대한수학교육학회지 <학교수학> 제 4 권 제 4 호
Journal of Korea Society of Educational Studies in Mathematics
<School Mathematics> Vol.4, No.4, 677-688. Dec 2002

연속확률변수 개념의 직관적 이해에 관한 고찰

박 영 희*

I. 서론

수학의 제반 영역 중에서 통계영역은 학생들이 식으로만 이해하기에 더 어렵다고 생각하는 경향이 있다. 즉, 통계는 실제 생활과 연관되어 발전해 왔으므로 맥락을 배제하고 수학적 표현으로만 이해하기에는 어려운 점이 많다. 물건의 개수를 세는 것과 관련하여 직관적으로 인식하기 쉬운 이산량에 비해 연속량은 'Zeno의 역설'이 의미하듯이, 무한 개념과 관련되어서 인간의 직관으로 이해하기가 쉽지 않다. 따라서, 이런 연속량에 대한 확률을 고려하는 연속확률변수의 성질은 학생들이 이해할 수 있도록 제시하기가 어려울 수 밖에 없다. 6차 교육과정에서는 연속확률변수의 확률을 적분을 이용하여 도입하고, 그 평균과 표준편차를 유효한 범위에 대하여 확률밀도함수의 적분값으로 구하였다. 그런데, 7차 수학과 교육과정의 수학I의 학습 지도상의 유의점으로 '연속확률변수의 평균과 표준편차를 적분을 이용하여 구하는 식은 지도하지 않는다'고 하였고, 선택과목인 확률과 통계에서 교수학습 방법으로 '확률변수는 직관적으로 이해할 수 있도록 도입하고, 연속확률변수의 도입에서 적분기호는 사용하지 않는다'고 하였다. 그러므로, 7차 교육과정에 의해 고등학교 수학I 교과서들에서는 연속확률변수 개념을 적분을 사용하지 않고 제시하였다.

본 연구에서는 이렇게 제시되는 연속확률변수 및 확률밀도함수에 대한 개념 제시가 어떻게 되고 있고 각 제시 방법에서 유의해야 할 점은 무엇인지를 본 연구에서 논의하고자 한다. 이를 위해 우선 연속확률변수와 그에 따른 확률밀도함수의 개념에 대하여 고찰하고, 이를 7차 교육과정의 수학I 교과서들 및 일부 대학 통계 교재에서 제시하는 방식을 분석하였다. 그 분석된 결과를 바탕으로 몇 가지 유형으로 분류하고 각 방법에 대한 학생들의 반응을 조사하였다.

II. 본론

1. 연속확률변수 개념

통계에서 확률변수란 나타날 수 있는 결과들에 대해 실수를 대응시키는 함수이고 이는 각 결과의 특성 중에서 관심있는 특성을 수치화해 주는 역할을 한다. 일상 생활에서 연속량으로서 통계적 대상이 되는 것은 키, 몸무게, 시간, 온도 등이 있는데 이를 확률변수로 나타낸 것이 연속확률변수가 된다. 이산확률변수는 초등학교나 중학교 확률 단원에서 주로 다루는 동전이나 주사위 던지기, 카드 뽑기 등과 관련

* 청주교육대학교

되어 있으며 그 개념을 학생들이 직관적으로 이해하기가 그리 어렵지 않다. 하지만, 연속확률변수의 대상이 되는 연속량은 그 양을 구성하는 성분이 무한하면서 세어 나갈 수 없기 때문에 이산확률변수의 대상이 되는 이산량의 특성과 차이가 있다. 이산확률변수 중에도 기하분포 또는 포아송분포는 확률변수값의 범위가 자연수집합과 같은 가부번 집합(denumerable set)이다. 하지만, 연속확률변수의 범위는 실수집합과 같이 비가부번 집합(nondenumerable set)이다. 따라서, 이산확률변수에 대한 설명이 연속확률변수에 적용될 수 없다. 연속확률변수의 범위는 실수 또는 구간으로 정해지고 어느 특정 구간에 연속확률변수가 속하는 확률값은 어떤 함수의 적분으로 구하게 된다. 이는 마치 속도에 대한 함수를 적분하여 어떤 특정 시간동안에 움직인 거리를 구하는 것과 같다. 즉, 연속확률변수에서 무한소 구간에서 확률의 변화량을 나타내는 함수를 생각할 수 있는데 이를 확률밀도함수라고 한다.

2. 연속확률변수의 직관적 이해를 위한 방법 고찰

(1) 상대도수 히스토그램의 점근적 접근

6차교육과정에 따른 교과서에서 범위가 [a, b]인 연속확률변수 X의 확률밀도함수를 $f(x)$라고 하고 이것은 다음과 같은 식을 만족한다고 제시하였다(김연식 외, 1996).

1) $f(x) \geq 0$
2) $\int_a^b f(x) = 1$
3) $P(\alpha \leq X \leq \beta) = \int_\alpha^\beta f(x)dx$
(단, $a \leq \alpha \leq \beta \leq b$).

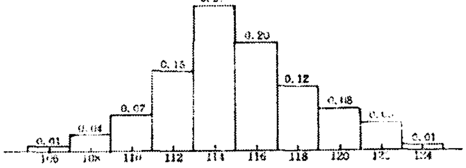
그리고, 정적분의 성질로부터

$P(X=c) = \int_c^c f(x)dx = 0$이 됨을 유도하였다.

