TR 85–019

ON THE EFFECT OF SELECTIVE STERILIZATION ON SEX–RATIO

JOHN H. HALTON

The University of North Carolina
Department of Computer Science
New West Hall 035 A
Chapel Hill, NC 27514 USA

JUNE 26, 1985

ON THE EFFECT OF SELECTIVE STERILIZATION ON SEX-RATIO

John H. Halton

Professor of Computer Science
The University of North Carolina
Chapel Hill, NC 27514 USA

ABSTRACT

This paper discusses the effect on the ratio of sons to daughters of an otherwise freely-procreating population of couples, of sterilizing one of the partners immediately after a preselected number of sons have been born. It is shown that the ratio of sons to daughters is, surprisingly, unaffected by this procedure. It is further demonstrated that, among the couples reaching the selected number of sons and therefore subjected to sterilization, the ratio of sons to daughters will exceed that in the general population by a specified amount.

Keywords: Demography; population statistics; sex-ratio; sterilization; population growth; population control.

June 1985

ON THE EFFECT OF SELECTIVE STERILIZATION ON SEX-RATIO

John H. Halton
Computer Science Department
The University of North Carolina
New West Hall
Chapel Hill, NC 27514

1. INTRODUCTION

In the January 1985 issue of *Natural History*, there appears an article entitled "*One Son is No Sons*", by Stanley A. and Ruth S. Freed. It deals with the effect on India's population growth of the government's efforts to encourage voluntary sterilization. The authors point out that the rural Indian families whom they have studied tend, traditionally, to value sons, as laborers on the family farm and as providers for their parents' old age, and therefore plan to have "between two and three sons" before opting for sterilization; but "almost no one desperately wants more than one daughter". They observe that the average couple that has undergone sterilization has "about three sons and two daughters".

The authors state that "a startling consequence of the sterilization program is its effect on the proportion of males and females" in the population of 'Shanti Nagar', the 'nom de science' of the typical North Indian village which they studied. "In 1958, for every 100 females there were 104 males; about twenty years later, there were 111 males. Some attribute such a preponderance of males to the suspected mistreatment of female children. They are said to be relatively deprived of such necessities as food, a practice deriving partly from poverty. Because the village was more prosperous in the late 1970s than in the 1950s, we expected to find a more equal balance of males and females. That, on the contrary, the proportion of males had substantially increased was a surprise. The reason is clearly the practice of sterilization, which offers couples some opportunity for controlling the sex of their children. They can schedule the operation after having a desired number of sons, regardless of the number of daughters."

In the March 1985 issue of Natural History, there appears a letter from James F. Crow, arguing that the attribution of the increase in the proportion of sons to daughters to selective sterilization is erroneous. Crow adds that "the same false argument, based on stopping further births after an heir is born, has been invoked to explain a reputed excess of males in royal households. It is necessary to look elsewhere for an explanation of the increased proportion of males in Shanti Nagar."

The Freeds confirmed Crow's statement in their reply, thanking other readers, too, for the correction. It is this correspondence which has prompted the present analysis, in an attempt to produce a rigorous treatment and a clearer understanding of this counter-intuitive situation. We need to adopt certain simplifying assumptions; but these are ones with which few statisticians or demographers would quarrel. The result of a rather laborious derivation is indeed that the ratio of the number of male and female children in the population is not affected by the practice of sterilization immediately after a fixed number of sons have been produced.

2. THE STOCHASTIC MODEL

We assume a simplified model, in which each couple has a large number N of opportunities to conceive and carry a child to live birth. It is assumed that such events are independent, with a rather small probability c of success. The complementary event (no conception or no live birth) then has the probability $d = 1 - c$; so that

2.1) $$c + d = 1.$$

The sample-space $W = \{S, D, o\}^N$ (where "S" denotes conception and live birth of a son, "D" denotes conception and live birth of a daughter, and "o" denotes no conception or no live birth) consists of finite sequences

2.2) $$w = [z(1), z(2), \ldots, z(N)],$$

in which each z(i) is either "S", "D", or "o". A given family will have a particular sequence w corresponding to it, and the number $t = t(w)$ of those z(i) which equal either "S" or "D" is the number of children produced by the parents. For example, with $N = 400$, the sequence w could have $z(12) = z(25) = z(65) =$ "S" and $z(36) = z(50) =$ "D", with all other $z(i) =$ "o"; and then we would have $t = t(w) = 5$. By the independence hypothesis, since the probability of an "S" or "D" is c, and that of an "o" is d, the probability of any sequence w with just t children ("S" or "D") is

2.3) $$\prod_{i=1}^{N} \text{Prob}\{z(i)\} = c^t d^{N-t} .$$

If we are given only that t children are born to a given couple (without the order of "S", "D", and "o" being specified), there will be as many sequences corresponding to this event as there are ways of selecting t of the N opportunities as leading to the live birth of a child. This number is

