

레이아웃 시뮬레이터를 위한 통계처리 모듈의 개발

신 호섭, 정 한일, 김 상균, 김 기태, 이 경석, 박 진우

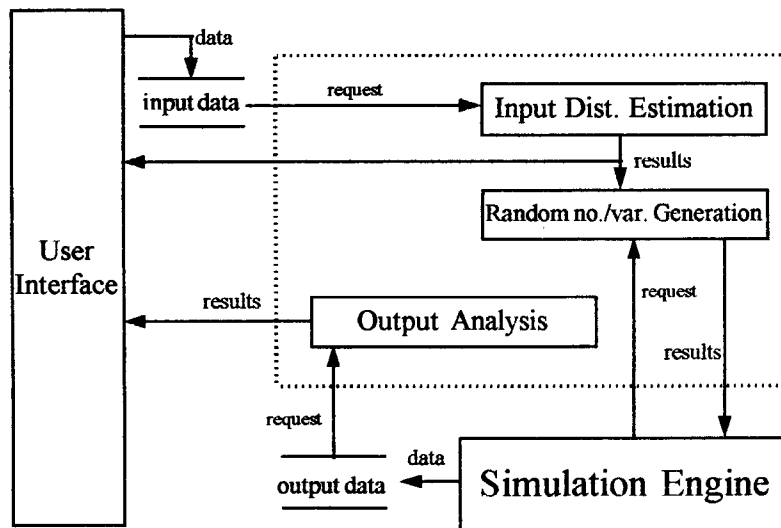
서울대학교 산업공학과

본 연구는 레이아웃 시뮬레이터의 근간이 되는 통계처리 모듈 개발에 관한 것이다. 통계모듈의 구성은 확률분포 추정 부모모듈, 난수 및 확률변수 발생 부모모듈 그리고 결과 분석 부모모듈로 구성된다. 전체적인 모듈의 구현은 객체지향의 개념으로 추진하였다. 이는 통계모듈의 개발에 독립성과 모듈관리 및 운영의 효율성을 제고시킬 것이다.

주요어 : 시뮬레이션, 입력분포추정, 결과분석, 난수 및 확률변수, 객체지향개념

1. 서 론

본 연구는 시뮬레이터의 근간이 되는 통계처리 모듈 개발에 관한 것이다. 레이아웃 시뮬레이터를 구성할 통계처리 모듈은 표본자료로부터 표본의 확률분포를 추정하는 확률분포 추정 부모모듈, 주어진 확률분포로부터 난수(random number)발생기를 이용하여 확률 변수를 발생시키는 난수/확률변수 발생 부모모듈, 시뮬레이션 결과를 통계처리하여 사용자가 쉽게 정보를 도출하게 하는 결과분석 부모모듈로 구성된다.



[그림1] 통계처리 모듈의 구성

기존의 시뮬레이션 패키지나 통계처리 패키지 사용시, 통계적 지식이 부족한 사용자는 출력결과로부터 유용한 정보를 얻어내는 것이 어려웠다. 그래서 본 연구에서는 초보적인 수준의 통계적 지식만을 갖춘 사용자라 하더라도 쉽게 입력분포를 추정할 수 있고, 출력결과로부터 유용한 정보를 도출해 낼 수 있도록 의미해석에 주안점을 둔 통계처리 모듈을 개발하고자 한다.

