남아선호사상에 기반한 출생 성비에 관한 확률론적 고찰

한양대학교 수학과 **김윤수**
arccschx@hanmail.net

한양대학교 수학과 **최은선**
oops5@dreamwiz.com

한양대학교 수학과 **차경준**
kjcha@hanyang.ac.kr

우리나라의 출생 성비는 자연 상태에서의 출생 성비(natural sex ratio at birth)로 추정되는 105를 초과하고 있는데, 그 원인· 중 하나로 남아선호사상으로 인한 인위적인 출산이 있다(e.g. 임신중절 등). 본 연구에서는 임신 중절 없이 남아를 낳을 때까지만 출산을 하려는 가정(family)들의 존재가 출생 성비를 높이는 데에 기여하는지를 알아보기 위하여, 이러한 가정들과 출생 성비(sex ratio at birth)의 관계를 분석하였다. 그 결과, 남아 한 명을 낳을 때까지만 출산을 하려는 가정의 수가 무한으로 발산할수록 출생 성비는 자연 상태에서의 출생 성비에 확률적으로 수렴하는 것을 알 수 있었다. 즉, 남아를 낳을 때까지만 출산을 하려는 가정들의 존재는 출생 성비에 영향을 미치지 않음을 확인하였다.

주제어 : 확률, 대수의 법칙, 남아선호, 출생 성비

0. 서론

가능성(chance)의 개념은 사회과학, 자연과학 및 공학 등의 분야와 일상생활에 광범위하게 이용되고 있는데, 이는 대부분의 현상이 불확실성을 포함하고 있기 때문이다. 인류는 불확실한 상황에서 어떤 사건이 발생할 가능성의 정도를 표현할 수 있는 언어로 확률(probability) 이론을 발전시켜 왔다.

1494년에 파촐리(Pacioli, 1450-1520)는 "Summa de arithmetica, geometrica, proportioni et proportionalita" 라는 저서에서 게임이 중단되었을 경우의 상금의 분배 문제를 언급하고 있다. 또한, 우연의 사실을 법칙으로 수학화 하려고 노력한 사람들 중에 파스칼(Pascal, 1623-1662)은 주사위 문제와 분배의 문제를 1654년에 숙고하였으

며, 이를 페르마(Fermat, 1601-1655)에게 전하여 두 사람은 이 문제를 명쾌하게 해결하였다. 이 사건을 계기로 확률이 수학적 이론으로 자리 잡기 시작했다([5]).

그 이후 호이겐스(Huygens, 1629-1695), 베르누이(Bernoulli, 1667-1748), 오일러(Euler, 1707-1783) 등의 노력으로 확률론은 급속히 발전하였으나 확률의 정의는 불충분하였다. 20세기에 들어와 콜모고로프(Kolmogorov, 1903-1987)는 "Foundations of Probability Theory"라는 책에서 공리적 확률을 정의하였다([5]). 이후 다양한 연구와 발전을 거쳐서 지금의 확률론은 측도론(measure theory)의 한 갈래로 자리 잡게 되었다([2]).

본 연구에서는 확률을 출생 성비[1)]에 적용시켜서 남아선호사상(男兒選好思想)과 남초 현상 간의 관계를 연구하고자 하였다.

1980년대 이후 우리나라의 출생 성비는 자연 상태에서의 출생 성비로 추정되는 105([9])보다 높아지는 현상이 일어났다([10]). 1960년대 이후 시행된 가족계획에 의하여 소가족이 선호되는 추세가 되면서([1,6]) 적은 수의 자녀를 추구하게 된 것과, 전통적으로 남아를 선호하는 것을([3,7]) 동시에 이룩하려다 보니 일어난 현상으로 보인다. Kim(2004)은 이것을 의학의 발달로 인하여 원치 않는 여아를 임신 중절할 수 있게 되었기 때문으로 분석하였다([6]).

