Maxima를 이용한 이론통계학 교육

이긍희(한국방송통신대학교 통계·데이터과학과)

1. Maxima는 무엇인가?

Maxima 프로그램은 미적분, 행렬 연산 등 기호 연산을 할 수 있는 수학 소프트웨어이다. 이 프로그램은 Maple, Mathematica와 같은 *computer algebra system* (CAS)이지만 무료로 이용할 수 있는 공개 소프트웨어이다. Maxima는 1960년대 MIT에서 만든 Macsyma를 바탕으로 1998년 미국 텍사스 주립대 Professor W. F. Schelter가 만든 소프트웨어이다. Maxima는 Windows, Mac OSX와 Linux에서 수행할 수 있으며 R 프로그램과 같이 다양한 contributed packages가 있다. Maxima를 이용하기 쉽게 graphical user interfaces (GUI)형태로 만든 소프트웨어가 wxMaxima이다. 실제 설명에서는 wxMaxima 중심으로 설명하도록 하겠다.

2. Maxima를 이용한 배경

한국방송통신대학교에서 대학원이 2012년 개설되면서 이론통계학 교과목을 운영하게 되었다. 대학원 입학 정원은 40명이며 온라인으로 진행되며 통계과학 전공과 바이오통계학 전공으로 구성되어 진행되고 있다. 학생은 전업학생보다는 의료계(의사, 약사 등), 회사(금융, 통계 작성기관 등) 등에 근무하는 직장인으로 구성되어 있다. 이들 학생 중에는 수학교사와 같이 수학에 대한 지식 배경이 좋은 학생도 있으나 많은 학생들이 수학 학습의 공백으로 미분, 적분 등에 익숙하지 못한 상황이다. 따라서 어떻게 이들에게 이론통계학의 핵심을 한 학기 동안 알려줄까 고민하면서 색다른 방식의 이론 통계학 교육을 진행하게 되었다.

이론통계학은 통계학의 추론 개념을 이해하는데 매우 중요한 교과목이다. 그러나 이론 통계학을 쉽게 접근하는 데에는 제약이 있다. 주요한 이유는 이론 통계학이 수학을 바탕으로 전개되기 때문이다. 수학을 학습해야만 진행할 수 있는 이론통계학 교육은 수학이란 장벽으로 인해 통계학의 핵심에 접근하지 못하게 하고 있다. 이러한 문제를 어떻게 해결할 수 있을까? 그 방법중 하나는 컴퓨터를 이용하여 기호 수학 연산을 하는 것이다. 이를 통해 이론통계학을 어렵게 만드는 수학의 어려움에서 벗어나 통계학 자체를 대면하게 된다. 이런 이유로 공개 소프트웨어인 Maxima를 이용하게 되었다.

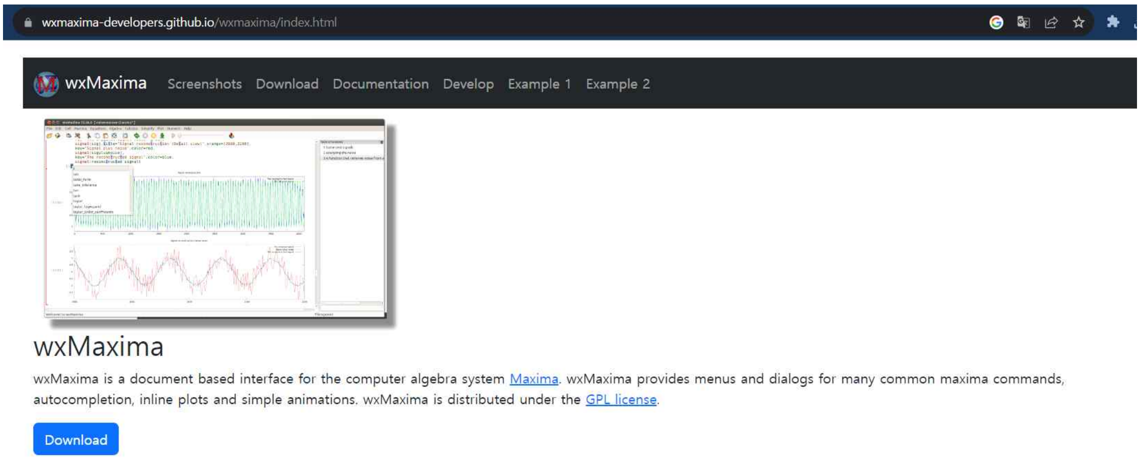
3. Maxima를 이용한 수학과 통계학

여기서는 Maxima의 GUI 버전인 wxMaxima의 이용하여 Maxima의 활용에 대해 살펴보겠다.