전체적인 모듈의 구현은 객체지향의 개념(object-oriented concept)을 바탕으로 추진하였다. 이는 소프트웨어 공학적 측면에서 객체지향의 일반적 장점을 활용하기 위한 것이다. 통계모듈의 경우 자료를 그래프로 보여주기 위한 인터페이스 부분 이외에는 거의 대부분 수치연산을 다루는 함수들로 구성되기 때문에 기존의 function library 형태를 취하는 것도 좋은 방법이며, 이 경우 객체지향방식보다 수행속도 면에서 미세한 우세가 예상된다. 본 연구에서는 통계모듈의 개발환경 및 다른 모듈과의 통합이라는 측면을 고려하여 function library와 class library라는 두 개념의 접목을 시도하였다. 즉, 통계모듈을 구성하는 function들을 그 기능에 따라 몇 개의 category로 분류하고, 이와 같이 결정된 category를 하나의 object class로 정의하였다. 그리고 각 class에는 그에 해당하는 function들을 member로 정의함으로써 library적 성격을 갖게 하였다. 결국 세밀하고 복잡하게 class계층을 구성하지 않고, 논리적으로 서로 연관된 함수들로 이루어진 소규모의 library class들을 이용하여 간단한 구조의 계층을 만들어서, 통계모듈을 모델링함으로써 통계모듈의 수정, 확장과 관련하여 다른 모듈과의 독립성을 높일 수 있고 모듈관리를 용이하게 하는 장점을 취할 수 있다.

2. 통계모듈의 구성

2.1 난수 및 확률변수 발생 부모들

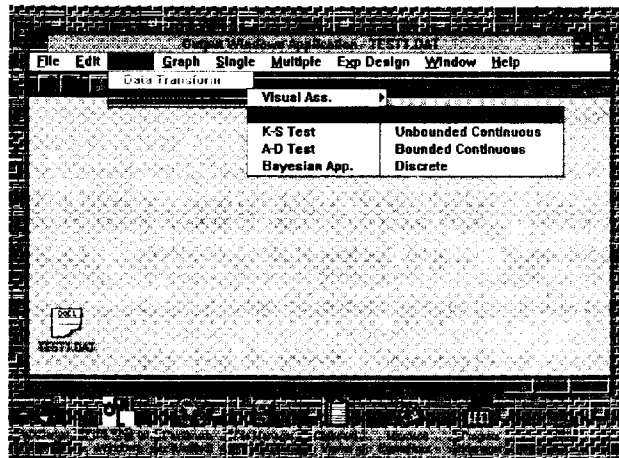
확률변수를 사용하는 시뮬레이션은 주어진 확률분포들로부터 확률변수를 발생하는 방법을 갖고 있어야 한다. 이러한 확률변수의 발생은 먼저 난수의 생성이 우선되어야 한다. 이러한 확률변수의 발생에는 무작위한 특성(randomness), 긴 주기, 재생성과 유일성(reproducibility & uniqueness), 계산상 효율성 등을 만족해야 한다. 현재 PC에서 널리 쓰이는 시뮬레이션 패키지들은 주로 Multiplicative Generator를 사용하여 난수를 발생시키고 있다. SIMAN, SLAM은 Schrage가 제안한 방법(multiplier는 16,807, modulus는 $2^{31}-1$)을 사용하고 있으며 SIMSCRIPT의 경우는 Marse와 Roberts가 제안한 방법(multiplier는 630,360,016, modulus는 $2^{31}-1$)을 채택하고 있다. 또한 GPSS에서는 Fishman과 Moore가 제안한 방법(multiplier는 397,204,094, modulus는 $2^{31}-1$)을 사용하고 있다. 본 연구에서는 Schrage, Fishman이 제안한 방법 중 난수발생기의 특성에 부

합하는 것이 선택될 것이다.

확률변수의 발생방법에는 ① Inverse Transform Method, ② Accept-Rejection Method, ③ Convolution 등이 있다. 그리고 확률변수 생성에서 가져야 되는 특징 중에서 시뮬레이션에서의 응용을 위해서는 계산상의 효율성 즉 생성에 소요되는 시간이 중요하다. 그러므로 Accept-Rejection Method는 여러 번 생성해야 되는 경우가 발생할 수 있으므로 본 생성모듈에서는 사용하지 않았다. 또한 정규분포의 생성의 경우에는 Numerical Method를 사용하는 등 생성에 소요되는 시간의 단축에 많은 노력을 기울였다. 그리고 본 모듈에서 발생시키는 확률분포는 확률분포 추정 부모들에서 추정대상이 되는 모든 분포를 포괄하고 있다.