2.4) $$\binom{N}{t} = \frac{N(N-1)(N-2)\ldots(N-t+1)}{t!} = \frac{N!}{(N-t)!\ t!},$$

and each of the sequences will have the same probability (2.3). Thus, the probability that a given couple will have t children is given by the binomial distribution,

2.5) $$X_t = \binom{N}{t} c^t d^{N-t} .$$

Clearly, by (2.1) and the Binomial Theorem,

2.6) $$\sum_{t=0}^{N} X_t = \sum_{t=0}^{N} \binom{N}{t} c^t d^{N-t} = (c + d)^N = 1.$$

By similar reasoning, the expected number of children per couple is

$$2.7) \quad C = E[t] = \sum_{t=0}^{N} t X_t = \sum_{t=1}^{N} t \binom{N}{t} c^t d^{N-t}$$

$$= N c \sum_{t=1}^{N} \binom{N-1}{t-1} c^{t-1} d^{N-t}$$

$$= N c \sum_{h=0}^{N-1} \binom{N-1}{h} c^h d^{N-1-h} = N c (c + d)^{N-1}$$

$$= N c,$$

where we have used the fact (easily verified from (2.4)) that

$$2.8) \quad t \binom{N}{t} = N \binom{N-1}{t-1}.$$

By a double application of (2.8), we see similarly that the variance of the number of children per couple is

$$2.9) \quad Var[t] = E[(t - E[t])^2] = E[t(t - 1)] + E[t] - (E[t])^2$$

$$= \sum_{t=2}^{N} t(t - 1) \binom{N}{t} c^t d^{N-t} + Nc - N^2 c^2$$

$$= N(N - 1)c^2 \sum_{t=2}^{N} \binom{N-2}{t-2} c^{t-2} d^{N-t} + Nc - N^2 c^2$$

$$= N(N - 1)c^2 \sum_{h=0}^{N-2} \binom{N-2}{h} c^h d^{N-2-h} + Nc - N^2 c^2$$

$$= N(N - 1)c^2 (c + d)^{N-2} + Nc - N^2 c^2$$

$$= N^2 c^2 - Nc^2 + Nc - N^2 c^2 = Nc(1 - c)$$

$$= C d.$$

Now, if b of a couple's children are boys and g are girls; then

$$b + g = t; \tag{2.10}$$

and if we assume that the the relative probabilities of boys and girls are p and q, fixed throughout, so that

$$p + q = 1; \tag{2.11}$$

then the conditional probability that a couple will have b boys, given that they have t children, is (by reasoning analogous to that employed above)

$$Y_{b|t} = \binom{t}{b} p^b q^{t-b} . \tag{2.12}$$

The probability of a couple having t children, just b of which are boys, is therefore

$$P_{b,t} = X_t Y_{b|t} = \binom{N}{t} c^t d^{N-t} \binom{t}{b} p^b q^{t-b} . \tag{2.13}$$

As we would expect, we observe that, by (2.1), (2.11), and the Binomial Theorem,

$$\begin{aligned} \sum_{t=0}^{N} \sum_{b=0}^{t} P_{b,t} &= \sum_{t=0}^{N} \binom{N}{t} c^t d^{N-t} \sum_{b=0}^{t} \binom{t}{b} p^b q^{t-b} \\ &= \sum_{t=0}^{N} \binom{N}{t} c^t d^{N-t} (p + q)^t \\ &= \sum_{t=0}^{N} \binom{N}{t} c^t d^{N-t} = (c + d)^N = 1. \end{aligned} \tag{2.14}$$

As before, we obtain, using (2.8) and the Binomial Theorem, that the expected number of male children is

$$\begin{aligned} B = E[b] &= \sum_{t=0}^{N} \sum_{b=0}^{t} b P_{b,t} \\ &= \sum_{t=0}^{N} \sum_{b=0}^{t} \binom{N}{t} c^t d^{N-t} b \binom{t}{b} p^b q^{t-b} \\ &= \sum_{t=1}^{N} \binom{N}{t} c^t d^{N-t} t p \sum_{b=1}^{t} \binom{t-1}{b-1} p^{b-1} q^{t-b} \\ &= \sum_{t=1}^{N} \binom{N}{t} c^t d^{N-t} t p \sum_{h=0}^{t-1} \binom{t-1}{h} p^h q^{t-1-h} \end{aligned} \tag{2.15}$$

$$= \sum_{t=1}^{N} \binom{N}{t} c^{t} d^{N-t} \, t \, p \, (p + q)^{t-1}$$

$$= p \sum_{t=1}^{N} t \binom{N}{t} c^{t} d^{N-t}$$

$$= N c p \sum_{t=1}^{N} \binom{N-1}{t-1} c^{t-1} d^{N-t}$$

$$= N c p \sum_{h=0}^{N} \binom{N-1}{h} c^{h} d^{N-1-h} \quad = N c p (c + d)^{N-1}$$

