韓國 OA學會 論文誌
第4卷 第2號, 1999. 6.

99-4-2-2-3

# 통계계산을 위한 Fortran과 C라이브러리의 구현*

신 봉 섭**, 박 춘 성***

# Implementation of Fortran and C Libraries for Statistical computing

Bongsup Shin**, Choonsung Park***

## 요 약

본 연구에서는 여러 응용분야에서 자주 사용되는 통계적 모의실험이나 통계계산에 유용하게 사용될 루틴들을 Fortran과 C 언어의 Subroutine이나 함수 형태로 작성하여 라이브러리로 구현하였다. 여기에는 일반적으로 자주 사용되는 확률분포들에 따르는 확률변수들의 난수생성기와 대표적인 확률분포들의 확률 계산이나 상위확률 및 상위백분위수의 계산 등에 유용한 루틴들을 포함하고 있다.

## Abstract

In this thesis Fortran and C Libraries are implemented and applied to the real situation of statistical simulation. They contain the routines of random number generators and for various statistical distributions which often used in stochastic simulation. They also contain the routines for calculating probabilities and upper quantiles of various statistical distributions Each routine of them was tested by the various procedures and proved to be very stable

---

* 이 연구는 1996년도 안양대학교 학술연구비의 지원에 의한 것임
** 안양대학교 정보통계학과 조교수
*** 경원전문대학 교양과 대우 부교수
논문접수 : 1999. 5.19. 심사완료 : 1999. 6.19.

# Ⅰ. 서 론

통계학분야 뿐만 아니라 다양한 응용분야의 많은 연구들에서 소위 '몬테칼로 모의실험(Monte Carlo simulation)'이라 불리는 통계적인 모의실험이 이루어지고 있다. 이러한 종류의 모의실험에서는 일반적으로 어떤 확률분포에 따르는 확률변수값들을 생성하는 과정이 반드시 따르게 된다. SAS, S-plus, MINITAB 등의 대부분 통계패키지에는 여러 확률분포들에 대한 난수생성기가 포함되어 있으나 이들 통계패키지는 통계학자가 아닌 연구자들에게는 잘 알려져 있지 않으며 이들은 모두 인터프리터(interpreter) 방식으로 작동하기 때문에 동일한 작업을 여러 번 반복하는 모의실험에서는 수행시간이 심각하게 오래 걸린다는 단점을 지니고 있다. 또한, 최근 들어 통계학계에서도 교육을 시작하고 있는 MS(Microsoft)사의 Excel에도 난수생성에 관련된 항목이 포함되어 있으나 이 역시 동일한 작업을 여러 번 반복하는 모의실험에는 적당한 도구가 되지 못한다. 이러한 단점을 피하기 위해서 일반적으로 통계적 모의실험에서는 컴파일러(compiler) 방식의 언어인 Fortran이나 C를 이용하는 것이 일반화되어 있다. 그러나 이들 언어에는 여러 확률분포에 대한 난수발생기들이 포함되어 있지 않기 때문에 이미 상용화되어 세계적으로 널리 이용되고 있는 IMSL(International Mathematical and Statistical Library)이나 Press et al. (1995a/b)의 NR(Numerical Recipes in Fortran/C) 등을 이용하여 왔다. 최근 MS사의 FORTRAN PowerStation에는 IMSL과 NR이 모두 포함되어 있으나 소프트웨어의 가격이 매우 높은 것이 흠이라고 할 수 있다.

여러 확률분포들의 수표를 필요할 때마다 찾는 일도 번거로운 작업일 뿐만 아니라 어떤 검정의 유의확률(p-value)을 계산하기 위해서는 일반적인 통계수표로는 충분하지 못한 경우도 있다. 따라서 본 연구에서는 일반적으로 자주 사용되는 확률분포들에 따르는 확률변수들의 난수생성기와 대표적인 확률분포들의 확률계산이나 상위확률 및 상위백분위수의 계산 등에 유용한 루틴들을 포함하고 있는 라이브러리를 구현하고자 한다.