이와 같이 6차교육과정에서는 연속확률변수의 확률계산에 적분을 사용할 수 있었으므로 확률밀도함수에 대한 성질을 공리처럼 제시하였다. 그래서 확률밀도함수의 곡선아래의 특정 구간의 넓이가 확률과 같음을 보이기 위해 정적분의 성질로부터 개념을 이끌어와서 제시하였다. 따라서, 정적분을 이해한 학생이라면 어느 정도 연속확률변수와 확률밀도함수 개념을 정적분 개념과 관련하여 기존의 인지구조에 동화시킬 수 있었을 것이다.

하지만 7차교육과정의 규정대로 적분을 사용하지 않고 직관적으로 연속확률변수를 학생들이 이해할 수 있는 방법으로 김응환 외(1996)가 제시하는 히스토그램과 확률밀도함수를 통계적 모형으로 통합하여 두 개념을 연결하여 교육하는 방법을 이용할 수 있다. 이와 비슷하게, 김우철 외(1997)도 연속확률변수에 대한 확률분포를 나타내는 방법으로 상대도수 히스토그램에서 자료의 수를 점점 늘이면서 계급의 간격을 점점 줄일 때에 모집단의 분포에 가까워지며 그 가까워지는 모양을 곡선으로 나타낸다. 그리고 이 곡선이 모집단의 분포를 나타내는 연속확률변수의 확률분포로 제시한다.

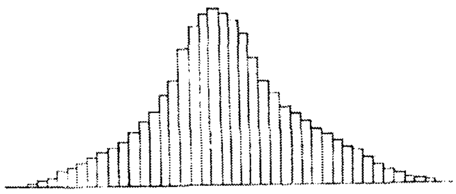
다음에 김우철 외(1997)가 제시하는 히스토그램의 점근적 방법을 그림과 함께 요약해서 제시한다.



<그림 1> 계급의 폭이 2인 상대도수 히스토그램

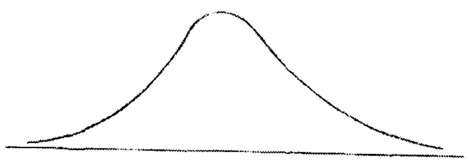
100개의 자료에 대한 상대도수 히스토그램을 다음 그림1과 같이 제시하였다.

이제 같은 특성에 대한 1000개의 자료에 대하여 계급의 폭을 좀 더 세분화한 상대도수 히스토그램을 그려서 다음 그림2와 같이 나타내었다.



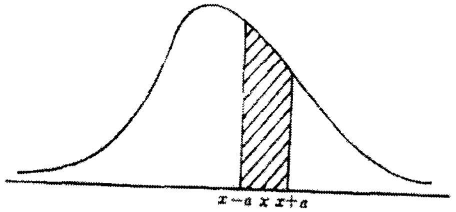
<그림 2> 계급의 폭이 0.5인 상대도수 히스토그램

이와 같이 자료의 수를 점점 늘이면서 계급간격을 점점 줄이면 더욱 더 모집단의 분포에 가까워지며 이는 다음 그림3과 같은 부드러운 곡선에 가까워지게 된다. 이 곡선을 연속확률변수의 확률분포로 생각할 수 있다.



<그림 3> 상대도수 히스토그램의 극한

연속확률변수 X가 x주위로 x-a와 x+a사이의 값을 취할 확률은 그림 4의 빗금친 부분의 넓이와 같다. 이는 연속확률분포가 상대도수 히스토그램의 극한적 개념이기 때문이다(pp. 112-113).



<그림 4> P[x-a ≤ X ≤ x+a]

김응환 외(1996)도 거의 같은 방식으로 연속확률변수를 점근적인 상대도수 히스토그램의 의미로서 제시하여 지도하는 접근방법을 사용한다.

연속형 통계자료에 대한 히스토그램을 그리기 위해 각 자료값의 도수를 세는 것이 의미가 없으므로 유사한 자료값들을 집단화하여 범주형 자료의 형태로 변환시킨다. 이때 자료는 몇 개의 계급으로 나누어져 그 계급의 도수 또는 상대도수는 계급경계에서 연결되어진 기둥의 넓이와 비례한다. 연속확률변수가 처음 나오는 고등학교 수학I이 10단계까지의 수학을 이수한 후 배우는 과목이다. 따라서 히스토그램에서 직사각형의 넓이는 각 계급의 도수에 비례하며, 각 직사각형의 윗변의 중점을 선분으로 이은 것이 도수분포다각형이 됨은 7차교육과정의 7-나 단계에 알게되므로 김우철 외(1977)의 설명이 이해하기 힘들지 않을 것이다.

이런 방법은 적분기호를 사용하지 않으면서, 시각적인 효과가 높기 때문에 학생들에게 연속확률변수의 확률밀도함수를 직관적으로 이해하도록 하고, 연속확률변수를 설명하는데 도입될 수 있다. 그런데, 김우철 외(1997)의 설명 중에 '자료의 수를 점점 늘이면서 계급간격을 점점 줄이면 더욱 더 모집단의 분포에 가까워지며...'는 그림1과 그림2를 비교하면서 보면 직관적으로 그렇게 될 것 같지만 더 정밀한 설명이 필요하다.

그림2에서 세로축의 높이에 대한 수치와 각 계급의 직사각형 높이를 결정하는 수식이 생략되어 있다. 그래서, '자료의 수를 점점 늘이면서 계급간격을 점점 줄이면...'대로 히스토그램을 그릴 때에,. 자료수가 많더라도 계급간격을 줄여서 계급수가 많아지면 각 계급의 상대도수는 작아질 수 밖에 없다. 세로축을 상대도수의 수치를 표시하는 것으로 보고 히스토그램

을 그린다면 다음과 같은 상대도수 히스토그램의 성질 때문에 상대도수 히스토그램가 접근적으로 확률밀도함수가 접근할 수 있다는 관점에 오류가 발생한다.

