Journal of Integer Sequences, Vol. 16 (2013),
Article 13.1.1

Expectations of Family Sizes Subject to Minimum Numbers of Each Gender

Martin Griffiths
Mathematical Institute
University of Oxford
Oxford OX1 3LB
United Kingdom
martin.griffiths@maths.ox.ac.uk

Alexander Bramham
Wadham College
University of Oxford
Oxford OX1 3PN
United Kingdom
alexander.bramham@wadh.ox.ac.uk

Abstract

We consider the infinite families of sequences that arise from a scenario involving the expected number of children born to a married couple when certain conditions are placed on the minimum number of each gender. We obtain both exact and asymptotic formulas for these expectations, and go on to consider the long-term behavior of sequences derived from them. For the situations studied here, it is shown that the long-term behavior of each of the resulting sequences always falls into one of two distinct categories.

1 Introduction

Suppose that a married couple plan to keep having children until the first instance at which they have at least one boy and one girl. Assuming that boys and girls are equally likely to be born, and excluding the possibility of multiple births, it is a straightforward calculation to show that that the expected number of children in such a family is 3. In [2] we considered a generalization of this scenario, and derived an expression for the expected number of children our couple would need to have in order that their family contains at least n boys and n girls.

In this article we generalize things yet further by studying the considerably more complicated scenario in which the parents keep having children until the first instance at which they have at least m boys and n girls. It is assumed throughout that only one child is born at a time (no twins, triplets, and so on) and that on the birth of each child we have $\mathbb{P}[\text{boy}] = \mathbb{P}[\text{girl}] = \frac{1}{2}$. Under these conditions, we obtain expressions for the expected number of children that would result.

Our expressions give rise to infinite families of integer sequences, and we consider here the contrasting long-term behavior of these sequences under various different scenarios. Indeed, the following definitions enable us to categorize these behaviors:

Definition 1. A sequence of integers $(x_n)_{n\geq 1}$ is termed *linear* if there exists $a, b \in \mathbb{Z}$ such that $x_n = an + b$ for all $n \in \mathbb{N}$.

Definition 2. A sequence of integers $(y_n)_{n\geq 1}$ is termed *eventually-linear* if there exists some $N \in \mathbb{N}$ and $a, b \in \mathbb{Z}$ such that $y_n = an + b$ for all $n \geq N$.

Definition 3. An increasing sequence of integers $(z_n)_{n\geq 1}$ is termed *quasi-linear* if it is not eventually-linear but, for any $n \in \mathbb{N}$, it is the case that either $z_{n+1} - z_n = a$ or $z_{n+1} - z_n = a+1$ for some fixed $a \in \mathbb{N}$.

It is worth noting here that the quasi-linear sequences we shall encounter do in fact satisfy slightly more stringent conditions than those given in Definition 3. This will be highlighted and explained in Section 6. In evaluating the expectations we obtain both exact and asymptotic formulas.

2 The expectation as a finite sum

Let the random variable $X_{m,n}$ denote the number of children born such that at least m boys and n girls result. Observe that over the first $m + n - 1$ children, it is not possible to have m boys and n girls, so $X_{m,n} > m + n - 1$. Let the random variable Y represent the number of boys born out of the first $m + n - 1$ children. It is clear that

$$\mathbb{P}[Y = k] = \frac{1}{2^{m+n-1}} \binom{m+n-1}{k}. \tag{1}$$

We obtain an expression for the expectation of $X_{m,n}$, denoted $\mathbb{E}[X_{m,n}]$, by conditioning on Y. To this end, consider the random variable $\mathbb{E}[X_{m,n} \mid Y = k]$. if there are k boys, the remaining $l = (m + n - 1) - k$ children are girls. Now, for $k < m$, we have $l \geq n$, and the couple have all the girls they desire. Otherwise, $k \geq m$ and the couple have all the boys they desire.

