

The Probability of Pairwise Shared Ancestry and the Expected Number of Pairs of k -th Cousins in a Population Sample

Philip M. Service

Department of Biological Sciences, Northern Arizona University, Flagstaff, USA

Correspondence to: Philip M. Service, Philip.Service@nau.edu

Keywords: Pairwise Shared Ancestry, Simulation, Denmark, Number of Cousins, Population Sample, Population Structure

Received: October 13, 2022

Accepted: November 6, 2022

Published: November 9, 2022

Copyright © 2022 by author(s) 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/>



Open Access

ABSTRACT

An analytical solution is derived for the probability that a random pair of individuals from a panmictic population of size N will share ancestors who lived G generations previously. The analysis is extended to obtain 1) the probability that a sample of size s will contain at least one pair of $(G - 1)^{\text{th}}$ cousins; and 2) the expected number of pairs of $(G - 1)^{\text{th}}$ cousins in that sample. Solutions are given for both monogamous and promiscuous (non-monogamous) cases. Simulation results for a population size of $N = 20,000$ closely approximate the analytical expectations. Simulation results also agree very well with previously derived expectations for the proportion of unrelated individuals in a sample. The analysis is broadly consistent with genetic estimates of relatedness among a sample of 406 Danish school children, but suggests that a different genetic study of a heterogenous sample of Europeans overestimates the frequency of cousin pairs by as much as one order of magnitude.

1. INTRODUCTION

The ability to rapidly and inexpensively genotype humans at thousands or tens of thousands of polymorphic loci makes possible identification of relatives in data bases [1-3], assessment of population structure [4, 5], and estimation of the frequency of relative pairs in large samples [6, 7]. In the latter case, it has been suggested that commonly used coalescent-based methods overestimate the number of cousins in samples [8]. In this paper, I derive expectations for the probability that a random pair of individuals will share at least one ancestor, or one pair of ancestors, who lived G generations in the past—that is, that they are $(G - 1)^{\text{th}}$ cousins. The analysis is extended to derive the expected number of $(G - 1)^{\text{th}}$ cousin pairs in a sample. These results complement the analysis of Shchur and Nielsen [9] which gives the expected number of individuals in a sample who are unrelated as $(G - 1)^{\text{th}}$ cousins. The present analysis assumes a constant-size panmictic population with discrete generations. While that describes no actual human popula-

tion, this analysis may provide a useful metric for comparison with the results from genetic analyses of real populations.

In a previous paper, pairwise shared ancestry in random-mating, constant-sized population was investigated by simulation [10]. A principal result was that there is an approximately 50% probability that random pairs of individuals will share at least one ancestor who lived $0.5 \log_2 N$ generations previously, or more recently, where N is population size. The probability of pairwise shared ancestry increases rapidly with additional generations, and approaches 1.0 for ancestors who lived about $0.65 \log_2 N$ generations in the past, or more recently. In this paper, I show that comparable results can be obtained analytically by assuming that a present-day individual represents a sample of ancestors who lived in each earlier generation. The probability of shared ancestry is then analogous to the probability that two samples drawn with replacement share at least one element in common. The analysis is extended to estimate the probability that a sample will not contain any pairs of individuals of specified relationship, *i.e.*, no sib pairs, no 1st cousins, etc.; and to derive the expectation of the number of pairs of $(G - 1)$ th cousins in a sample. The present analytical results, as well as those of Shchur and Nielsen [9], are confirmed by simulation.

2. METHODS

2.1. Population Simulation

A constant population of size $N = 20,000$ with non-overlapping generations was simulated. Reproduction was either monogamous or promiscuous, and reproductive success was random (Poisson distributed). Sib mating was permitted and presumably occurred at a frequency expected by chance, although with promiscuous reproduction virtually all sib mating would have been between half-sibs. Each simulation started at generation 0 and proceeded for 15, 12 or 8 generations (see below). For computational convenience, complete genealogical information was not recorded for individuals in descendant generations. Instead, only information about generation 0 ancestors was retained. Additional details about the simulation procedure can be found in [10].

In order to simulate the probability of pairwise shared ancestry, simulations were run for 15 generations. In each generation, a sample of 5000 randomly selected pairs was evaluated for shared generation 0 ancestors. Five hundred replicate simulations were performed for each mode of reproduction (monogamous or promiscuous). In order to simulate the likelihood that a sample of more than two will contain no related individuals, the same simulation procedure was used, except that samples of size s were taken without replacement in each generation, and all members of the sample examined for shared generation 0 ancestors. Sample size, s , was either 20 or 200. The number of samples in each generation was adjusted according to sample size so that the total number of individuals sampled was $2000 = 20 \times 100$ or 200×10 . One hundred replicate simulations were performed for each sample size and type of reproduction, and each replicate was run for eight generations. A similar procedure was used to determine the number of pairs of $(G - 1)$ th cousins in a sample, except that each of the 100 replicates was run for 12 generations.

With monogamous reproduction, each individual in generation G had 2^{G-1} ancestor pairs in generation 0. There is no prohibition against a given ancestor pair occurring more than once in the genealogy of a later-generation individual. In fact, with a sufficient number of generations, the number of “ancestors” will exceed the past population size and multiple occurrences are guaranteed. With $G = 15$, for example, $2^{G-1} = 16,384$, which is greater than the number of pairs in the simulated population. In this sense, then, descendant individuals represent samples, with replacement, of generation 0 pairs. A similar argument applies to the case of promiscuous reproduction: every individual in generation G had 2^G ancestors (rather than ancestor pairs) in generation 0. Thus, for $G = 15$, $2^G = 32,768$, which, again, is larger than the total population size. This “theoretical” sampling of ancestors should not be confused with the actual sampling of individuals during the course of simulations, as described in the preceding paragraph.