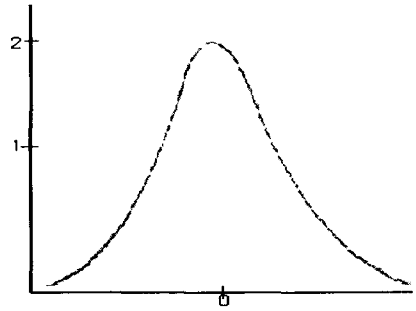
첫째, 확률밀도함수는 1보다 클 수 있지만 세로눈금이 상대도수일 때에 상대도수 히스토그램의 기둥높이는 1보다 클 수 없다.

상대도수 히스토그램을 이용하면 그림2의 상대도수 히스토그램에서 자료수와 계급수를 늘이면서 분포다각형을 그리면 어떤 곡선으로 접근하는 것을 설명하기가 쉽다. 그리고, 그림4와 같이 곡선 아래의 넓이가 그 부분의 확률과 같음을 설명하기가 용이하다. 그렇지만 확률밀도함수값이 1이 넘을 때는 이 확률밀도함수가 나타내는 곡선에 접근하는 상대도수 히스토그램을 생각할 수 없음이 문제가 된다. 예를 들면, 평균이 0이고 표준편차가 $\frac{1}{2\sqrt{2\pi}}$인 정규분포의 확률밀도함수는 다음과 같이 나타낸다.

$f(x) = 2e^{-4\pi x^2}, \quad -\infty < x < \infty.$

따라서, 그림 5처럼 $f(0)=2$가 되는데, 이를 모집단으로 하고 표본을 n개 뽑아서 세로눈금이 상대도수인 상대도수 히스토그램에서 0을 포함한 계급의 기둥이 1을 넘을 수 없다. 그러므로, 이런 확률밀도함수의 곡선이 존재하지만 이 곡선에 어떻게 상대도수 히스토그램으로 접근할 수 있는지를 보여줄 수는 없다. 확률밀도함수 $f(x)$는 그에 해당하는 연속확률변수가 실수값 x를 취할 확률을 의미하는 것이 아니라 연속확률변수의 값이 어떤 실수값 x의 전후에서 순간적으로 변화할 때 확률의 변화 정도를 나타내기 때문에 이런 문제점이 발생한다.

둘째, 상대도수 히스토그램에서 계급간격 및 계급 수가 전체적 모양에 영향을 주므로 같은 자료에 대해서 다른 모양의 상대도수 히스토그램이 그려질 수 있다.

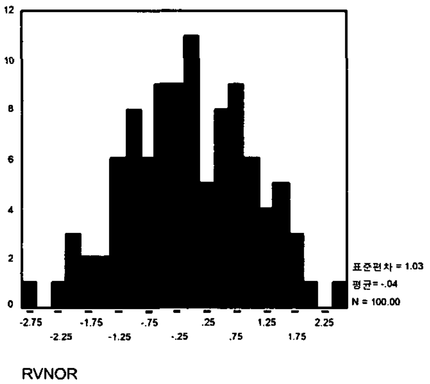


<그림 5> f(0)>1인 확률밀도함수 곡선

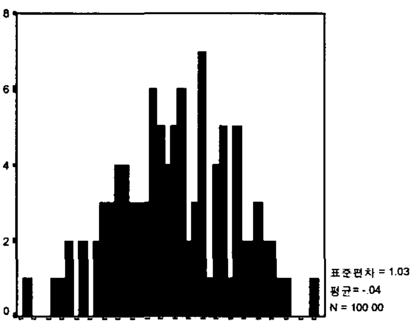
현대통계기법 중의 하나인 탐색적 자료분석은 주어진 자료에 대하여 드러나지 않는 정보를 다각도로 접근하여 파악하는 현시성(revelation)을 중요시한다(허명회, 1993). 그에 대한 구체적 방법중의 하나가 히스토그램의 계급간격을 달리하면서 나타나는 분포의 모양을 통해 숨겨진 정보를 알아내는 것이다. 7차 교육과정의 7-나 단계에서 학습지도상의 유의점으로 '가능하면 구간의 크기를 달리 하는 히스토그램을 그리고 분포의 경향을 관찰한다'를 제시하였는데 이는 탐색적 자료분석의 영향을 받은 것으로 보인다.

그림6과 그림7은 SPSS에서 표준정규분포를 따르는 자료 100개를 난수발생시키고, 이에 대해 계급간격을 달리하여 도수 히스토그램으로 나타낸 것이다. 그림6은 계급구간이 22개로 하여 나타내었고, 그림7은 계급구간을 44개로 하여 나타내었다. 도수 히스토그램이지만 자료수가 100개이므로 세로축 눈금에 0.01을 곱하여 보면 상대도수 히스토그램과 동일하게 된다. 두 그래프의 세로축을 동일한 척도에 맞춰서 비교해 보면 그림 6은 기둥높이의 최대값이 11 정도인데 그림7에서는 최대 기둥높이가 7정도이다. 즉, 계급구간이 많아질수록 도수 히스토그램의 기둥 높이가 낮아짐을 알 수 있다. 그러므로, 동일 자료에 대해 계급구간을 많게 하

면 각 계급의 간격이 작아지고 따라서 계급에 속하는 도수가 작아지므로 상대도수도 적게 된다. 따라서, 상대도수 히스토그램의 계급구간을 크게 하면서 점근적으로 연속확률밀도곡선으로 접근한다고 하면 곡선의 전체 모양이 가로축에 비해 곡선 높이가 낮은 모양으로 보일 것이고 계급구간을 상대적으로 작게 하면서 접근하면 앞에 보다 위로 솟은 모양으로 확률밀도함수의 곡선이 연상될 것이다.

