# On the Evolution of Consanguinities 

Paul G. Bamberg<br>Department of Mathematics, Harvard University, Cambridge, Massachusetts, USA<br>Email: bamberg@math.harvard.edu

How to cite this paper: Bamberg, P.G. (2017) On the Evolution of Consanguinities. Applied Mathematics, 8, 550-561.
https://doi.org/10.4236/am.2017.84044

Received: January 29, 2017
Accepted: April 27, 2017
Published: April 30, 2017

Copyright © 2017 by author and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/


#### Abstract

"Consanguinity" is a gender-neutral term for "fraternity" or "sorority." Initially a consanguinity includes $M$ male members and $F$ female members. Each week a member, chosen at random, selects a new member, always of the same gender as the member making the selection. This model for evolution is isomorphic to the classic Pólya's urn. The male and female members play the same roles as the red and black balls in the urn, and the procedure for selecting a new member is equivalent to drawing a ball from the urn, then replacing it and adding a new ball of the same color. It is well known that for Pólya's urn, the proportion of red balls in the urn is a martingale. It follows that for a consanguinity, the proportion of the membership that is male is a martingale. Furthermore, being bounded, this martingale converges to a limit. For a martingale that is the sum of independent random variables, such as a symmetric random walk, there is also a well-known second-degree martingale from which the variance of the limiting distribution can be deduced. What the author discovered, in the process of solving his own examination problem, is that a similar martingale exists also for Pólya's urn, even though in this case the number of red balls is the sum of random variables that are not independent. This new martingale can be used to calculate the variance of the limiting distribution. Traditionally, the probability that $r$ red balls will be drawn from Pólya's urn in $n$ trials is derived by a rather tricky argument involving conditional probability. This article uses an obvious but overlooked simpler approach. Pólya's formula for the probability that $m$ male members will be chosen in $n$ weeks is derived, without any mention of conditional probability, by an elementary counting argument, and its limit is shown to be a beta distribution.


## Keywords

[^0]
## 1. Genesis of the Problem

There is a long tradition of setting unsolved problems on mathematics examinations. The most famous example is Stokes's Theorem, a three-dimensional version of which appeared on the 1854 Smith's Prize Exam at Cambridge University [1], several years before a proof was published by Hankel.

The following problem appeared on the final examination in the author's course Mathematics 117, "Probability and Random Processes with Economic Applications," in December 2016.
"The initial membership of a newly founded MGSO (mixed gender social organization) is three women and one man. At the start of week $n$ (starting with week 0 ), there are $n+4$ members, of whom $M_{n}$ are male. A member is chosen at random, and he or she recruits a new member of the same gender as himself/herself.

Show that the fraction of the membership that is male, the random variable

$$
\mu_{n}=\frac{M_{n}}{n+4},
$$

is a martingale. State what property of this martingale guarantees that it converges almost surely, and describe the probability distribution for $\mu_{\infty}$."

An equivalent problem was stated and solved by Eggenberger and Pólya in 1923 [2]. Instead of a consanguinity with male and female members, they considered an urn which initially, for the special case in the exam problem, contained $R=1$ red balls and $S=3$ ( $S$ for "Schwarz") black balls. A ball was repeatedly chosen at random from the urn and then replaced, along with another ball of the same color.

Since the students had done examples based on Pólya's urn and had studied the martingale convergence theorem, everything was straightforward except for "describe the probability distribution for $\mu_{\infty}$," which the author of the exam question had not taken the trouble to work out!

This variant of Pólya's urn is of political interest at Harvard because of sanctions against single-gender social organizations that have been proposed for the 2017-18 academic year [3]. An existing single-gender organization can presumably avoid the sanctions by adding a single member of the opposite gender, and the proposed model for expanding the membership is not unreasonable.