2.2. The Probability That Two Random Samples Will Share Elements

Consider two random samples, A and B, both of size a , taken with replacement from a population of

size N , the members of which are numbered 1 through N . What is the probability that the two samples will contain at least one of those numbered elements in common? I proceed by comparing each element of sample B with each element of sample A. For convenience, the elements are labeled: B_1 denotes the first element of sample B, and so on.

The probability that the first element of sample B is the same as the first element of sample A is:

$$\Pr(B_1 = A_1) = \frac{1}{N}$$

And the probability that the first element of sample B is not the same as the first element of sample A is:

$$\Pr(B_1 \neq A_1) = 1 - \left(\frac{1}{N}\right) = \frac{N-1}{N}$$

Similarly, the probability that the first element of B is not the same as the i^{th} element of A is:

$$\Pr(B_1 \neq A_i) = \frac{N-1}{N}$$

Sample A has a elements, therefore the probability that the first element of B is not the same as any element of A is:

$$\Pr(B_1 \neq \text{all } A_i) = \left(\frac{N-1}{N}\right)^a \quad (1)$$

But sample B also has a elements. The probability that element 2 of B is not the same as any element of A is also given by the right-hand side of Equation (1). It follows, therefore, that the probability that all elements of B are different from all elements of A is:

$$\Pr(\text{all } B_i \neq \text{all } A_i) = \left(\left(\frac{N-1}{N}\right)^a\right)^a = \left(\frac{N-1}{N}\right)^{a^2}$$

Lastly, the probability that at least one element of B is the same as at least one element of A is:

$$1 - \Pr(\text{all } B_i \neq \text{all } A_i) = 1 - \left(\frac{N-1}{N}\right)^{a^2} \quad (2)$$

2.3. The Probability That Random Pairs of Individuals Will Share at Least One Pair of Ancestors Who Lived G Generations in the Past

2.3.1. Monogamous Reproduction

The preceding analysis can be adapted to random-mating, constant-size, monogamous populations as follows. With Poisson-distributed reproductive success, approximately 20% of each generation's lineages become extinct, with a mean time to extinction of slightly more than 1.5 generations [10-14]. Thus, all present-day individuals will be descended from only 80% of an ancestral population that existed two or more generations in the past. In terms of the population modeled by simulation, a "census" size of $N = 20,000$ (or 10,000 monogamous pairs) means that approximately 8000 generation 0 pairs gave rise to persistent lineages. To avoid confusion, this "effective monogamous pair population size" will be referred to as N_m . An individual in generation G is descended from 2^{G-1} ancestor pairs in generation 0, and that is the size of the ancestor-pair sample represented by that individual. Let C_G be the probability that a random pair of individuals share at least one pair of ancestors who lived G generations previously; that is, the probability that they are $(G-1)^{\text{th}}$ cousins. C_G is obtained from Equation (2) by substituting N_m for N and 2^{G-1} for a :

$$C_G = 1 - \left(\frac{N_m - 1}{N_m} \right)^{(2^{G-1})^2} \quad (3)$$

2.3.2. Promiscuous Reproduction

In the case of promiscuous (non-monogamous) reproduction, each offspring is produced by a newly created random pairing of male and female parents. As long as N is reasonably large, almost all sibs will be half-sibs and almost all 1st cousins, half-1st cousins, etc. For simplicity, I will generally not use the “half” modifier for the remainder of this paper, with the understanding that it is implied when discussing results for promiscuous reproduction. With promiscuous reproduction, every individual is descended from 2^G ancestors who lived G generations previously. Approximately 80% of each generation’s members give rise to persistent lineages. Thus, for a total population size $N = 20,000$, the “effective promiscuous population size” is $N_p = 16,000$. Let D_G be the probability that a random pair of individuals share at least one ancestor who lived G generations in the past. D_G is obtained from Equation (2) by substituting N_p for N and 2^G for a :

$$D_G = 1 - \left(\frac{N_p - 1}{N_p} \right)^{(2^G)^2} \quad (4)$$

If the relationship of interest is sibs, rather than cousins, Equations (3) and (4) are approximations to the probability of pairwise shared ancestry, given the suggested values for N_m and N_p . That is because sibs share ancestors who lived only one generation previously. Therefore, in that case, N_m and N_p should be about 86.5% of the census population size—that is, the “effective size” of the immediately preceding generation. This special case is of little consequence for the present investigation and is, therefore, ignored in the analytical calculations shown in the Results.

2.4. Shared Ancestry in Samples Larger than Two

2.4.1. The Probability That a Sample Will Not Contain Related Individuals

The analytical model is readily extended to consideration of relationships among individuals in samples larger than pairs. In a sample of s individuals, drawn without replacement, there are $s(s-1)/2$ possible pairwise comparisons. For monogamous reproduction, the probability that any one pair in the sample will not be $(G-1)$ th cousins is $(1 - C_G)$. The probability that all of the pairwise comparisons are not cousins (or sibs if $G = 1$) is:

$$\Pr(\text{no pairs are } (G-1)^{\text{th}} \text{ cousins}) = (1 - C_G)^{s(s-1)/2} \quad (5)$$

Therefore, the probability that there is at least one pair of $(G-1)$ th cousins in the sample is:

$$\Pr(\text{at least one pair of } (G-1)^{\text{th}} \text{ cousins}) = 1 - (1 - C_G)^{s(s-1)/2} \quad (6)$$