An inductive argument on the expectation shows that the couple must expect $2t$ children for t to be of a specified gender. In the first case $t = m - k$ for the remaining boys, giving an expectation of $(m+n-1)+2(m-k)$. In the second case $t = n-(m+n-1-k) = k-m+1$

for the remaining girls, giving an expectation of $(m+n-1)+2(k-m+1)$. This may be summarized as follows:

$$\mathbb{E}[X_{m,n} \mid Y=k] = m+n-1+\begin{cases} 2(m-k), & \text{if } k<m; \\ 2(k-m+1), & \text{if } k \geq m. \end{cases}$$

The above result, in conjunction with (1) and the formula for conditional expectation [3], leads to

$$\begin{aligned} \mathbb{E}\left[X_{m,n}\right] &= \sum_{k=0}^{m+n-1} \mathbb{E}\left[X_{m,n} \mid Y=k\right] \mathbb{P}[Y=k] \\ &= \frac{1}{2^{m+n-1}}\left(\sum_{k=0}^{m-1}(n+3m-2k-1)\binom{m+n-1}{k}\right. \\ &\qquad\qquad \left. + \sum_{k=m}^{m+n-1}(n-m+2k+1)\binom{m+n-1}{k}\right). \end{aligned}$$

This may be simplified by rearranging and using the well-known results $\sum_{k=0}^{r}\binom{r}{k}=2^r$ and $\sum_{k=0}^{r} k\binom{r}{k} = r2^{r-1}$ to give

$$\begin{aligned} \mathbb{E}\left[X_{m,n}\right] &= \frac{1}{2^{m+n-1}}\left(\sum_{k=0}^{m+n-1}(n-m+2k+1)\binom{m+n-1}{k}\right. \\ &\qquad \left. + \sum_{k=0}^{m-1}\big((n+3m-2k-1)-(n-m+2k+1)\big)\binom{m+n-1}{k}\right) \\ &= \frac{1}{2^{m+n-1}}\left(2\sum_{k=0}^{m+n-1} k\binom{m+n-1}{k} + (n-m+1)\sum_{k=0}^{m+n-1}\binom{m+n-1}{k}\right. \\ &\qquad \left. + 2\sum_{k=0}^{m-1}\big(2(m-k)-1\big)\binom{m+n-1}{k}\right) \\ &= \frac{1}{2^{m+n-1}}\left(2\cdot 2^{m+n-2}(m+n-1)+2^{m+n-1}(n-m+1)\right. \\ &\qquad \left. + 2\sum_{k=0}^{m-1}\big(2(m-k)-1\big)\binom{m+n-1}{k}\right) \\ &= 2n + \frac{1}{2^{m+n-2}}\sum_{k=0}^{m-1}\big(2(m-k)-1\big)\binom{m+n-1}{k}. \end{aligned} \tag{2}$$

3 Scenario 1

We first consider the behavior of $\mathbb{E}[X_{m,n}]$ as n increases without limit whilst keeping m fixed. To this end, let us set

$$f(m,n) = \frac{1}{2^{m+n-2}} \sum_{k=0}^{m-1} \big(2(m-k)-1\big)\binom{m+n-1}{k},$$

and examine what happens to $f(m,n)$ as n tends to infinity for any fixed $m \in \mathbb{N}$. First, $1 \leq 2(m-k)-1 < 2m$ when $0 \leq k \leq m-1$. Next, consider the binomial coefficient $\binom{m+n-1}{k}$ for fixed m and some fixed value of k such that $0 \leq k \leq m-1$. This is a polynomial in n of degree k, which we denote by $g_{k,m}(n)$. Note that there exists some $N \in \mathbb{N}$ such that $g_{m-1,m}(n) \geq g_{k,m}(n)$ for each k, $0 \leq k \leq m-1$, and for all $n > N$. Therefore, when $n > N$, it is the case that

$$\begin{aligned} f(m,n) &< \frac{2m}{2^{m+n-2}} \sum_{k=0}^{m-1} g_{k,m}(n) \\ &\leq \frac{2m^2 g_{m-1,m}(n)}{2^{m+n-2}} \\ &= \frac{m^2}{2^{m-3}} \cdot \frac{g_{m-1,m}(n)}{2^n}. \end{aligned}$$

Since the numerator and denominator of the fraction on the right exhibit polynomial and exponential growth, respectively, it follows that $f(m,n) \to 0$ as $n \to \infty$.