The term "consanguinity" for an MGSO whose name consists of Greek capital letters was suggested to the author by Prof. Richard Thomas of the Harvard Classics Department. Instead of saying " $X$ is my fraternity brother" or " $Y$ is my sorority sister," one can say " $Z$ is my consanguinity sibling."

None of the nine students who took the exam correctly guessed the describe the probability distribution for $\mu_{\infty}$. Perhaps James Clerk Maxwell, who tied for first on the 1854 Smith's Prize Exam, would have fared better.

Pólya's urn has developed since 1923 into an entire branch of mathematics. It is the subject of a recent book [4]. An excellent online resource that contains most of the key results is [5].

The rest of the paper is organized as follows. The well-known proof that the proportion of the membership $\mu_{n}$ that is male is a convergent martingale is
reviewed in Section 2 and is supplemented by a second martingale that involves $M_{n}^{2}$. In Section 3, the probability distribution for $M_{n}$ is derived by an elementary counting argument, and it is shown that for large $n$, the probability of adding $m$ male members is proportional to $(1-m / n)^{2}$, consistent with the fact that a beta distribution with parameters 1 and 3 has the density function $3(1-x)^{2}$. In Section 4, it is shown that the expected number of weeks before a second male member is added is infinite. In Section 5, the general case of starting with $M$ male members and $F$ female members is analyzed.

## 2. Martingale Analysis

We first specify the sample space $\Omega$ and probability measure for an experiment in which a consanguinity evolves for $n$ weeks. An element $\omega$ of this sample space is a sequence $\left(\omega_{1}, \omega_{2}, \cdots, \omega_{n}\right)$ where $\omega_{1}$ specifies which of the four original members selects the member added in week $1, \omega_{2}$ specifies which of the five members present after week 1 selects the member added in week 2 , and so forth. We specify not merely whether the new member added in week $i$ is male or female, but also which existing member made the selection. Our probability model is then that each $\omega$ is equally likely.

Let the random variable $M_{n}$ denote the number of males in the consanguinity after $n$ weeks. $M_{n}$ is a function on the sample space, whose value for each element $\omega$ can be determined by identifying the gender of each member who makes a selection. After $n$ weeks the total membership is $n+4$. The ratio

$$
\mu_{n}=\frac{M_{n}}{n+4}
$$

is also a random variable, the proportion of the membership that is male after $n$ members have been added.
We next review briefly the fundamentals of conditional expectation and martingales, a simplification of sections 8.2, 9.1 and 9.2 of the textbook used in Math117. [6]

Let $\Omega$ be a sample space, which in the case at hand is finite. Let $A$ be an event (a subset of $\Omega$ ), which in our case might be "after four weeks there are three male members." Then the conditional expectation $\mathbb{E}[X \mid A]$ is computed by restricting the sample space to the subset $A$ and calculating expectation in the usual way for this restricted sample space.

The random variable $M_{n}$ generates a finite partition of the sample space into events of the form $M_{n}=k$, i.e. "after $n$ weeks, the consanguinity includes $k$ male members." Since these are the only events of interest to us, we need make no mention of sigma-fields and can use the simplified notation

$$
\mathbb{E}\left[X \mid M_{n}\right](\omega)
$$

Since the number of males $M_{n}$ and the proportional of males $\mu_{n}$ both generate the same partition of the sample space, "conditioning on $M_{n}$ " and "conditioning on $\mu_{n}$ " are equivalent.

This conditional expectation defines a new random variable as follows: given an element $\omega$ of the sample space, we determine the value $M_{n}(\omega)$, then compute the expectation of $X$ using only those elements $\omega$ that lead to that value.