그런데 임신 중절은 하지 않지만, 남아를 출산하면 더 이상 자녀를 가지려 하지 않고 여아를 출산한 경우에만 남아를 얻기 위하여 다시 자녀를 가지려는 가정(family)들의 존재 또한 남초 현상에 기여할 수도 있다고 사료된다.

따라서 본 연구에서는 외부 간섭 없이 남아선호사상만으로 출생 성비를 높일 수 있는지를 알아보기 위하여, 임신 중절 없이 남아 1명을 낳을 때까지만 출산을 하려는 가정들이 존재할 때 그들이 성비에 미치는 영향을 알아보고자 하였다.

1. 방법론

남아를 낳을 때까지만 출산을 하려는 가정들의 존재가 남초 현상에 기여하는지를 확인하기 위해서, 자연 상태에서 출생아가 남아일 확률과 남아를 낳을 때까지만 출산을 하려는 가정들이 존재할 때 그 가정들의 출생아 중 남아가 차지하는 비율을 비교해서 이러한 가정들이 출생 성비에 미치는 영향을 알아보고자 한다.

1.1에서는 출생아 중 남아가 차지하는 비율을 구하기 위하여 여러 가지 변수들을 정의하고, 1.2에서는 남아를 낳을 때까지만 출산을 하려는 가정의 수가 무한으로 발산하는 경우의 출생 성비에 대하여 알아보았다.

---

1) $\frac{\text{남자의 수}}{\text{여자의 수}} \times 100$, 즉 여자 100명에 대한 남자의 수를 나타낸다.

1.1 출생아 중 남아가 차지하는 비율을 구하기 위한 변수 정의

남아를 한 명을 낳을 때까지만 출산을 하려는 가정들이 존재하고, 이 가정들이 남아를 낳을 때까지 출산을 시도하는 데에는 한계가 있다고 가정한다. 또한 각 가정의 출생아의 성별은 서로 독립이라고 가정한다.
이 가정들의 출생아 중 남아가 차지하는 비율을 구하기 위하여 변수 m, n, p, q와 확률변수 X, Y를 다음과 같이 쓰기로 한다.

m을 남아를 낳을 때까지만 출산을 하려는 가정의 수, n을 m개의 가정들 중 특정한 한 가정에서 남아를 낳을 때까지 시도하는 출산의 횟수로 정의한다(n번째에도 여아를 출산한 경우에는 아 출산을 포기).
p를 자연 상태에서 태어나는 출생아가 남아일 확률로, q를 자연 상태에서 태어나는 출생아가 여아일 확률, 즉 $1-p$로 정의한다.
X를 특정한 한 가정에서 태어나는 남아의 수, Y를 특정한 한 가정에서 태어나는 출생아의 수로 정의하고, m개의 가정들 중 i번째 가정에서 태어나는 남아의 수와 전체 출생아의 수는 각각 X_i, Y_i로 표기한다.

또한 확률변수 $X^{(T)}$, $Y^{(T)}$, $\overline{X_m}$, $\overline{Y_m}$, D을 다음과 같이 정의한다.

정의 1. $X^{(T)}$를 전체 가정의 남아의 수, 즉 $\sum_{i=1}^{m} X_i$로 정의한다.

정의 2. $Y^{(T)}$를 전체 가정의 출생아의 수, 즉 $\sum_{i=1}^{m} Y_i$로 정의한다.

정의 3. $\overline{X_m}$를 남아를 낳을 때까지만 출산을 하려는 가정의 수가 m일 때 한 가정당 평균 남아의 수, 즉 $\frac{1}{m}\sum_{i=1}^{m} X_i$로 정의한다($\sum_{i=1}^{m} X_i = m\overline{X_m}$).

정의 4. $\overline{Y_m}$를 남아를 낳을 때까지만 출산을 하려는 가정의 수가 m일 때 한 가정당 평균 출생아의 수, 즉 $\frac{1}{m}\sum_{i=1}^{m} Y_i$로 정의한다($\sum_{i=1}^{m} Y_i = m\overline{Y_m}$).

정의 5. D를 전체 가정의 출생아 중 남아가 차지하는 비율, 즉 $\frac{X^{(T)}}{Y^{(T)}}$로 정의한다.

