Historia Mathematica Vol. 13, No. 1, 2000, 1-14

피어슨 곡선족에서 온 표본분포들에 관한 소고

인하대학교 수학통계학부 **구자흥**
항공대학교 명예교수 **유동선**

Abstract

The first part of this thesis discusses the Pearson's Curve Family which gives β -distribution, Γ-distribution, χ^2-distribution and t-distribution.

The second part of this thesis gives some brief process of calculations for normal distribution density and t-distribution density by the 7-th type Curve of Pearson's Curve Family.

Finally, a conclusion arrives that Student(Gosset) could not find out his famous 'Student's t-distribution' without his attending of 'Pearson's Differential Equation' class taught by Pearson himself when he was a senior student.

However, if he had got a professorship at the Pearson Statistics Laboratory, the University of London, then he could not have found 'Student's t-distribution' for small sampling technique of modern statistics.

0. 서론

현대 물리학의 이론적 발전이 고전 역학(뉴턴 물리학)의 이론적 토대 위에서 직 · 간접적으로 의존하면서 조성되었듯이, 현대 통계학의 소표본분포론(small sampling distribution)의 유도에는 대표본론의 완성에 지대한 공적을 남긴 피어슨(Karl Pearson, 1857-1936)의 '피어슨 미분방정식'(Pearson's differential equation)에서 오래 전에 시작하였다.

즉 현대 통계학의 방법론 중 소표본론의 토대를 이루는 스튜던트(Student)의 t-分布는 물론 Γ-분포, β-분포, χ^2-분포, F-분포 등의 확률 밀도 함수(p.d.f.)들은 피어슨 미분 방정식의 해곡선 함수(function of the solution)들로서 이미 제시되었다는 사실이다.

뿐만 아니라, 고셋(William Sealy. Gosset, 1876-1937)은 일찍이 런던 대학교(University of London) 대학에서 피어슨 교수로부터 피어슨 미분 방정식을 수학하였으며, 그 이론적 배경이 없었다면 고셋이 경험적으로 그의 t-분포를 꾸미고 찾아내는데 확신이 서지 않았을 것이라 전해진다.

그러나 아이로니컬(ironical)하게도 고셋이 당시 통계학을 대표하던 런던 대학교의 피어슨 연구실에 오래 안주하면서 강의와 연구를 하였다면 당시 대표본 이론을 바탕으로 하는 고전 통계학에 심취되어 t-분포의 창작은 더욱 늦어졌을 것이며, 특히 기네스(Guinness) 맥주회사 양조 화학기사로 일하면서 소맥(小麥)의 재배와 양조공정에서 수집된 생생한 원자료(raw data)가 스튜던트의 t-분포의 개발에 촉진제가 되었다는 사실이다.

본 연구에서는 피어슨 미분 방정식의 조성과 그 해법을 추적함으로써 소표본론(표본분포)의 이론적 배경의 발전을 재조명하고자 하는 데 그 목적이 있다.

1. 소표본론과 현대통계학의 발전

1-0 통계학 발전 단계의 시대적 구분

휜베르크(Fienberg)의 논문 "Brief History of Statistics"에 의하면 1654년과 1750년 사이의 약 100년간을 '확률론과 통계학의 태동기'로서 당시 대표자로는 영국의 정치산술파의 그라운트(Graunt)와 프랑스의 고전확률론의 페르마(Fermat)와 파스칼(Pascal)이 이들의 공적과 함께 거명하고 있다.

다음 1750년부터 1820년까지의 70년간을 '해석적 확률론의 시대'로 고전확률론의 완성에 공적이 큰 라플라스(Laplace)와 가우스(Gauss) 등과 그들의 연구업적을 거명하였다.

그리고 그 뒤 80년간(1820-1900)은 기술통계학(Descriptive Statistics)의 체계화에 공적이 큰 피어슨이 대표하던 시기이다.

그리고 1900년에서 1950년까지 약 50년간은 확률론의 본질을 측도론적 바탕 위에 도입함으로써 현대 확률론(modern probability theory)을 확립한 러시아의 응용수학자 콜모고로프(A.N. Kolmogorov)와 측정과 검정이론 등 소표본론적 기법(small sampling technique)의 이론적 골격을 건설하는데 시조(始祖)격이 된 고셋의 '스튜던트의 t-분포'의 발견 도입과 분산분석, 실험계획법에 이르기까지 불후의 공적을 남겨 현대통계학의 비조(鼻祖)가 된 피셔(R.A. Fisher)가 혁혁한 업적을 발휘한 시기로 구분된다.