We cite the three laws of conditional expectation, a simplified version of Proposition 8.8 from [6]

1) Taking out what is known (TOWIK):

If $X$ is constant on each level set of $\mu_{n}$, then

$$
\begin{equation*}
\mathbb{E}\left[X \cdot Y \mid \mu_{n}\right]=X \mathbb{E}\left[Y \mid \mu_{n}\right] \tag{1}
\end{equation*}
$$

2) Independence drops out (IDO):

If $X$ is independent of $\mu_{n}$, then

$$
\begin{equation*}
\mathbb{E}\left[X \mid \mu_{n}\right]=\mathbb{E}[X] \tag{2}
\end{equation*}
$$

3) Tower Law:

If $m<n$, then

$$
\begin{equation*}
\mathbb{E}\left[\mathbb{E}\left[X \mid \mu_{n}\right] \mid \mu_{m}\right]=\mathbb{E}\left[X \mid \mu_{m}\right] \tag{3}
\end{equation*}
$$

The term "martingale" was in use by roulette players long before it found its way into mathematics. Imagine that you are playing roulette in a fair casino where there is no zero on the wheel. If after $n$ plays you have $k$ chips and bet one chip on red, you are equally likely to end up with $k+1$ chips or $k-1$ chips, for an expectation of $k$ chips. If random variable $Y_{n}$ is the size of your stack of chips after $n$ plays, then the sequence $Y_{n}$ has the martingale property $\mathbb{E}\left[Y_{n+1} \mid Y_{n}\right]=Y_{n}$.

For a consanguinity, the number of male members $M_{n}$, which can never decrease, is not a martingale. However, even someone who is unfamiliar with the term "martingale" is likely to guess correctly that the average number of male members is going to remain one-fourth, for the simple reason that the probability of choosing a new male member is equal to the probability that the new member is selected by a male, which in turn is equal to the proportion of the membership that is male.

More formally, we need to prove that the sequence of random variables $\mu_{n}$ has the martingale property: the expectation of $\mu_{n+1}$, conditioned on $\mu_{n}$, is equal to $\mu_{n}$, or

$$
\mathbb{E}\left[\mu_{n+1} \mid \mu_{n}\right]=\mu_{n}
$$

First we consider the case $n=0$. The random variable $\mu_{0}$ has the constant value $\frac{1}{4}$. For $\mu_{1}$ there are two possible values. With probability $\frac{1}{4}$, a new male member is selected, and the proportion of males rises to $\frac{2}{5}$. With probability $\frac{3}{4}$, a new female member is selected, and the proportion of males drops to $\frac{1}{5}$. The expectation of $\mu_{1}$ is therefore

$$
\left(\frac{1}{4}\right)\left(\frac{2}{5}\right)+\left(\frac{3}{4}\right)\left(\frac{1}{5}\right)=\frac{2+3}{20}=\frac{1}{4}
$$

The general proof is scarcely more difficult. Introduce a new random variable $X_{n}$, the number of male members chosen in week $n$. This random variable has the value 1 with probability $\mu_{n}$, the value 0 with probability $1-\mu_{n}$.

Clearly $M_{n+1}=M_{n}+X_{n}$, and since expectation is linear,

$$
\mathbb{E}\left[M_{n+1} \mid \mu_{n}\right]=\mathbb{E}\left[M_{n} \mid \mu_{n}\right]+\mathbb{E}\left[X_{n} \mid \mu_{n}\right]
$$

For the first term, conditioning $M_{n}$ on a function of itself changes nothing (TOWIK) (1). For the second term, the expectation of $X_{n}$ is simply the probability that it equals 1 rather than 0 , namely $\mu_{n}$, so

$$
\mathbb{E}\left[M_{n+1} \mid \mu_{n}\right]=M_{n}+\mu_{n}=(n+4) \mu_{n}+\mu_{n}
$$

Dividing by $n+5$ we find

$$
\mathbb{E}\left[\mu_{n+1} \mid \mu_{n}\right]=\mu_{n}
$$

which is the martingale property.
By the Tower $\operatorname{Law}(3)$, we can condition $\mu_{n+1}$ first on $\mu_{n}$, then on $\mu_{n-1}$, and eventually on $\mu_{0}$ and conclude that

$$
\begin{equation*}
E\left[\mu_{n+1} \mid \mu_{0}\right]=\mu_{0}=\frac{1}{4} \tag{4}
\end{equation*}
$$

reaching a conclusion that requires half a page of computation in [2], page 281.