As an example of a prediction by Equation (5) and Equation (6), consider the case of sampling 20 individuals from a monogamous population of size $N = 20,000$. There are 190 possible pairwise comparisons. The probability that any one pair will not share an ancestor who lived three generations previously (*i.e.*, will not be 2nd cousins) is $(1 - C_3) = 0.998$. The probability that none of the 190 possible pairs will be 2nd cousins is $0.998^{190} \approx 0.684$ (Equation (5)). Or, conversely, there is about a 32% chance that the sample will contain at least one pair of 2nd cousins (Equation (6)); and by similar calculation, there is a nearly 100% chance that the same sample will contain at least one pair of 4th cousins. If reproduction is promiscuous, Equations (5) and (6) are modified by substituting D_G for C_G . Then, with $N = 20,000$ and $s = 20$, the probability that the sample will contain at least one pair of (half) 2nd cousins is about 53.2%, and there is about a 95% chance that a sample will contain at least one pair of 3rd cousins.

2.4.2. The Expected Number of Pairs of $(G - 1)^{\text{th}}$ Cousins in a Sample

Equation (3) gives the probability that a random pair of individuals will be $(G - 1)^{\text{th}}$ cousins (in the monogamous case). As already noted, in a sample of s individuals, there are $s(s - 1)/2$ possible pairwise comparisons. Thus, the expected number of pairs of cousins in the sample is simply the product of (1) the probability that any one pair are cousins and (2) the number of possible pairs. Let P_G be the number of pairs of $(G - 1)^{\text{th}}$ cousins in a sample from a population with monogamous reproduction; and Q_G the number of pairs of (mostly half) cousins if reproduction is promiscuous. Then:

$$P_G = C_G \left(\frac{s(s-1)}{2} \right) \quad (7)$$

and

$$Q_G = D_G \left(\frac{s(s-1)}{2} \right) \quad (8)$$

In a sample of 20 individuals taken from a monogamous population of 20,000, the expected number of pairs of 2nd cousins is 0.38, which is consistent with the 32% probability that such a sample will contain at least one pair of 2nd cousins. That same sample would be expected to contain 5.98 pairs of 4th cousins.

2.4.3. The Expected Number of Related Individuals in a Sample

Shchur and Nielsen [9] derive expectations for the quantities, U_p and V_p , which are the number of individuals in a sample which do not have $(p - 1)^{\text{th}}$ cousins in that same sample. U_p applies to a monogamous population and V_p to a non-monogamous (promiscuous reproduction) population. They depend upon both sample size and effective population size, as expected from the preceding discussion. U_p , for example, and P_G (Equation (7)) convey related information. $1 - U_p$ is the expected number of individuals in a sample who are $(p - 1)^{\text{th}}$ cousins of at least one other member of the sample. That is not the same as the number of pairs of $(p - 1)^{\text{th}}$ cousins, but it sets upper and lower bounds on the number of pairs. For example, if there are r related individuals, the minimum number of pairs is $r/2$, and the maximum number of cousin pairs is $r(r - 1)/2$. Thus, there is an approximate test for consistency between the Shchur-Nielsen expectations for the number of related individuals in a sample and the expected number of cousin pairs in a sample (Equations (7) and (8)). Given $N = 20,000$ and monogamy, the expected number of 4th cousins, $(1 - U_5)$, in a sample of 20 individuals is 9.45. The minimum number of pairs expected by Shchur-Nielsen is, therefore, 4.73, which is consistent with the 5.98 pairs obtained with Equation (7). (As explained below, Sec. 3.2.3, the effective population size used for the Shchur-Nielsen calculations in this example is 8000.)

3. RESULTS

3.1. The Probability of Pairwise Shared Ancestry

The simulation results and analytical calculations for pairwise shared ancestry are similar for monogamous (Figure 1(a)) and promiscuous (Figure 1(b)) reproduction. There is, however, a slight, but consistent, difference: in both cases, for generations 3 through 7, the probabilities of pairwise shared ancestry obtained by simulation are 3% - 6% higher than those obtained by the analytical model. One possible explanation is that two individuals in a genealogical simulation may not represent truly independent samples of generation 0 ancestors. The two may be related by more recent shared ancestors. In which case, they would necessarily also share generation 0 ancestors. For $N = 20,000$, the effect is not large, and might be expected to diminish with increasing population size. Conversely, for generation 1 (and to a lesser extent for generation 2), the analytical model appears to overstate the probability of shared ancestry. As explained in Sec. 2.3.2, a likely explanation is that the analytical model assumes that the number of persistent generation 0 ancestors has already been reduced by 20%, although that reduction actually requires about two generations in the simulations. In any event, by either simulation or analysis, the probability of shared ancestry in the first two generations is close to zero (Figure 1).

