The carry-on-until-one-girl proportion

MARTIN GRIFFITHS

Introduction

An imaginary society adopts the rule that every couple has children until they have a girl (and then have no more). What is the expected proportion of females in the population comprising the first-generation offspring? Under certain assumptions, the answer to this depends on how many couples belong to the society. Throughout this note we will assume the following:

1. All couples are able to produce children, and remain fertile for as long as it takes to for a daughter to be born (we are therefore not imposing an upper limit on the age of people in the society).
2. Only one child is born at a time (no twins, triplets, and so on).
3. On the conception of each child we have P(boy) = P(girl) = $\frac{1}{2}$.

In addition to obtaining a general solution, we highlight several misconceptions that can arise when students tackle such problems.

This note is related to several interesting pieces that have appeared in the *Gazette* recently [1, 2, 3] in connection with 'stopping policies' of the type described above. In these articles the authors were, amongst other things, concerned with the expected number of boys, the expected number of girls and the ratio of these expected values in a society subject to such policies. One might, however, also be interested in the expected values of the ratios or the proportions, the latter of which are dealt with here.

Some initial calculations

Suppose that the society comprises n couples. Let M_n and F_n denote the total number of males and females, respectively, arising as first-generation offspring from these n couples. We then define P_n to be the random variable representing the proportion of females in this offspring population. It is therefore the case that

$$P_n = \frac{F_n}{F_n + M_n}.$$

Furthermore, let p_n denote $\mathrm{E}(P_n)$, the expected proportion of females in the offspring population.

We start by considering the situation for a population comprising just one couple. First, note that

$$\mathrm{E}(M_1) = 0 \times \frac{1}{2} + 1 \times \frac{1}{2^2} + 2 \times \frac{1}{2^3} + \dots = 1$$

and

$$\mathrm{E}(F_1) = 1 \times \frac{1}{2} + 1 \times \frac{1}{2^2} + 1 \times \frac{1}{2^3} + \dots = 1.$$

https://doi.org/10.1017/mag.2015.82 Published online by Cambridge University Press

Might we infer from this that the expected proportion of females arising from one family is 50%? Well, this would be the case if it were generally true that, for random variables X and Y,

$$\mathrm{E}\left(\frac{X}{X + Y}\right) = \frac{\mathrm{E}(X)}{\mathrm{E}(X) + \mathrm{E}(Y)}.$$

However, this statement is false in general. Whilst expectation is certainly a linear operator, this property cannot be extended in general to deal with non-linear functions of random variables in the manner shown above.

It is reasonably straightforward though to calculate p_1. With B and G denoting the arrival of a boy and a girl, respectively, we see that the possible sequences of births are given by G, BG, BBG, BBBG, BBBBG, and so on. The probability that the kth sequence occurs is equal to $\frac{1}{2^k}$. Note in fact that $F_n = n$ for each $n \in \mathbb{N}$, and that M_1 follows a geometric distribution given by $\mathrm{P}(M_1 = j) = \frac{1}{2^{j+1}}$ for $j \geqslant 0$. The expected proportion of females in the offspring population is given by

$$p_1 = \mathrm{E}\left(\frac{1}{1 + M_1}\right) = \sum_{k=1}^{\infty} \frac{1}{2^k k} = \log 2 \approx 69\%.$$

Students may be tempted to use the random variable

$$\frac{1}{n}\left(P_1(1) + P_1(2) + \ldots + P_1(n)\right) \tag{1}$$

as an estimator for p_n, where $P_1(1) + P_1(2) + \ldots + P_1(n)$ is a sum of n random variables, each identically and independently distributed as P_1. If this was an unbiased estimator, then, since $p_1 = \log 2$, it would be the case that $p_n = \log 2$ for each $n \in \mathbb{N}$. However, for $n \geqslant 2$, (1) is in fact a biased estimator for p_n, by which we mean that

$$\mathrm{E}\left(P_1(1) + P_1(2) + \ldots + P_1(n)\right) \neq p_n.$$

That this is the case will be shown in due course.