Thus, for any fixed $m \in \mathbb{N}$, $2n$ is an asymptotic formula for $\mathbb{E}[X_{m,n}]$ in a particularly strong sense. Not only is it true that

$$\frac{\mathbb{E}[X_{m,n}]}{2n} \to 1 \quad \text{as} \quad n \to \infty,$$

but also $\mathbb{E}[X_{m,n}] - 2n \to 0$. With $\langle x \rangle$ denoting the nearest integer to x, we are able to deduce from this that $\big(\langle \mathbb{E}[X_{m,n}] \rangle\big)_{n\geq 1}$ is an eventually-linear sequence, a result that may in fact be argued heuristically. As an example, the first 14 terms of $\big(\langle \mathbb{E}[X_{5,n}] \rangle\big)_{n\geq 1}$ are given by

$$10, 10, 11, 11, 12, 14, 15, 17, 19, 20, 22, 24, 26, 28, \ldots,$$

and is linear from the 10th term onwards.

Furthermore, by considering the expression $f(m,n)$, it may be shown that, for fixed m,

$$\mathbb{E}[X_{m,n}] = 2n + \frac{h(n)}{2^{n+m-2}(m-1)!},$$

where $h(n)$ is a monic polynomial of degree $m-1$. By way of an example,

$$\mathbb{E}[X_{5,n}] = 2n + \frac{n^4 + 22n^3 + 203n^2 + 950n + 1920}{2^{n+3}4!}.$$

4 Scenario 2

Now let a and b be fixed positive integers such that, without loss of generality, $a < b$. We consider the behavior, as r tends to infinity, of $\mathbb{E}\left[X_{m,n}\right]$ for the situation in which $(m, n) = (ra, rb)$; in other words, we look at what happens in the limit as m and n remain in fixed proportion. From (2) we have

$$\begin{aligned}
\mathbb{E}[X_{ra,rb}] &= 2rb + \frac{1}{2^{r(a+b)-2}} \sum_{k=0}^{ra-1} \big(2(ra-k)-1\big) \binom{r(a+b)-1}{k} \\
&< 2rb + \frac{2ra}{2^{r(a+b)-2}} \sum_{k=0}^{ra-1} \binom{r(a+b)-1}{k} \\
&= 2rb + 4ra \left(\frac{1}{2^{r(a+b)-1}} \sum_{k=0}^{ra-1} \binom{r(a+b)-1}{k} \right). \qquad (3)
\end{aligned}$$

Hoeffding's inequality [4, 7] gives us information concerning the cumulative size of the tails of probability distributions. For the case of a binomial distribution $S \sim \mathrm{B}(n, p)$ and $\epsilon > 0$ we have the following result [7]:

$$\mathbb{P}[S \leq n(p-\epsilon)] \leq \exp\left(-2\epsilon^2 n\right). \qquad (4)$$

On setting $p = \frac{1}{2}$, $n(p-\epsilon) = ra - 1$ and $n = r(a+b) - 1$, we obtain

$$\epsilon = \frac{r(b-a)+1}{2\big(r(a+b)-1\big)}$$

(noting that $b > a$, so that ϵ is indeed positive). Thus, from (4),

$$\begin{aligned}
\mathbb{P}[S \leq ra-1] &\leq \exp\left(-2 \left(\frac{r(b-a)+1}{2\big(r(a+b)-1\big)} \right)^2 \big(r(a+b)-1\big) \right) \\
&= \exp\left(-\frac{\big(r(b-a)+1\big)^2}{2\big(r(a+b)-1\big)} \right).
\end{aligned}$$

Since, for fixed a and b,

$$ra \exp\left(-\frac{\big(r(b-a)+1\big)^2}{2\big(r(a+b)-1\big)} \right) \to 0 \quad \text{as} \quad r \to \infty,$$

it may be seen, from (3), that when m and n remain in proportion we have both

$$\frac{\mathbb{E}[X_{ra,rb}]}{2rb} \to 1 \quad \text{and} \quad \mathbb{E}[X_{ra,rb}] - 2rb \to 0 \quad \text{as} \quad r \to \infty,$$

emphasizing here that the assumption $a < b$ is indeed necessary for this result to hold. Thus $\big(\langle \mathbb{E}[X_{ra,rb}] \rangle\big)_{n \geq 1}$ is an eventually-linear sequence, exhibiting qualitatively similar behavior to the case in which m was kept fixed. To give an example, $\big(\langle \mathbb{E}[X_{20r,21r}] \rangle\big)_{n \geq 1}$ is linear from the 242nd term onwards.