This analysis appears in almost every introduction to martingales; for example, as exercise 27 on page 258 of [7]. What appears not to be well known is that there is also a second-degree martingale that makes the computation of variance equally easy.

The inspiration for this second martingale is a symmetric random walk, as described, for example, in Example 9.6 of [6].

Let $X_{i}$ denote the $i$ th step in the walk: it is a random variable that has values +1 and -1 , each with probability $1 / 2$.

Let

$$
Y_{n}=\sum_{i=1}^{n} X_{i}
$$

Then

$$
\mathbb{E}\left[Y_{n+1} \mid Y_{n}\right]=\mathbb{E}\left[\left(Y_{n}+X_{n+1}\right) \mid Y_{n}\right]=\mathbb{E}\left[Y_{n} \mid Y_{n}\right]+\mathbb{E}\left[X_{n+1} \mid Y_{n}\right]
$$

By TOWIK (1),

$$
\mathbb{E}\left[Y_{n} \mid Y_{n}\right]=Y_{n}
$$

By IDO (2),

$$
\mathbb{E}\left[X_{n+1} \mid Y_{n}\right]=\mathbb{E}\left[X_{n+1}\right]=0
$$

and we conclude that

$$
\mathbb{E}\left[Y_{n+1} \mid Y_{n}\right]=Y_{n} .
$$

For a second-degree martingale, define a new random variable, $Z_{n}=Y_{n}^{2}-n$ Then

$$
Z_{n+1}^{2}=\left(Y_{n}+X_{n+1}\right)^{2}-(n+1)=Y_{n}^{2}+2 Y_{n} X_{n+1}+X_{n+1}^{2}-n-1
$$

Again condition on $Y_{n}$.
By TOWIK(1),

$$
\mathbb{E}\left[Y_{n}^{2} \mid Y_{n}\right]=Y_{n}^{2}
$$

By IDO (2),

$$
\mathbb{E}\left[X_{n+1} Y_{n} \mid Y_{n}\right]=\mathbb{E}\left[X_{n+1} \mid Y_{n}\right]=0
$$

Finally, $X_{n+1}^{2}$ is the constant random variable 1.
So

$$
\mathbb{E}\left[Z_{n+1} \mid Y_{n}\right]=Y_{n}^{2}+1-n-1=Z_{n}
$$

which is the martingale property.
In the case of the consanguinity we can write

$$
M_{n}=M_{0}+\sum_{i=1}^{n} X_{i}
$$

where $X_{i}$ is the number of male members ( 0 or 1 ) added in week $i$. However, in this case $X_{n+1}$ is no longer independent of $M_{n}$-indeed this lack of independence is precisely why Eggenberger and Pólya [2] found the urn problem interesting.

It took a bit of experimentation for the author to discover a second degree martingale even for the simple special case that appeared on the examination.

The martingale turns out to be

$$
\begin{equation*}
Z_{n}=\frac{1}{(n+4)(n+5)}\left(M_{n}^{2}+M_{n}\right) \tag{5}
\end{equation*}
$$

roof: consider

$$
Z_{n+1}=\frac{1}{(n+5)(n+6)}\left(M_{n+1}^{2}+M_{n+1}\right)
$$

Since

$$
\begin{gathered}
M_{n+1}=M_{n}+X_{n+1} \\
Z_{n+1}=\frac{1}{(n+5)(n+6)}\left[M_{n}^{2}+2 M_{n} X_{n+1}+X_{n+1}^{2}+M_{n}+X_{n+1}\right]
\end{gathered}
$$

This time it will be more difficult to deal with the second term, because $X_{n+1}$ is not independent of $M_{n}$.