3.1 wxMaxima의 설치

wxMaxima는 다음 사이트에서 파일을 내려받아 설치할 수 있다.

https://wxmaxima-developers.github.io/wxmaxima/index.html



<그림 1> wxMxaima 관련 홈페이지

설치를 마치면 바탕화면에 wxMaxima 아이콘이 나타나며 이를 더블클릭하여 wxMaxima를 시작한다. 생성된 화면에서 Enter를 클릭하면 입력선(%i)이 나타나며 여기에 명령어를 입력한 후 한줄 씩 "shift+Enter"를 클릭하면 프로그램이 수행된다. 물론 상단 메뉴를 이용하여 메뉴방식으로 프로그램을 수행할 수 있다. 수행된 결과는 (%o)에 나타난다. 물론 프로그램을 일괄 작성하여 수행할 수도 있다.

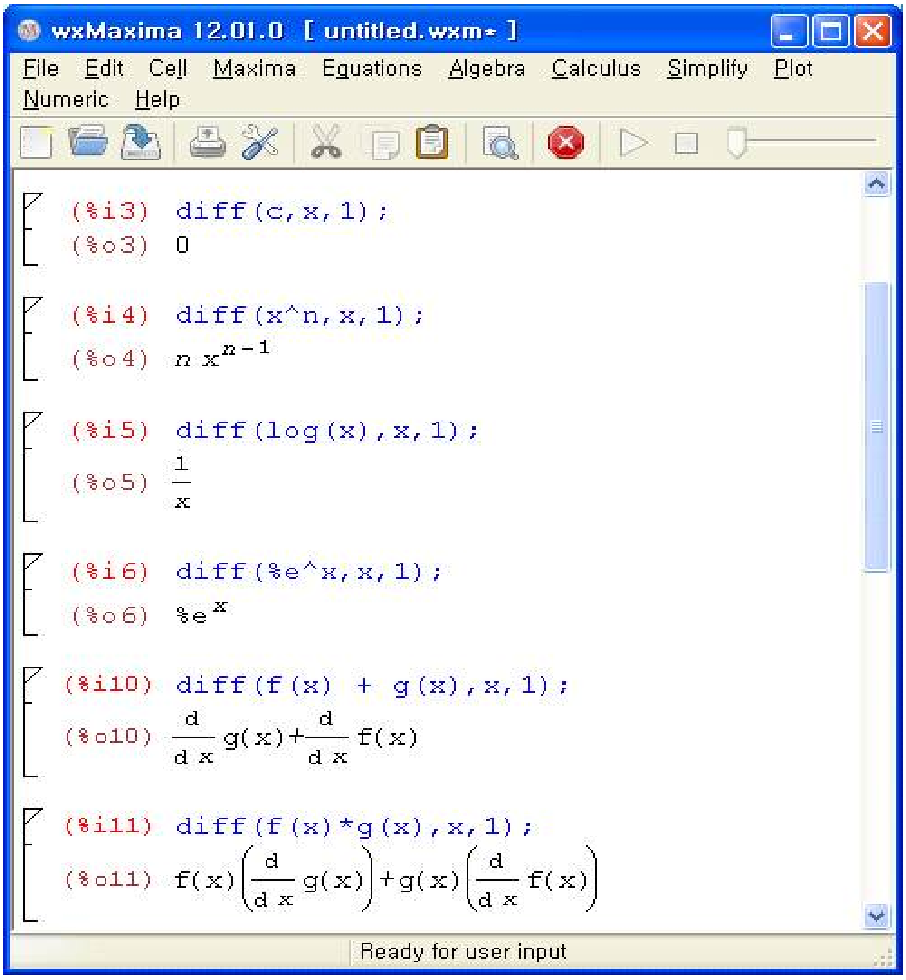
3.2 Maxima를 이용한 미분과 적분

Maxima를 이용하여 다양한 수학 기호 연산을 하고 그래프로 표현할 수 있다. 다음과 같은 다양한 미분 관련 연산 결과를 <그림 2>와 같이 Maxima로 표현하여 연산할 수 있다. 여기서 diff는 미분함수이며 (%i)는 <그림 2>의 입력명령어를 나타낸다.