1-1 고셋과 소표본론

현대통계학은 고셋에 의한 소위 스튜던트의 t-분포의 발상에서 시작되었다.

그 이전까지는 영국 런던 대학교의 피어슨이 이끄는, 통계학 연구실이 중심이던 기술통계학이 통계학을 대표하였으며, 대표본론적 기법(large sampling technique)이 그 주류를 이루고 있었다.

일찍이 고셋은 옥스퍼드 대학교(Oxford University) 뉴 칼리지(New College)에서 化學과

數學을 공부하였으며 1899년 기네스 맥주회사의 농예화학기사로 입사하여 봉직했으며, 주로 온도에 예민하고 시간에 따라 변화가 많은 실험과 그 결과를 다루는 연구에 몰두하였다.

그때 그가 사용한 실험결과들의 자료(data)는 그 소계열(소표본)으로서 그 당시 통계적 분석방법인 대표본론적 기법만으로는 고셋이 그 한계성을 누구보다도 먼저 부딪히게 되었다.

이것이 고셋이 스튜던트의 t-분포를 탄생시킨 계기가 되었다.

만일 그가 런던 대학교의 통계학연구실에 근무하였더라면 t-분포를 경험적으로 탄생시키지 못하였을 것이다.

그는 매번 같은 크기의 표본을 추출하여도 그때마다 다른 표본 평균값과 표본 분산값이 얻어지는 것에 대응하는 확률변수로 표본평균 $\overline{X}$와 표본분산 S^2을 착안하게 되었다.

또 2차원 정규분포에서 표본크기 n쌍인 표본 (x_i, y_i) $(i=1, 2, \cdots, n)$에 관한 다음과 같은 표본 상관계수에 대해서도 그 표본분포(r의 분포)를 구하기 위하여, 표본크기 $n=4$인 표본 745개와 표본크기 $n=8$인 표본 750개를 사용하여 r의 경험 도수분포곡선을 구하였고, 이들을 비교하여 표본상관계수 r의 분포가 표본크기 n에 의존하지 않는다는 것을 밝혀내었다.

$$R=\frac{\mathrm{cov}(x,\ y)}{S_x S_y}$$

그러나 고셋의 현대통계학에 대한 가장 큰 업적은 t-분포에 관한 아이디어의 창출이다.

그리고 고셋이 t-분포를 찾아내게 된 배경에는 피어슨으로부터 런던 대학교에서 배운 피어슨 분포 곡선족이 큰 도움이 되었다는 사실을 그냥 넘겨서는 안 된다.

1-2 피어슨과 피어슨 분포족

피어슨은 1866년에 런던 대학교 칼리지 스쿨(College School)에 입학하였고, 1875년에는 케임브리지 대학교(Cambridge University) 킹스 칼리지(King's College)에서 저명한 수학교수들인 스토크스(Stokes), 케일리(Cayley), 맥스웰(Maxwell) 등에게서 수학을 배웠고, 베를린(Berlin)에서 몸젠(T. Mommsen)에게 Rome법을, 피셔(K.Fisher)에게 철학 그리고 퀸시(Quinche)에게서 물리학을 배웠다.

그리고 1884년에는 런던 유니버시티 칼리지(University College)의 응용수학과 역학 교수가 되었다.

그 뒤 우생학자 골턴(Galton)의 간청으로 통계학의 기초를 확립했으며 골턴의 회귀직선을 토대로 상관이론(correlation)을 완성하였다. 또 1890년에는 동물학자 웰던(Weldon)과 협력하여 '생물계측학'(Biometrics)을 시작하였다.

그의 저술로서는 다음과 같은 저명한 논문들을 *Philosophical Transaction of the Royal Society of London*에 발표하였다.

1894년 "진화론의 수학적 이론"(Contribution to Math'l Theory of Evolution),
1985년 "고른 자료의 기운 변동"(Skew variation in homogeneous material),

1886년 "회귀, 유전 그리고 잡혼번식"(Regression, Heredity and Panmixia)

그리고 또 이들 연구를 통하여 유전법칙의 타당성을 검증하기 위하여 새로운 통계적 방법을 착안하였고, 1900년에는 자료의 타당성검정을 위한 χ^2-검정(chi-square test)을 시도하여 생물통계학에 크게 공헌하였다.