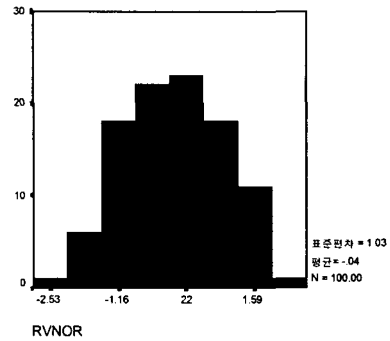


<그림 6> 계급 구간이 22개



<그림 7> 계급구간이 44개

또한, 그림6과 그림7에서 연상되는 전체 분포에 대한 곡선은 이봉(bimodal)형태를 띠는데 그림8처럼 계급구간을 8개로 하면 일봉(unimodal)형태로 다르게 보인다.



<그림 8> 계급구간이 8개

이와 같이 동일한 자료에 대해 계급구간 및 계급간격을 달리하면 다른 모양의 곡선이 연상된다. 따라서, 방법에 따라 다른 확률밀도함수의 곡선에 접근하게 될 수도 있다.

따라서 상대도수 히스토그램의 점근적 방법으로 연속확률밀도함수 곡선에 접근할 때는 세로축이 상대도수로 표시된 상대도수 히스토그램의 기둥 높이가 연속확률밀도함수의 값에 접근하는 것이 불가능하다. 세로축이 상대도수를 의미할 때 상대도수 히스토그램은 연속확률밀도함수에 접근하는 것이 아니라 연속확률밀도함수의 곡선 모양-일봉 분포냐, 이봉 분포냐, 꼬리가 오른쪽이 두꺼운가, 꼬리가 왼쪽이 두꺼운가 등-에 점진적으로 닮아간다고만 설명해야 할 것이다.

따라서 김우철 외(1997)의 설명에서 '자료의 수를 점점 늘이면서 계급간격을 점점 줄이면 더욱 더 모집단의 분포에 가까워지며...'는 다음과 같은 방식으로 더 명확하게 되어야 할 것이다.

유동선(2000)은 확률밀도함수에 대한 설명을 다음과 같이 하고 있다.

확률밀도함수 $f(x)$는 x의 주변에서 확률의 밀도를 나타내는 것이지 확률을 나타내는 것이 아니며, 작은 구간 Δx 안에 들 확률은 근

사적으로 $\Delta x f(x)$가 된다. 따라서 전구간을 세분하고, 각 세분된 구간의 중심을 차례로 $x_0, x_1, x_2, \cdots$이라 하고 세분된 구간의 폭을 $\Delta x = C$라고 하고, 각 세분된 계급의 도수를 $f_0, f_1, f_2, \cdots$이라면 $f(x_i) \fallingdotseq (f_i/N)/C$ 이므로

$$\sum_i \Delta x \cdot \frac{f_i/N}{C} = \frac{1}{CN}\sum_i \Delta x f_i = 1.$$

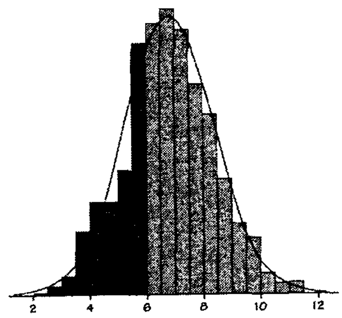
여기서 $\sum_i \Delta x f_i$는 전 히스토그램의 면적이므로 CN이다(C: 계급의 폭, N: 총도수). 따라서 확률밀도함수를 정하려면, 상대도수 히스토그램의 세로눈금을 계급의 폭으로 나눈 값으로 변환하고, 계급의 폭을 세분하고, 도수를 늘려서 도수다각형을 그리면, 그의 극한은 분포곡선이 되고, 확률밀도함수가 정해진다(pp. 247 - 248).

유동선(2000)의 설명과 김우철 외(1997)의 설명의 차이는 위의 '상대도수 히스토그램의 세로눈금을 계급의 폭으로 나눈 값으로 변환...'의 유무에 있다. 세로눈금을 (상대도수/계급폭)으로 한 것은 전체 면적이 1이 되도록 히스토그램을 상대도수 밀도로 재척도화 한 것이며 이런 재척도화가 안되면 전체 높이가 1인 것에만 초점이 맞춰서 확률밀도함수에 점근적으로 접근할 수 없다(Wonnacott etc, 1982). 상대도수 히스토그램의 세로눈금이 상대도수가 아니라 (상대도수/계급폭)이라고 할 때에 계급폭이 줄어들게 되면 상대도수도 줄어들지만 상대도수와 계급폭의 비율은, 자료수가 증가할 때에 모집단과 유사한 비율로 접근하게 된다. 그렇게 세로눈금을 정의할 때에 김우철 외(1997)의 '자료의 수를 점점 늘이면서 계급간격을 점점 줄이면 더욱 더 모집단의 분포에 가까워지며...'란 설명이 올바르게 된다.