Again we condition everything on $M_{n}$. The key is that

$$
\mathbb{E}\left[X_{n+1} \mid M_{n}\right]=\mu_{n}=\frac{M_{n}}{n+4}
$$

By TOWIK (1),

$$
\mathbb{E}\left[M_{n}^{2} \mid M_{n}\right]=M_{n}^{2}
$$

By TOWIK (1),

$$
\mathbb{E}\left[M_{n} X_{n+1} \mid M_{n}\right]=M_{n} \mathbb{E}\left[X_{n+1} \mid M_{n}\right]=M_{n} \mu_{n}=\frac{M_{n}^{2}}{n+4}
$$

Since $X_{n+1}$ equals 0 or 1 , its square is equal to itself, and

$$
\mathbb{E}\left[X_{n+1}^{2} \mid M_{n}\right]=\mathbb{E}\left[X_{n+1} \mid M_{n}\right]=\mu_{n}=\frac{M_{n}}{n+4}
$$

Combining all the terms we find

$$
\mathbb{E}\left[Z_{n+1} \mid M_{n}\right]=\frac{1}{(n+5)(n+6)}\left(M_{n}^{2}\left(1+\frac{2}{n+4}\right)+M_{n}\left(1+\frac{2}{n+4}\right)\right)
$$

or

$$
\begin{equation*}
\mathbb{E}\left[Z_{n+1} \mid M_{n}\right]=\frac{1}{(n+5)(n+6)} \frac{n+6}{n+4}\left(M_{n}^{2}+M_{n}\right)=Z_{n} \tag{5}
\end{equation*}
$$

which establishes the martingale property for the random variable $Z_{n}$
By the Tower Law (3), therefore,

$$
\mathbb{E}\left[Z_{n}\right]=\mathbb{E}\left[Z_{n} \mid M_{0}\right]=Z_{0}=\frac{1}{20}(1+1)=\frac{1}{10}
$$

It follows that

$$
\frac{1}{(n+4)(n+5)}\left(\mathbb{E}\left[M_{n}^{2}\right]+\mathbb{E}\left[M_{n}\right]\right)=\frac{1}{10}
$$

But we already know (4) that

$$
\mathbb{E}\left[M_{n}\right]=\frac{n+4}{4}
$$

and so

$$
\mathbb{E}\left[M_{n}^{2}\right]=\frac{(n+4)(n+5)}{10}-\frac{n+4}{4}
$$

Thus the variance is

$$
\begin{equation*}
\operatorname{Var} M_{n}=\mathbb{E}\left[M_{n}^{2}\right]-\mathbb{E}\left[M_{n}\right]^{2}=\frac{3 n(n+4)}{80} \tag{6}
\end{equation*}
$$

in agreement with the value that Eggenberger and Pólya calculate from the probability mass function [2].

The Martingale Convergence Theorem, stated and proved, for example, in section 9.2 of [6], asserts that if a sequence of random variables like $\mu_{n}$ is uniformly integrable, it converges almost surely to a random variable $\mu_{\infty}$. Since each $\mu_{n}$ is bounded below by 0 and above by 1 , the requirement of uniform integrability is trivially satisfied.

Although we have still not determined the probability mass function, we now know enough to determine the expectation and variance of $\mu_{\infty}$. Since $\mu_{n}$ has an expectation of $\frac{1}{4}$ for all $n$, so does $\mu_{\infty}$. From the results that we have just
obtained, combined with the fact that $\mu_{n}=\frac{M_{n}}{n+4}$, it follows that

$$
\operatorname{Var} \mu_{\infty}=\lim _{n \rightarrow \infty} \frac{3 n}{80(n+4)}=\frac{3}{80}
$$

At this point we might conjecture an answer to the original exam question, "describe the probability distribution for $\mu_{\infty}$." Perhaps the limit is a beta distribution with parameters $\alpha=1$ and $\beta=3$, for which the density function is $f(x)=3(1-x)^{2}$. Any book on mathematical statistics, e.g. [8] has the formulas for the expectation and variance of the beta distribution. The expectation is

$$
\frac{\alpha}{\alpha+\beta}=\frac{1}{4}
$$

while the variance is

$$
\frac{\alpha \beta}{(\alpha+\beta)^{2}(\alpha+\beta+1)}=\frac{3}{80} .
$$

So the conjectured beta distribution has the right expectation and variance. Rather than attempt to compare higher moments, though, it will simpler just to work out the probability mass function and take its limit.

