는 역시 감마분포임을 확인할 수 있다.

표4.3 감마 분포를 따르는 두변수의 중첩된 결과로 나온 감마분포의 모수

반 복 횟 수	α	β	검정통계량(자유도)	확 률 값(%)
1	7.52	5.44	96.71(85)	81.80
2	7.50	5.43	81.43(86)	37.63
3	7.52	5.41	85.49(85)	37.63
4	7.57	5.37	106.28(85)	53.02
5	7.68	5.32	76.31(85)	94.29
6	7.57	5.42	95.03(85)	25.92
7	7.67	5.38	78.56(86)	78.42
8	7.54	5.40	102.38(86)	29.38
9	7.63	5.34	89.94(84)	89.12
10	7.66	5.35	95.95(86)	78.11

3. 이론적인 방법과의 비교

이론적으로 분포함수의 pdf를 구하려면 상황에 맞는 적절한 방법을 사용하지만 일반적으로 다음과 같은 단계를 거쳐야 한다. n개의 독립변수를 갖는 변수변환을 생각해 보자.

〈1 단계〉 n개의 독립변수와 1 : 1 대응이 되도록 하는 (n－1)개의 가상변수를 설정하여 모두 n개의 종속변수를 만든다.

〈2 단계〉 역함수를 구한다. 즉 각 독립변수를 n개 종속변수의 함수로 나타낸다.

〈3 단계〉 역함수식을 종속변수로 편미분하여 Jacobian을 구한다.

〈4 단계〉 변환함수식에 따른 결합밀도함수를 구하고, 이 식의 독립변수를 종속변수로 치환한다.

〈5 단계〉 치환된 함수식과 Jacobian을 곱하여 n개 종속변수들만의 결합밀도함수를 구한다.

〈6 단계〉 새로운 결합밀도함수를 (n－1)개의 가상 종속변수들에 대해서 (n－1)번 적분을 하여 변환된 최종 pdf를 구한다.

이상과 같은 절차는 변환 함수식이 복잡하고 독립변수가 많은 경우에는 위의 어느 단계라도 용이치가 않고 수치 해석적으로도 해결하기 어렵다고 본다. 그러나 시뮬레이션 방법은 이같은 복잡한 절차를 거치지 않아도 된다. 또한 독립변수의 분포가 수식으로 표시안되는 경험분포라도 상관이 없으며, 분포함수의 수식자체가 특이한 경우라도 해결이 용이하다 하겠다.

시뮬레이션 결과로써 나온 분포함수의 형태가 독특하여 어떠한 방법으로도 이론분포식을 구할 수 없는 경우라도 평균, 분산, 데이타의 범위 분포함수의 형태등 여러가지 정보를 제공하여 주는 장점이 있다.

V. 결 론

지금까지 살펴본 바와 같이, 확률변수의 변환을 이론적인 방법으로 해결하고자 할때에 수학적인 계산절차상의 난점으로 인하여 적용하기가 어려운 경우에는 시뮬레이션 방법으로 대부분 극복할 수가 있다고 본다. 물론 시뮬레이션의 특성상 정확한 pdf를 유도할 수가 없다는 단점은 있으나, 이 모델의 응용은 일반적인 함수식의 유도보다는 그때 그때마다 발생되는 특수한 상황의 변환문제 또는 결과 분포함수의 유도를 실용적인 면에서 처리하여 준다는 점에서 의의를 찾을 수가 있다. 예를 들면 어떤 다른 시뮬레이션 문제를 처리하여 출력되어 나온 확률변수의 분포함수를 알고자 할 때에, 이 모델을 응용하여 그 분포에 맞는 pdf를 찾고자 할 경우이다.

기존의 분포함수로는 pdf의 도출이 어려운 경우라면 다항식으로 처리하여 대부분의 경우 적용이 가능하다. 동일한 문제라도 매번의 독립적인 시뮬레이션에서 유도된 다항식이 서로 약간씩 다르다고 하더라도 실제로 함수를 그래프로 그려보면 거의 일치함을 알 수가 있었으므로 이 다항식들을 이론적인 pdf로 삼아도 실무적인 관점에서는 별 문제가 없다고 사료된다. 아울러 본 모델은 다른 여러방면으로도 응용할 수가 있으므로 시뮬레이션 및 수리통계학의 교육용으로도 사용될 수가 있고, 특히 FBI 루틴은 연산식이 자주 바뀌어야 될 다른 프로그램에서도 유용하게 사용될 수가 있다.

References

1. Burden, R.L., and Faires, J.D., and Reynolds, A.C., *Numerical Analysis*, Weber & Shmidt, 1981.
2. Conte, S.D., and Boor, D., *Elementary numerical Analysis*, 3rd Ed., McGraw-Hill, 1980.
3. Draper, N.R., and Smith, H., *Applied Regression Analysis*, John Wiley & Sons, 1966.

4. Gordon, G., *Systems Simulation*, 2nd Ed., Prentice Hall, 1978.
5. Hogg & Craig, *Introduction to Mathematical Statistics*, 4th Ed., Collier Macmiian, 1978.
6. Kennendy, W.J., and Gentle, J.E., *Statistical Computing*, Marcel dekker, 1980.
7. Kreyszig, E., *Engineering Mathematics*, 4th Ed., John Wiley & Sons, 1980.