$$= N c p = C p.$$

Likewise, the variance of the number of sons in a family is

2.16) $$\mathrm{Var}[b] = E[b(b - 1)] + E[b] - (E[b])^{2}$$

$$= \sum_{t=0}^{N} \sum_{b=0}^{t} \binom{N}{t} c^{t} d^{N-t} \, b(b - 1) \binom{t}{b} p^{b} q^{t-b} + Ncp - N^{2} c^{2} p^{2}$$

$$= \sum_{t=2}^{N} \binom{N}{t} c^{t} d^{N-t} \, t(t - 1)p^{2} \sum_{b=2}^{t} \binom{t-2}{b-2} p^{b-2} q^{t-b} + Ncp - N^{2} c^{2} p^{2}$$

$$= N(N - 1)c^{2} p^{2} \sum_{t=2}^{N} \binom{N-2}{t-2} c^{t-2} d^{N-t} (p + q)^{t-2} + Ncp - N^{2} c^{2} p^{2}$$

$$= N(N - 1)c^{2} p^{2} (c + d)^{N-2} + Ncp - N^{2} c^{2} p^{2}$$

$$= N^{2} c^{2} p^{2} - Nc^{2} p^{2} + Ncp - N^{2} c^{2} p^{2}$$

$$= Cp(1 - cp).$$

3. THE EFFECT OF SELECTIVE STERILIZATION IMMEDIATELY AFTER A GIVEN NUMBER OF SONS

The response of many families to the Indian government's voluntary sterilization program has been to have one of the parents sterilized immediately after a desired number, say n (usually, 2 or 3), of sons have been born. Since the natural processes involved prior to the sterilization are the same as those in the absence of any sterilization, the situation is one in which the families choosing a specific value of n effectively enforce restriction of the sample space to a subspace H(n) of W, such that H(n) is the union of two disjoint sets F(n) and G(n):

3.1) $$H(n) = F(n) \cup G(n);$$

where

3.2) $$F(n) = \{w \mid b(w) < n\}$$

is the set of sequences corresponding to families having less than n sons, and therefore not opting for sterilization, and

3.3) $$G(n) = \{w \mid b(w) = n, \text{ AND there is a } k, \text{ such that } z(k) = \text{"S" AND (for all } i > k) \; z(i) = \text{"o"}\}$$

is the set of sequences corresponding to families with just n sons, after which sterilization was performed.

The probability of each sequence in F(n) remains the same as that of the same sequence in W; while the probability of a sequence

3.4) $$w = [z(1), z(2), \ldots, z(k), o, o, \ldots, o]$$

in G(n), where z(k) = "S" and just t-1 of the previous z(i) = "S" or "D" (with just n-1 of these previous z(i) = "S"), is

3.5) $$(cp)^{n} (cq)^{t-n} d^{k-t},$$

since the z(i) = "o" for i > k have probability 1 (not d), because of the sterilization enforced by the family after the n'th son is born. The number of sequences of this form, all with the same probability (3.5), is equal to the number of ways of selecting n-1 "S" and t-n "D" out of k-1 symbols; namely,

3.6) $$\frac{(k-1)!}{(n-1)!\,(t-n)!\,(k-t)!} = \binom{k-1}{n-1}\binom{k-n}{t-n}.$$

Thus, the probability that a sequence w in G(n) has parameters k and t as above (without regard to order of the first k-1 symbols), is

3.7) $$Q_{k,t}(n) = \binom{k-1}{n-1}\binom{k-n}{t-n}(cp)^{n}(cq)^{t-n}d^{k-t}$$

Summing the probabilities of all sequences in F(n), we obtain by (2.13) that

3.8) $$R_n = \mathrm{Prob}\{F(n)\} = \sum_{b=0}^{n-1} \sum_{t=b}^{N} P_{b,t}$$

$$= \sum_{b=0}^{n-1} \sum_{t=b}^{N} \frac{N!}{(N-t)!\,(t-b)!\,b!} c^t d^{N-t} p^b q^{t-b}$$

$$= \sum_{b=0}^{n-1} \binom{N}{b} (cp)^b \sum_{t=b}^{N} \binom{N-b}{t-b} (cq)^{t-b} d^{N-t}$$

$$= \sum_{b=0}^{n-1} \binom{N}{b} (cp)^b (cq + d)^{N-b} ;$$

and, similarly, by (3.7), for the sequences involving sterilization,

3.9) $$S_n = \mathrm{Prob}\{G(n)\} = \sum_{k=n}^{N} \sum_{t=n}^{k} Q_{k,t}(n)$$

$$= \sum_{k=n}^{N} \sum_{t=n}^{k} \binom{k-1}{n-1} \binom{k-n}{t-n} (cp)^n (cq)^{t-n} d^{k-t}$$

$$= \sum_{k=n}^{N} \binom{k-1}{n-1} (cp)^n \sum_{t=n}^{k} \binom{k-n}{t-n} (cq)^{t-n} d^{k-t}$$

$$= \sum_{k=n}^{N} \binom{k-1}{n-1} (cp)^n (cq + d)^{k-n} .$$

To proceed, we require the following lemma.