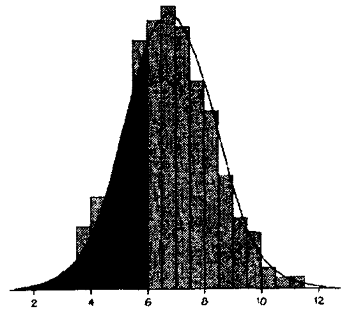
(2) 밀도곡선을 이용한 연속확률변수 개념 도입

Moore & McCabe(2000)는 밀도곡선(density curve)을 이용하여 상대도수 히스토그램과 확률밀도함수의 곡선 개념을 연결한다.

밀도곡선은 수평축상 또는 그 위에 위치하며 그 곡선과 수평축이 이루는 면적이 1이라고 정의하고 분포의 대략적인 모양을 나타낸다. 그림9에서 6이하의 상대도수 히스토그램의 합은 0.303이다. 그림10에서 밀도곡선의 6이하의 넓이는 0.293이다. 따라서 이 부분에 대한 상대도수 히스토그램과 밀도곡선의 넓이 차이는 0.01로 아주 작다. Moore & McCabe(2000)는 이를 이용하여 수평축의 어떤 구간에 대한 확률은 그 구간상의 상대도수 히스토그램의 넓이를 합한 것과 같듯이 그 부분의 밀도곡선 아래쪽의 넓이로서 연속적인 확률변수의 넓이를 나타내



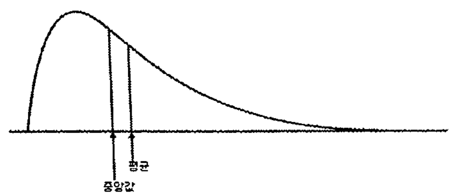
<그림 9> 히스토그램과 밀도곡선
(Moore & McCabe, 2000, p. 66)



<그림 10> 6이하 부분의
히스토그램의 상대도수와
밀도곡선의 넓이비교
(Moore & McCabe, 2000, p. 67)

는 것으로 볼 수 있다고 본다.

또한, 이런 밀도곡선 중에서 대칭이고, 일봉이며, 종 모양을 정규곡선이라 하고, 정규분포에 대한 모양을 나타낸다고 보았다(Moore & McCabe, 2000).



<그림 11> 비대칭분포에서 중앙값과 평균의 위치

<그림 12> 무게중심으로서 평균

그리고, 이 밀도곡선을 이용하여 시각적으로 중앙값은 수평축에서 그림 11와 같이 양쪽으로 같은 넓이를 갖는 곳을 의미하며, 그림 12에서 평균은 무게 중심을 나타낸다고 한다.

밀도곡선은 연속확률변수가 따르는 모집단분포를 나타내며 상대도수 히스토그램은 그 모집단에서 얻은 표본에 대한 분포를 나타낸다. 상대도수 히스토그램의 점근적 방법이 자료수 및 계급구간을 늘리면서 동적으로 확률밀도함수에 접근하는 것으로 설명하는 것에 비해 밀도곡선에 의한 방법은 정적인 접근이라고 할 수 있다.

그런데, 밀도곡선은 정지된 히스토그램에서 마치 정적분에서 x축과 함수 곡선 아래가 이루는 면적과 그에 대하여 잘게 쪼갠 한가지 분할(partition) 집합 상에 구성된 직사각형과의 면적이 거의 차이가 없음을 보이듯이 계산하여 히스토그램의 특정 구간의 넓이가 밀도곡선의 그 구간의 넓이와 같음을 보인다. 따라서, 상대도수 히스토그램의 점근적 접근 방법에 비해 구성이 단순하다. 그렇지만 부드러운 밀도곡선과 상대도수 히스토그램이 면적이 비슷하다고 그 특성도 같다고 보아야 할 지에 대한 의문이 남는다. 즉, 모집단을 의미하는 밀도곡선의 성질과 표본의 분포를 의미하는 상대도수 히스토그램의 성질 사이에 간격이 존재한다.

(3) 적분없이 $P(X=c)=0$이 되는 이유를 설명하기

$P(X=c)=0$은 확률밀도함수에 대한 정적분의 성질에서 바로 유도되지만, 적분기호를 사용하지 않는다고 하면 이에 대해 학생들이 쉽게 이해하기 어렵다. 이산확률변수에서 확률이 0임은 그 사실이 절대로 일어나지 않음을 의미한다. 그런데 시간이나, 키, 몸무게 등의 연속량에서 연속확률변수가 어떤 한 가지 값을 반드시 나타내게 된다. 예를 들어 어떤 아기의 몸무게가 3.5Kg인 확률은 0이다. 그런데 이 3.5Kg의 몸무게를 가진 아기에 대한 예를 찾을 수 있다. 그리고, 그런 식으로 따진다면 아기들은 어떤 몸무게도 가질 수 없다는 식으로 얘기된다. 이렇게, 이산확률변수에서 성립하던 확률 0에 대한 개념이 연속확률변수 개념에 와서는 혼란을 일으키게 된다.

이와 같이 생각하면 X가 연속확률변수일 때 $P(X=c)=0$이 되는 것은 적분을 사용하여 설명한다고 하더라도 쉽게 이해하기 어렵다.

이에 대해 R. Johnson과 G. Bhatacharyya의 'Statistics: Principles and Method'를 참고하여 구자흥 외(1992)는 다음과 같이 설명한다.