2장에서는 구현된 라이브러리에 포함된 루틴들의 종류와 알고리즘 등에 대하여 다루었다. 3장에서는 구현된 라이브러리에 포함된 루틴들을 이용하여 일표본 위치문제에서 모수적 t-검정과 비모수적인 윌콕슨 부호순위검정의 효율을 비교하는 모의실험을 실시하였으며, 마지막으로 4장에서는 결론 및 향후 연구과제에 대하여 다루었다.

# Ⅱ. 구현된 라이브러리의 루틴들

구현된 라이브러리는 Fortran 버전인 sbsf.lib와 C 버전인 sbsc.lib로 작성되었으며, 이들은 각각 DOS용인 MS Fortran 5.0과 Turbo C 2.0에서 테스트되었다. 테스트에 사용된 컴파일러들이 조금 오래된 것들이지만 아직도 많은 사용자들에게서 이용되고 있다. 이는 아마도 윈도우즈용 컴파일러의 높은 가격 때문일 것이다. 특히 소프트웨어의 구입에 예산이 거의 없는 대학들에서는 더욱 그러하다.

구현된 라이브러리의 사용법이나 상세한 설명은 안양대학교 정보통계학과의 자료실(http://stat.anyang.ac.kr)을 참고하기 바라며, 여기에서는 각 라이브러리에 포함된 루틴들의 종류와 알고리즘 등에 대하여만 간단히 소개하도록 하겠다.

## 2.1 확률분포들에 대한 난수생성에 관련된 루틴들

확률분포들에 대한 난수생성기의 각 루틴들은 Devroye(1986), Forsythe et al.(1977), Kennedy, Jr and Gentle(1980), Press et al.(1995a/b), Ripley(1987), Rubinstein(1981) 등에서 추천하는 알고리즘을 이용하였으며 이들 알고리즘은 이미 많은 연구자들에 의하여 충분히 비교되고 논의되어 안정성을 인정받은 것들이다.

여러 확률분포들에 대한 난수의 생성에서 가장 기본이 되는 것은 0과 1사이의 균일분포 U(0,1)에 따르

는 의사난수(pseudo-random number)의 제작이다. 왜냐하면, U(0,1)에 따르는 의사난수는 역변환법을 이용하여 다른 확률분포들(예를 들면 지수분포, 코오쉬분포, 이중지수분포(Laplace분포) 등)에 대한 난수를 생성하는 단계에 중요한 역할을 하기 때문이다. 본 연구에서 사용한 U(0,1) 의사난수는 IMSL의 GGUBS에서 채택하고 있는 승수(multiplier) $7^5$, 법수(modulus) $(2^{31}-1)$인 승산합동법(multiplicative congruential method)을 이용하였다.

생성된 의사난수들이 U(0,1)를 따르는 지를 알아보기 위해 $\chi^2$-분포를 이용한 적합도 검정, Kolmogorov-Smirnov의 적합도 검정, Cramer-von Mises의 적합도 검정 등을 실시하여 모두 만족한 결과를 얻었다. 또 생성된 난수들의 임의성(randomness)을 확인하기 위해 run-검정도 실시하였으며 여러 상황에서 모두 만족한 결과를 얻었다. 승산합동법을 포함한 여러 의사난수발생 알고리즘에 대한 최근의 연구로는 박경렬 외(1998)를 들 수 있다.

여기에 포함된 Fortran과 C 루틴들을 정리하면 다음의 〔표 2-1〕과 같으며, 각 루틴들은 모두 실수형 배열에 N개의 해당 난수들을 생성한다.

