Family planning

A. C. PASEAU

Without doing the morally unspeakable or the medically infeasible, can a preference for daughters rather than sons increase their relative number? If, to be more precise, the only variable over which you have control is your number of children, can you increase the ratio*

Expected value (no. of daughters) : Expected value (no. of sons) ?

Naïvely, you might think so. If for example you adopt the policy 'stop procreating as soon as a girl is born' won't you bear more girls compared to boys than you would otherwise? No, in fact. Suppose the chances of a girl's being born are g and the chances of a boy's being born are b (so that $g + b = 1$, $g \neq 0$, $b \neq 0$), each birth being an independent event. The ratio of the expected number of daughters to the expected number of sons for families that follow the stop-when-we-have-a-daughter policy is the same as those that don't, namely:

$$\frac{g}{b}.$$

Think of a society in which everyone adopts the stop-when-I-have-a-daughter policy. Suppose there are N families for some large N and that procreation happens at numbered stages. At stage 1, $g \times N$ families have girls and $b \times N$ have boys. The ratio of girls to boys in the population is now g/b. Since they've had their girl, the $g \times N$ families stop procreating. The families that continue procreating are the $b \times N$ with boys. At stage 2, g of the $b \times N$ families have a girl and b of the $b \times N$ have a second boy. The ratio of girls to boys among the children added to the population at stage 2 is g/b, so the overall ratio of girls to boys in the population remains g/b. And so it goes on, the ratio of girls to boys in the population staying constant at g/b. This reasoning is of course informal, but it is easy to make it rigorous.

In other words, you can't cheat nature! A preference for daughters won't result in more daughters on average. In particular, under the given assumptions, one cannot tell whether a society has a preference for boys or girls by looking at their mean numbers. That doesn't mean you can't read off such a preference from the society's demographics. For example, if the predominant patterns of children are G, BG, BBG, BBBG, ..., evidently the society has a bias in favour of girls.

Similarly, following the stop-when-I-have-k-daughters-policy won't on average increase the number of your daughters compared to the number of your sons. The reasoning is the same as in the case $k = 1$.

* Not to be confused with the expected value of the ratio of the number of daughters to the number of sons.

https://doi.org/10.1017/S0025557200002916 Published online by Cambridge University Press

Mathematicians are in general aware of this fact. The interest of this note is to see how it links up with the combinatorial equation

$$\sum_{i=0}^{\infty} \left({}^{k+i-1}C_i\right) ix^i = \frac{kx}{(1-x)^{k+1}}$$

and thereby provides an exercise for able sixth-formers combining probability, combinatorics and the calculus.

The first way to convince yourself of the last equation is to consider the expected value of your number of boys if you follow the stop-when-I-have-k-daughters policy. There are infinitely many ways in which you can have k daughters:

k daughters and 0 sons

k daughters and 1 son

k daughters and 2 sons

If you end up with $k + i$ children then you will have had k daughters and i sons, your last child being a daughter. Now there are $\left({}^{k+i-1}C_i\right)$ possible ways in which i boys can be distributed among $k + i - 1$ children of whom i are boys and $k - 1$ are girls. Since the chances of a girl being born are g and of a boy being born are b, your expected number of sons is therefore

$$\sum_{i=0}^{\infty} \left({}^{k+i-1}C_i\right) i\, g^k b^i.$$

Your expected number of daughters on the other hand is obviously k. As argued earlier, the ratio of the expected number of girls to the expected number of boys is g/b. Thus

$$\frac{g}{b} = \frac{k}{\sum_{i=0}^{\infty}\left({}^{k+i-1}C_i\right) i\, g^k b^i}$$

hence

$$\sum_{i=0}^{\infty} \left({}^{k+i-1}C_i\right) i\, g^k b^i = \frac{kb}{g}$$

and replacing g with $1 - b$ and dividing through by $(1 - b)^k$,

$$\sum_{i=0}^{\infty} \left({}^{k+i-1}C_i\right) i\, b^i = \frac{kb}{(1-b)^{k+1}}.$$

