ANCESTORS AT THE NORMAN CONQUEST

Kenneth W. Wachter

There are few illustrations of the inevitability of math in history so apt as the popular puzzle about the number of ancestors anyone had during some great historical event like the Norman Conquest. Each of us has two parents. If we reason that two parents mean four grandparents, eight great grandparents, and twice as many ancestors more each generation back, then each of us should have had a thousand million ancestors alive when William the Conqueror stumbled up the sands toward Hastings. Carrying the fallacy a step further, the ancestors of an English person of today would have been so numerous at the time the Emperor Hadrian visited his northern province that there would have been one ancestor for every square inch of surface area of the globe -- land and ocean.

The flaw in this reasoning, saving us from having such a myriad of ancestors, is not hard to pinpoint. People can marry their cousins, and so the distinct positions in a pedigree need not be filled by distinct men and women. But the question remains, how many distinct ancestors is a person likely to have had at any time? For how many generations back does the number of distinct ancestors grow larger? When does it start to shrink? Can a typical person expect to be descended, back beyond some date in the past, from the whole population of his country of origin, or at least from everyone in the population who had descendants? In this chapter we offer answers to these questions on the basis of a simple probability model. The exercise is an amusing one, and it exemplifies points of contact between model building and less formal varieties of speculation.

As a genealogist constructs a family tree, working backwards through time, he is effectively choosing for each individual in the tree two people out of some pool of possible candidates to record as the individual's parents. If the individual is generation $g = 0$, then at generation $g = 1$ there is one father and one mother. At generation $g = 2$ there are $2^2 = 4$ positions in the family tree, 2 for males, 2 for females. At generation $g = 3$ there are $2^3$ positions in all in the tree, $2^2$ of them for males and $2^2$ of them for females. Thus $g$ generations back there are $2^g$ positions in the tree, $2^{g-1}$ for males.
We shall suppose that the \(2^g-1\) positions for males are filled by \(m(g)\) distinct male ancestors and that the \(2^g-1\) positions for females are filled by \(f(g)\) females. To fill the \(2^g\) slots in the tree for generation \(g+1\), the genealogically must find \(m(g) + f(g)\) fathers for all the distinct males and \(m(g) + f(g)\) mothers out of some collection of \(x(g+1)\) men. He must also find these fathers out of some collection of \(y(g+1)\) women. Of the \(m(g) + f(g)\) mothers some \(m(g+1)\) will be distinct men, and of the \(m(g) + f(g)\) fathers some \(f(g+1)\) will be distinct women. What we want is a model which allows us to calculate expected values for the numbers \(m(1), f(1), m(2), f(2)\), ..., going back in time.

Whenever two slots in a family tree are filled by the same person, it means that further down the tree a couple who married were in fact cousins. It seems sensible to distinguish between two sorts of cousin marriage, close and distant, and this distinction forms an essential step in our analysis. Marriages of close cousins, on the one hand, presumably involve a conscious choice of a cousin as a marriage partner, and they may reflect social preference for cousin marriage, desires to keep property in families, special opportunities for acquaintance, and similar social factors. Such choice would be systematic rather than random. Marriages of distant cousins, on the other hand, must often arise from accident, without the couple knowing of the kin relationship. If a bride's and groom's ancestors come from the same country, there being only a limited number of people alive at any point in the past in that country, then chances are that the couple have ancestors in common. Duplicate people in the slots in the family tree are likely to occur at random, with more of these random duplications the larger the number of slots to be filled and the smaller the pool of people in the population from which these slots must be filled. A probability model ought to be able to predict chances of various numbers of duplications. On these grounds we propose to regard close cousin marriage as a systematic process and distant cousin marriage as a random process.

We should mention that there are arguments for opposite points of view, emphasizing the random character of close cousin marriage or the systematic character of distant cousin marriage. For close cousin marriage, Gilbert and Hammel (1966) have demonstrated by microsimulation that in one society where elaborate cousin-preference systems have been supposed, a simple random model based on geographic propinquity would explain a large fraction of the observed close cousin marriage. For distant cousin marriage, it is easy to hypothesize patterns of social geographical stratification which would severely restrict the pool of plausible candidates from among whom not only each person's found. Genetic duplicative factors, societally instead of over loose some groupings, and perhaps even cousin in the majority.