## 3. The Probability Mass Function

The following derivation is equivalent to that of Eggenberger and Pólya [2], but it is strictly an elementary counting argument, making no mention of the fact that the number of males $M_{n}$ is the sum of a set of random variables which, although not independent, are nonetheless "exchangeable."

We have already identified a finite sample space whose elements all have equal probability. The number of elements in the sample space is the product of the number of equally likely choices that can be made in successive weeks: namely

$$
4 \times 5 \times 6 \cdots(3+n)=\frac{(n+3)!}{3!}
$$

Now we must count the number of elements of the sample space whose effect is to add $m$ males and $f=n-m$ females to the membership.

For the first male added there is only one candidate for the member who did the choosing; for the second male added, there are two candidates, and so forth. Given a fixed set of $m$ weeks in which males are chosen, there are in all $(m-1)$ ! alternatives. Similarly, for the weeks in which females are chosen, there are $3 \times 4 \times 5 \cdots(2+f)$ alternatives for the female choosers.

There are $\binom{n}{m}$ ways to select the set of $m$ weeks in which a male is chosen. So the number of elements of the sample space that contribute to the event " $m$ males are added" is

$$
\binom{n}{m}(m-1)!\frac{(f+2)!}{2!}
$$

The probability of adding $m$ males is simply the proportion of elements in the sample space that contribute to this event: namely
$\frac{\frac{n!}{m!(n-m)!}(m-1)!\frac{(n-m+2)!}{2!}}{\frac{(n+3)!}{3!}}=\frac{3(n-m)(n-m-1)}{n(n+1)(n+2)}=\frac{3\left(1-\frac{m}{n}\right)\left(1-\frac{m}{n}-\frac{1}{n}\right)}{n\left(1+\frac{1}{n}\right)\left(1+\frac{2}{n}\right)}$.
This formula agrees with Equation (6) of Eggenberger and Pólya [2], where the authors multiply conditional probabilities, then permute the numerators of the fractions that are multiplied.

We now set $\frac{m}{n}=x$ and take the limit as $n \rightarrow \infty$ by dropping all terms that approach 0 , obtaining

$$
\mathbb{P}\left(\mu_{\infty}=x\right)=\frac{3}{n}(1-x)^{2}
$$

which corresponds to the density function for the beta distribution with parameters 1 and 3,

$$
\begin{equation*}
f(x)=3(1-x)^{2} \tag{8}
\end{equation*}
$$

## 4. Waiting for the Second Male Member

The probability that only females are added over a period of $n$ weeks is simply the ratio

$$
\frac{3 \times 4 \times 5 \cdots(2+n)}{4 \times 5 \times 6 \cdots(3+n)}=\frac{3}{n+3}
$$

Suppose that the consanguinity agrees to pay a fine of 1 dollar for each week that its membership includes only one male. What is the expected amount of the fine?

Let $1_{n}$ be the indicator function of the event "in the first $n$ weeks, all new members are female." The total fine is the random variable

$$
F=1_{1}+1_{2}+1_{3}+\cdots
$$

For example, if the first male is added in week 3, the first two of these indicator functions equal 1 , all the others are 0 , and the fine is 2 dollars.