〔표 2-1〕 확률분포들에 대한 난수생성에 관련된 Fortran과 C 루틴들

| 루틴명 | 확률분포 | 설명 |
|---|---|---|
| RUNI | 일양분포 U(0,1) | |
| RBIN | 이항분포 B(n,p) | n : 시행횟수, p : 성공확률 |
| RNBI | 음이항분포 NB(k,p) | k : 성공횟수, p : 성공확률 |
| RGEO | 기하분포 Geo(p) | p : 성공확률 |
| RHGE | 초기하분포 HG(n,l,m) | n : 표본수, l : 로트수<br>m : 로트 중 불량품의 수 |
| RPOI | 포아송분포 P(m) | m : 평균 |
| RGAM | 감마분포 Gam(a,1) | |
| RBET | 베타분포 Be(a,b) | |
| RCHI | 카이제곱분포 $\chi^2(n)$ | n : 자유도 |
| REXP | 지수분포 Exp(m) | |
| RDEX | 이중지수분포 Laplace(0,1) | |
| RLNO | 대수정규분포 LN(m,s) | m, s : 각각 대응되는 정규분포의 평균과 표준편차 |
| RNOR | 표준정규분포 N(0,1) | |
| RWEI | 와이블분포 W(a,1) | |
| RCAU | 코오쉬분포 Cau(0,1) | |

### 2.2 확률분포들의 확률계산에 관련된 루틴들

확률분포들의 확률계산에 관련된 루틴들에는 가장 많이 사용되는 이산형 분포인 이항분포와 포아송분포의 모든 확률값들을 배열에 담아주는 루틴과 대표적인 연속형 분포인 표준정규분포, t-분포, $\chi^2$-분포, F-분포들의 상위확률 및 상위백분위수를 계산해 출력해 주는 루틴들을 포함하고 있다. 이들에 포함된 Fortran과 C 루틴들을 정리하면 다음의 〔표 2-2〕와 같다.

〔표 2-2〕 확률분포들의 확률계산에 관련된 Fortran과 C 루틴들

| 루틴명 | 설 명 |
|---|---|
| PBIN | 이항분포 B(n,p)의 모든 확률을 배열에 출력 |
| PPOI | 포아송분포 P(m)의 모든 확률을 배열에 출력 |
| PNOR | 표준정규분포 N(0,1)에서 상위확률을 계산 |
| QNOR | 표준정규분포 N(0,1)에서 상위백분위수를 계산 |
| PT | 자유도 n인 t-분포 t(n)에서 상위확률을 계산 |
| QT | 자유도 n인 t-분포 t(n)에서 상위백분위수를 계산 |
| PCHI | 카이제곱분포 $\chi^2(n)$에서 상위확률을 계산 |
| QCHI | 카이제곱분포 $\chi^2(n)$에서 상위백분위수를 계산 |
| PF | 자유도 $n_1$, $n_2$인 F-분포 $F(n_1, n_2)$에서 상위확률을 계산 |
| QF | 자유도 $n_1$, $n_2$인 F-분포 $F(n_1, n_2)$에서 상위백분위수를 계산 |

## Ⅲ. 모의실험에의 활용

본 장에서는 일표본위치문제에서 모수적 검정인 t-검정과 비모수적 검정인 윌콕슨 부호순위검정의 검정력이 모집단의 분포가 바뀜에 따라 어떻게 변화하는지를 구현된 라이브러리의 루틴들을 이용하여 모의실험을 실시해보고자 한다. 모집단의 분포로는 정규분포, 이중지수분포, $\varepsilon$-오염정규분포, 코오쉬분포가 고려되었다. $\varepsilon$-오염정규분포의 분포함수는

$F(x) = (1-\varepsilon)\,\Phi(x) + \varepsilon\,\Phi(x/\sigma)$

와 같이 주어지며, $\Phi(x)$ 는 표준정규분포의 분포함수를 $\varepsilon$ 은 오염의 정도를 각각 나타낸다.

귀무가설 $H_0 : \theta = 0$ 에 대하여 대립가설 $H_1 : \theta > 0$ 을 설계하기 위하여 다음의 관계식을 이용하였다.

$\theta_m = m \cdot \delta/\sqrt{n},\ m = 0, 1, 2, 3.$

즉, $m=0$ 이면 귀무가설을 의미하며 $m$ 의 증가는 더욱 두드러진 대립가설을 의미한다. $n$ 은 표본의 개수로서 본 모의실험에서는 10을 사용하였고, $\delta$ 는 모의실험의 결과를 잘 살필 수 있도록 각 분포에 따라 적절하게 선택되었다.