A choice of factors is the work of one of kinsmen, and one effect, that choice is on the one hand, and the other hand, must often arise from accident, without the couple knowing of the kin relationship. If a bride's and groom's ancestors come from the same country, there being only a limited number of people alive at any point in the past in that country, then chances are that the couple have ancestors in common. Duplicate people in the slots in the family tree are likely to occur at random, with more of these random duplications the larger the number of slots to be filled and the smaller the pool of people in the population from which these slots must be filled. A probability model ought to be able to predict chances of various numbers of duplications. On these grounds we propose to regard close cousin marriage as a systematic process and distant cousin marriage as a random process.
Genealogical Demography

Our parents but his or her more distant forebears would be more interested in coefficients of inbreeding more complicated than our simple measure, the number of distinct ancestors, have collected some data on consanguinity in western societies. This work is summarized in Chapters 7 and 8 of Bodmer and Cavalli-Sforza (1971), but strictness of stratification over long time periods is not readily estimable from it. For some groups, stratification may have been so rigid, pervasive and prolonged as to be an important determinant of distant cousin marriage, but it is hard to imagine this being true for a majority of any national population.

A genealogist filling in a family tree who talks of a choice of parents or a pool of candidates eligible to be ancestors is adopting an artificial but a useful point of view. In the world where clocks run forward, brides choose grooms and grooms choose brides or parents choose them for them, but the effect, as we should see it if we ran the clock backward, is that children acquire parents, and more and more branches grow on the genealogical tree.

It may be right not only to think of the branches of the tree as spreading out through the country and through society. Even if the chance of mobility at any one generation is small, over many generations it can add up to a lot of movement. Furthermore, evidence is coming to light that geographical mobility may not have been infrequent, even in rural villages centuries ago. Chapter 2 of Laslett (1977) contains interesting material along these lines. Immigration and emigration do complicate matters, and we shall restrict our treatment of ancestors to the case of an English person of wholly English ancestry. For such a person, we shall assume that over the span of time back to 1500 ancestors diffuse through most regions of the country and most social classes. This assumption may be wrong, but it is a good one for a first attack.

In our model we measure time in generations, each generation lasting 30 years. This is a handy figure. It would be elementary to change our calculations to allow for some change in the span of a generation as we go back in time, if we could agree on some account of how the span has changed. More effort would be involved in allowing for random spans of generation, letting the same person appear not only at different positions in the genealogical tree but at different levels. Such duplications between levels imply marriages between distant cousins "several times removed," as the phrase goes in some dialects. These certainly occur, although it is not obvious that the total number of distant cousin marriages is much altered by letting them occur between cousins several times removed.
For times when the number of ancestors is large enough for a random model of "parent choice" to be appropriate, we formulate our model in the following way: Consider \( m(g) \) distinct males and \( f(g) \) distinct females in the tree at the \( g \)-th generation. At generation \( g+1 \) we have \( x(g+1) \) males and \( y(g+1) \) females from whom the parents of these people must be drawn. Think of distributing \( m(g) + f(g) \) "fatherhood markers" among the \( x(g+1) \) males, just as we might distribute \( m + f \) balls among \( x \) bins. More than one ball may go into the same bin. The first ball may land in any of \( x \) bins, the second again in any of the \( x \) bins, and so there are \( x^m x^f \) possible combinations of balls in bins or fatherhood markers on eligible males. If we distribute balls among bins at random, giving each of the \( x^m x^f \) combinations an equal chance, the probability of obtaining exactly \( u \) nonempty bins, or exactly \( u \) distinct fathers, is given by a formula on pages 60 and 102 of Feller (1968), namely

\[
\frac{x!}{u!(x-u)!} \sum_{x=0}^{u} \frac{(-1)^k u! (u-k)^{m+f}}{k!(u-k)! x^{m+f}}
\]

This is the probability that \( m(g+1) \) be equal to \( u \) given \( m(g) \) and \( f(g) \) in our model, in other words, the probability of having exactly \( u \) distinct male ancestors at the next generation back. The same formula applies to motherhood markers and female ancestors.