그 당시까지만 해도 도수분포는 모두 정규분포형이라고 알려져 있었으나 엄밀히 연구해본 결과 정규형이 아닌 도수분포곡선(frequence curve)도 상당수 존재한다는 사실이 밝혀졌으며 이들 도수곡선들은 다음과 같은 피어슨 미분방정식(Pearson's differential equation)의 해(solution)로 구해낼 수 있다고 주장하고, '피어슨 곡선족'(Pearson's Curve Family)을 통하여 12종류의 분포함수를 제시하였다.[1)]

$$\frac{dy}{dx} = \frac{(a_1 x + a_0)y}{px^2 + qx + r} \quad \cdots \ (1)$$

그런데 위 미분방정식의 유도방법은 다음 두 가지 방법이 있다.

『제 1 방법』 경험적으로 볼 때 통계자료가 동질성(homogeneity)을 지니기만 한다면 그 도수분포 곡선은 일반적으로 다음과 같은 가정 하에서 유도된다.

즉 X가 주어진 특성을 나타내는 계량적 변수라 하고, X의 실현치의 함수 $y=g(x)$에 의하여 X의 밀도함수를 나타내었다.

【가정 1】 만일 $|x| \to \infty$이면 $y \to 0$이다.

【가정 2】 함수 $y=g(x)$에 대한 최빈수(mode)는 하나이다.

【가정 3】 모든 x에 대하여 $g(x) \geq 0$이다.

【가정 4】 $y=g(x)$는 x축에 고도의 접촉을 한다.

즉 위의 가정들을 만족하는 밀도곡선 $y=g(x)$는 식 (1)을 만족한다.

『제 2 방법』 (초기하분포에서의 유도)

밀폐된 상자 속에 흰 공이 k개, 검은 공이 h개 잘 섞여 있다 하자.

이제 비복원 추출법으로 m개의 공을 추출할 때, 그 중 흰 공이 x개 들어 있을 확률을 $P(x)$라 하면, 초기하 분포의 밀도함수에 의하여 다음과 같이 쓸 수 있다.

$$P(X) = \frac{\binom{k}{X}\binom{h}{m-X}}{\binom{h+k}{m}}, \quad 0 \leq X \leq m \leq h+k \quad \cdots \ (2)$$

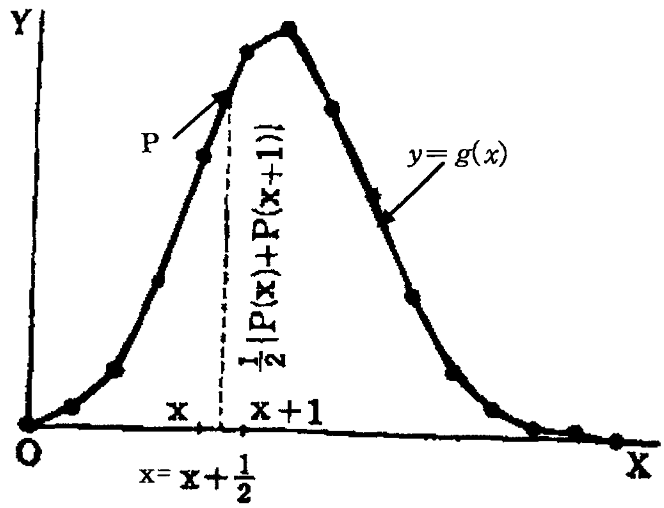
그러면 $P(x)$에 대한 다음 정차방정식(difference equation)이 성립한다.

1) 정한영 저, **통계학사**, 한림대학교 출판부, 1995, pp. 204-208.

$$P(X+1)=\frac{(k-X)(m-X)}{(X+1)(h-m+1+X)}P(X) \quad \cdots (3)$$

식 (2)의 절선도에 대하여 다음과 같은 하나의 부드러운 곡선을 상정하기로 한다.

$$y=g(x) \quad \cdots (4)$$



<도 1-1> $y=g(x)$: 접점 P를 지나는 부드러운 곡선

그러면 식 (4)로 주어진 곡선상의 점

$$x=X+\frac{1}{2} \ ; \ y=g(x)=\frac{1}{2}[P(X)+P(X+1)] \quad \cdots (5)$$

을 좌표로 가지는 점 P를 통과하는 곡선상의 기울기는 다음과 같다.