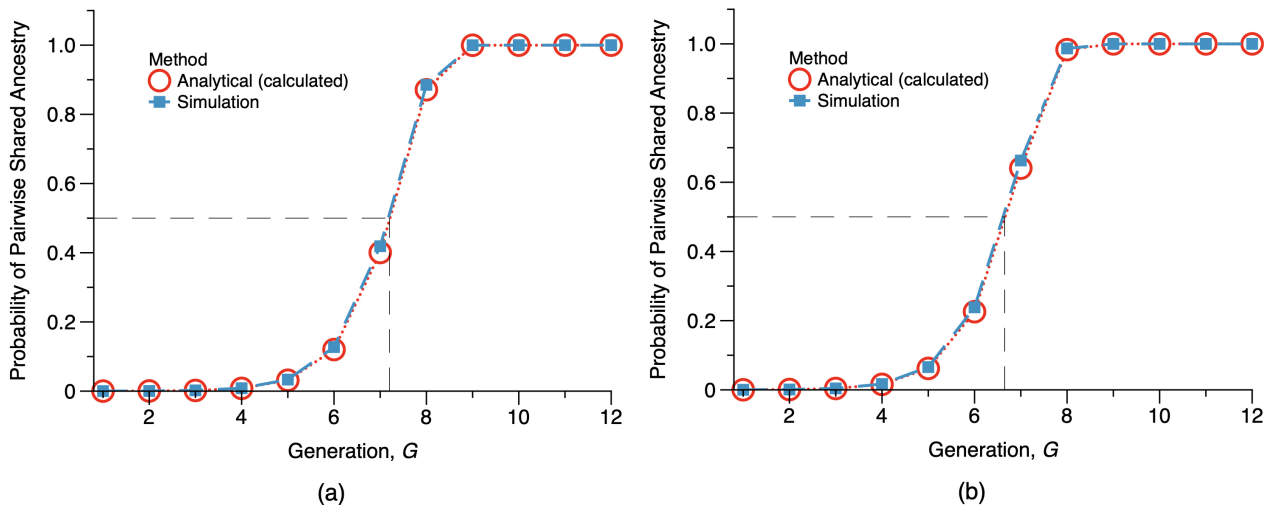


Figure 1. The probability that random pairs will share G^{th} generation ancestors. (a) Monogamous reproduction; (b) promiscuous reproduction. $N = 20,000$, simulation results are the means of 500 replicates. See text for details of analytical calculations.

For $N = 20,000$ and monogamy, the time required for the probability of pairwise shared ancestry to reach 50% is about 7.2 generations (Figure 1(a)); that is $0.504 \log_2 N$, in close agreement with previous results [10]. By generation 9, the probability of pairwise shared ancestry is very nearly 100%; that is $0.630 \log_2 N$ generations, which is also consistent with earlier results. Promiscuous mating increases the probability of pairwise shared ancestry, although as already noted the relationships are half degree.

The present results can be extended in a number of ways. For example, a statement such as “random pairs of individuals have an $x\%$ chance of sharing ancestors who lived G generations previously” is equivalent to “on average, every individual is related to $x\%$ of the other individuals in the population by ancestors who lived G generations in the past”. Or, which is the same thing, “an average individual’s $(G - 1)^{\text{th}}$ cousins will number $x\%$ of the population”.

3.2. The Probability That Samples Larger than Two Will Contain Related Individuals

3.2.1. The Probability that a Sample of Size, s , Will Contain No Pairs of $(G - 1)^{\text{th}}$ Cousins

There is very close agreement between the analytical model (Equation (5)) and the simulation results (Figure 2). It is noteworthy that, at least for panmictic populations, a very small sampling fraction is sufficient to guarantee that even closely related individuals will be included in a sample. For example, with $N = 20,000$ and $s = 200$, there is an approximately 100% probability that at least one pair of 1st cousins will be included (Figure 2(b), Figure 2(d)). Furthermore, required sample sizes increase much more slowly than the population size. As noted above, with $N = 20,000$, $s = 20$, and monogamy, there is about a 32% chance that a sample will contain one or more pairs of 2nd cousins (Figure 2(a)). Similar calculations can be performed for any population and sample size. If N is one million, a sample of 140 individuals would also have about a 32% chance of including at least one pair of 2nd cousins (from Equation (6)). In other words, a $50\times$ increase in population size requires only a $7\times$ increase in sample size to produce a similar probability that a sample will include one or more pairs of $(G - 1)^{\text{th}}$ cousins.

Very small sample sizes are sufficient to guarantee that a sample will include at least one pair of relatively distantly related cousins. For $N = 20,000$ and $s = 20$, there is an essentially 100% probability that the sample will contain at least one pair of 4th, and more distantly related, cousins (Figure 2(a), Figure 2(c)).

3.2.2. The Number of Pairs of $(G - 1)^{\text{th}}$ Cousins in a Sample

There is close agreement between the simulation results and analytical expectation (Figure 3). With