For this probability distribution the mean of \( u \) equals the expression

\[ x \cdot x(1 - 1/x)^{m+f} \]

Using \( \exp[z] \) to stand for the constant \( e \) raised to the \( z \)-th power, the quantity \( x(1 - \exp[(-m-f)/x]) \) is an excellent approximation to this mean when \( x \) is large. Thus, given \( m(g) \) and \( f(g) \), the expectations of \( m(g+1) \) and \( f(g+1) \) are the following functions of the sizes \( x(g+1) \) and \( y(g+1) \) of the pools of eligible fathers and mothers:

\[ E_m(g+1) = x(g+1)(1 - \exp[(-m(g) - f(g))/x(g+1)]) \]

\[ E_f(g+1) = y(g+1)(1 - \exp[(-m(g) - f(g))/y(g+1)]) \]

Under the model just cited for \( m + f \) balls in \( x \) bins, the number of empty bins \( x - u \) has an approximately Poisson probability distribution when \( m + f \) and \( x \) are large. Its variance is therefore close to its mean. Its standard deviation, being close to the square root of its mean, is very small relative to the mean.
the mean less than one thousandth of the mean for x in the million, the range of values of x that we shall have. Thus for our purposes it is fair to identify the random variables m(g) and f(g) with their mean values. Then assuming a sex ratio close to 1.00 for the sizes of the pools x(g) and y(g), we have approximate equality at every generation between the number of distinct male ancestors m(g) and the number of distinct female ancestors f(g), and we may replace m(g) + f(g) by 2m(g) and write our recurrence equation in the form

\[
\frac{m(g+1)}{x(g+1)} = 1 - \exp[-2(m(g)/x(g))(x(g)/x(g+1))]
\]

The quantity m(g)/x(g) is the proportion that actual ancestors put out of the pool of those eligible to be ancestors at the g-th generation. The equation expresses this proportion as a function of its previous value and the growth rate x(g+1)/x(g) of the pool.

We might be concerned that the probability distribution from Feller puts no constraint on the ratio of children m(g) + f(g) to parents m(g+1) + f(g+1). Our model might be unrealistic in producing arbitrarily large family sizes. It is a relief to notice, therefore, that the small standard deviation relative to the mean insures us that the probability of ratios out of line with the growth rate of the pool itself is negligible. Were we not using total population or some other pool that grew primarily by natural increase, the average family size in the model might stray from the plausible range [1].

We have now formulated our model for the effects of distant cousin marriage. We still need some assumptions about the effects of close cousin marriage. In the absence of better information on this subject, the author has consulted his own family tree researched by his mother. The first place where two sons are filled by the same person occurs six generations back, where a man named Ebenezer appears twice, being the grandfather of both Tabitha and Thomas Bishop, who are man and wife. Each of Ebenezer's parents of course then occurs twice, so there are only 2^5 - 1 or 63 distinct males and 2^6 - 1 or 63 distinct females, making 2^7 - 2 = 126 distinct people at the seventh generation due to this close cousin marriage. Whether two duplications after a stretch of six generations back from the first are too many or too few is hard to say. Other duplications occur further back in this tree. Pending better evidence, we shall allow 63 instead of 64 male ancestors at the seventh generation, and again allow 63 instead of 64 distinct male ancestors for each of these 63 males back another six generations to the thirteenth. We do the same for females. These are our fixed allowances for close cousin marriage. Beyond the thirteenth generation back, the total number of ancestors is large enough that the number of duplications predicted by the random model for
We shift to the random model at $g = 13$ with $m(g) = 63 \times 63 = 3969$ and $f(g) = 63 \times 63 = 3969$, proceeding on from there.

We begin our calculations with a child born in England of wholly English ancestry around 1947 who reaches the age of 30 in 1977. We trace the values of $m(g)$ back at intervals of 30 years. Assuming wide diffusion of ancestors throughout the society and the country by 1600, we take $x(g) + y(g)$ to be the total population of England for all earlier years. If desired, these values could be reduced by guesses at the numbers of infertile men and women, members of religious orders, and others ineligible to be the parents of each generation, is wide uncertainty in the total population figures themselves, much guesses are not likely to improve our answers very much, since not the population at each 30-year interval instead of the total population. Taking total population, however, gives some leeway for the differing ages of marriage and childbirth among different members of the family tree.