(성비 $= \dfrac{\text{전체 가정의 남아의 수}}{\text{전체 가정의 여아의 수}} = \dfrac{X^{(T)}}{Y^{(T)} - X^{(T)}} = \dfrac{D}{1-D}$ 인데, 여아가 한 명도 태어나지 않은 경우, 즉 $X^{(T)} = Y^{(T)}$ 인 경우 분모가 0이 되어서 성비가 정의되지 않으므로 성비 대신 D라는 변수를 정의한다.)

정의에 의하여 X, Y는 n, p를 모수(parameter)로 갖는 확률변수이며, $X^{(T)}$, $Y^{(T)}$, $\overline{X_m}$, $\overline{Y_m}$, D는 m, n, p를 모수로 갖는 확률변수임을 알 수 있다.

1.2 $m \to \infty$, n이 모든 가정에서 동일한 경우 D와 p의 비교

가정의 수는 ∞으로 발산하고, 남아를 낳을 때까지 출산을 시도하는 횟수 n은 모든 가정에서 자연수 c로 같은 경우(c번째 출생아도 여아이면 남아 생산을 포기함)의 D를 p와 비교하고자 한다.

정리 1. $n=c$인 한 가정에서 X와 Y는 <표 1>와 같은 결합확률분포를 갖는다.

<표 1> $n=c$인 한 가정에서의 X와 Y의 결합확률분포표

X \ Y	1	2	3	$\cdots$	$c-1$	c	$P_X(X)$
0	0	0	0	$\cdots$	0	q^c	q^c
1	p	pq	pq^2	$\cdots$	pq^{c-2}	pq^{c-1}	$1-q^c$
$P_Y(Y)$	p	pq	pq^2	$\cdots$	pq^{c-2}	q^{c-1}	

<표 2> $n=c$인 한 가정에서의 사건들과 확률

발생할 수 있는 사건	확률	X	Y
남	p	1	1
여 남	pq	1	2
여 여 남	pq^2	1	3
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\overbrace{\text{여 여} \cdots \text{여}}^{c-1\text{명}}$ 남	pq^{c-1}	1	c
$\overbrace{\text{여 여} \cdots \text{여}}^{c\text{명}}$	q^c	0	c

증명. 한 가정에서 발생할 수 있는 사건들과 그 확률은 <표 2>와 같다.

X와 Y가 취하는 값의 각 쌍에 대응하는 확률을 구하면 <표 1>과 같이 된다. ■

보조정리 1. (대수의 강 법칙) $\{A_n\}$이 독립동일분포(independently and identically distributed)를 가지는 확률변수의 수열이라고 하자. $E(|A_1|) < \infty$이면 $n \to \infty$에 따라 표본평균 $\overline{A_n} = \dfrac{A_1 + A_2 + \cdots + A_n}{n}$는 $E(A_1)$에 *almost surely*$(a.s)$로 수렴한다.

증명. ([4] p.193, [8] p.566)참조. ■

$$D = \frac{X^{(T)}}{Y^{(T)}} = \frac{\sum_{i=1}^{m} X_i}{\sum_{i=1}^{m} Y_i} = \frac{m\overline{X_m}}{m\overline{Y_m}} = \frac{\overline{X_m}}{\overline{Y_m}}$$ 인데, $E(X)$과 $E(Y)$이 존재하면 보조정리 1. 에

따라서 $m \to \infty \Rightarrow \overline{X_m} \xrightarrow{a.s} E(X)$, $\overline{Y_m} \xrightarrow{a.s} E(Y)$ 이다.

먼저 $E(X)$, $E(Y)$를 구해 보자.

정리 2. $n = c$인 한 가정에 대하여 $E(X) = 1-(1-p)^c$, $E(Y) = \dfrac{1}{p}\{1-(1-p)^c\}$이다.