LEMMA 1: Let

3.10) $$K_N = K_N(n, x) = \sum_{k=n}^{N} \binom{k-1}{n-1} x^n (1 - x)^{k-n}$$

and

3.11) $$L_N = L_N(n, x) = \sum_{b=n}^{N} \binom{N}{b} x^b (1 - x)^{N-b} .$$

Then, for all integers n and N, and for all real x, such that

3.12) $$1 \le n \le N \quad \text{and} \quad 0 \le x \le 1,$$

3.13) $$K_N(n, x) = L_N(n, x).$$

Proof. We first note that, for all n and N satisfying (3.12),

3.14) $$K_N(n, 0) = L_N(n, 0) = 0$$

and

3.15) $$K_N(n, 1) = L_N(n, 1) = 1.$$

Further, we see that, for fixed $0 < x < 1$ and $n >= 1$,

3.16) $$K_n(n, x) = L_n(n, x) = x^n .$$

establishing the identity when $N = n$. It therefore only remains to prove that (3.13) holds for fixed $0 < x < 1$ and $n >= 1$, when $N > n$. We proceed by induction. Suppose that (3.13) has been established for $N = m$. Then

3.17) $$K_{m+1} = K_{m+1}(n, x) = \mathrm{sum}_{k=n}^{m+1} \binom{k-1}{n-1} x^n (1 - x)^{k-n}$$

$$= \mathrm{sum}_{k=n}^{m} \binom{k-1}{n-1} x^n (1 - x)^{k-n} + \binom{m}{n-1} x^n (1 - x)^{m-n+1}$$

$$= K_m + \binom{m}{n-1} x^n (1 - x)^{m-n+1} ,$$

and, since (as is easily verified from (2.4)), for all $1 <= b <= m$,

3.18) $$\binom{m+1}{b} = \binom{m}{b} + \binom{m}{b-1},$$

3.19) $$L_{m+1} = L_{m+1}(n, x) = \mathrm{sum}_{b=n}^{m+1} \binom{m+1}{b} x^b (1 - x)^{m-b+1} .$$

$$= \mathrm{sum}_{b=n}^{m} [\binom{m}{b} + \binom{m}{b-1}] x^b (1 - x)^{m-b+1} + x^{m+1}$$

$$= (1 - x) \mathrm{sum}_{b=n}^{m} \binom{m}{b} x^b (1 - x)^{m-b}$$

$$+ x \mathrm{sum}_{b=n}^{m+1} \binom{m}{b-1} x^{b-1} (1 - x)^{m-b+1}$$

$$= (1 - x) \sum_{b=n}^{m} \binom{m}{b} x^{b} (1 - x)^{m-b}$$

$$+ x \sum_{h=n-1}^{m} \binom{m}{h} x^{h} (1 - x)^{m-h}$$

$$= [(1 - x) + x] \sum_{b=n}^{m} \binom{m}{b} x^{b} (1 - x)^{m-b}$$

$$+ \binom{m}{n-1} x^{n} (1 - x)^{m-n+1}$$

$$= L_{m} + \binom{m}{n-1} x^{n} (1 - x)^{m-n+1} .$$

Since our hypothesis states that $K_m = L_m$, we see that we have established that $K_{m+1} = L_{m+1}$, completing the inductive proof of our lemma: Q.E.D.

Taking cp = x and noting that cq + d = 1 - cp = 1 - x, we see that the last sum in (3.9) is simply $K_N(n, cp)$, which equals $L_N(n, cp)$, by the Lemma. Therefore,

3.20) $$S_n = \text{Prob}\{G(n)\} = \sum_{b=n}^{N} \binom{N}{b} (cp)^{b} (1 - cp)^{N-b} ;$$

and so, by (3.8) and (3.20),

3.21) $$\text{Prob}\{H(n)\} = \text{Prob}\{F(n)\} + \text{Prob}\{G(n)\} = R_n + S_n$$

$$= \sum_{b=0}^{N} \binom{N}{b} (cp)^{b} (1 - cp)^{N-b} = [cp + (1 - cp)]^{N} = 1,$$

as we would expect.