2.2 확률 분포 추정 부모들

표본 자료로부터 확률분포를 추정하는 이론적인 접근 방법은 아직 연구중인 상태로 그 중요도만큼이나 어렵다. 현재까지 널리 이용되는 확률분포의 추정방법은 크게 3가지가 있는데, 확률 플롯(probability plot), 진동수 비교(frequency comparison), 박스 플롯(box-plot) 등을 이용하는 시각적 판정법(visual assessment), 카이 제곱 테스트(chi-square test), 콜모고로프-스미르노프 테스트(Kolmogorov-Smirnov test), 앤더슨-달링 테스트(Anderson-Darling test) 등을 이용하는 적합도 판정법, 그리고 베이저안 접근법(Bayesian approach) 등 3가지로 요약될 수 있다. 본 연구에서 전체적으로 지원되는 분포로는 연속분포 16가지, 이산분포 5가지가 제공될 수 있다.



[그림2] 입력분포 추정 부모들 수행의 예

그러나 각 판정법 별로 지원될 수 있는 분포에는 차이가 있는데 시각적 판정법과 카이제곱 테스트에는 21개 분포 모두가 지원되고, 콜모고로프-스미르노프 테스트, 그리

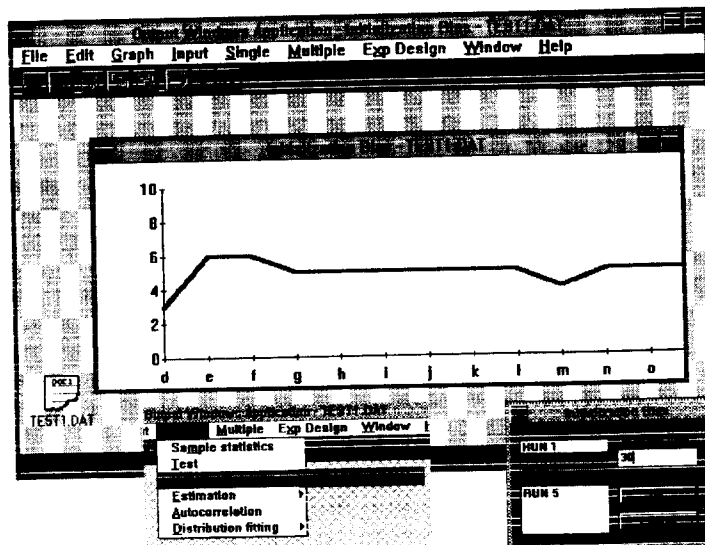
고 앤더슨-달링 테스트에는 지수 분포(Exponential distribution), 정규 분포(Normal distribution), 대수 정규 분포(Lognormal distribution), 와이블 분포(Weibull distribution)가 지원될 수 있으며 베이저안 접근법에는 지수분포, 정규분포, 일양분포(Uniform distribution), 포아송 분포(Poisson distribution) 등이 지원된다. 또한 본 연구에서는 통계적 지식이 부족한 사람들도 각각의 방법을 쉽게 사용하여 분포를 추정할 수 있도록 하였다.

추정 분포 선택에서는 사용자가 검정 방법 그리고 추정분포를 선택하여 검정을 실행하면, 실행의 결과로부터 각 분포의 적합도의 값에 의해 순위를 부여함으로써 사용자가 분포를 선택할 수 있도록 한다. 또한 추정 분포 선택시 종류별로 나열된 분포들에 대해 사용자가 원하는 만큼 다수개의 분포를 선택할 수 있으며, 각 분포 선택시 분포의 모수에 대한 정보도 다양하게 제공될 수 있다.

검정모들은 앞에서 제시한 3가지 추정방법을 지원하는 방법들로 구성될 것이다.

2.3 결과 분석 부모들

기존의 시뮬레이션 패키지에서 제공하는 통계처리기능은 사용자 편의와 실험계획지원이라는 점에서 부족한 바가 많았다. 통계적 지식이 부족한 사용자는 출력 결과로부터 유용한 정보를 얻어내는 것이 어려웠고, 또한 필요한 정보를 경제적으로 얻기 위한 사전실험계획은 전적으로 분석자의 손에 맡겨졌었다. 이러한 이유로 본 연구에서는 기초적인 통계적 지식만을 갖춘 사용자라 할지라도 각종 정보를 쉽게 도출해 낼 수 있는, 의미해석에 주안점을 둔 결과 분석 부모들을 개발하고자 한다.