There is another way of proving this equation. First, divide both sides by b:

$$\sum_{i=0}^{\infty} \left({}^{k+i-1}C_i\right) i\, b^{i-1} = \frac{k}{(1-b)^{k+1}}.$$

Notice that the resulting equation's right-hand side is

$$\frac{d}{db}\left(\frac{1}{(1-b)^k}\right),$$

https://doi.org/10.1017/S0025557200002916 Published online by Cambridge University Press

and that its left-hand side is

$$\frac{d}{db}\sum_{i=0}^{\infty}\left({}^{k+i-1}C_i\right)b^i.$$

To prove the equation it is therefore sufficient to show that for positive i the coefficient of b^i in $(1 - b)^{-k}$, that is, the coefficient of b^i in

$$\underbrace{\left(1 + b + b^2 + \dots\right)\left(1 + b + b^2 + \dots\right)\dots\left(1 + b + b^2 + \dots\right)}_{k\ \text{brackets}}$$

is

$$\left({}^{k+i-1}C_i\right).$$

This last fact follows from a combinatorial argument about partitions. We must find the number of ways of picking a power of b from the first parenthesis, a power of b from the second parenthesis, and so on, with the constraint that the sums of these powers is i. Think of $k + i - 1$ white balls in a row. Colour $k - 1$ of them black, so that there are i remaining white balls. The number of white balls to the left of the first black ball (which could be the leftmost of the balls) can be thought of as the power of b chosen from the first parenthesis. The number of white balls to the right of the first black ball and the left of the next black ball (again, this number could be 0) can be thought of as the power of b chosen from the second parenthesis, and so on. Since there are $\left({}^{k+i-1}C_{k-1}\right) = \left({}^{k+i-1}C_i\right)$ ways of colouring $k - 1$ of the $k + i - 1$ balls black, there are $\left({}^{k+i-1}C_i\right)$ ways of picking powers of b from the respective k multiplicands so that their sum is i. Hence for positive i the coefficient of b^i in the k-fold product of $\left(1 + b + b^2 + \dots\right)$ is indeed $\left({}^{k+i-1}C_i\right)$.

There is a third, equally interesting way to prove our combinatorial identity, which we set out more briefly. It is easy to show that the probability generating function $G(s)$ of the number of children up to and including the first girl is $gs/(1 - bs)$, the probability generating function of the geometric distribution. Then by independence of births, the probability generating function of the number of children up to and including the kth girl is $(G(s))^k = \left(\frac{gs}{1 - bs}\right)^k$. The expected number of children is $\frac{d}{ds}(G(s))^k = \frac{d}{ds}\left(\frac{gs}{1 - bs}\right)^k$ evaluated at 1, i.e., the value at $s = 1$ of

$$\frac{kg^k s^{k-1}}{(1 - bs)^k} + \frac{kb(gs)^k}{(1 - bs)^{k+1}}.$$

Given that $g + b = 1$, this value is

$$k + \frac{kb}{(1 - b)}.$$

Equating this to the expected number of girls, k, plus the expected number of boys, $\sum_{i=0}^{\infty}\left({}^{k+i-1}C_i\right) i\, g^k b^i$, yields the combinatorial identity.

The connection with family planning yields a more general equation. Suppose there are m sexes, born with probability p_i for $1 \leqslant i \leqslant m$, where $p_1 + p_2 + \dots + p_m = 1$ and each $p_i > 0$. Suppose a family adopts the policy of stopping procreation when they have produced k offspring of the first sex. The expected value of the number of children of the first sex is therefore k. By a reapplication of the earlier reasoning, the expected value of the number of children of the ith sex is $c(p_i/p_1)$, where c is some constant, since the expected value of (no. of children of the ith sex) divided by the expected value of (no. of children of the j th sex) is

$$\frac{p_i}{p_j}.$$

Finally, since the expected value of the number of children of the first sex is k, $c = k$.