각 분포에서 주어진 크기의 표본들을 생성하고, t-검정과 윌콕슨 부호순위 검정의 검정통계량 T와 W를 계산한 후 이 값들을 유의수준 10%, 5%, 1%에서의 기각값들과 비교한다. 이러한 실험을 2000번 반복하여 각 검정법이 기각한 횟수를 반복수 2000으로 나눈 값이경험적검정력(empirical power)이며 $m=0$ 인 경우는 경험적 유의수준이라 한다. 한편, 윌콕슨 부호순위검정의 기각값들은 이산적이므로 정해진 유의수준을 만족하도록 확률화검정(randomized test)을 실시하였다. 이상의 결과를 〔표 3-1〕에 요약 정리하였고, 이를 위한 FORTRAN Code가 부록에 수록되어 있다.

〔표 3-1〕의 결과를 요약하면 다음과 같다. 모수적 검정인 T는 모집단이 정규분포를 따를 때에는 정확한 경험적 유의수준을 보여주고 있으며, 비모수적 검정인 W에 비해 조금 높은 검정력을 나타내고 있다. 그러나 모집단의 분포가 오염정규분포이거나 코오쉬분포와 같이 꼬리가 두터운 분포에서는 실제 유의수준을 밑도는 보수적인 성향을 보여주고 있다. 반면에 분포무관검정인 W는 모집단의 분포와는 무관하게 비교적 정확한 경험적 유의수준을 유지하며, 정규분포의 경우를 제외한 모든 분포에서 T보다 높은 검정력을 나타내고 있다. 특히 극단적으로 두터운 꼬리를 갖는 코오쉬분포에서는 W의 검정력이 T의 검정력에 비해 월등함을 알 수 있다. 이러한 결과들은 우리가 예상한 결과와 일치하는 것이다.

## Ⅳ. 결론 및 향후과제

본 논문에서는 통계적 모의실험에서 자주 사용되는

〔표 3-1〕 경험적 유의수준과 검정력

| 분 | $\alpha$ | | .10 | | .05 | | .01 |
|---|---|---|---|---|---|---|---|
| 포 | $m$ | T | W | T | W | T | W |
| 정규 | 0 | .1050 | .1095 | .0525 | .0500 | .0090 | .0110 |
| | 1 | .3740 | .3625 | .2375 | .2300 | .0685 | .0690 |
| | 2 | .7275 | .7090 | .5605 | .5455 | .2660 | .2455 |
| | 3 | .9465 | .9405 | .8890 | .8670 | .5890 | .5590 |
| 이중지수 | 0 | .1055 | .1020 | .0470 | .0540 | .0110 | .0125 |
| | 1 | .4060 | .4350 | .2690 | .3015 | .0920 | .1180 |
| | 2 | .7535 | .7845 | .6195 | .6540 | .3260 | .3405 |
| | 3 | .9290 | .9425 | .8640 | .8765 | .6340 | .6185 |
| 오염정규 ($\varepsilon$=0.1, $\sigma$=5) | 0 | .0885 | .0950 | .0410 | .0500 | .0040 | .0090 |
| | 1 | .4305 | .4600 | .2800 | .3110 | .0885 | .1135 |
| | 2 | .7475 | .8245 | .6320 | .7040 | .3650 | .3995 |
| | 3 | .8750 | .9565 | .8150 | .9000 | .6535 | .7085 |
| 오염정규 ($\varepsilon$=0.2, $\sigma$=5) | 0 | .0955 | .0960 | .0380 | .0455 | .0030 | .0070 |
| | 1 | .4255 | .4925 | .2980 | .3470 | .0855 | .1390 |
| | 2 | .6770 | .7975 | .5560 | .6815 | .3095 | .3905 |
| | 3 | .8155 | .9050 | .7300 | .8400 | .5390 | .5795 |
| Cauchy | 0 | .0915 | .1010 | .0300 | .0495 | .0025 | .0105 |
| | 1 | .3690 | .5430 | .2445 | .4050 | .0885 | .1745 |
| | 2 | .5960 | .8070 | .4845 | .7130 | .3010 | .4155 |
| | 3 | .7060 | .8955 | .6285 | .8365 | .4535 | .5595 |