어느 신생아의 몸무게를 X라 하자. 이때 X가 3.5Kg이 될 확률이 0 이라고 하면 놀라운 일일 것이다. 왜냐하면 우리는 주위에서 X가 3.5Kg의 신생아의 예를 많이 보아 왔기 때문이다. 그러나 X가 3.35468965Kg이 될 확률이 0 이라는 사실은 당연하게 여길 텐데, 그것은 측

정기기의 불완전성으로 그와 같은 몸무게의 신생아의 예를 본 적이 없기 때문이다. 이와 같이 3.5Kg과 근접한 즉 3.495Kg부터 3.505Kg정도의 무게는 그 구분이 불가능하기 때문에 보통 3.5Kg으로 여겨지게 되는 것이다. 그러므로 3.5Kg이 될 확률이란 실제로는 X가 구간 (3.495, 3.505)에 속할 확률이 되므로 그 값이 0이 아니다(pp. 106-107).

구자홍 외(1992)의 설명은 측정기기의 정밀도의 한계로 모든 연속량의 값은 어떤 구간에 속하는 것으로 보고 따라서 그 확률값은 0이 아니라고 한다. 이런 방식으로 설명하면 실제 우리가 관측하는 연속량의 값은 오차범위 내에서 어떤 구간에 속하는 것으로 보는 것이고 $X=c$가 아니라 아주 작은 $\varepsilon(>0)$에 대하여 $c-\varepsilon<X<c+\varepsilon$인 결과를 본 것이 된다. 즉, 연속량은 아무리 많이 쪼개어도 연속량이라는 의미이다. 또한 시간의 멈춤이나 키의 측정 등에서 그 순간이나 경계를 한가지 실수값으로 정확히 결정할 수 없음을 의미하며 우리가 지각할 수 없는 아주 작은 크기의 구간에 속하는 것으로만 보는 것이다.

다른 관점으로 김효석(1988)은 '예를 들어 신장이 175cm라고 할 때 정확한 값은 175.000 1020…와 같이 소숫점 이하 자리수를 무한하게 생각할 수 있는데 여기에 해당하는 사람은 거의 없을 것이다'로 연속확률변수가 어떤 특정 값을 가질 확률이 0인 이유를 설명한다. 이는 마치 무한자리를 가진 복권의 1등 당첨번호를 맞출 확률을 0으로 보는 것과 같다.

이와 관련하여 더 간단한 설명으로 강금식(1999)은 '연속확률변수는 취할 수 있는 실수의 수가 무한히 많은데 이중 어떤 한 실수를 취할 확률은 0에 가깝다....예를 들면 체중이 정확하게 60.2kg이 될 확률이라든가 온도가 정확하게 30℃가 될 확률은 0이다'라고 설명한다. 그런데 이런 설명은 연속이 아닌 이산분포중에도 포아송분포는 취할 수 있는 값이 무한이지만 각 값에 대한 확률은 0이 아닌 사실로 볼 때 적절하지 못하다.

이렇게 X가 연속확률변수일 때 적분개념 없이 $P(X=c)=0$이 되는 것의 논의는 수학 외적인 논의가 된다. 사실 적분을 이용하여도 $P(X=c)=0$이 됨을 수학적인 설명을 제공한 것뿐이지 연속과 무한의 관련성에 관한 그 근본을 해결한 것은 아니다. Zeno의 역설도 실수체의 완비성 공리로 설명되지만 그것이 수학적인 설명만 제공한 것이지 그 역설의 근본적인 문제를 해결한 것은 아닌 것으로 본다. 그것은 수학으로 대답될 수 없는 문제이며 실무한의 존재성과 관련된 철학의 문제로 본다. 하지만 철학도 이 문제에 명확한 해답을 주지는 못한다고 한다(박선화, 1998). 그렇지만 이러한 논의들을 통해 수학의 대상이 되는 연속, 무한의 개념에 대한 학생들의 이해를 수학 외적으로 넓히는 계기가 될 수 있을 것이다.

3. 7차교육과정의 수학I 교과서들의 연속확률변수 설명 분석

7차교육과정에 따른 수학I 교과서들에서 연속확률변수를 설명하는 방법은 크게 두 가지로 분류할 수 있다.

첫째는 회전판에서 바늘이 특정 구간에 정지할 때 바늘 끝이 가리키는 눈금을 X라 하여 특정 구간에 속할 확률이 호의 길이에 비례함을 직관적으로 인식하도록 한다. 그래서, 균등분포(uniform distribution) 상황에서 특정구간에 속할 확률이 직선아래의 넓이와 같다고 설명한다. 이를 확장하여 X가 특정 구간에 속할 확률은 곡선아래의 넓이와 같다고 약속한다. 세 가지 교과서가 이런 방식으로 연속확률변수를

설명하는데 이런 설명에서는 균등분포와 같은 아주 단순한 상황에서 성립한 사실이 비균등한 분포에서도 성립한다고 유도하는 데에 비약이 있다. 이에 대한 보완은 교사의 몫으로 남겨지게 된다.

그리고, 이런 방식의 설명에서 $X=c$의 확률은 c점에 해당하는 호의 길이가 0이므로 $P(X=c)=0$이 된다고 본다. 하지만 한 교과서에서는 '사람의 키가 175cm인 확률은 이 값이 반올림한 값이므로 그 확률은 0이 아니고 $P(174.5 \leq X < 175.5)$를 의미한다'고 참고란에서 설명하였다. 이는 구자홍 외(1992)의 설명과 비슷한 관점이다.