$$\frac{dy}{dx}=P(X+1)-P(X) \quad \cdots (6)$$

그런데 식 (2)에 의하면 다음이 성립한다.

$$\frac{P(X+1)}{P(X)}=\frac{\binom{k}{X+1}\binom{h}{m-X-1}}{\binom{k}{X}\binom{h}{m-X}}$$

$$= \frac{\dfrac{k!}{(X+1)!(k-X-1)!} \cdot \dfrac{h!}{(m-X-1)!(h-m+X+1)!}}{\dfrac{k!}{X!(k-X)!} \cdot \dfrac{h!}{(m-X)!(h-m+X)!}}$$

$$= \frac{(k-X)(m-X)}{(X+1)(h-m+1+X)}$$

$$\therefore P(X+1) = \frac{(k-X)(m-X)}{(X+1)(h-m+1+X)} P(X) \quad \cdots (7)$$

그러므로 식 (6)과 식 (7)에 의하여 다음을 얻는다.

$$\frac{dy}{dx} = P(X+1) - P(X)$$

$$= \frac{(k-X)(m-X)}{(X+1)(h-m+1+X)} P(X) - P(X)$$

$$= \left\{ \frac{(k-X)(m-X)}{(X+1)(h-m+1+X)} - 1 \right\} P(X)$$

$$= \frac{(m+km-h-1)-(k+h+2)X}{(X+1)(h-m+1+X)} P(X) \quad \cdots (8)$$

한편 식 (5)와 식 (7)에서 다음이 성립한다.

$$y = \frac{1}{2}\{P(X) + P(X+1)\}$$

$$= \frac{1}{2}\left\{1 + \frac{(k-X)(m-X)}{(X+1)(h-m+1+X)}\right\} P(X)$$

$$= \frac{1}{2}\left\{ \frac{h+km-m+1+(h-k+2-2m)X}{(X+1)(h-m+1+X)} + 2X^2 \right\} P(X)$$

그러므로 다음을 얻는다.

$$\frac{1}{y}\frac{dy}{dx} = \left\{ \frac{2(m+km-h-1)-2(h+k+2)X}{km+h+1-m+(h-k+2-2m)X+2X^2} \right\}$$

또 $X = x - ½$로 두면 다음을 얻는다.

$$\frac{1}{y}\frac{dy}{dx} = 2\left\{ \frac{(m+km-h-1)-(h+k+2)(x-½)}{km+h+1-m+(h-k+2-2m)(x-½)+2(x-½)^2} \right\}$$

$$= \frac{x + \dfrac{(h-k-2m-2km)}{2(k+h+2)}}{\dfrac{1}{k+h+2}x^2 + \dfrac{2m+k-h}{2(k+h+2)}x + \dfrac{h+k+1+2km}{4(k+h+2)}} \quad \cdots (9)$$

식 (9)로 주어진 미분방정식을 일반형으로 다시 쓰면 다음 피어슨 미분방정식을 얻는다.

$$\frac{1}{y}\frac{dy}{dx} = \frac{x+a}{px^2+qx+r} \quad \cdots (10)$$

1-3 피어슨 곡선족의 분류

위 피어슨 미분방정식 (10)의 해곡선(solution function curve)은 그 분모(分母)의 2차 방정식 $px^2+qx+r=0$의 근(根) α와 β에 의하여 분류된다. 즉

$$\begin{aligned} px^2+qx+r &= p\left[\left(x+\frac{q}{2p}\right)^2 - \left(\frac{q^2}{4p^2} - \frac{r}{p}\right)\right] \\ &= p\left[\left(x+\frac{q}{2p}\right)^2 - \frac{4r^2}{q^2}K(K-1)\right] \end{aligned}$$

이때,

$$K = \frac{q^2}{4pr} \quad \cdots (11)$$

이고, 또 두 근 α와 β의 곱은 다음과 같다.

$$\begin{aligned} \alpha \cdot \beta &= \frac{r}{p} = \frac{4r^2}{q^2} \cdot \frac{q^2}{4pr} \\ &= \frac{4r^2}{q^2}K \quad \cdots (12) \end{aligned}$$

따라서 두 근(根) α와 β는 K의 값에 따라 정해진다.