5 The expectation as an infinite sum

We use the standard definition for expectation:

$$\mathbb{E}[X] = \sum_x x\mathbb{P}[X = x]$$

in order to obtain an expression for $\mathbb{E}[X_{m,n}]$ comprising an infinite series. Note that $X_{m,n}$ can take any integer value greater than or equal to $m+n$. To calculate the probability $\mathbb{P}[X_{m,n} = k]$ we first count the number of possible strings of length k in an alphabet comprising two symbols that satisfy our requirements. This number is then multiplied by $\frac{1}{2^k}$, as each outcome occurs with this probability. To this end, consider the last child born. If this was a boy, we require precisely $m-1$ boys in the first $k-1$, and if a girl, we require precisely $n-1$ girls in the first $k-1$. Thus the number we seek is $\binom{k-1}{m-1} + \binom{k-1}{n-1}$ and the expectation can be expressed as:

$$\mathbb{E}[X_{m,n}] = \sum_{k=m+n}^{\infty} \frac{k}{2^k}\left(\binom{k-1}{m-1} + \binom{k-1}{n-1}\right).$$

Applying the identity $\binom{k-1}{m-1} = \frac{m}{k}\binom{k}{m}$ provides us with the alternative form:

$$\mathbb{E}[X_{m,n}] = \sum_{k=m+n}^{\infty} \frac{1}{2^k}\left(m\binom{k}{m} + n\binom{k}{n}\right). \tag{5}$$

In Section 6 we go on to consider the case for which $n = m+j$ for some fixed $j \in \mathbb{N}$. The formulation for $\mathbb{E}[X_{m,n}]$ as an infinite sum given above is more convenient for studying the expectation's behavior when m and n are kept a fixed distance apart. We shall also make use of a result from [2]:

$$\sum_{k=2r}^{\infty} \frac{1}{2^k}\binom{k}{r} = r\left(1 + \frac{1}{2^{2r}}\binom{2r}{r}\right), \tag{6}$$

a consequence of Stirling's formula [1, 5]:

$$\binom{2m}{m} \sim \frac{2^{2m}}{\sqrt{m\pi}}, \tag{7}$$

and a result which follows from Stirling's formula:

$$\frac{1}{2^{2m}}\binom{2m}{m} \to 0 \quad \text{as} \quad m \to \infty. \tag{8}$$

6 Scenario 3

In this final section we set $n = m + j$, where j is a fixed non-negative integer, and consider what happens to the expectation as m increases without limit, noting that a formula for the special case $j = 0$ is derived in [2], leading to sequence A159061 in the *On-line Encyclopedia of Integer Sequences* [6]. From (5) we have

$$\begin{aligned}\mathbb{E}\left[X_{m,m+j}\right] &= \sum_{k=2m+j}^{\infty} \frac{1}{2^k}\left(m\binom{k}{m} + (m+j)\binom{k}{m+j}\right) \\ &= \sum_{k=2m}^{\infty} \frac{m}{2^k}\binom{k}{m} - \sum_{k=2m}^{2m+j-1} \frac{m}{2^k}\binom{k}{m} \\ &\qquad + \sum_{k=2(m+j)}^{\infty} \frac{m+j}{2^k}\binom{k}{m+j} + \sum_{k=2m+j}^{2m+2j-1} \frac{m+j}{2^k}\binom{k}{m+j}. \end{aligned} \tag{9}$$