둘째는 김우철 외(1997)와 유동선(2000)의 설명처럼 상대도수 히스토그램의 점근적 접근으로 확률밀도함수를 설명하여 연속확률변수가 특정구간에 속할 확률이 그 구간의 곡선아래의 면적과 같다고 설명한다. 다섯 종류의 교과서가 여기에 속하는데 이중 네 교과서는 자료수를 늘리고 계급나비를 작게 하면 히스토그램이나 그 위에 그려진 도수분포다각형이 곡선에 가까워 질 것으로 설명한다. 그런데 이 중 한 교과서는 첨부된 그림의 세로눈금이 상대도수라고 되어 있으며, 한 교과서는 세로축에 설명없이 y라고만 표시되어 있다. 나머지 세 교과서는 세로축의 눈금에 대한 설명이 아예 생략되어 있다. 세로축의 눈금이 상대도수를 계급의 크기로 나눈 값이라는 표시를 명확히 해야 할 것이다. 그리고, 일부 교과서 중에서 그림에서 히스토그램만 그려서 확률밀도함수 곡선에 가까워진다고 설명하고 있는데 엄밀하게 얘기하면 도수분포다각형을 그 위에 그려서 그 도수분포다각형과 히스토그램의 면적차이가 없음을 보이고, 도수분포다각형이 확률밀도함수 곡선에 가까워진다고 설명해야 더 옳을 것이다. 이에 대한 보완은 결국 교사의 몫으로 남겨진다.

하지만, 한 교과서는 설명하는 표에서 (상대도수/계급의 크기)를 계산하였고, 히스토그램의 세로눈금을 (상대도수/계급의 크기)로 표시하여 전체 면적이 1이 되도록 히스토그램을 상대도수 밀도로 재척도화 한 것임을 명확하게 표시하였다. 하지만 이 교과서도 도수분포다각형은 표시하지 않았다.

4. 수업 적용 및 결과 분석

청주의 A고등학교 2학년 한 반 32명을 대상으로 확률밀도함수 개념에 대한 수업을 진행하였다. 이 학생들은 이산확률변수에 대한 평균 및 표준편차까지 배운 상태였다. 이 학생들에게 교과서에 제시된 방법인 바늘을 회전시키다 멈추는 상황을 I방법, 상대도수 히스토그램의 점근적 접근으로 설명하는데 세로축이 무엇인지는 설명하지 않는 II방법, Moore & McCabe (2000)가 제시한 밀도곡선에 의한 III방법, 히스토그램의 세로눈금을 (상대도수/계급의 크기)로 표시하여 II방법을 수정한 IV방법 네 가지에 대하여 각 방법을 적용한 수업자료를 준비하였다. I방법, II방법, IV방법은 해당되는 교과서 중에서 연속확률변수 단원에 관한 수업 자료를 준비했고, 교과서에 이용되지 않은 III방법은 Moore & McCabe(2000)의 교재를 번역한 수업자료를 준비했다. 각 방법들에 각각 비슷한 시간을 할애하여 전체 2시간에 걸친 수업을 하였다.

수업 후에 네 가지 방법 중에서 연속확률변수의 확률계산을 이해하는 데에 가장 도움이 된 방법을 선택하도록 하였다. 무응답을 제외한 25명의 응답자 중에서 III방법이 9명으로 가장 많았고, II방법이 8명, I방법과 IV방법이 각각 4명이었다. 한 학생은 'III방법이 가장 쉬운데, IV방법은 심화로, II방법은 일반으로 했으면

좋겠다'는 의견을 적었다. 이런 결과에서 밀도 곡선을 이용한 III방법이 단순하여 이해하기 쉬웠던 것 같다. 그리고, 상대도수 히스토그램의 점근적 접근 방식인 II방법이 학생들의 흥미를 일으키고 직관적인 이해에 도움을 준 것 같다. 하지만 이에 대하여 올바로 수정한 방법인 IV 방법은 히스토그램의 세로눈금을 (상대도수/계급의 크기)로 하여야만 하는 이해가 더 필요하게 되어 II방법보다 학생들이 직관적으로 이해하기 어려운 것 같다.

X가 연속확률변수일 때 적분개념 없이 $P(X=c)=0$이 되는 이유에 대해 생각나는 대로 적도록 하였는데 위의 네 가지를 배운 직후라서 인지 무응답이나 전혀 맞지 않는 응답을 한 학생이 13명이 되었다. 이는 설문 응답자 25명의 거의 50%에 해당한다. 응답 중에서 '점차 좁아져 나중엔 밑변이 0이 되므로', '범위를 점점 세분화하면 그 면적이 0이 되므로', 는 II나 IV의 방법에 영향을 받은 것으로 보인다. 그리고, '면적을 나타내지 못해서', '넓이가 나타나지 않는다'는 응답은 I과 III의 영향을 받은 것으로 보인다. 그리고, '1/∞=0이라서'라는 응답은 연속확률변수가 무한개의 연속량을 다루는 것으로 본데서 나온 응답이고 이는 강금식(1999)의 관점과 같다. 이 응답을 한 학생 들은 모두 II방법이 가장 이해가 잘 된다고 하였다. 그리고, 한 학생이 '실제로 바늘이 c를 가리킬 때에 바늘 두께와 c라는 눈금과 일치한다는 기준을 세밀하게 계속 나눈다면 거의 없으므로' 라고 응답한 것은 김효석(1988)의 설명 방식과 거의 비슷하다고 볼 수 있다.