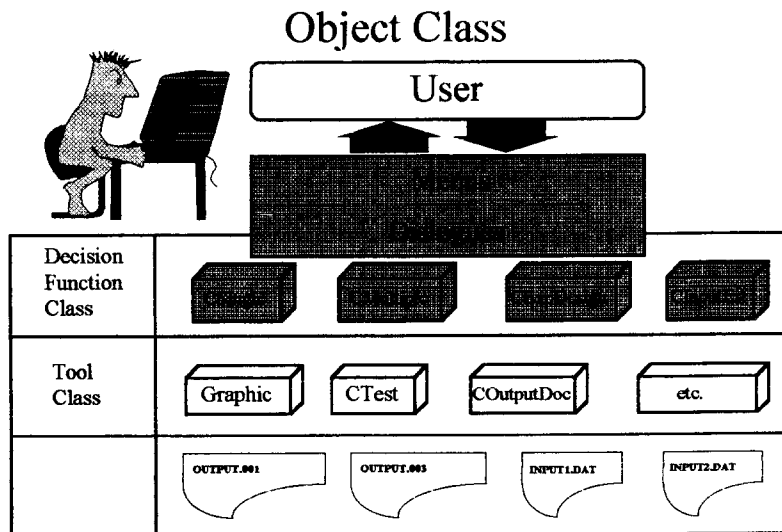


[그림3] 결과분석 부모들 수행의 예

결과분석의 관점에서 볼 때 시뮬레이션은 단일대안에 대한 복수의 실험과 다수대안 간의 비교를 위한 실험으로 구분될 수 있으며, 분석기능은 각각의 실험종류에 따라 모수추정(점추정, 구간추정), 회귀분석, 가설검증(대안비교, 분포추정)으로 대별할 수 있다. 개발된 결과분석 부모들은 위의 모든 기능을 지원하며, 다양한 그래픽양식을 통한 시각적 분석 및 통계적 기법을 활용한 정량적 분석을 모두 시행할 수 있다. 예를 들어, bar chart, line graph 등의 기능이 메뉴수준에서 제공되며, 입력분포추정 부모들에서 사용되는 각종 검정방법, 그리고 batch means, autoregressive, standardized time series 등의 평균추정기법이 간단한 사용자 대화를 통해 구현된다. 또한 실험계획과 관련하여 실험횟수 및 정밀도에 관한 의사결정 지원기능도 갖추고 있다.

3. 결론

앞에서 제시한 각종 부모들은 객체 지향 언어(C++)를 이용하여 구현되었다. 통계 모듈을 구성하는 각각의 function들을 몇 개의 category로 분류하여 object로 정의하였다. 전체 object의 구조를 tool class와 function class로 분류하여, tool class가 function class를 지원하는 library적 성격을 띄게 하였다. 이런 구조를 채택함으로써 얻을 수 있는 장점으로는 각 부모들간의 독립성을 제고하여 추후의 수정이나 확장과 관련한 모듈의 관리에 효율을 기할 수 있을 것이다.