The expectation of the fine, by linearity of expectation, is

$$
\mathbb{E}[F]=\mathbb{E}\left[1_{1}\right]+\mathbb{E}\left[1_{2}\right]+\mathbb{E}\left[1_{3}\right]+\cdots
$$

But the expectation of $1_{n}$ is just the probability of the event that only females are added over a period of $n$ weeks, so

$$
\mathbb{E}[F]=\sum_{n=1}^{\infty} \frac{3}{n+3}
$$

This series is divergent, and the expected fine is infinite!

## 5. Martingales for General Starting Conditions

If the consanguinity begins with $N$ members, $M$ males and $F$ females, the
proportion of male members is still a martingale:

$$
\mu_{n}=\frac{M_{n}}{n+N}
$$

Furthermore, the second-degree martingale is the simplest imaginable generalization: 4 gets replaced by $N$.

$$
\begin{equation*}
Z_{n}=\frac{1}{(n+N)(n+N+1)}\left(M_{n}^{2}+M_{n}\right) \tag{9}
\end{equation*}
$$

We repeat the proof, considering

$$
Z_{n+1}=\frac{1}{(n+N+1)(n+N+2)}\left(M_{n+1}^{2}+M_{n+1}\right)
$$

Since

$$
\begin{gathered}
M_{n+1}=M_{n}+X_{n+1} \\
Z_{n+1}=\frac{1}{(n+N+1)(n+M+2)}\left[M_{n}^{2}+2 M_{n} X_{n+1}+X_{n+1}^{2}+M_{n}+X_{n+1}\right]
\end{gathered}
$$

Again we condition everything on $M_{n}$ and use

$$
\mathbb{E}\left[X_{n+1} \mid M_{n}\right]=\mu_{n}=\frac{M_{n}}{n+N}
$$

By TOWIK (1),

$$
\mathbb{E}\left[M_{n}^{2} \mid M_{n}\right]=M_{n}^{2} .
$$

By TOWIK (1),

$$
\mathbb{E}\left[M_{n} X_{n+1} \mid M_{n}\right]=M_{n} \mathbb{E}\left[X_{n+1} \mid M_{n}\right]=M_{n} \mu_{n}=\frac{M_{n}^{2}}{n+N} .
$$

Since $X_{n+1}$ equals 0 or 1 , its square is equal to itself, and

$$
\mathbb{E}\left[X_{n+1}^{2} \mid M_{n}\right]=\mathbb{E}\left[X_{n+1} \mid M_{n}\right]=\mu_{n}=\frac{M_{n}}{n+N}
$$

Combining all the terms we find

$$
\mathbb{E}\left[Z_{n+1} \mid M_{n}\right]=\frac{1}{(n+N+1)(n+N+2)}\left(M_{n}^{2}\left(1+\frac{2}{n+N}\right)+M_{n}\left(1+\frac{2}{n+N}\right)\right)
$$

or

$$
\mathbb{E}\left[Z_{n+1} \mid M_{n}\right]=\frac{1}{(n+N+1)(n+N+2)} \frac{n+N+2}{n+N}\left(M_{n}^{2}+M_{n}\right)=Z_{n}
$$

which establishes the martingale property for $Z_{n}$ (10).
With more than one initial male member, the calculation of variance becomes slightly more complicated, since $Z_{0}$ depends on the number of initial male members. By the Tower Law (3),

$$
\mathbb{E}\left[Z_{n}\right]=\mathbb{E}\left[Z_{n} \mid M_{0}\right]=Z_{0}=\frac{1}{N(N+1)}\left(M^{2}+M\right)=\frac{M(M+1)}{N(N+1)}
$$

It follows that

$$
\frac{1}{(n+N)(n+N+1)}\left(\mathbb{E}\left[M_{n}^{2}\right]+\mathbb{E}\left[M_{n}\right]\right)=\frac{M(M+1)}{N(N+1)}
$$