<표 2-1> K의 값에 따른 2차 방정식의 근 α와 β의 분류

K의 값	근 α와 β의 값
$K<0$	부호가 상반된 두 실근($\alpha<0$, $\beta>0$; $\alpha<0$, $\beta>0$)
$0<K<1$	허근
$K>1$	부호가 같은 두 실근($\alpha>0$, $\beta>0$; $\alpha<0$, $\beta<0$)
$K=0$, $q=0$	$p\neq0$, 부호가 상반되고 절대값이 같은 두 실근 ($\lvert\alpha\rvert=\lvert\beta\rvert$, $\alpha<0$, $\beta>0$; $\alpha<0$, $\beta>0$)
	$p=0$, 두 근은 ∞($\alpha\to-\infty$, $\beta\to\infty$)
$K=1$	실의 등근($\alpha=\beta$)
$K=\infty$	$p=0(p\to0)$, 하나의 근은 ∞이다($\alpha\to\infty$, $\beta\to\infty$). $r=0$, 하나의 근은 0이다($\alpha=0$ 또는 $\beta=0$).

그리고 K의 값은 위 2차 방정식의 계수 p, q와 상수항 r의 값에 따라 피어슨 분포족은 제 1형의 경우 β-분포, 제 3형인 경우 Γ-분포와 χ^2-분포, 그리고 제 7형인 경우 정규분포와 t-분포의 형태로 다음과 같이 유도된다.

(A) 제 1형 피어슨 곡선

이때 식 (12)에서 $K<0$인 경우이므로 피어슨 미분방정식의 분모에 관한 2차 방정식 $px^2+qx+r=0$은 부호가 서로 다른 실근 α', β'을 가지므로 다음과 같이 쓸 수 있다.

$$px^2+qx+r=p(x+\alpha')(x-\beta') \quad \cdots (13)$$

또 피어슨 미분방정식이 변수 분리형이므로 다음과 같이 다시 쓸 수 있다.

$$\int \frac{dy}{y}+\int \frac{(x+a)dx}{p(x+\alpha')(\beta'-x)}=0 \quad \cdots (14)$$

한편 원점을 $(-a, 0)$으로 평행이동 시키고 $\alpha=\alpha'-a$: $\beta=\beta'+a$라 두면 다음과 같다.

$$\int \frac{dy}{y}+\int \frac{xdx}{p(\alpha+x)(\beta-x)}=0 \quad \cdots (15)$$

식 (15)의 둘째 항의 피적분함수를 부분분수로

$$\frac{x}{(\alpha+x)(\beta-x)}=\frac{A}{\alpha+x}+\frac{B}{\beta-x}$$

로 두고, A와 B의 값들을 결정하면 다음과 같다.

$$A=-\frac{\alpha}{\alpha+\beta}, \quad B=\frac{\beta}{\alpha+\beta}$$

그러므로 다음을 얻는다.

$$\int \frac{dy}{y}-\frac{1}{p}\int \frac{\alpha}{\alpha+\beta}\frac{dx}{\alpha+x}+\frac{1}{p}\int \frac{\beta}{\alpha+\beta}\frac{dx}{\beta-x}=0,$$

$$\ln y-\frac{\alpha}{p(\alpha+\beta)}\ln(\alpha+x)+\frac{\beta}{p(\alpha+\beta)}\ln(\beta-x)=\ln c,$$

$$\ln\frac{y}{c}=\ln(\alpha+x)^{\alpha\nu}(\beta-x)^{\beta\nu}$$

$$\therefore\ y=c(\alpha+x)^{\alpha\nu}(\beta-x)^{\beta\nu} \quad \cdots (16)$$

여기서 $\nu = \dfrac{1}{p(\alpha+\beta)}$ 이고 c는 적분상수이다. 이 분포의 표준형은 다음과 같이 다시 쓸 수 있다.

$$y = y_0\left(1+\frac{x}{\alpha}\right)^{\nu\alpha}\left(1-\frac{x}{\beta}\right)^{\nu\beta} \quad \cdots (17)$$

여기서 y_0은 상수이다.

식 (16)에서 $\nu\alpha = s$, $\nu\beta = t$로 두고, 특히 $\alpha=0$, $\beta=1$인 경우 $0<x<1$에서 식 (16)으로 주어진 x의 함수 $f(x)$가 확률 밀도 함수(p.d.f)가 되도록 적분상수 c의 값을 결정하면 β-분포(β-distribution)의 p.d.f.를 얻게 된다.