$$(\%i3)\ f(x)=C(\text{상수}) \rightarrow f'(x)=0 \qquad (\%i4)\ \frac{d}{dx}x^n = nx^{n-1},\ \frac{d}{dx}e^x = e^x$$

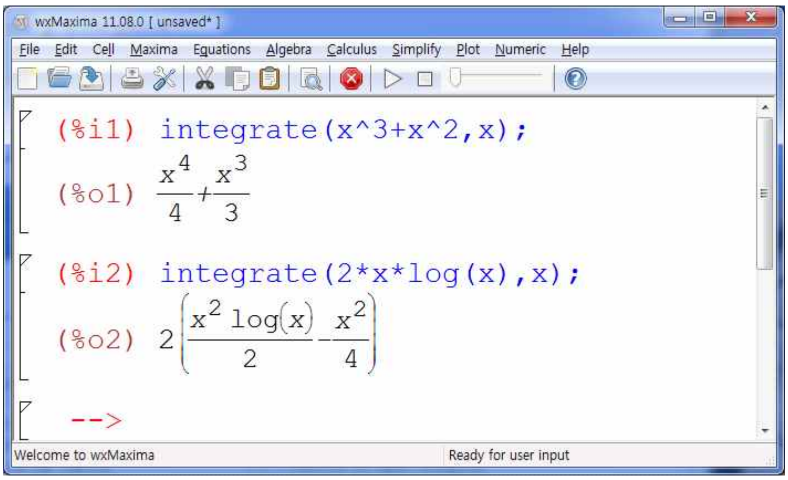
$$(\%i5)\ \frac{d}{dx}\ln x = \frac{1}{x} \qquad (\%i10)\ \frac{d}{dx}(f(x)+g(x)) = f'(x)+g'(x)$$

$$(\%i11)\ \frac{d}{dx}(f(x)g(x)) = f'(x)g(x)+f(x)g'(x)$$



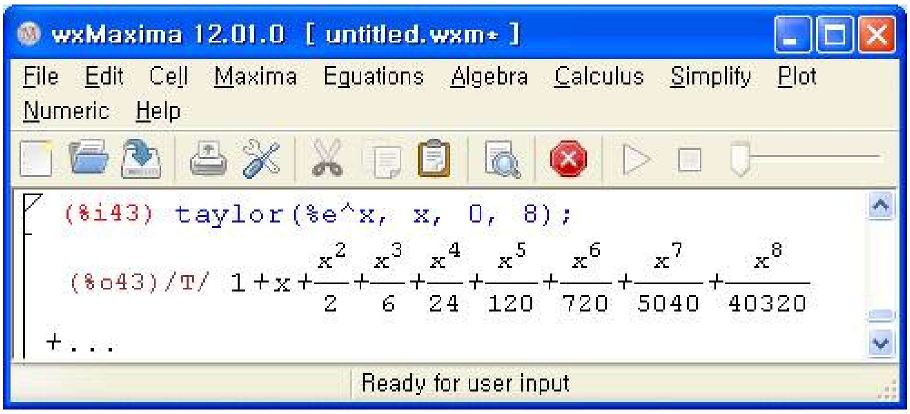
<그림 2> Maxima를 이용한 미분 연산

<그림 3>은 x^3+x^2과 $2x \cdot \log(x)$을 Maxima로 부정적분한 결과이다.