The expected number of children per couple is

3.22) $$C' = E_n[t] = \sum_{b=0}^{n-1} \sum_{t=b}^{N} t\, P_{b,t} + \sum_{k=n}^{N} \sum_{t=n}^{k} t\, Q_{k,t}(n) = C_R + C_S ;$$

where, as in (3.8),

$$
\begin{aligned}
(3.23)\qquad C_R &= \sum_{b=0}^{n-1}\sum_{t=b}^{N} t\, P_{b,t} \\
&= \sum_{b=0}^{n-1}\binom{N}{b}(cp)^{b}\sum_{t=b}^{N} t\binom{N-b}{t-b}(cq)^{t-b} d^{N-t} \\
&= \sum_{b=0}^{n-1}\binom{N}{b}(cp)^{b}\Big[(N-b)cq\sum_{t=b+1}^{N}\binom{N-b-1}{t-b-1}(cq)^{t-b-1} d^{N-t} \\
&\qquad\qquad + b\sum_{t=b}^{N}\binom{N-b}{t-b}(cq)^{t-b} d^{N-t}\Big] \\
&= \sum_{b=0}^{n-1}\binom{N}{b}(cp)^{b}\Big[(N-b)cq\,(cq+d)^{N-b-1} + b\,(cq+d)^{N-b}\Big] \\
&= Ncq\sum_{b=0}^{n-1}\binom{N-1}{b}(cp)^{b}(1-cp)^{N-b-1} \\
&\qquad + Ncp\sum_{b=1}^{n-1}\binom{N-1}{b-1}(cp)^{b-1}(1-cp)^{N-b} \\
&= Ncq\sum_{b=0}^{N-1}\binom{N-1}{b}(cp)^{b}(1-cp)^{N-b-1} \\
&\qquad - Ncq\sum_{b=n}^{N-1}\binom{N-1}{b}(cp)^{b}(1-cp)^{N-b-1} \\
&\qquad + Ncp\sum_{b=1}^{N}\binom{N-1}{b-1}(cp)^{b-1}(1-cp)^{N-b} \\
&\qquad - Ncp\sum_{b=n}^{N}\binom{N-1}{b-1}(cp)^{b-1}(1-cp)^{N-b} \\
&= Ncq\,[cp+(1-cp)]^{N-1} - Ncq\sum_{b=n}^{N-1}\binom{N-1}{b}(cp)^{b}(1-cp)^{N-b-1} \\
&\qquad + Ncp\,[cp+(1-cp)]^{N-1} - Ncp\sum_{b=n}^{N}\binom{N-1}{b-1}(cp)^{b-1}(1-cp)^{N-b}
\end{aligned}
$$

$$= Ncq - Ncq \sum_{b=n}^{N-1} \binom{N-1}{b} (cp)^{b} (1 - cp)^{N-b-1}$$

$$+ Ncp - Ncp \sum_{b=n}^{N} \binom{N-1}{b-1} (cp)^{b-1} (1 - cp)^{N-b}$$

$$= Cq - \frac{q}{p} \sum_{b=n}^{N-1} (b + 1) \binom{N}{b+1} (cp)^{b+1} (1 - cp)^{N-b-1}$$

$$+ Cp - Ncp \sum_{b=n}^{N} \binom{N-1}{b-1} (cp)^{b-1} (1 - cp)^{N-b}$$

$$= Cq - \frac{q}{p} \sum_{b=n+1}^{N} b \binom{N}{b} (cp)^{b} (1 - cp)^{N-b}$$

$$+ Cp - \sum_{b=n}^{N} b \binom{N}{b} (cp)^{b} (1 - cp)^{N-b}$$

$$= C - \frac{1}{p} \sum_{b=n+1}^{N} b \binom{N}{b} (cp)^{b} (1 - cp)^{N-b}$$

$$- n \binom{N}{n} (cp)^{n} (1 - cp)^{N-n}$$

and, as in (3.9),

3.24) $$C_S = \sum_{k=n}^{N} \sum_{t=n}^{k} t \, Q_{k,t}$$

$$= \sum_{k=n}^{N} \binom{k-1}{n-1} (cp)^{n} \sum_{t=n}^{k} t \binom{k-n}{t-n} (cq)^{t-n} d^{k-t}$$

$$= \sum_{k=n}^{N} \binom{k-1}{n-1} (cp)^{n} \Big[(k - n)cq \sum_{t=n+1}^{k} \binom{k-n-1}{t-n-1} (cq)^{t-n-1} d^{k-t}$$

$$+ n \sum_{t=n}^{k} \binom{k-n}{t-n} (cq)^{t-n} d^{k-t} \Big]$$

$$= \sum_{k=n}^{N} \binom{k-1}{n-1} (cp)^{n} \Big[(k - n)cq \, (cq + d)^{k-n-1} + n \, (cq + d)^{k-n} \Big]$$

$$= n \frac{q}{p} \sum_{k=n+1}^{N} \binom{k-1}{n} (cp)^{n+1} (1 - cp)^{k-n-1}$$

$$+ n \sum_{k=n}^{N} \binom{k-1}{n-1} (cp)^{n} (1 - cp)^{k-n}$$

$$= n \frac{q}{p} \sum_{b=n+1}^{N} \binom{N}{b} (cp)^{b} (1 - cp)^{N-b}$$

$$+ n \sum_{b=n}^{N} \binom{N}{b} (cp)^{b} (1 - cp)^{N-b}$$

$$= n \frac{1}{p} \sum_{b=n+1}^{N} \binom{N}{b} (cp)^{b} (1 - cp)^{N-b}$$

$$+ n \binom{N}{n} (cp)^{n} (1 - cp)^{N-n} .$$

Gathering (3.22), (3.23), and (3.24) together, we obtain, as we would expect, that

3.25) $$C' = C_R + C_S$$

$$= C - \frac{1}{p} \sum_{b=n+1}^{N} (b - n) \binom{N}{b} (cp)^{b} (1 - cp)^{N-b} < C;$$

since every term in the sum subtracted from C in (3.25) is positive.