On considering just the finite sums above, we obtain

$$\begin{aligned} &-\sum_{k=2m}^{2m+j-1} \frac{m}{2^k}\binom{k}{m} + \sum_{k=2m+j}^{2m+2j-1} \frac{m+j}{2^k}\binom{k}{m+j} \\ &\qquad = -\frac{1}{2^{2m}} \sum_{k=0}^{j-1} \frac{m(2m+1)_k}{2^k(m+1)_k}\binom{2m}{m} + \frac{1}{2^{2m}} \sum_{k=j}^{2j-1} \frac{m(2m+1)_k}{2^k(m+1)_k}\binom{2m}{m} \\ &\qquad\qquad + \frac{j}{2^{2m}} \sum_{k=2m+j}^{2m+2j-1} \frac{(2m+1)_k}{2^k(m+1)_k}\binom{2m}{m} \\ &\qquad = \frac{m}{2^{2m}}\binom{2m}{m}\left(\sum_{k=j}^{2j-1} \frac{(2m+1)_k}{2^k(m+1)_k} - \sum_{k=0}^{j-1} \frac{(2m+1)_k}{2^k(m+1)_k}\right) \\ &\qquad\qquad + \frac{j}{2^{2m}}\binom{2m}{m} \sum_{k=2m+j}^{2m+2j-1} \frac{(2m+1)_k}{2^k(m+1)_k}, \end{aligned} \tag{10}$$

where the *Pochhammer symbol* $(t)_n$ is defined by $t(t+1)\cdots(t+n-1)$, which is also known as the *rising factorial*. Note that $(t)_0 = 1$ by definition.

Since

$$\frac{(2m+1)_k}{2^k(m+1)_k} \le 1 \tag{11}$$

for all non-negative integers k, it is the case that

$$j \sum_{k=2m+j}^{2m+2j-1} \frac{(2m+1)_k}{2^k(m+1)_k} \le j^2.$$

Then, for fixed j, it follows from (8) that

$$\frac{j}{2^{2m}}\binom{2m}{m} \sum_{k=2m+j}^{2m+2j-1} \frac{(2m+1)_k}{2^k(m+1)_k} \to 0 \quad \text{as} \quad m \to \infty. \tag{12}$$

Next, for k such that $0 \le k \le j-1$,

$$\frac{(2m+1)_{j+k}}{2^{j+k}(m+1)_{j+k}} - \frac{(2m+1)_k}{2^k(m+1)_k} = \frac{(2m+1)_k}{2^k(m+1)_k}\left(\frac{(2m+k+1)_j}{2^j(m+k+1)_j} - 1\right)$$
$$= \frac{(2m+1)_k}{2^k(m+1)_k}\left(-\frac{G_{j,k}(m)}{H_{j,k}(m)}\right), \tag{13}$$

where, for fixed j and k, $G_{j,k}(m)$ and $H_{j,k}(m)$ are polynomials in m of degrees $j-1$ and j, respectively, and, without loss of generality, $H_{j,k}(m)$ is monic. The leading coefficient of $G_{j,k}(m)$ is in fact a function of j and k, and, out of all the values of k from 0 to $j-1$, is largest when $k = j-1$ (as is easily shown to be true). The leading coefficient of $G_{j,j-1}(m)$ is given by $\frac{j(3j-1)}{4}$. From this it follows that

$$m\frac{G_{j,j-1}(m)}{H_{j,j-1}(m)} \to \frac{j(3j-1)}{4} \quad \text{as} \quad m \to \infty. \tag{14}$$

Now, using (11) and (13), we have

$$\frac{m}{2^{2m}}\binom{2m}{m}\left|\sum_{k=j}^{2j-1}\frac{(2m+1)_k}{2^k(m+1)_k} - \sum_{k=0}^{j-1}\frac{(2m+1)_k}{2^k(m+1)_k}\right|$$
$$= \frac{m}{2^{2m}}\binom{2m}{m}\left|\sum_{k=0}^{j-1}\frac{(2m+1)_k}{2^k(m+1)_k}\left(-\frac{G_{j,k}(m)}{H_{j,k}(m)}\right)\right|$$
$$\le \frac{1}{2^{2m}}\binom{2m}{m}\left|\sum_{k=0}^{j-1} m\frac{G_{j,k}(m)}{H_{j,k}(m)}\right| \tag{15}$$

We know from (14) that

$$\left|\sum_{k=0}^{j-1} m\frac{G_{j,k}(m)}{H_{j,k}(m)}\right| \to c \quad \text{as} \quad m \to \infty \tag{16}$$

for some some $c \in \mathbb{R}$ which depends on j. From (8), (15) and (16), it then follows that

$$\frac{m}{2^{2m}}\binom{2m}{m}\left(\sum_{k=j}^{2j-1}\frac{(2m+1)_k}{2^k(m+1)_k} - \sum_{k=0}^{j-1}\frac{(2m+1)_k}{2^k(m+1)_k}\right) \to 0 \quad \text{as} \quad m \to \infty. \tag{17}$$