<그림 3> Maxima를 이용한 적분 연산

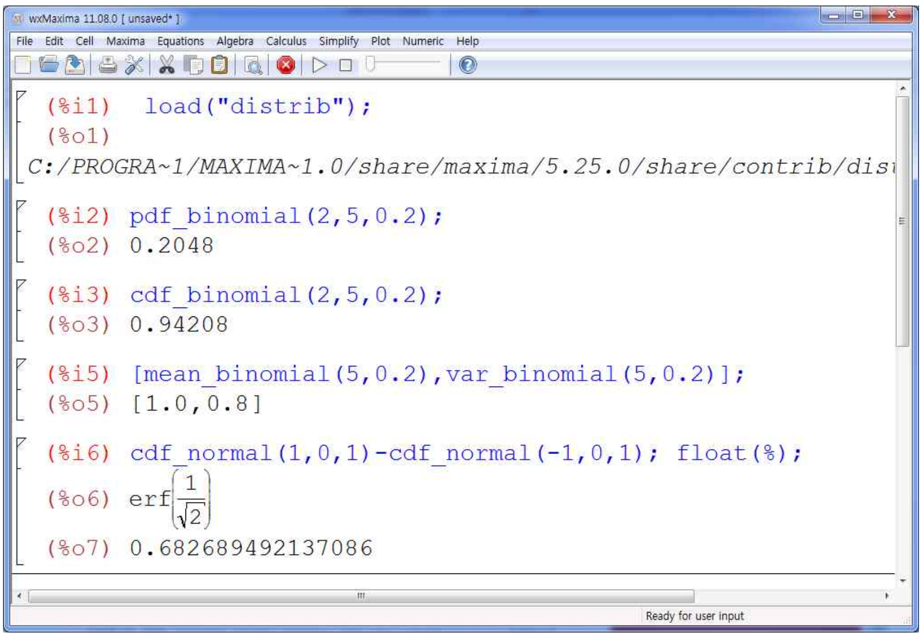
$f(x)=e^x$의 테일러급수는 $e^x=\sum_{k=0}^{\infty}\frac{1}{k!}x^k$ 인데 Maxima를 이용해서도 동일한 결과를 구할 수 있다 (<그림 4>).



<그림 4> Maxima를 이용한 테일러 급수 연산

3.3 Maxima를 이용한 확률분포의 계산

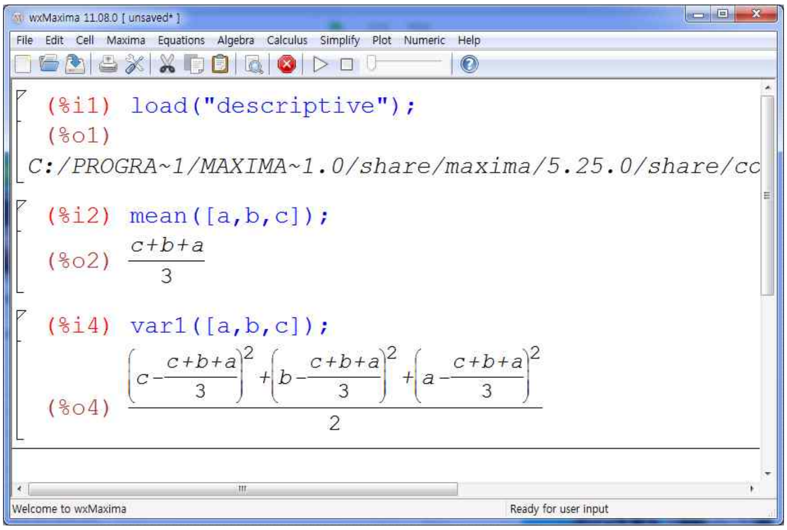
Maxima를 이용하여 이항분포, 정규분포 등 확률분포 관련 확률값 등을 R 프로그램과 마찬가지로 구할 수 있다. 이를 위해서는 distrib 패키지를 불러와야 한다(%i1). <그림 5>에서 Maxima를 이용하여 이항분포 B(5,0.2)를 따르는 확률변수 X에 대해 $P(X=2)$(%i2), $P(X\le 2)$(%i3)와 기대값, 분산을 구하고(%i5), 표준정규분포(Z)의 누적분포함수를 이용하여 $P(-1<Z<1)$를 구했다(%i6).