Similarly, the expected number of sons per couple is

3.26) $$B' = E_n[b] = \sum_{b=0}^{n-1} \sum_{t=b}^{N} b P_{b,t} + \sum_{k=n}^{N} \sum_{t=n}^{k} b Q_{k,t}(n) = B_R + B_S ;$$

where

3.27) $$B_R = \sum_{b=0}^{n-1} \sum_{t=b}^{N} b P_{b,t}$$

$$= \sum_{b=1}^{n-1} b \binom{N}{b} (cp)^{b} \sum_{t=b}^{N} \binom{N-b}{t-b} (cq)^{t-b} d^{N-t}$$

$$= \sum_{b=1}^{n-1} b \binom{N}{b} (cp)^{b} (1 - cp)^{N-b}$$

$$= \sum_{b=1}^{N} b \binom{N}{b} (cp)^{b} (1 - cp)^{N-b}$$

$$- \sum_{b=n}^{N} b \binom{N}{b} (cp)^{b} (1 - cp)^{N-b}$$

$$= Ncp \sum_{b=1}^{N} \binom{N-1}{b-1} (cp)^{b-1} (1 - cp)^{N-b}$$

$$- \sum_{b=n}^{N} b \binom{N}{b} (cp)^{b} (1 - cp)^{N-b}$$

$$= Ncp\,[cp + (1 - cp)]^{N-1} - \sum_{b=n}^{N} b \binom{N}{b} (cp)^{b} (1 - cp)^{N-b}$$

$$= Cp - \sum_{b=n}^{N} b \binom{N}{b} (cp)^{b} (1 - cp)^{N-b}$$

and

3.28) $$B_S = n\,C_S = \sum_{b=n}^{N} n \binom{N}{b} (cp)^{b} (1 - cp)^{N-b} .$$

Again, we see that

3.29) $$B' = B_R + B_S$$

$$= B - \sum_{b=n+1}^{N} (b - n) \binom{N}{b} (cp)^{b} (1 - cp)^{N-b} < B.$$

Finally, we observe that

3.30) $$B' - p\,C' = B - \sum_{b=n+1}^{N} (b - n) \binom{N}{b} (cp)^{b} (1 - cp)^{N-b}$$

$$- p\,C + \sum_{b=n+1}^{N} (b - n) \binom{N}{b} (cp)^{b} (1 - cp)^{N-b}$$

$$= B - p\,C;$$

and, since B = pC, we arrive at the result that

3.31) $$B' = p\,C'.$$

4. CONCLUSIONS

First, we note that (3.30) and (3.31) establish that the ratio of the expected numbers of children and of sons per couple is undisturbed by the selective sterilization practiced by the villagers of 'Shanti Nagar'. Yet, the Freeds observe that, for every 100 females in the village, there were 104 males in 1958 and 111 males about twenty years later. As Crow remarks, "it is necessary to look elsewhere for an explanation".

The Central Limit Theorem tells us that, in a population of Q independent samples (for instance, in a village containing Q post-pubertal married women), the observed average value of a random variable X (such as t or b) will be approximately normally distributed (the closeness of the approximation improving as Q increases), with mean E[X] and variance Var[X] / Q. In particular, the Freeds' statement that "there were 68 sterilizations ..., tantamount to 26 percent of the women of childbearing age (15 to 45 years)" suggests that the value of Q in 'Shanti Nagar' was certainly less than 500. By (2.7), (2.9), (2.15), and (2.16), we see that E[t] = C (which, we are told, was about 5), Var[t] / Q = C d / Q (more than 0.01), E[b] = C p (about 2.5), and Var[b] = C p (1 - cp) / Q (more than 0.005).

We are told that, in 1958, there were 104 males to every 100 females; this means that $(Cp)^\sim$ / $(Cq)^\sim$ = 1.04; where $(Cp)^\sim$ is the observed average number of sons per couple (whose mathematical expectation is E[b] = Cp, by (2.15)), and $(Cq)^\sim$ is the observed average number of daughters per couple (whose expectation is E[g] = E[t - b] = E[t] - E[b] = C - Cp = Cq, by (2.7), (2.10), (2.11), and (2.15)). The relative standard deviation of each of these quantities is more than 0.025 (square root of 0.005, divided by 2.5); so that, with good accuracy, we may estimate that their ratio (which is nearly 1) will have a standard deviation of more than 0.05. We are also told that, twenty years later, there were 111 males to every 100 females; and this means that the ratio was then 1.11. If we make the hypothesis that, in fact, both samples were drawn from the same sample-space; that is, that the statistics were unaffected by the sterilizations; then we know that the difference, 1.11 - 1.04 = 0.07, of the two observed ratios is a sample taken from a normal distribution with mean 0 and standard deviation greater than 0.07 (square root of twice the square of 0.05). The probability of a difference at most as large as the standard deviation is more than 31% --- hardly a significant result.