More ploddingly, there are infinitely many ways to have k children of the 1st sex under this policy:

k children of the 1st sex and 0 children of other sexes

k children of the 1st sex and 1 child of other sexes

k children of the 1st sex and 2 children of other sexes

Suppose you have k children of the first sex and m other children, j of which are of the ith sex (where $i \neq 1$ and $0 \leqslant j \leqslant m$). As your last child must be of the first sex, there are $\left({}^{k+m-1}C_{k-1}\right)$ ways in which the children of the first sex can be distributed among the m children not of the first sex. Any such sequence of children has probability

$$p_i^k(1 - p_i - p_1)^{m-j}p_i^j,$$

p_1^k being the probability of having the k children of the first sex, $(1 - p_i - p_1)^{m-j}$ the probability of having $m - j$ children of neither the first nor the ith sex, and p_i^j the probability of having j children of the ith sex. The expected value of the number of children of the ith sex is thus

$$\sum_{m=0}^{\infty}\sum_{j=0}^{m} j\left({}^{k+m-1}C_{k-1}\right).\left({}^{m}C_j\right)p_i^j(1 - p_i - p_1)^{m-j}p_1^k.$$

Putting together the two different ways of calculating the expected values yields the following equation:

$$\sum_{m=0}^{\infty}\sum_{j=0}^{m} j\left({}^{k+m-1}C_{k-1}\right).\left({}^{m}C_j\right)p_i^j(1 - p_i - p_1)^{m-j}p_1^k = k\left(\frac{p_i}{p_1}\right).$$

A similar proof to the earlier one may be given by considering the joint probability generating function of the numbers of children of other sexes before the arrival of k of the first sex. We conclude not with this but with a derivation of a more general identity from the special case in which there are two sexes. Rewrite the left-hand side of the last equation as

$$\sum_{m=0}^{\infty} p_1^k \left({}^{k+m-1}C_{k-1} \right) \left[\sum_{j=0}^{m} j \left({}^{m}C_j \right) p_i^j (1 - p_i - p_1)^{m-j} \right].$$

The rightmost bracket is the expected number of children of the i th sex in a set of m children none of whom is of the first sex multiplied by the probability of there being no children of the first sex among these m children. By the general argument given at the start, the expected number of children of the i th sex in a set of m children none of whom is of the first sex must be $mp_i / (1 - p_1)$ since the ratio of the expected number of children of the i th sex to the expected number of children of the j th sex ($i \neq 1, j \neq 1$) in such a set must be p_i / p_j and must also sum to m. Given that the probability of there being no children of the first sex among m children is $(1 - p_1)^m$, the expression in the square brackets may be written as $mp_i (1 - p_1)^{-1} (1 - p_1)^m = mp_i (1 - p_1)^{m-1}$. Hence the left-hand side of the equation may be written more simply as

$$\sum_{m=0}^{\infty} p_1^k \left({}^{k+m-1}C_{k-1} \right) mp_i (1 - p_1)^{m-1}$$

which is equal to

$$p_1^k p_i (1 - p_1)^{-1} \left[\sum_{m=0}^{\infty} \left({}^{k+m-1}C_m \right) m (1 - p_1)^m \right].$$

Since by our first combinatorial identity, $\sum_{i=0}^{\infty} \left({}^{k+i-1}C_i \right) i\, b^i = kb / (1 - b)^{k+1}$, this equals

$$p_1^k p_i (1 - p_1)^{-1} k \frac{(1 - p_1)}{(1 - (1 - p_1))^{k+1}},$$

which simplifies to

$$\frac{p_1^k p_i k}{p_1^{k+1}}$$

which, as expected, is

$$\frac{kp_i}{p_1}.$$

We leave the exploration of the more general case of the policy ‘stop when k_i children of the i th sex are born’ to the interested reader.

Acknowledgement

Thanks to the referee for comments.

A. C. PASEAU

Wadham College, University of Oxford, Oxford OX1 3PN

e-mail: *alexander.paseau@wadh.ox.ac.uk*