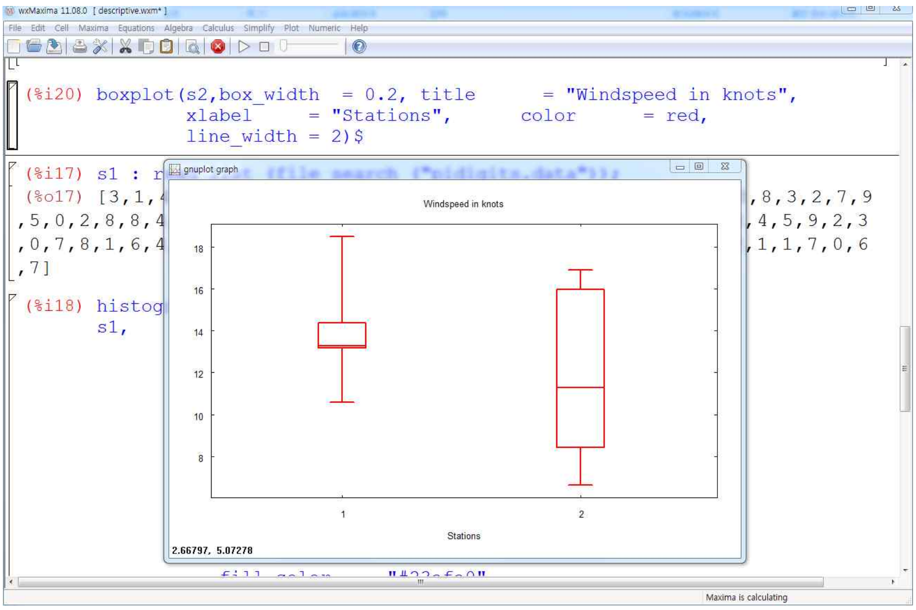
<그림 5> Maxima를 이용한 확률 계산

3.4 Maxima를 이용한 통계분석

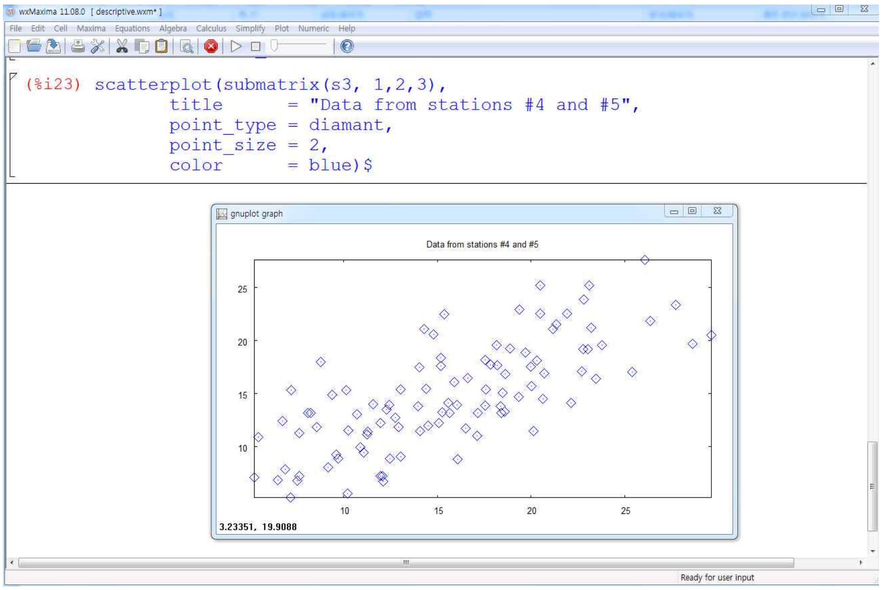
Maxima를 이용하여 통계패키지와 마찬가지로 통계분석을 할 수 있다. 통계 패키지와의 차이는 평균과 분산 등에 대해 기호 연산이 가능하다는 점이다(<그림 6>). <그림 7>과 <그림 8>을 보면 Maxima를 이용하여 데이터에 대한 상자그림이나 산점도 등의 작성할 수 있으며 회귀분석, 신뢰구간 구하기, 다양한 검정도 실시할 수 있다. Maxima에서 기초통계 분석을 하려면 descriptive패키지를 불러와야 한다.