The rather bathetic result seems to be, therefore, that there is nothing significant to explain --- unless the same trend is observed in other North Indian villages. Further investigation is clearly indicated.

5. THE STATISTICS OF THE STERILIZED SUB-POPULATION

It is interesting to look at the relevant statistics, limited to only those couples which have undergone sterilization. The corresponding sequences all have exactly n sons and belong to the set G(n) defined in (3.3). Then the probability of G(n) is given by (3.20), and the conditional probability of any sequence w in G(n) with parameters k and t (as defined in (3.3) - (3.7)) is

5.1) $$Q'_{k,t}(n) = Q_{k,t}(n) / S_n \, .$$

The expected number of children per sterilized couple is, by (3.24),

5.2) $$C'' = \sum_{k=n}^{N} \sum_{t=n}^{k} t\, Q'_{k,t}(n) = C_S / S_n$$

$$= \frac{n}{p} - n \frac{q}{p} \binom{N}{n} (cp)^{n} (1 - cp)^{N-n} / S_n \, .$$

Since the number B'' of sons per sterilized couple is exactly n, we see that

5.3) $$B'' = n = p\, C'' + n\, q \binom{N}{n} (cp)^{n} (1 - cp)^{N-n} / S_n$$

$$> p\, C'',$$

since something positive is added to pC'' to get B''.

By (3.20) and (5.2),

5.4) $$C'' = \frac{n}{p} - \frac{n \frac{q}{p} \binom{N}{n} (cp)^{n} (1 - cp)^{N-n}}{\sum_{b=n}^{N} \binom{N}{b} (cp)^{b} (1 - cp)^{N-b}} \, .$$

Therefore, the expected number of daughters per sterilized couple will be

5.5) $$D'' = C'' - n = n \left(\frac{q}{p} \right) \frac{\sum_{b=n+1}^{N} \binom{N}{b} (cp)^{b} (1 - cp)^{N-b}}{\sum_{b=n}^{N} \binom{N}{b} (cp)^{b} (1 - cp)^{N-b}}$$

$$= n \left(\frac{q}{p} \right) \frac{S_{n+1}}{S_n} \, .$$

Note that, by (3.20),

$$5.6) \qquad S_n = \sum_{b=n}^{N} \binom{N}{b} (cp)^b (1 - cp)^{N-b}$$

$$= \left[\sum_{b=0}^{N} - \sum_{b=0}^{n-1} \right] \binom{N}{b} (cp)^b (1 - cp)^{N-b}$$

$$= 1 - \sum_{b=0}^{n-1} \binom{N}{b} (cp)^b (1 - cp)^{N-b} .$$

We know that $N \sim 400$, $n < 6$, $p \sim q \sim 0.5$, and $C \sim 5$; so that $q/p \sim 1.0$ and $Ncp = Cp \sim 2.5$. Thus we are justified in using the approximation,

$$5.7) \qquad \binom{N}{b} (cp)^b (1 - cp)^{N-b} \sim e^{-Ncp} \frac{1}{b!} (Ncp)^b ,$$

for $b <= n$. With this approximation, we have from (5.6) that

$$5.8) \qquad S_n \sim 1 - e^{-Ncp} \sum_{b=0}^{n-1} \frac{1}{b!} (Ncp)^b .$$

We note that (5.8) has only one parameter, namely, Ncp. Though the data at our disposal are somewhat limited, it seems reasonable to assume that Ncp is known within 10%, lying between 2.25 and 2.75. The resulting values of S_n are shown in Table I below.

Similarly, (5.6) shows dependence only on q / p and Ncp (the latter through (5.8)). Again, it would seem reasonable to assume that q / p is within 10% of its estimated value: between 0.9 and 1.1. The resulting values of D'' and n / D'' are shown in Table II.

Recalling the Freeds' remark that, among couples that had undergone sterilization, there were "about three sons and two daughters", we observe that, on the basis of the average number of sons, the typical value of n is 3; while, on the basis of son-to-daughter ratio in the table above, n seems to be between 1 and 3. The agreement is quite good, considering the uncertainty of the parameters and the size of the sample used.