확률분포들의 난수생성기와 확률계산 및 상위백분위수를 계산하기 위한 라이브러리의 구현과 응용 예를 다루었다. 구현된 라이브러리를 이용한 모의실험의 결과로부터 굳이 상용화된 고가의 패키지를 이용하지 않고도 통상적으로 기대되는 결과를 얻을 수 있었다. 물론 앞으로 더 세심한 검증절차를 거쳐 안정성을 확보해야 하겠다.

구현된 라이브러리에 포함된 루틴들이 통계적 모의실험에서 가장 기본적인 것들이긴 하지만 아주 일부분에 지나지 않는다. 따라서 앞으로 필요한 많은 부분들이 추가되어야 할 것이며, 윈도우즈용 컴파일러에서 사용될 수 있도록 DLL(Dynamic Linking Library)로도 개발이 되어야 할 것이라 생각된다.

## 부록 : FORTRAN code

```
C234567
      PARAMETER (NREP=2000, N=10)
      REAL X(N),Y(N), U(1)
      DOUBLE PRECISION DSEED
      DATA DSEED/12345679.D0/
C
      OPEN(6,FILE='MONTE.RES')
C
      Q10=QT(N-1,0.05)
      Q5=QT(N-1,0.025)
      Q1=QT(N-1,0.005)
C
C*** LOOP OF DISTRIBUTION CHAGE ***
C


      DO 588 ID=1,5
C
        WRITE(6,800) N,NREP,ID
        WRITE(6,802)
        WRITE(6,803)
C
C*** ALTERNATIVE GENERATING LOOP ***
C
      DO 1300 M=0,3
C
        RM=M*SM/SQRT(N)
C
      KA=0
      KB=0
      KC=0
      NA=0
      NB=0
      NC=0
C
C******* REPLICATION LOOP **********
C
        DO 1200 IT=1,NREP
        IF (ID .EQ. 1) THEN
          SM=1.
          CALL RNOR(DSEED,N,X)
        ELSEIF (ID .EQ. 2) THEN
          SM=1.4
          CALL RDEX(DSEED,N,X)
        ELSEIF (ID .EQ. 3) THEN
          SM=1.5
       CALL RCNOR(DSEED,N,X,Y,5.,0.1)
       ELSEIF (ID .EQ. 4) THEN
          SM=1.8
       CALL RCNOR(DSEED,N,X,Y,5.,0.2)
        ELSE
          SM=3.
        CALL RCAU(DSEED,N,X)
        ENDIF
C
      DO 48 I=1,N
        X(I)=X(I)+RM
```

```
 48   CONTINUE
C
C==== PARAMETRIC T-TEST ======
C
      SX=0.
      SX2=0.
     DO 50 I=1,N
      SX=SX+X(I)
      SX2=SX2+X(I)*X(I)
 50   CONTINUE
     XBAR=SX/FLOAT(N)
     SXX=SX2-FLOAT(N)*XBAR*XBAR
     S=SQRT(SXX/FLOAT(N-1))
     T=SQRT(FLOAT(N))*XBAR/S
     IF (T .GE. Q10) KA=KA+1
     IF (T .GE. Q5) KB=KB+1
     IF (T .GE. Q1) KC=KC+1
C
C== WILCOXON'S SIGNED RANK TEST ==
C
        NWILCO=0
     DO 150 I=1,N
      DO 150 J=I,N
 IF (X(J)+X(I) .GT. 0.) NWILCO=NWILCO+1
 150  CONTINUE
C
C———————— RANDOMIZED TEST FOR
C
    CALL RUNI(DSEED,1,U)
     IF ((NWILCO .GE. 41).or.((NWILCO
     .EQ. 40).and.(RND .LE. 3./19.)))
    +      NA=NA+1
     IF ((NWILCO .GE. 45).or.((NWILCO
     .EQ. 44).and.(RND .LE. 8./11.)))
    +      NB=NB+1
    IF (NWILCO .GE. 50)   NC=NC+1
C
C***** END OF REPLICATION LOOP *****
C
 1200   CONTINUE
C
WRITE(6,901) M,FLOAT(KA) /FLOAT(NREP),
FLOAT(NA)/FLOAT(NREP)
WRITE(6,902) M,FLOAT(KB)/FLOAT(NREP),
FLOAT(NB)/FLOAT(NREP)
WRITE(6,903) M,FLOAT(KC)/FLOAT(NREP),
FLOAT(NC)/FLOAT(NREP)
WRITE(6,904)
C
C***END OF ALTERNATIVE GENERATING LOOP***
C
 1300  CONTINUE
C
C***END OF DISTRIBUTION CHANGE LOOP***
C
  588 CONTINUE
C
800  FORMAT(/,1X,'N  =',I3,5X,'REPLICATION=
'I5,5X,'DISTRIBUTION =',I3)
802 FORMAT(/,1X,' M. ALPHA T-TEST
WILCOXON ')
803   FORMAT   (1X,'============
==========================')
901  FORMAT(1X,I3,5X,'0.10',5X,2F10.4)
902  FORMAT(1X,I3,5X,'0.05',5X,2F10.4)
903  FORMAT(1X,I3,5X,'0.01',5X,2F10.4)
904 FORMAT(1X,'————————————————')
      STOP
      END
C
C***SUBROUTINES FOR SPECIAL DISTRIBUTIONS ***
C
C——((( CONTAMINATED NORMAL  )))——
C
      SUBROUTINE RCNOR
       (DSEED,N,R,R2,SIG,CON)
      REAL R(N)
      REAL R2(N)
```

```
      DOUBLE PRECISION DSEED
      CALL RUNI(DSEED,N,R2)
      CALL RNOR(DSEED,N,R)
      DO 7 I=1,N
      IF (R2(I) .LT. CON) R(I)=SIG*R(I)
7     CONTINUE
      RETURN
      END
```