<그림 6> Maxima를 이용한 평균 및 분산 계산



<그림 7> Maxima를 이용한 상자그림 그리기



<그림 8> Maxima를 이용한 산점도 그리기

4.5 Maxima를 이용한 이론 통계학 교육

Maxima를 활용해서 이용하여 이론통계학 문제를 풀어보자. <그림 8>은 정규분포와 관련된 다음 문제를 Maxima로 푼 결과이다.

(1) $N(\mu,\sigma^2)$를 따르는 확률변수의 확률밀도함수 전체면적이 1임을 보이시오.

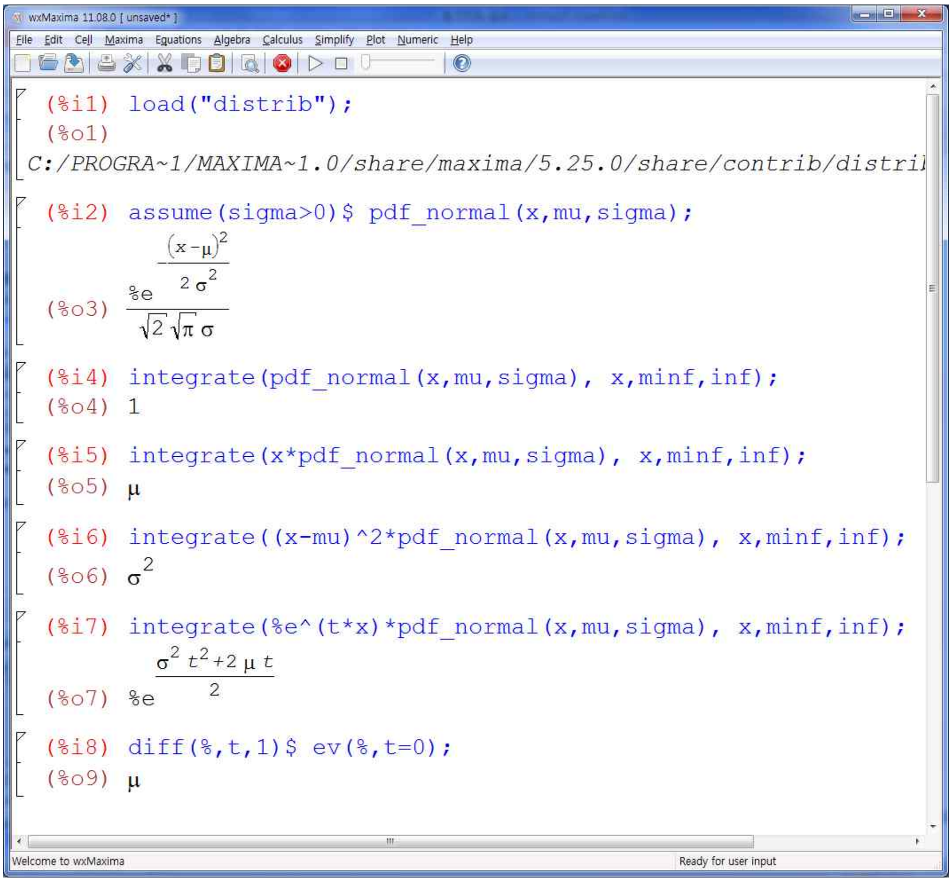
(2) $N(\mu,\sigma^2)$를 따르는 확률변수의 기대값과 분산을 구하시오.

(3) $N(\mu,\sigma^2)$를 따르는 확률변수의 적률생성함수를 구하시오.