We now calculate p_2. Note that the outcomes for two couples occur in pairs, and that M_n no longer follows a geometric distribution for $n \geqslant 2$. We need to consider all possible outcomes, the first few of which are given below:

(G)(G)
(G)(BG), (BG)(G)
(G)(BBG), (BG)(BG), (BBG)(G)
(G)(BBBG), (BG)(BBG), (BBG)(BG), (BBBG)(G)
⋮

There is exactly one outcome comprising two Gs, two outcomes comprising two Gs and one B, and, in general, k outcomes consisting of two Gs and $k - 1$ Bs, each of which has a probability of $\frac{1}{2^{k+1}}$ of occurring. We therefore have

https://doi.org/10.1017/mag.2015.82 Published online by Cambridge University Press

$$\begin{aligned} p_2 &= \mathrm{E}\left(\frac{2}{2 + M_2}\right) \\ &= \sum_{k=1}^{\infty} \frac{2k}{2^{k+1}(k + 1)} \\ &= \sum_{k=1}^{\infty} \frac{1}{2^k} - 2\sum_{k=1}^{\infty} \frac{1}{2^{k+1}(k + 1)} \\ &= \sum_{k=1}^{\infty} \frac{1}{2^k} - 2\left(\sum_{k=1}^{\infty} \frac{1}{2^k k} - \frac{1}{2}\right) \\ &= 2 - 2\log 2 \\ &\approx 61\%. \end{aligned}$$

Note that from this result we may infer that (1) is not in general an unbiased estimator for p_n.

For three couples the outcomes arise as triplets. There is exactly one outcome comprising three Gs, three outcomes comprising three Gs and one B, and, in general, $\binom{k + 1}{2}$ outcomes consisting of three Gs and $k - 1$ Bs. Therefore

$$p_3 = \mathrm{E}\left(\frac{3}{3 + M_3}\right) = \sum_{k=1}^{\infty} \frac{3\binom{k + 1}{2}}{2^{k+2}(k + 2)} = 3\log 2 - \frac{3}{2} \approx 58\%.$$

Generalising

Let us suppose now that there are n couples. A particular outcome will comprise n girls and m boys for some non-negative integer m. We are thus left with the task of enumerating the ways in which the m boys may be distributed amongst the n couples. This corresponds to the number of ways in which m may be written as an ordered sum of n non-negative integers, which is in turn equal to $\binom{m + n - 1}{m}$. The expected proportion is thus given by

$$p_n = \mathrm{E}\left(\frac{n}{n + M_n}\right) = \sum_{k=1}^{\infty} \frac{n\binom{k + n - 2}{n - 1}}{2^{k+n-1}(k + n - 1)}.$$

It is in fact the case that p_n tends to 50% as n increases without limit, and we provide here a heuristic explanation of this. Suppose that there are n couples, where n is a very large positive integer, and that each has had just their first child. Although the number of boys born will not necessarily be

https://doi.org/10.1017/mag.2015.82 Published online by Cambridge University Press

close to the number of girls born in an absolute sense, since n is large, we would expect the proportion of girls amongst these n children to be very close to one half. Thus, approximately $n/2$ of the couples stop having children at this point whilst the remaining couples, of which there will be roughly $n/2$, continue in their quest for a girl. Then, since $n/2$ might still be regarded as large, we would expect the proportion of girls amongst the second-born children to be close to one half once more. Thus, approximately at least, a further $n/4$ couples stop having children at this point. The remaining $n/4$ couples have a third child, and so on. From this it follows that we would expect the total number of girls born to be roughly

$$\frac{n}{2} + \frac{n}{4} + \frac{n}{8} + \ldots = n,$$

and similarly for the number of boys born. This gives the expected proportion of girls born as approximately one half.

Finally, in [1, 2, 3] the authors are interested in society-wide ratios, and hence in situations for which there are many families. We show here that in this case the proportion of expected values provides a good approximation to the expected value of the proportion. From the definition of M_n, it is clear that

$$M_n = M_1(1) + M_1(2) + \ldots + M_1(n),$$

where the right-hand side is a sum of n random variables, each identically and independently distributed as M_1. Then, since $\mathrm{E}(M_1) = 1$, it follows that $\mathrm{E}(M_n) = n$. Therefore the proportion of expected values is given by

$$\frac{\mathrm{E}(F_n)}{\mathrm{E}(F_n + M_n)} = \frac{\mathrm{E}(F_n)}{\mathrm{E}(F_n) + \mathrm{E}(M_n)} = \frac{n}{2n} = \frac{1}{2}.$$

From this and our observations above, we see that the expected value of the proportion tends to the proportion of the expected values as n increases without limit.

Acknowledgement

The author would like to thank an anonymous referee for providing a number of helpful comments and suggestions.

References

1. H. Christian and B. Trustrum, If everyone wanted girls more boys might be born, *Math. Gaz.* **98** (November 2014) pp. 396-401.
2. A. C. Paseau, Family planning, *Math. Gaz.* **95** (July 2011) pp. 213-217.
3. A. C. Paseau, The 'stop after k girls or N children' policy, *Math. Gaz.* **98** (November 2014) pp. 402-413.

doi:10.1017/mag.2015.82

MARTIN GRIFFITHS

Dept. of Mathematical Sciences, University of Essex, Colchester CO4 3SQ