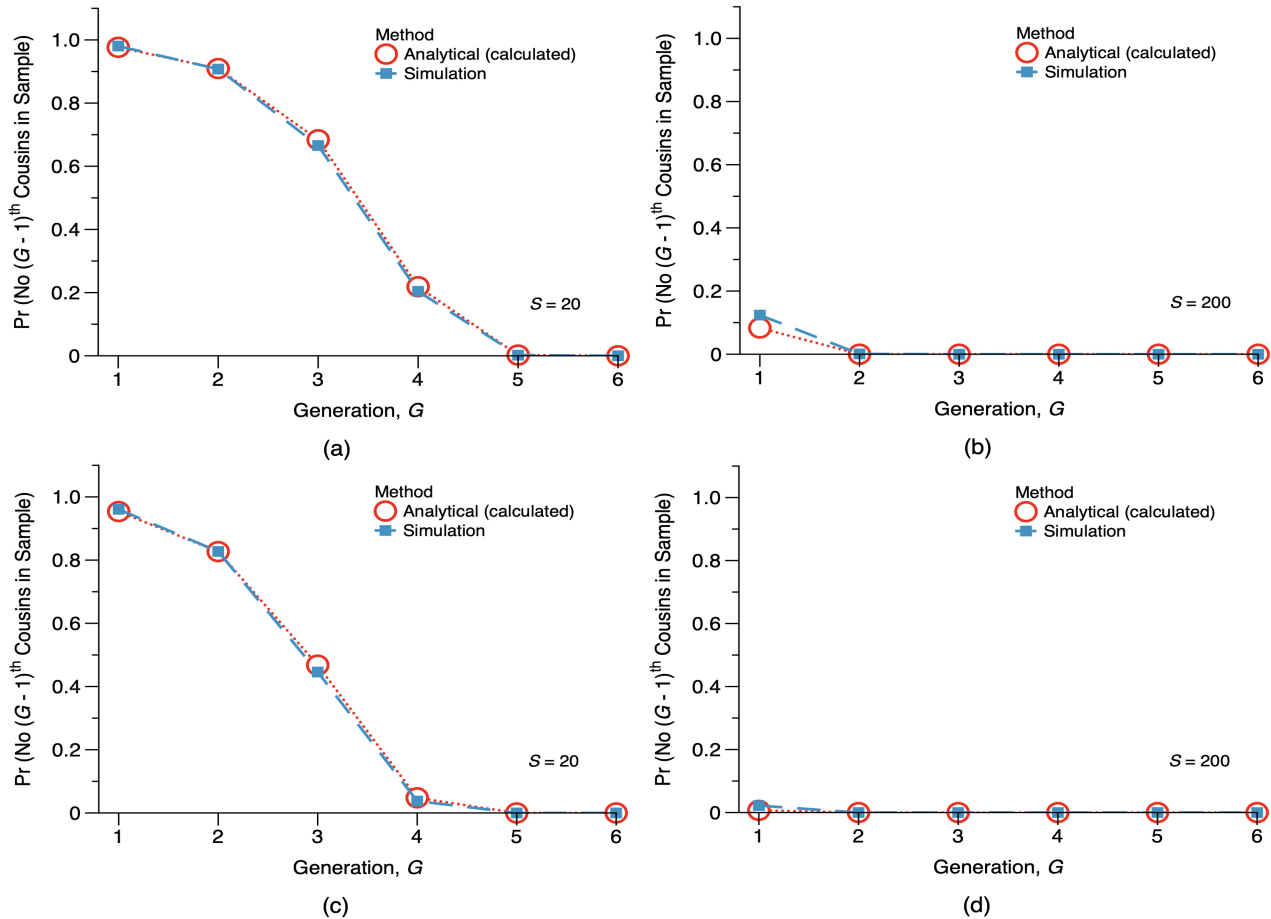


Figure 2. The probability that a sample of size s will contain no pairs of $(G - 1)^{\text{th}}$ cousins. (a) and (b) monogamous reproduction; (c) and (d) promiscuous reproduction. (a) and (c) $s = 20$; (b) and (d) $s = 200$. Simulation results are means of 100 replicates with $N = 20,000$. See text for effective population sizes used for analytical calculations.

monogamous reproduction, and $N = 20,000$, every member of a sample is related to every other member of a sample as 8^{th} cousins ($G = 9$, [Figure 3\(a\)](#), [Figure 3\(b\)](#)). That result is independent of sample size, and depends only on the population size (effective number of monogamous pairs, N_m). It reflects the fact that random pairs of individuals, drawn from the population, will be 8^{th} cousins with probability approaching 1.0 ([Figure 1\(a\)](#)). To be sure, the actual number of pairs of cousins of a given degree in a sample does depend on sample size, because there are more possible pairwise combinations with larger samples. The analytical plots in [Figure 3\(a\)](#) and [Figure 3\(b\)](#) are identical except that every point in [Figure 3\(b\)](#) is $104.74\times$ higher than the corresponding point in [Figure 3\(a\)](#). With promiscuous reproduction the number of cousin pairs of given degree is greater than for monogamous reproduction, as expected ([Figure 3\(c\)](#) and [Figure 3\(d\)](#)). Almost all possible pairs in a sample are 7^{th} cousins ($G = 8$).

3.2.3. The Proportion of Unrelated Individuals in a Sample of Size, s

For consistency with the preceding, I will replace U_p and V_p of Shchur and Nielsen [9] with U_G and V_G , which are the expected number of individuals in a sample that do not have $(G - 1)^{\text{th}}$ cousins, where G explicitly references generations. In order to compare expectations for different sample sizes, they divide U_G and V_G by sample size to obtain the expected proportion of unrelated individuals in the sample. A possible source of confusion in this analysis has to do with “effective population size”. Shchur and Nielsen