III. 결론

이산확률변수에 대해서는 막대그래프로 쉽게 분포를 나타낼 수 있으며, 특정 사건의 확률은 해당되는 막대그래프들의 높이를 합하여 구할 수 있다. 그렇지만, 연속확률변수에 대한 확률 밀도함수에서 왜 곡선의 넓이가 확률이 되는지를 학생들이 이해하기가 어렵다.

7차 교육과정에서는 연속확률변수를 적분을 사용하지 않고 직관적으로 도입하도록 하였다. 따라서 학생들이 적분없이 직관적으로 연속확률변수 및 그에 따른 확률밀도함수의 개념과 그 함수 그래프에서 특정 구간의 넓이와 그 구간의 확률이 같음에 대하여 학생들이 직관적으로 이해할 수 있도록 교과서 구성이 되어야 한다.

이에 대한 방법으로 상대도수 히스토그램으로 점근적 접근 방법이 기존에 제시되었는데, 애매하게 진술되어 있으면 오류를 일으킬 수도 있으므로 이를 명확하게 설명할 필요성이 있음을 본 연구에서 지적하였다. 그리고, Moore & McCabe(2000)가 제시한 밀도곡선의 장단점을 분석하였다. 그리고, 연속확률변수에서 한 값에 대한 확률이 0이 되는 것과 실제 현상과의 괴리를 직관적인 의미로 설명할 수 있는 방법을 찾아보았다.

이 접근방법을 이용한 7차교육과정의 수학I 교과서를 분석해 보면 상대도수 히스토그램의 접근 방법을 사용하는 교과서들에서 정확한 설명이 결여되어 있음을 알 수 있었다 이에 대한 보완이 필요하다고 생각한다. 그리고, 원판 위에 바늘이 돌다가 멈추는 상황을 사용한 교과서들은 균등분포라는 아주 단순한 상황에서 바로 비균등하면서 확률밀도함수가 곡선이 되는 상황에 적용할 수 있다고 유도한다. 이는 학생들이 다양한 상황 속에서 그 의미를 충분히 음미하도록 하지 못하고 교과서의 내용만을 추종하도록 하는 결과를 낳게 할 수 있다.

연속확률밀도함수 곡선을 이용하여 연속확률

변수의 확률계산을 하는 방법에 대한 이해를 알기 위해 네 가지 방식으로 현장에 소규모 적용을 해 본 결과, 밀도곡선에 의한 방법이 가장 이해가 잘 된다고 조사되었고, 상대도수 히스토그램의 점근적 접근이 그 다음이었다.

추후에 이와 같은 접근 방법들을 더 많은 조사대상에 대해 적용하여 보고 그 효과를 보다 깊이 분석할 필요가 있다. 그래서 연속확률변수 및 확률밀도함수 개념에 대해 학생들에게 직관적으로 이해되기 쉬우면서 논리적 비약이 비교적 없는 방법을 찾기 위한 연구가 지속되어야 할 것이다.

참고문헌

강금식(1999). EXCEL활용 현대통계학. 서울: 박영사.

구자흥 외 (1992). 통계학-원리와 방법-. 서울: 자유아카데미.

김연식 · 김홍기(1996). 고등학교 수학 I. 서울: (주) 두산.

김우철 외 (1997). 통계학개론. 서울: 영지문화사.

김응환 · 김승동 · 오후진(1996). 확률밀도함수의 지도를 위한 고등학교 교과서내용의 재구성. 수학교육, 35(2), 117-123.

김효석(1988). 상경계열을 위한 통계학. 서울: 형설출판사.

박규홍 외 (2002). 고등학교 수학I. 서울: 교학사.

박선화(1998). 수학적 극한개념의 이해에 관한 연구. 서울대학교 대학원 박사학위논문.

우정호 외 (2002). 고등학교 수학I. 서울: (주)대한교과서.

유동선(2000). 통계학해법 대사전. 서울: 교우사.

이강섭 외 (2002). 고등학교 수학I. 서울: 지학사.

임재훈 외 (2002). 고등학교 수학I. 서울: (주) 두산.

조태근 외 (2002). 고등학교 수학I. 서울: (주)금성출판사.

최봉대 외 (2002). 고등학교 수학I. 서울: (주) 중앙교육진흥연구소.

최상기 외 (2002). 고등학교 수학I. 서울: (주) 고려출판.

최용준 · 신현성(2002). 고등학교 수학I. 서울: (주) 천재교육.

허명회(1993). 탐색적 방법에 의한 통계자료분석론. 서울: 자유아카데미.

Moore & McCabe (2000). *Introduction to the practice of statistics*. New York: W. H. Freeman and Company.

Ronald J. Wonnacott & Thomas H. Wonnacott (1993). 현대사회통계학. 차종천 역. 서울: 나남출판. (영어원작은 1982년에 출판)

A Study on the Intuitive Understanding Concept of Continuous Random Variable

Park, Young Hee (Cheongju National University of Education)

The context and intuitive understanding is very important in Statistics Education. Especially, there is a need to mitigate student's difficulty in studying probability density function. One of teaching method this concept is to using relative frequency histogram. But, as using this method, we should know several problems included in that. This study investigate problems in the method for teaching probability density function as gradual meaning of histogram. Also, as alternative approach, this thesis introduce the density curve concept. The application of four methods to teach the concept of the probability density function and analysis of the survey result is done in this research.

key words: continuous random variable, probability density function, histogram, density curve, statistics education

e-mail: yhpark@cje.ac.kr