그러므로 '제 1형의 피어슨 곡선족'에서 β-분포의 p.d.f.가 유도된다.
즉 식 (16)에서 $\alpha=0$, $\beta=1$로 두고,

$$f(x) = cx^s(1-x)^t \quad (0<x<1) \quad \cdots (18)$$

에 대하여 $f(x)$가 p.d.f.가 되도록 c의 값을 결정해 보면 다음과 같다.

$$\begin{aligned}\int_0^1 f(x)\,dx &= c\int_0^1 x^s(1-x)^t\,dx \\ &= cB(s+1,\ t+1) \\ &= c\frac{\Gamma(s+1)\Gamma(t+1)}{\Gamma(s+t+2)} \\ &= 1,\end{aligned}$$

$$c = \frac{\Gamma(s+t+2)}{\Gamma(s+1)\Gamma(t+1)},$$

$$\therefore\ f(x) = \frac{\Gamma(s+t+2)}{\Gamma(s+1)\Gamma(t+1)}x^s(1-x)^t \quad \cdots (19)$$

여기서 $0<x<1$이다. 그리고 위에서 얻어낸 식 (19)는 β-분포의 p.d.f.이다.

(B) 제 Ⅲ형 피어슨 곡선

이 곡선은 피어슨 미분방정식에서 $K=\infty$, $p=0$, $r\neq0$인 경우이다. 그러므로 식 (16)에서 어느 한쪽 근은 무한근($\beta'=\infty$)이라야 한다. 그래서 $\alpha=\alpha'-a$ $(<\infty)$ 이고 $\beta=\beta'+a$라 두기로 하면 $\beta'\to\infty$이면 $\beta\to\infty$가 된다.

그러므로 식 (17)에서 다음을 얻는다.

$$f(x)=c\left(1+\frac{x}{a}\right)^{\nu a}\lim_{\beta\to\infty}\left(1-\frac{x}{\beta}\right)^{\nu\beta} \quad \cdots\ (20)$$

그런데 식 (20)의 우변 둘째 항에 대하여 $\lambda=-\beta/x$로 두면 다음과 같이 된다.

$$\begin{aligned}f(x)&=c\left(1+\frac{x}{a}\right)^{\nu a}\lim_{\lambda\to\infty}\left(a+\frac{1}{\lambda}\right)^{-\lambda\nu x}\\&=c\left(1+\frac{x}{a}\right)^{\nu a}\left\{\lim_{\lambda\to\infty}\left(1+\frac{1}{\lambda}\right)^{\lambda}\right\}^{-\nu x}\\&=c\left(1+\frac{x}{a}\right)^{\nu a}e^{-\nu x} \quad \cdots\ (21)\end{aligned}$$

여기서 $\nu=-1/q$이다. 이것을 '제 III형 피어슨 곡선'이라 한다.

이제 식 (21)에서 Γ-분포의 p.d.f.를 유도해보면 다음과 같다.

우선, 위 (21)식에 대하여 $1+x/a=X$, $\nu a=w$로 두면 다음을 얻는다.

$$\begin{aligned}f(x)&=c\left(1+\frac{x}{a}\right)^{\nu a}e^{-\nu x}\\&=cX^{w}e^{-\nu[a(X-1)]}a\\&=cae^{\nu a}X^{w}e^{-a\nu X}\\&=c'X^{w}e^{-\frac{X}{\beta}} \quad \cdots\ (22)\end{aligned}$$

여기서 $a\nu=1/\beta$이다. 이제 식 (22)에서 Γ-분포의 p.d.f.를 얻기 위하여 $X=\beta t$로 두고, 다음과 같이 상수 c'을 결정하자.

$$\begin{aligned}\int_0^{\infty}f(x)\,dx&=c'\int_0^{\infty}(\beta t)^{w}e^{-t}\beta dt\\&=c'\beta^{w+1}\int_0^{\infty}t^{w}e^{-t}dt\\&=c'\beta^{w+1}\Gamma(w+1)=1,\end{aligned}$$

$$c'=\frac{1}{\Gamma(w+1)\beta^{a+1}}$$

$$\therefore\ f(x)=\frac{1}{\Gamma(w+1)\beta^{w+1}}X^{w}e^{-\frac{X}{\beta}} \quad \cdots\ (23)$$

여기서 $0<x<\infty$이다. 즉 Γ-분포의 p.d.f가 식 (23)과 같이 유도된다.