TABLE I.

n	Ncp	e^{-Ncp}	$\sum_{b=0}^{n-1} \frac{1}{b!} (Ncp)^b$	$1 - e^{-Ncp} \sum_{b=0}^{n-1} \frac{1}{b!} (Ncp)^b$
1	2.25	0.10540	1.00000	0.89460
2			3.25000	0.65745
3			5.78125	0.39066
4			7.67969	0.19057
5			8.74756	0.07801
6			9.22810	0.02737
1	2.75	0.06393	1.00000	0.93607
2			3.75000	0.76027
3			7.53125	0.51854
4			10.99740	0.29696
5			13.38037	0.14462
6			14.69101	0.06084

TABLE II.

n	$\frac{q}{p}$	Ncp	D''	$\frac{n}{D''}$	$\frac{q}{p}$	Ncp	D''	$\frac{n}{D''}$
1	0.9	2.25	0.66142	1.512	1.1	2.25	0.80840	1.237
2			1.06957	1.870			1.30725	1.530
3			1.31708	2.278			1.60972	1.864
4			1.47377	2.714			1.80127	2.221
5			1.57849	3.168			1.92926	2.592
1	0.9	2.75	0.73097	1.368	1.1	2.75	0.89341	1.119
2			1.22769	1.629			1.50051	1.333
3			1.54624	1.940			1.88985	1.587
4			1.75322	2.282			2.14283	1.867
5			1.89293	2.641			2.31359	2.161

Perusal of Table II suggests that, at least for small enough values of n, the ratio n / D'' of sons to daughters is an increasing function of the cut-off number n of sons. Returning to (5.5), let us define

5.9) $$T_n = \frac{n}{D''} / \frac{p}{q} = \frac{S_n}{S_{n+1}} .$$

This is the result of dividing the sex-ratio in G(n) by that in the general population. It is the approximation to $(p / q) T_n$ which is seen to increase in Table II.

The result we have conjectured on the basis of the tabulated figures above is established mathematically in the following lemma.

LEMMA 2: If we define

5.10) $$V_n = V_n(x) = U_n(x) / U_{n+1}(x)$$

and

5.11) $$U_n = U_n(x) = \sum_{b=n}^{N} \binom{N}{b} x^b (1-x)^{N-b} ,$$

where $0 < x < 1$; then V_n increases with n; that is,

5.12) $$V_{n+1} > V_n$$

for all $0 < n < N$.

Proof. By (5.11),

5.13) $$0 < U_n < 1;$$

and (5.12) is true if

5.14) $$\frac{U_{n+1}}{U_{n+2}} > \frac{U_n}{U_{n+1}} .$$

We note that, by (5.11),

5.15) $$U_n = U_{n+1} + \binom{N}{n} x^n (1 - x)^{N-n}$$

and

5.16) $$U_{n+2} = U_{n+1} - \binom{N}{n+1} x^{n+1} (1 - x)^{N-n-1} ;$$

so that (5.14) is true if

$$U_{n+1}^2 > [U_{n+1} + \binom{N}{n} x^n (1 - x)^{N-n}] [U_{n+1} - \binom{N}{n+1} x^{n+1} (1 - x)^{N-n-1}],$$

or, equivalently,

$$U_{n+1} [\binom{N}{n+1} - \binom{N}{n} \frac{1 - x}{x}] x^{n+1} (1 - x)^{N-n-1} + \binom{N}{n} \binom{N}{n+1} x^{2n+1} (1 - x)^{2N-2n-1} > 0;$$

which reduces on division by the positive quantity $\binom{N}{n+1} x^{n+1} (1 - x)^{N-n-1}$ to

5.17) $$U_{n+1} [1 - \frac{n + 1}{N - n} \frac{1 - x}{x}] + \binom{N}{n} x^n (1 - x)^{N-n} > 0.$$

Finally, by (5.11), we see that (5.17) (and therefore (5.12)) holds if

5.18) $$\sum_{b=n}^{N} \binom{N}{b} x^b (1 - x)^{N-b} > \frac{n + 1}{N - n} \sum_{b=n+1}^{N} \binom{N}{b} x^{b-1} (1 - x)^{N-b+1} .$$

Now observe that

$$5.19)\qquad \sum_{b=n}^{N} \binom{N}{b} x^{b} (1-x)^{N-b} > \sum_{b=n}^{N-1} \binom{N}{b} x^{b} (1-x)^{N-b}$$

$$= \sum_{h=n+1}^{N} \binom{N}{h-1} x^{h-1} (1-x)^{N-h+1}$$

$$>= \sum_{b=n+1}^{N} \binom{N}{b-1} x^{b-1} (1-x)^{N-b+1} \frac{n+1}{b} \frac{N-b+1}{N-n}$$

$$= \frac{n+1}{N-n} \sum_{b=n+1}^{N} \binom{N}{b} x^{b-1} (1-x)^{N-b+1} \quad ;$$

which establishes (5.18), and hence proves our lemma: Q.E.D.

Noting that (by (3.20), (5.9), (5.10), and (5.11))

$$5.20)\qquad S_n = U_n(cp) \quad \text{and} \quad T_n = V_n(cp),$$

we conclude that, for all $0 < n < N$,

$$5.21)\qquad T_{n+1} > T_n .$$