In order to carry out our exercise, we need guesses at the population of England in medieval times. Unfortunately, all such guesses are speculative. We have taken the values back to 1230 from the graph on page 386 and the remarks on page 387 of Hollingsworth (1969) and for dates before that we have interpolated between estimates of Josiah Russell (1948). The most accurate estimates pertain to 1377 and 1347. Although we should not trust any of the figures too far, they suffice for our purposes.

The results of our calculations appear in Table 1. The last column shows the generation $g$. The first column shows an average date when the ancestors in this generation are reaching the age of 30, and the second column shows our guess at the total English population $x(g) + y(g)$ at this year. The third column shows our estimate of the number of distinct ancestors $m(g) + f(g) = 2m(g)$. The proportion of ancestors in the pool, $m(g)/x(g)$, occupies the fourth column and the rate of growth of the pool, $x(g+1)/x(g)$ the fifth column. Raising $e$ to the power given by minus twice the fourth column divided by the fifth column and subtracting the result from unity gives the entry in the next row in the fourth column. In this way, given the population figures, we generate the next row of the table from the previous row. The seventh column shows $2^g$, the total number of distinct slots in the family tree.

The outcomes of our model are intriguing. Around the discovery of America, our individual has more than 60,000 distinct ancestors. Some 95% of the slots in the family tree at this time of Wycliffe and the Peasants' Revolt, at the twentieth 600,000, and nearly a third of the slots in the tree are filled
Table II: Predictions of numbers of distinct ancestors at each generation.

<table>
<thead>
<tr>
<th>Year</th>
<th>m(g) + p(g)</th>
<th>m(g) + f(g)</th>
<th>m(g) / x(g)</th>
<th>x(g+1) / x(g)</th>
<th>2h</th>
<th>3h</th>
</tr>
</thead>
<tbody>
<tr>
<td>1047</td>
<td>2,500,000</td>
<td>7,938</td>
<td>.0022</td>
<td>.829</td>
<td>2h</td>
<td>1h</td>
</tr>
<tr>
<td>1047</td>
<td>2,900,000</td>
<td>15,633</td>
<td>.0054</td>
<td>.759</td>
<td>1h</td>
<td>2h</td>
</tr>
<tr>
<td>1047</td>
<td>2,300,000</td>
<td>31,438</td>
<td>.0142</td>
<td>.773</td>
<td>1h</td>
<td>3h</td>
</tr>
<tr>
<td>1047</td>
<td>1,700,000</td>
<td>61,728</td>
<td>.0363</td>
<td>.765</td>
<td>1h</td>
<td>3h</td>
</tr>
<tr>
<td>1047</td>
<td>1,300,000</td>
<td>117,776</td>
<td>.0905</td>
<td>.923</td>
<td>1h</td>
<td>3h</td>
</tr>
<tr>
<td>1047</td>
<td>1,200,000</td>
<td>213,874</td>
<td>.1782</td>
<td>1.167</td>
<td>2h</td>
<td>4h</td>
</tr>
<tr>
<td>1047</td>
<td>1,200,000</td>
<td>213,874</td>
<td>.1782</td>
<td>1.167</td>
<td>2h</td>
<td>4h</td>
</tr>
<tr>
<td>1047</td>
<td>1,100,000</td>
<td>368,579</td>
<td>.2632</td>
<td>1.607</td>
<td>1h</td>
<td>3h</td>
</tr>
<tr>
<td>1047</td>
<td>1,350,000</td>
<td>628,576</td>
<td>.2793</td>
<td>1.622</td>
<td>1h</td>
<td>3h</td>
</tr>
<tr>
<td>1047</td>
<td>3,850,000</td>
<td>1,053,510</td>
<td>.2913</td>
<td>1.877</td>
<td>1h</td>
<td>3h</td>
</tr>
<tr>
<td>1047</td>
<td>2,300,000</td>
<td>1,553,820</td>
<td>.4855</td>
<td>1.906</td>
<td>1h</td>
<td>3h</td>
</tr>
<tr>
<td>1047</td>
<td>3,200,000</td>
<td>1,906,866</td>
<td>.6575</td>
<td>.931</td>
<td>1h</td>
<td>3h</td>
</tr>
<tr>
<td>1047</td>
<td>2,900,000</td>
<td>2,042,455</td>
<td>.7564</td>
<td>.926</td>
<td>1h</td>
<td>3h</td>
</tr>
<tr>
<td>1047</td>
<td>2,500,000</td>
<td>2,012,114</td>
<td>.8048</td>
<td>.800</td>
<td>1h</td>
<td>3h</td>
</tr>
<tr>
<td>1047</td>
<td>2,300,000</td>
<td>1,732,588</td>
<td>.8662</td>
<td>.850</td>
<td>1h</td>
<td>3h</td>
</tr>
<tr>
<td>1047</td>
<td>1,700,000</td>
<td>1,478,584</td>
<td>.8697</td>
<td>.824</td>
<td>1h</td>
<td>3h</td>
</tr>
<tr>
<td>1047</td>
<td>1,300,000</td>
<td>1,230,650</td>
<td>.8790</td>
<td>.929</td>
<td>1h</td>
<td>3h</td>
</tr>
<tr>
<td>1047</td>
<td>1,300,000</td>
<td>1,104,255</td>
<td>.8494</td>
<td>.845</td>
<td>1h</td>
<td>3h</td>
</tr>
<tr>
<td>1047</td>
<td>1,000,000</td>
<td>952,279</td>
<td>.8657</td>
<td>1,073,741</td>
<td>1h</td>
<td>3h</td>
</tr>
</tbody>
</table>