We may now combine results (6), (7), (9), (10), (12) and (17) to give

$$\mathbb{E}\left[X_{m,m+j}\right] \sim \sum_{k=2m}^{\infty}\frac{m}{2^k}\binom{k}{m} + \sum_{k=2(m+j)}^{\infty}\frac{m+j}{2^k}\binom{k}{m+j}$$
$$= m\left(1 + \frac{1}{2^{2m}}\binom{2m}{m}\right) + (m+j)\left(1 + \frac{1}{2^{2(m+j)}}\binom{2(m+j)}{m+j}\right)$$
$$\sim m\left(1 + \frac{1}{\sqrt{m\pi}}\right) + (m+j)\left(1 + \frac{1}{\sqrt{(m+j)\pi}}\right).$$

Finally, on tidying up (remembering that j is fixed), we obtain the following asymptotic formula:

$$\mathbb{E}\left[X_{m,m+j}\right] \sim 2m\left(1+\frac{1}{\sqrt{m\pi}}\right).$$

Notice that from this result we may deduce, for fixed j, that $\left(\langle\mathbb{E}\left[X_{m,m+j}\right]\rangle\right)_{n\geq 1}$ is not an eventually-linear sequence. These sequences do, however, turn out to be quasi-linear. By way of an example we obtain, using the exact formula (5) with $j=3$, the following first few terms of $\left(\langle\mathbb{E}\left[X_{m,m+3}\right]\rangle\right)_{n\geq 1}$:

$$8, 10, 12, 15, 17, 19, 21, 23, 25, 28, 30, 32, 34, 36, 38, 40, 42, 45, \ldots$$

Such sequences are in fact quasi-linear in a particularly strong sense, as we now describe:

Definition 4. Let $(x_n)_{n\geq 1}$ be a sequence of positive integers, and suppose that the sequence $(d_n)_{n\geq 1}$ has the following property. The first x_1 terms of $(d_n)_{n\geq 1}$ comprise a finite linear sequence with common difference a for some $a \in \mathbb{N}$, then $d_{x_1+1} - d_{x_1} = a+1$, and the terms of $(d_n)_{n\geq 1}$ numbered x_1+1 to x_2 form a finite linear sequence with common difference a, and so on. Then we term $(d_n)_{n\geq 1}$ *strongly quasi-linear* if $(x_n)_{n\geq 1}$ is strictly increasing.

For example, the sequence $\left(\langle\mathbb{E}\left[X_{m,m+3}\right]\rangle\right)_{n\geq 1}$ given above is strongly quasi-linear. We have $x_1 = 3$, $x_2 = 6$, $x_3 = 8$, and so on.

7 Acknowledgement

The authors would like to thank an anonymous referee for making suggestions that have helped improved the clarity of this article.

References

[1] M. Griffiths, *The Backbone of Pascal's Triangle*, United Kingdom Mathematics Trust, 2008.

[2] M. Griffiths, How many children? *Math. Gaz.* **90** (2006), 146–149.

[3] G. Grimmett and D. Stirzaker, *Probability and Random Processes*, 3rd edition, Oxford University Press, 2001.

[4] W. Hoeffding, Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.* **58** (1963), 13–30.

[5] D. E. Knuth, *The Art of Computer Programming, Vol. 1: Fundamental Algorithms*, third edition, Addison-Wesley, 1997.

[6] N. J. A. Sloane, The On-Line Encyclopedia of Integer Sequences, `http://oeis.org/`.

[7] Wikipedia contributors, *Hoeffding's inequality*, Wikipedia, The Free Encyclopedia (2012). `http://en.wikipedia.org/wiki/Hoeffding's_inequality`.

2010 *Mathematics Subject Classification*: Primary 11B25; Secondary 05A10, 40A05.
Keywords: arithmetic progression, binomial coefficient, discrete random variable, expected value.

(Concerned with sequence A159061.)

Received December 14 2012; revised version received January 1 2013. Published in *Journal of Integer Sequences*, January 1 2013.

Return to Journal of Integer Sequences home page.