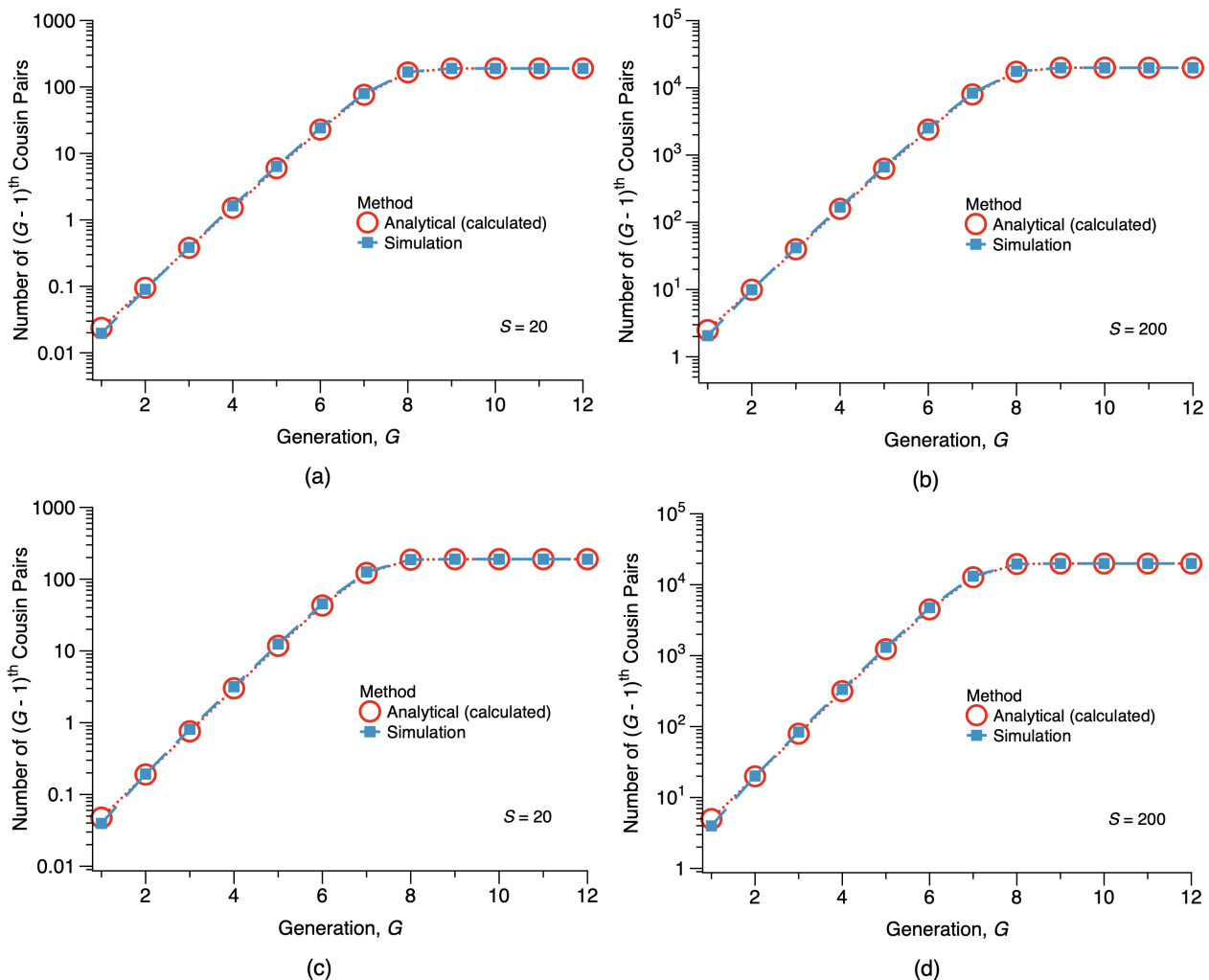


Figure 3. The number of pairs of $(G - 1)^{\text{th}}$ cousins in a sample of size s . (a) and (b) monogamous reproduction; (c) and (d) promiscuous reproduction. (a) and (c) $s = 20$; (b) and (d) $s = 200$. Simulation results are means of 100 replicates with $N = 20,000$. Scale of y -axis differs between left- and right-hand graphs.

refer to N as effective population size and later state “we assume that there are exactly N male and N female individuals” ([9], p. 1281). From that, it would appear that the appropriate value for N in their expressions for the expectations of U_G and V_G would be $N = 8000$, the number of generation 0 males or females with persistent genealogies, given a total population size of 20,000. Results for the calculated Shchur-Nielsen expectation and the simulations are shown in Figure 4. The correspondence between expectation and simulation is high, especially given that there is some stochasticity in the simulations. The correspondence appears to confirm that the interpretation of N in the Shchur-Nielsen formulas was correct.

4. SUMMARY AND DISCUSSION

A simple probabilistic model that treats each present-day individual as a sample of ancestors from previous generations was used to derive the probability that random pairs of present-day individuals will share G^{th} generation ancestors. The model was extended to obtain the expected number of pairs of $(G - 1)^{\text{th}}$ cousins (or sibs, if $G = 1$) in a sample of size s . The predictions of the analytical model are in close agreement with simulations of a panmictic population of size $N = 20,000$, with either monogamous or