[그림4] Object의 계층

4. 참고문헌

- [1] Aggarwal, O P. "Some minimax invariant procedures for estimating a cumulative distribution function. Ann. Math. Statist. VOL 26, pp.450-462, 1955
- [2] Averill M. Law and W. David Kelton, Simulation Modelling and Analysis 1st ed., McGraw-Hill, pp.155-332, 1982
- [3] Averill M. Law and W. David Kelton, Simulation Modelling and Analysis 2nd ed., McGraw-Hill, 1991, pp.420-457
- [4] Fishman, G.S., Concepts and Methods in Discrete Event Digital Simulation, John Wiley, 1973
- [5] Fishman, G.S., Principles of Discrete Event Digital Simulation, John Wiley, 1978
- [6] Fishman, G.S. and L. R. Moore, "An Exhaustive Analysis of Multiplicative Congruential Random Number Generators with Modulus $2^{31}-1$ " SIAM J. Sci. Stat. Comput., 7, pp.24-25, 1986
- [7] Gideon f. Inbar, Tsur Ginat, "Parallel Nonlinear Neuronal Channels: The Generalized Steady-State Analysis" IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, VOL. SMC-13, NO.5, pp.1033-1037, 1983
- [8] Johnson Kotz, Distributions in statistics, John Wiley, 1970
- [9] Marse, K., S.D. Roberts, "Implementing a Portable FORTRAN Uniform (0, 1) Generator," Simulation, 41, pp.135-139, 1983
- [10] Morris H. DeGroot, Optimal statistical decisions, McGraw-Hill Book Company, 1970
- [11] PHADIA, E.G. "Best invariant confidence bands for a continuous cumulative distribution function" Austral. J. Statist. VOL 16, pp.148-152, 1974
- [12] PHADIA, E.G. "Minimax estimation of a cumulative distribution function." Ann. Statist. VOL 1, pp.1149-1157, 1973
- [13] Roberts, A., Verberg, D., Convex Functions, Academic, New York, 1973
- [14] Schruben, L., H. Singh, and L. Tierney, "Optimal tests for initialization bias in simulation output," O.R., Vol.31, No.6, 1983

- [15] William H. Press, Saul A. Teukolsky, William T. Vetterling and Brian P. Flannery, Numerical Recipes in C 2nd ed. Cambridge University Press, 1992
- [16] Yaakov Friedman, Alexander Gelman, Eswar Phadia, "Best invariant estimation of a distribution function under the KOLMOGOROV-SMIRNOV loss function" The Annals of Statistics, VOL. 16, NO. 3, pp.1254-1261, 1988

4. 참고문헌

- [1] Aggarwal, O P. "Some minimax invariant procedures for estimating a cumulative distribution function. Ann. Math. Statist. VOL 26, pp.450-462, 1955
- [2] Averill M. Law and W. David Kelton, Simulation Modelling and Analysis 1st ed., McGraw-Hill, pp.155-332, 1982
- [3] Averill M. Law and W. David Kelton, Simulation Modelling and Analysis 2nd ed., McGraw-Hill, 1991, pp.420-457
- [4] Fishman, G.S., Concepts and Methods in Discrete Event Digital Simulation, John Wiley, 1973
- [5] Fishman, G.S., Principles of Discrete Event Digital Simulation, John Wiley, 1978
- [6] Fishman, G.S. and L. R. Moore, "An Exhaustive Analysis of Multiplicative Congruential Random Number Generators with Modulus $2^{31}-1$," SIAM J. Sci. Stat. Comput., 7, pp.24-25, 1986
- [7] Gideon f. Inbar, Tsur Ginat, "Parallel Nonlinear Neuronal Channels: The Generalized Steady-State Analysis" IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, VOL. SMC-13, NO.5, pp.1033-1037, 1983
- [8] Johnson Kotz, Distributions in statistics, John Wiley, 1970
- [9] Marse, K., S.D. Roberts, "Implementing a Portable FORTRAN Uniform (0, 1) Generator," Simulation, 41, pp.135-139, 1983
- [10] Morris H. DeGroot, Optimal statistical decisions, McGraw-Hill Book Company, 1970
- [11] PHADIA, E.G. "Best invariant confidence bands for a continuous cumulative distribution function" Austral. J. Statist. VOL 16, pp.148-152, 1974
- [12] PHADIA, E.G. "Minimax estimation of a cumulative distribution function." Ann. Statist. VOL 1, pp.1149-1157, 1973
- [13] Roberts, A., Verberg, D., Convex Functions, Academic, New York, 1973
- [14] Schruben, L., H. Singh, and L. Tierney, "Optimal tests for initialization bias in simulation output," O.R., Vol.31, No.6, 1983