증명. 정리 1.에서 구한 X와 Y의 주변확률분포를 이용하면

$E(X) = \sum_{j=0}^{1} j \cdot P_X(X=j) = 0 \cdot q^c + 1 \cdot (1-q^c) = 1 - q^c = 1-(1-p)^c$ 임을 계산할 수 있다.

마찬가지 방법으로 $E(Y)$를 구하면 다음과 같이 표현된다.

$$E(Y) = \sum_{j=1}^{c} j \cdot P_Y(Y=j) = \sum_{j=1}^{c} j \cdot pq^{j-1} + c \cdot q^c = p + 2pq + 3pq^2 + \cdots + (c-1)pq^{c-2} + cpq^{c-1} + cq^c$$

$E(Y)$를 계산하기 위해서 A를 $\sum_{j=1}^{c-1} j \cdot pq^{j-1}$로 정의하고, A에 $1-p$를 곱하면 다음과 같이 된다.

$$\begin{aligned} A &= p + 2p(1-p) + 3p(1-p)^2 + \cdots + (c-2)p(1-p)^{c-3} + (c-1)(1-p)^{c-2} \\ (1-p)A &= p(1-p) + 2p(1-p)^2 + \cdots + (c-3)p(1-p)^{c-3} + (c-2)p(1-p)^{c-2} \\ &\quad + (c-1)p(1-p)^{c-1}. \end{aligned}$$

따라서,

$$1-(1-p)A = pA = p + p(1-p) + p(1-p)^2 + \cdots + p(1-p)^{c-2} - (c-1)p(1-p)^{c-1},$$

$$A = 1 + (1-p) + (1-p)^2 + \cdots + (1-p)^{c-2} - (c-1)(1-p)^{c-1}.$$

그러므로,

$$
\begin{aligned}
E(Y) &= \sum_{j=1}^{c} j \cdot pq^{j-1} + c \cdot q^{c-1} \\
&= 1 + (1-p) + (1-p)^2 + \cdots + (1-p)^{c-2} - (c-1)(1-p)^{c-1} + c(1-p)^{c-1} \\
&= 1 + (1-p) + (1-p)^2 + \cdots + (1-p)^{c-1} = \frac{1}{p}\{1-(1-p)^c\}.
\end{aligned}
$$ ■

보조정리 2. $n \ge 1$에 대하여 A_n, A, B_n, B가 확률변수들이고, $n \to \infty$에 따라 $A_n \xrightarrow{a.s} A$, $B_n \xrightarrow{a.s} B$이라고 하자. $P(B=0)=0$이라면 $\frac{A_n}{B_n} \xrightarrow{a.s} \frac{A}{B}$이다.

증명. ([8] p.557)참조. ■

정리 3. $m \to \infty$, $n = c$이면 $D \xrightarrow{a.s} p$이다.

증명. $D = \frac{\overline{X_m}}{\overline{Y_m}}$인데, $m \to \infty \Rightarrow \overline{X_m} \xrightarrow{a.s} E(X)$, $\overline{Y_m} \xrightarrow{a.s} E(Y)$ 이다.

또한 $P(E(Y)=0)=0$이므로, $D = \frac{\overline{X_m}}{\overline{Y_m}} \xrightarrow{a.s} \frac{E(X)}{E(Y)} = \frac{1-(1-p)^c}{\frac{1}{p}\{1-(1-p)^c\}} = p$이다. ■

정리 4. $m \to \infty$, $n \to \infty$이면 $D \xrightarrow{a.s} p$ 이다.

증명. 정리 2.에 의하여 $n = c$인 한 가정에 대하여 $E(X) = 1-(1-p)^c$, $E(Y) = \frac{1}{p}\{1-(1-p)^c\}$ 이므로, $c \to \infty$이면 $E(X)=1$, $E(Y)=\frac{1}{p}$가 된다. 이것을 이용하면 정리 3.을 증명한 것과 같은 방법으로 증명할 수 있다. ■

2. 결론

본 연구를 요약하면, 남아 한 명을 낳을 때까지만 출산을 c회 시도하는 한 가정에서 남아의 수 대한 기대값은 $E(X) = 1-(1-p)^c$, 출생아의 수에 대한 기대값은 $E(Y) = \frac{1}{p}\{1-(1-p)^c\}$이며, 그러한 가정들이 m개 존재할 때, $m \to \infty$ 이면 전체 출생아의 성비 $D \xrightarrow{a.s} \frac{E(X)}{E(Y)} = p$라고 할 수 있다.