But we already know that

$$
\mathbb{E}\left[M_{n}\right]=\frac{(n+N) M}{N}
$$

and so

$$
\begin{align*}
\mathbb{E}\left[M_{n}^{2}\right] & =\frac{(n+N)(n+N+1) M(M+1)}{N(N+1)}-\frac{(n+N) M}{N}  \tag{10}\\
& =\frac{(n+N) M}{N} \frac{n(M+1)+M(N+1)}{N+1}
\end{align*}
$$

Thus the variance is

$$
\operatorname{Var} M_{n}=\mathbb{E}\left[M_{n}^{2}\right]-\mathbb{E}\left[M_{n}\right]^{2}=\frac{(N-M) n M(n+N)}{N^{2}(N+1)}
$$

in agreement with the value that Eggenberger and Pólya [2] calculate from the probability mass function. Dividing by $n+N$ and taking the limit as the limit as $n \rightarrow \infty$, we find

$$
\operatorname{Var} \mu_{\infty}=\frac{(N-M) M}{N^{2}(N+1)},
$$

which is indeed the variance for a beta distribution with parameters $\alpha=M, \beta=F=N-M$.

## 6. Conclusions

The original examination question has been solved at last.

- The proportion of male members, $\mu_{n}$, is a martingale whose expectation is $\mathbb{E}\left(\mu_{n}\right)=1 / 4$ (4) for all $n$.
- Being bounded below by 0 and above by 1 , this martingale has a limit as $n \rightarrow \infty$.
- Without working out the distribution function for $\mu_{n}$, a second martingale (5) can be used to show that the variance of the limiting distribution is $\operatorname{Var} \mu_{\infty}=3 / 80$ (6), which is consistent with the conjecture that this limiting distribution is a beta distribution.
- The distribution of the number of male members (7) can be determined by an elementary counting argument, and its limit is a beta distribution(8).
- The newly found second-degree martingale generalizes easily (10) to arbitrary starting conditions.


## Acknowledgements

I thank the Editor and the referee for their comments. I also thank Dean Rakesh Khurana, who made me aware of the issue of fraternities and sororities at Harvard, my student Morgan Breitmeyer, with whom I had enlightening discussions about sororities on campus, and my student Rebecca Jarvis, who reads German and was able to confirm my interpretation, initially done with the aid of Google Translate, of Eggenberger and Pólya's original article.

## References

[1] From 1854: Smith's Prize Exam. http://www.clerkmaxwellfoundation.org/SmithsPrizeExam_Stokes.pdf.
[2] Eggenberger, F. and Pólya, G. (1923) Über die Statistik verketteter Vorgänge. Z. Angew. Math. Mech., 3, 279-289. https://doi.org/10.1002/zamm. 19230030407
[3] Letter on Single-Gender Social Organizations (2016) Cambridge, Mass. http://www.harvard.edu/president/news/2016/letter-on-single-gender-social-organi zations.
[4] Mahmoud, H. (2008) Polya Urn Models. Chapman and Hall/CRC. https://doi.org/10.1201/9781420059847
[5] Polya's Urn Process. http://www.math.uah.edu/stat/urn/Polya.html.
[6] Dineen, S. (2013) Probability Theory in Finance. Graduate Studies in Mathematics, 70, 305p.
[7] Grimmett, G. and Stirzaker, D. (2011) Probability and Random Processes. Oxford University Press, New York.
[8] Chirara, L. and Hesterberg, T. (2011) Mathematical Statistics with Resampling and R. Wiley, Hoboken, New Jersey.

## Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc. A wide selection of journals (inclusive of 9 subjects, more than 200 journals)
Providing 24-hour high-quality service
User-friendly online submission system
Fair and swift peer-review system
Efficient typesetting and proofreading procedure
Display of the result of downloads and visits, as well as the number of cited articles
Maximum dissemination of your research work
Submit your manuscript at: http://papersubmission.scirp.org/
Or contact am@scirp.org


[^0]:    Pólya's Urn, Martingale, Martingale Convergence, Beta Distribution