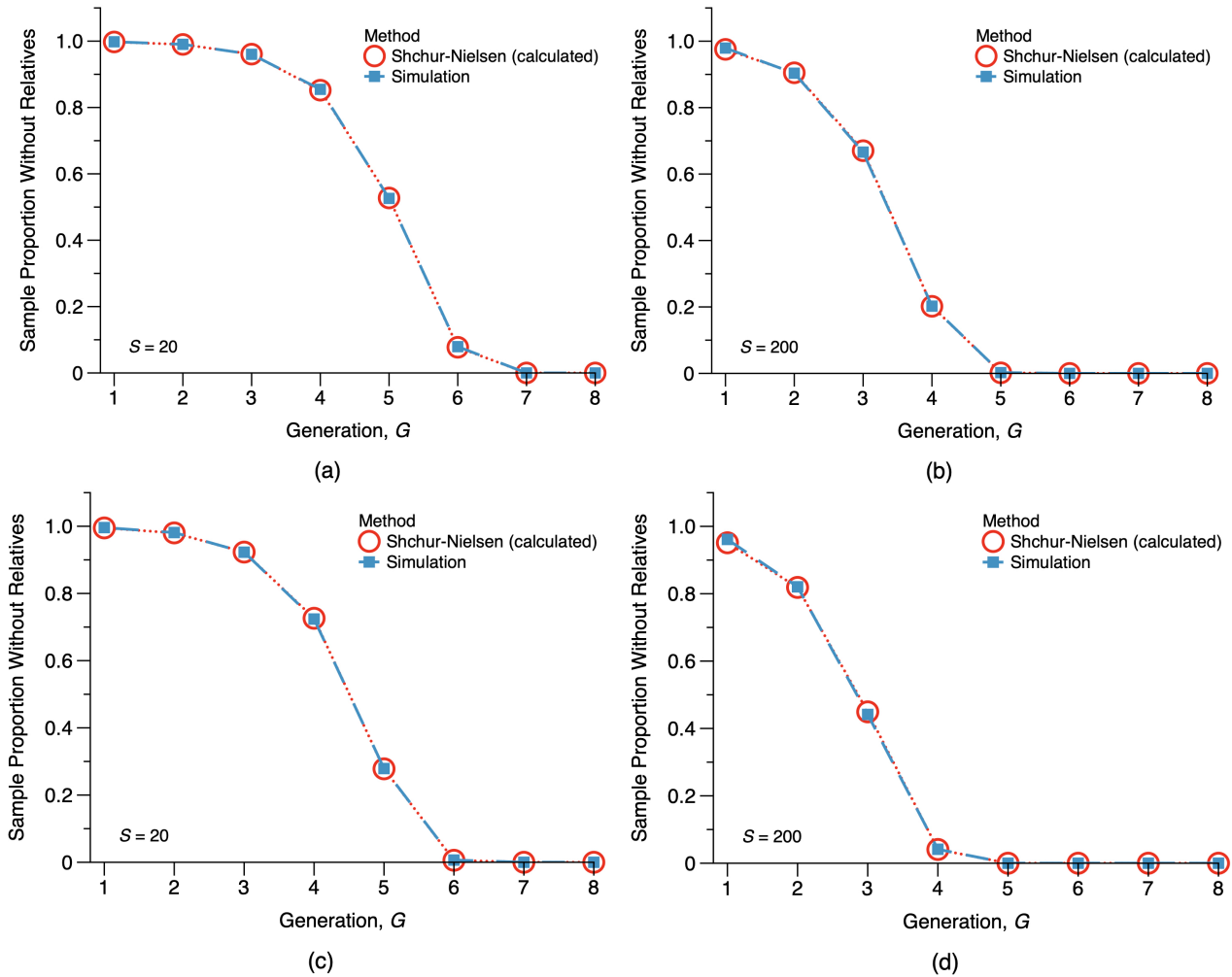


Figure 4. The proportion of a sample of size s that will consist of individuals who are unrelated as $(G - 1)^{\text{th}}$ cousins. (a) and (b) monogamous reproduction; (c) and (d) promiscuous reproduction. (a) and (c) $s = 20$; (b) and (d) $s = 200$. Simulation results are means of 100 replicates with $N = 20,000$. Effective population size used for calculated results was 8000 (see text for explanation). Simulation results are from the same simulations shown in Figure 2.

promiscuous reproduction. In addition, simulation results are in very close agreement with the calculated expectation of the number of individuals in a sample who are unrelated as $(G - 1)^{\text{th}}$ cousins [9].

Both the present results and those of Shchur and Nielsen [9] demonstrate that, for panmictic constant-size populations, surprisingly small sample sizes are sufficient to make it likely that relatively closely related individuals will be included in a sample. Although there may be a low probability that any one pair of individuals will share recent ancestors, the number of pairwise comparisons is essentially a function of the square of the sample size (divided by two). Thus, the opportunity for including relatives in a sample is larger than might first appear.

How these results apply to real populations is not obvious. The answer will depend on the degree of population subdivision and the sampling scheme. Even in highly subdivided populations, many statistics of qualitative pairwise shared ancestry may not be strongly affected, relative to a panmictic population of the same total size, provided that global migration rates are on the order of 5% per generation [11]. To be sure, global sampling in subdivided populations, reduces the probability of pairwise shared ancestry. On the other hand, sampling within restricted subpopulations can result in much higher probabilities of shared

ancestry, for the simple reason that the effective population sizes are smaller. To the extent that subpopulations are approximately panmictic, the analytical approaches discussed in this paper may be useful.

Athanasiadis *et al.* concluded that the current population of Denmark is only weakly genetically structured [5]. If so, it may be reasonable to assume that, at least until recently, it was approximately panmictic; and, therefore, that the analytical models discussed in the present paper may be reasonably applied to the Danish population and samples drawn from it. Three pairs of 2nd degree relatives were found in a sample of 406 school children [5]. The sample included only individuals for whom all four grandparents had been born in Denmark. Athanasiadis *et al.* estimated Denmark's recent effective population size to be about 500,000. If we assume monogamy, the Shchur-Nielsen expectation for $1 - U_2$, the number of 1st cousins (3rd degree relatives) in this sample, is 2.63. For promiscuous reproduction, the expectation of $1 - V_2$ is 5.24. Both numbers are more or less consistent with observation, given that three pairs could entail anywhere from three to six different individuals, and that we expect more 3rd than 2nd degree relatives in a sample. (N for the purposes of these calculations is taken to be 250,000, the effective number of individuals of either gender, as discussed in Sec. 3.2.3.) With either mode of reproduction, the Shchur-Nielsen expectation is that every member of the sample was related to at least one other member by shared ancestors who lived seven generations ago (6th cousins). Assuming 30 years for generation time, those pairs of 6th cousins shared ancestors who lived roughly 210 years ago. From Equation (6), this sample had about a 73% probability of including at least one pair of 1st cousins, a result that also appears to be consistent with the data.