뿐만 아니라, 위 Γ-분포에서 $w=k/2-1$, $\beta=2$인 특별한 경우가 '자유도'(degree of freedom)가 $\nu=k$인 χ^2-분포의 p.d.f.이다.

(C) 제 Ⅶ형 피어슨 곡선

(i) 이 곡선은 피어슨 미분방정식에서 $K=0$인 경우, 특히 $p=0$이고 $r\neq 0$인 경우이다. 피어슨 미분방정식에서 다음이 성립한다.

$$\int \frac{dy}{y} = \int \frac{x+a}{r}\,dx \quad \cdots\ (24)$$

또 $X=x+a$로 두면 다음을 얻는다.

$$\int \frac{dy}{y} = \int \frac{X}{r}\,dx,$$

$$\ln y = \frac{1}{r}\frac{X^2}{2} + \ln c,$$

$$\ln \frac{y}{c} = \frac{1}{2r}X^2,$$

$$\therefore\ f(X) = ce^{\frac{1}{2r}X^2} = ce^{-\frac{X^2}{2\sigma^2}} \quad \cdots\ (25)$$

여기서 $\sigma^2=-r$이다. 그리고 식 (25)의 함수 $f(x)$가 p.d.f가 되도록 상수 c를 결정하면 다음을 얻는다.

$$f(X) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{X^2}{2\sigma^2}} \quad \cdots\ (26)$$

여기서 $-\infty < X < \infty$이다. 이것은 기대값이 영(zero)이고 그 분산이 σ^2인 정규분포 $N(0, \sigma^2)$의 p.d.f.이다.

(ii) 만일 $K=0, p\neq 0, r\neq 0$이고, 또 $a=0$일 경우 피어슨 미분방정식에서 다음 등식이 성립한다.

$$\int \frac{dy}{y} = \int \frac{x}{r+px^2}\,dx,$$

$$y=\left[kr\left(1+\frac{p}{r}x^2\right)\right]^{\frac{1}{2p}}$$

이제 $\frac{p}{r}=\frac{1}{2\alpha}$, $\frac{1}{2p}=-\left(\lambda+\frac{1}{2}\right)$이라고 두면 다음과 같이 된다.

$$f(x)=c\left(1+\frac{x^2}{2\alpha}\right)^{-(\lambda+1/2)} \quad \cdots (27)$$

또 식 (27)의 함수 $f(x)$에서 $2\alpha=n$으로 두고, 함수 $f(x)$가 p.d.f.가 되도록 적분상수 c를 정해주면 다음과 같다.

$$\begin{aligned}\int_{-\infty}^{\infty}f(x)\,dx &= c\int_{-\infty}^{\infty}\left(1+\frac{x^2}{n}\right)^{-(\lambda+1/2)}dx \\ &= 2c\int_0^{\infty}\left(1+\frac{x^2}{n}\right)^{-(\lambda+1/2)}dx \qquad \left[\text{이제 } \left(\frac{n+x^2}{n}\right)^{-1}=T\text{라 두면}\right] \\ &= 2c\int_1^0 T^{\lambda+1/2}\left\{-\frac{\sqrt{n}}{2T^{3/2}}(1-T)^{-1/2}\right\}dT \\ &= c\sqrt{n}\int_0^1 T^{\lambda-1}(1-T)^{1/2-1}dT \\ &= c\sqrt{n}B(\lambda,\ 1/2)=1,\end{aligned}$$

$$c=\frac{1}{\sqrt{n}B(\lambda,\ 1/2)},$$

$$\therefore\ f(x)=\frac{\Gamma(\lambda+1/2)}{\sqrt{2\alpha\pi}\,\Gamma(\lambda)}\left(1+\frac{x^2}{2\alpha}\right)^{-\lambda+1/2} \quad \cdots (28)$$

특히, $\alpha=\lambda=n/2$로 두면 다음을 얻는다.

$$f(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2} \quad \cdots (29)$$

여기서 $-\infty < x < \infty$이다. 이것은 '자유도 $\nu=n$인 스튜던트의 t-분포'의 p.d.f.이다.

그러므로 피어슨 미분방정식의 해곡선들은 K의 값과 분모의 2차식의 계수 p, q 그리고 상수항 r에 따라 다음과 같이 분류된다.