결론적으로 남아를 낳을 때까지만 출산 하려는 가정들이 존재하더라도, 가정의 수가 발산하면 전체 출생아의 성비는 자연 상태에서 남아가 태어날 확률(p)에 *almost surely*로 수렴하게 되므로, 남아를 낳을 때까지만 출산을 하려는 가정들의 존재는 우리나라의 출생 성비에 전혀 영향을 미치지 않는다고 할 수 있다.

본 연구는 역사적으로 의미 있는 소재인 남아선호사상과 출생 성비의 관계를 다루었다는 데에 의의가 있으며, 이는 실생활에서 발생할 수 있는 현상을 수학적으로 다루어 특히, 확률통계학에서 다루는 기초 확률계산, 대수의 법칙 등 교육학적으로도 이용될 수 있을 것이다. 더욱이, 본 연구에서 계산된 결과가 대수의 법칙에 적용된 예로서 그 결과가 예상된다 하더라도 실제 계산을 통하여 증명한 데에 그 의의를 둘 수 있을 것이다.

향후, 남아를 낳을 때까지만 출산을 하려는 가정의 수가 무한으로 발산하는 것이 아니라 유한인 경우에 대한 출생 성비에 대해서도 연구할 예정이다.

참고 문헌

1. 한국 보건사회 연구원, *Low Fertility in Korea :Analysis on Socio-economic Factors*, 한국보건사회연구원, 2005.
2. Apostol, T. M., *Calculus*, Vol 2., 2nd Ed. John Wiley & Sons, 1969.
3. Cho, L.J., Fred. A, Kwon, T.H., *The Determinants of Fertility in the Republic of Korea*, National Research Council Report No.4, Washington, D.C., National Academy Press, 1982.
4. Grimmett, G.R., Stirzaker, *D.R., Probability and Random Processess,* Oxford University Press, c1982.
5. Katz, V.J, *A history of mathematics : an introduction*, 2nd Ed. Addison Wesley, 1998.
6. Kim, D.S., *Missing Girls in South Korea: Trends, Levels and Regional Variations*, Institut National Etudes Démographiques (59)2004, 865-878.
7. Kwon, T.H., Lee, H.Y., *The preference of number and sex of children in a Korean town*, Bulletin of the Population and Development Studies Center, 1976.
8. Port, S.C., *Theoretical Probability for Applications*. Wiley-Interscience, 1983.
9. http://en.wikipedia.org/wiki/Sex_ratio
10. http://www.kosis.kr/

A probabilistic study of the sex ratio at birth related to son preference

Department of Mathematics, Hanyang University **Yun Soo Kim**
Department of Mathematics, Hanyang University **Eun Sun Choi**
Department of Mathematics, Hanyang University **Kyung Joon Cha**

The sex ratio at birth of South Korea is exceeding the natural sex ratio at birth, which is estimated to be about 105. One of the reasons of high sex ratio at birth is due to sex-selective abortion which is caused by strong son preference. The main objective of this study is to identify whether the families which are trying to bear children only until they acquire one son contribute to high sex ratio at birth. As a result, we obtain the theorem that if the number of such families diverge, the sex ratio at birth converges to the natural sex ratio almost surely. Therefore, we conclude that the existence of the families which are trying to bear children only until they receive one son does not affect the sex ratio at birth.

Key Words : probability, the law of large numbers, son preference, high sex ratio at birth

2000 Mathematics Subject Classification : 60F15, 97A40

ZDM Subject Classification : A40, K60

접수일 : 2008년 9월 20일 수정일 : 2008년 10월 30일 게재확정일 : 2008년 11월 11일