* 밑줄친 부분이 구현된 라이브러리의 루틴을 이용하는 부분임.

## 참고문헌


[1] Devroye, L. (1986). Non-Uniform Random Variate Generation, New York, Springer-Verlag.

[2] Forsythe, G. E., Malcolm, M. A., and Moler, C. B.(1977). Computer Methods for Mathematical Computations, Englewood Cliffs, NJ, Prentice-Hall.

[3] Kennedy, Jr W. J. and Gentle, J. E. (1980). Statistical Computing. Marcel Dekker, Inc.

[4] Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P.(1994a). Numerical Recipes in Fortran, 2nd Edition, Cambridge University Press.

[5] Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P.(1994b). Numerical Recipes in C, 2nd Edition, Cambridge University Press.

[6] Ripley, B. D. (1981). Stochastic Simulation, John Wiley \& Sons, Inc.

[7] Rubinstein, R. Y. (1981). Simulation and Monte Carlo Method, Technion, Israel Institute of Technology.

[8] 박경렬, 권기창, 권영담 (1998). 의사난수 생성기의 일양성과 독립성 검정. 한국데이터정보과학회지, 9권 2호, 237-246.


## 저자소개


**신봉섭**

1984. 02 : 건국대학교 사범대학 수학교육학과 졸업

1987. 02 : 서울대학교 대학원 계산통계학과 이학석사

1993. 08 : 서울대학교 대학원 계산통계학과 이학박사

1994. 03 ~ 현재 : 안양대학교 정보통계학과 조교수로 재직중

관심분야 : 전산통계, 통계계산 및 그래픽스, 통계교육, 로버스트 회귀분석

**박춘성**

1973. 2 : 건국대학교 이과대학 수학과 졸업

1980. 2 : 건국대학교 대학원 수학과 이학석사

1988. 8 : 건국대학교 대학원 수학과 이학박사

1984. 3~94. 2 : 경원전문대학 교양과 재직

1998. 9~현재 : 경원전문대학 교양과 대우 부교수