<표-2> 피어슨 곡선의 분류

Type	K의 값	근	곡선
I	$K<0$	상반된 두 개의 실근 ($\alpha\beta<0$)	$y=y_0\left(1+\frac{x}{\alpha}\right)^{\nu\alpha}\left(1-\frac{x}{\beta}\right)^{\nu\beta}$ (β-분포) 여기서 $\nu=1/p(\alpha+\beta)$
II	$K=0,\ q=0,\ p\neq0$	$\lvert\alpha\rvert=\lvert\beta\rvert,\ \alpha\beta<0$ 상반된 두 개의 실근	
III	$K=\infty,\ p=0,\ r\neq0$	$\alpha\to\infty$ 또는 $\beta\to\infty$	$y=y_0\left(1+\frac{x}{\alpha}\right)^{\nu m}e^{-\nu x}$ (Γ-분포, χ^2-분포) 여기서 $\nu=-1/q$
IV	$0<K<1$	허근	
V	$K=1$	중근($\alpha=\beta$)	
VI	$1<K<\infty$	동부호인 두 개의 실근	
VII	$K=0,\ q=0,\ p=0$	$\alpha\to-\infty,\ \beta\to\infty$	$y=y_0e^{-\frac{x^2}{2\sigma^2}}$ (정규분포) 여기서 $\sigma^2=-r$
	$K=0,\ q=0,\ a=0$	$\alpha\to-\infty,\ \beta\to\infty$	$y=y_0\left(1+\frac{x_2}{2\omega}\right)^{-\lambda-½}$ (t-분포)

여기서 $K=\frac{q^2}{4pr}$, $\frac{dx}{dy}=\frac{(x+a)y}{r+qx+px^2}$ 이다.

2. 결론

소표본론의 통계적 기법은 고셋의 경험적인 표본분포, 특히 스튜던트의 t-분포에서 발단되었으며. 이 분포의 이론적 증명은 피셔에 의하여 완성되었다고 한다. 그러나 고셋이 t-분포 등 소표본론의 표본분포를 탄생시키는 데는 런던 대학교의 피어슨에게 배운 피어슨 곡선

족에 관한 공부가 아주 큰 도움이 되었다고 한다. 하지만 아이로니컬하게도 만일 고셋이 졸업 후 기네스 맥주회사의 양조화학기사가 되지 않고 런던 대학교의 피어슨 교수 연구실에 연구원이나 교수요원으로 남아있었더라면 대표본류의 고전통계학에 심취되어 소표본론의 시작의 계기를 준 '*t*-분포'의 경험적 발견이 늦어졌을지도 모른다고 전해진다. 이렇게 하여 탄생된 소표본론에 근거한 현대통계학의 이론과 방법론은 물리학을 위시한 자연과학은 물론 공학, 의학, 농학 등 응용 과학 내지 사회학, 경영학 사회과학의 발전에 크게 밑거름이 된 것이다. 특히 미국 1922년 벨(Bell) 연구소의 Walter A. Shewhart에 의한 통계적 품질관리(Statistical Quality Control)의 이론과 기법의 공학, 산업공학에의 적용은 품질개선 발전면에서 '제2의 산업혁명'을 가져오는 계기를 마련한 것이라 하겠다.

참고 문헌

1. 佐藤良一郎 著, **數理統計學** (增補版), 培風館, 1958, pp. 138-165.
2. ウイルクス 著, **數理統計學**, 小河原正已 譯, 1967, pp. 102-108.
3. 北川敏男 著, **統計學の認識**, 白揚社, 1958, pp. 105-108.
4. Daston, Lorranine J., *Classical Probability in the Enlightenment*, Princeton Univ. Press, 1988, p. 423
5. Fienberg, Stephan E., *Statistical Science* Vol. 7, No. 2. 1992, pp. 208-225.
6. Gigerenzer, Gerd etc., *The Empire of Chance How Probability Changed Science and Everyday*, Cambridge Univ. press, 1989, p. 340.
7. Hald, Anders, *A History of Probability and Statistics and Their Applications before 1750*, Wiley, 1990, p. 586.
8. Krüger, L. etc., *The Probability Revolution*, Vol. 1: Ideas in History, MIT Press, 1987, p. 449.
9. Krüger, L. etc., *The Probability Revolution*, Vol. 2: Ideas in History MIT Press, 1987, p. 459.
10. Porter, T.M., *The Rise in Statistical Thinking 1820-1900*, Princeton Univ. Press, 1986, p. 333.