Just before the Black Death, nearly 30% of the 5,650,000 inhabitants of England turn up as ancestors. How, when we go back through the reign of King John, we find the number of ancestors starting to decline from its high point of around 2 million. Each person in the tree is occupying an average of 16. The effect of distant cousin marriage on the numbers of distinct ancestors is becoming enormous. By then 80% of the population were ancestors of our single individual.

With 80% already ancestors under King John, we might suppose that the whole population would be turning up as ancestors a few generations further back. Surprisingly, the proportion in the fourth column of our table never rises to 100%. In 1047, it oscillates around 85% for all the generations of Hangan, and Norman kings. Our individual is likely to be descended from only about 85% of the population at the Norman Conquest. The relatively constant proportion around 85% in a state of equilibrium level, built into our formulae for m(g). The numbers of distinct ancestors get smaller as the population gets smaller. But the proportion around 85%, once achieved, remains, as we go back in time, in spite of the changes in population size and the perpetual rule of two parents for each generation.
Thus our model leads us to imagine the population of
Domesday England divided into two groups, the population
of ancestors of our individual. Five-sixths of the population
are relations, one-sixth are of no particular group of the population
the relations are not some identifiable group of the population
like the Saxons or the husbandmen and servants. The assumptions
of our model imply that the group is a random sample out of the
whole population. Some other breakdown between the ancestry of any
sixth of the population is defined by that if people are not
other English person. It seems likely that these groups they define are
sibling or bilateral cross cousins, the groups they define from the
most certain to be different, although we have not proved this
result in our model. Enough back in time, when the popula-
tion is very small, it must start to be common for ancestor
groups to coincide.

We have framed our discussion in terms of England and a
person of English ancestry back to the Norman Conquest. The
methods we have developed, of course, can be applied to the
ancestors of other nations and races, so far as an identifiable
pool of potential ancestors can be determined. It is strange to
think of ourselves within our population today as divided into
two groups insofar as we are or are not the ancestors of some
specific unknown person in the distant future. It is also
strange to think of the links which tie each of us by descent to
the past, to a group special in no way except in having one
of us as their common descendant. Each link between parent and
child is intimate and full of emotional associations, and only
30 such links back in time bring us to each ancestors at the Nor-
man Conquest. Yet, these ancestors together, a random five-
sixths of the population, seem almost beyond picturing. Each of
us is connected to the historical past in many ways -- through
the language we speak, the changed face of the earth around us,
the institutions we act within, our store of knowledge, images,
emotions, and preconceptions. The "ties of blood," that is, of
genes and lineal descent, in some sense the most real of all
connections, are also the most mystical. Always hard to grasp
imaginatively, they become still more elusive when a statistical
element enters, when they connect us with millions of people, a
part of the whole population that we can count but not identify.
Our exercise in modelling numbers of ancestors opens the way to
piquant speculations.