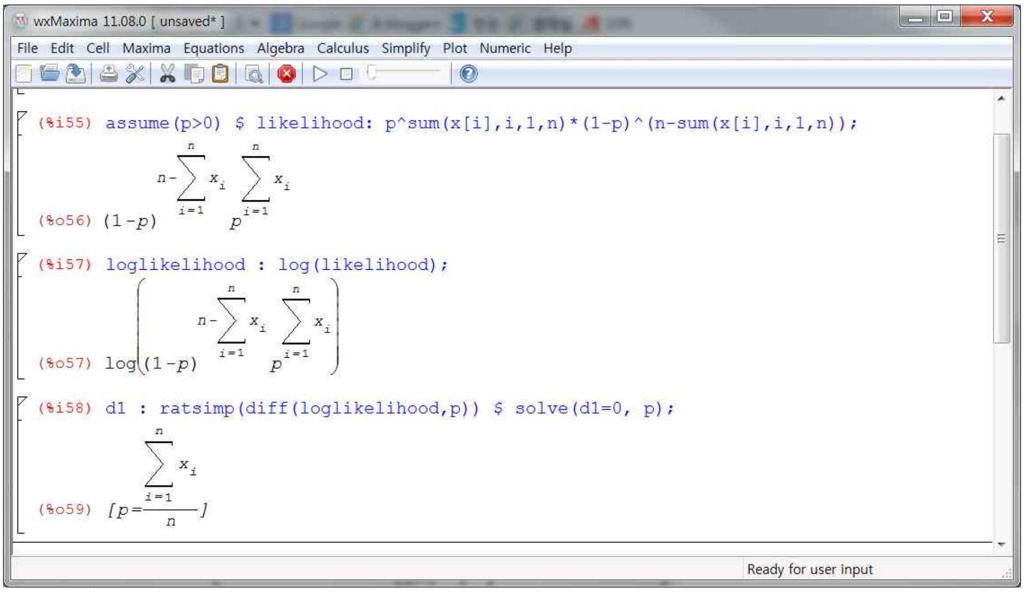
(4) $N(\mu,\sigma^2)$의 적률생성함수를 이용하여 확률변수의 기대값을 구하시오.

<그림 8>을 보면 먼저 distrib 패키지를 불러와서(%i1) 정규분포 $N(\mu,\sigma^2)$를 구해서 정규분포 확률밀도함수를 정의하고(%i2) 확률밀도함수 전체면적이 1임을 보였다(%i4). (%i5,%i6)에서 $N(\mu,\sigma^2)$를 따르는 확률변수의 기대값과 분산을 구했고 (%i7)에서 $N(\mu,\sigma^2)$를 따르는 확률변수의 적률생성함수를 구했다. (%i8)에서 적률생성함수를 미분한 후 t=0을 기대값을 구했다.

<그림 9>는 $X_1,X_2,\cdots,X_n \sim B(1,p)$(이항분포)일 때 이의 가능도함수(likelihood function)를 정의하고(%i55, %i57), 이를 이용하여 p의 최대가능도추정량(MLE)를 추정한 것이다(%58). 여기서 $는 명령어를 연결하여 수행한 후 최종결과를 표시하는 것이다.



<그림 8> Maxima를 이용한 정규분포의 기대값과 분산 구하기



<그림 9> Maxima를 최대가능도 추정량 구하기

4. 맺음말

Maxima를 이용한 미분과 적분 등 기호 연산에 유용할 뿐만 아니라 간단한 통계 연산도 가능한 공개 소프트웨어이다. 이 프로그램을 이용할 경우 이론 통계학 학습에 어려움으로 작용하고 있는 수학을 넘어서 통계학 추론에 집중할 수 있다는 장점이 있다. 향후 Maxima를 본격적으로 이론 통계학 교육에 활용하려면 Maxima를 이용한 이론통계분석 결과를 축적해야 한다. 이용자들은 Maxima가 기호 계산기 성격이 강하다는 점을 이해하고 무조건적으로 활용하기보다는 문제를 풀면서 부족한 기호 연산을 보완해야 겠다는 생각으로 Maxima를 접근할 필요가 있다.

5. 참고문헌

(1) Leon Q. Brin(2011) Maxima and the Calculus.

(2) Josef Leydold Martin Petry(2011) Introduction to Maxima for Economics, Institute for Statistics andMathematics, WUWien.