With the assumption of panmixis, and an effective population size of 500,000, we can estimate that every present-day Dane (excluding recent immigrants) shares at least one 12th generation, or more recent, ancestor with every other Dane: in other words, within approximately the past 360 years ($0.65 \log_2 500,000 = 12.3$). Furthermore, their most recent common ancestor (MRCA) may have lived as recently as 19 generations, or 570 years, ago [14]. In fact, these estimates may be conservative because the population of Denmark was smaller in the past.

Henn *et al.* [6], on the basis of genetic data, concluded that a “heterogenous” sample of 5000 Europeans contained tens of thousands of 2nd through 9th cousins—specifically that the sample included on the order of 30,000 pairs of 4th cousins. Equations (7) and (8) can be expanded and rearranged to calculate the effective population size implied by that result. It is consistent with $N_m = 106,500$ monogamous pairs or $N_p = 426,000$. Shchur and Nielsen calculated similar numbers [9]. Both population estimates would seem to be far too small, given the description of the sample. For comparison, if $N_m = 10$ million, the expected number of 4th cousin pairs in a sample of 5000 individuals would be about 320. The present results appear to confirm that estimates of relationship among sample members may be inflated when using coalescent-based methods to infer identity-by-descent [8].

ACKNOWLEDGEMENTS

I thank Kiisa Nishikawa for many helpful comments and suggestions.

CONFLICTS OF INTEREST

The author declares no conflicts of interest regarding the publication of this paper.

REFERENCES

1. Yang, X., Xu, S., The HUGO Pan-Asian SNP Consortium (2011) Identification of Close Relatives in the HUGO Pan-Asian SNP Database. *PLOS ONE*, **6**, e29502. <https://doi.org/10.1371/journal.pone.0029502>
2. Ehrlich, Y., Shor, T., Itsik, P. and Carmi, S. (2018) Identity Inference of Genomic Data Using Long-Range Familial Searches. *Science*, **362**, 690-694. <https://doi.org/10.1126/science.aau4832>
3. Kim, J., Edge, M.D., Algee-Hewitt, B.F.B., Li, J.Z. and Rosenberg, N.A. (2018) Statistical Detection of Relatives Typed with Disjoint Forensic and Biomedical Loci. *Cell*, **175**, 848-858. <https://doi.org/10.1016/j.cell.2018.09.008>

4. Moltke, I., Fumagalli, M., Korneliussen, T.S., Crawford, J.E., Bjerregaard, P., Jørgensen, M.E., Grarup, N., Gulløv, H.C., Linneberg, A., Pedersen, O., Hansen, T., Nielsen, R. and Albrechtsen, A. (2015) Uncovering the Genetic History of the Present-Day Greenlandic Population. *American Journal of Human Genetics*, **96**, 54-69. <https://doi.org/10.1016/j.ajhg.2014.11.012>
5. Athanasiadis, G., Cheng, J.Y., Vilhjálmsson, B.J., Jørgensen, F.G., Als, T.D., Le Hellard, S., Espeth, T., Sullivan, P.F., Hultman, C.M., Kjærgaard, P.C., Schierup, M.H. and Mailum, M. (2016) Nationwide Genomic Study in Denmark Reveals Remarkable Population Homogeneity. *Genetics*, **204**, 711-722. <https://doi.org/10.1534/genetics.116.189241>
6. Henn, B.M., Hon, L., Macpherson, J.M., Eriksson, N., Saxonov, S., Itsik, P. and Mountain, J.L. (2012) Cryptic Distant Relatives Are Common in Both Isolated and Cosmopolitan Genetic Samples. *PLOS ONE*, **7**, e24267. <https://doi.org/10.1371/journal.pone.0034267>
7. Seidman, D.N., Shenoy, S.A., Kim, M., Babu, R., Woods, I.G., Dyer, T.D., Lehman, D.M., Curran, J.E., Duggirala, R., Blangero, J. and Williams, A.L. (2020) Rapid, Phase-Free Detection of Long Identity-by-Descent Segments Enables Effective Relationship Classification. *American Journal of Human Genetics*, **106**, 453-466. <https://doi.org/10.1016/j.ajhg.2020.02.012>
8. Nelson, D., Kelleher, J., Ragsdale, A.P., Moreau, C., McVean, G. and Gravel, S. (2020) Accounting for Long-Range Correlations in Genome-Wide Simulations of Large Cohorts. *PLOS Genetics*, **16**, e1008619. <https://doi.org/10.1371/journal.pgen.1008619>
9. Shchur, V. and Nielsen, R. (2018) On the Number of Siblings and p-th Cousins in a Large Population Sample. *Journal of Mathematical Biology*, **77**, 1279-1298. <https://doi.org/10.1007/s00285-018-1252-8>
10. Service, P.M. (2022) Pairwise Shared Ancestry in Random-Mating Constant-Size Populations. *Natural Science*, **14**, 193-202. <https://doi.org/10.4236/ns.2022.145019>
11. Service, P.M. (2022) Pairwise Shared Genealogical Ancestry in Structured Populations. *Natural Science*, **14**, 305-321. <https://doi.org/10.4236/ns.2022.148028>
12. Service, P.M. (2021) The Future Common Ancestry of All Present-Day Humans. *Natural Science*, **13**, 117-132. <https://doi.org/10.4236/ns.2021.134011>
13. Derrida, B., Manrubia, S.C. and Zanette, D.H. (2000) On the Genealogy of a Population of Biparental Individuals. *Journal of Theoretical Biology*, **203**, 303-315. <https://doi.org/10.1006/jtbi.2000.1095>
14. Chang, J.T. (1999) Recent Common Ancestors of All Present-Day Individuals. *Advances in Applied Probability*, **31**, 1002-1026. <https://doi.org/10.1239/aap/1029955256>