Doubts about Isonymy

ALAN R. ROGERS[1]

Abstract   The method of isonymy, developed by Crow and Mange for estimating inbreeding from surname frequencies, requires an assumption that has not been appreciated: It is necessary to assume that all males in some ancestral generation, the founding stock, had unique surnames. Because this assumption is seldom justified in real populations, the applicability of the isonymy method is extremely limited. Even worse, the estimates it provides refer to an unspecified founding stock, and this implies that these estimates are devoid of information.

In many human populations, individuals of both sexes inherit the surnames of their fathers. The similarity of this mechanism to haploid genetic inheritance underlies Crow and Mange's (1965) method for estimating inbreeding coefficients from *isonymy*, that is, from the frequency of marriages between individuals with the same surname. Their method is easy to use and has been extended or employed in many subsequent publications [see Crow (1983) and references therein].

As Crow (1980, 1983) has emphasized, the method involves several assumptions that are unlikely to hold exactly. Nonetheless, the isonymy method is generally viewed optimistically as a cheap way to obtain crude but useful estimates of inbreeding. Occasionally, estimates of inbreeding inferred from surname data agree with those obtained in other ways (Rogers 1987). More often, the different methods give substantially different answers (Lasker 1985; Jorde 1989). In what follows I discuss two problems that may account for the poor match between these estimates.

Crow (1980, 1983) and others have emphasized that the isonymy method is limited in its applicability by a variety of assumptions. The most restrictive of these is usually thought to be the assumption that each surname had a monophyletic origin, that is, that each surname was introduced into the population by a single individual, the ancestor of all contemporary copies of that name. I argue, however, that an even more restrictive assumption is needed: One must assume that in some previous generation each male had a unique surname. This implies not only that each surname had a monophyletic origin but also that all these origins

---

occurred in the same generation. I refer to this as the *strong monophyletic assumption.* This assumption is stated explicitly by Holgate (1971) and by Crow (1983, p. 391), but neither of these researchers shows it to be necessary.

Let $\phi_\tau^{(t)}$ denote the probability that two genes, drawn at random without replacement from the population of generation t, are identical by descent from generation τ. In other words, $\phi_\tau^{(t)}$ is the probability that both genes are descendants of a single gene in generation τ, the founding stock. Crow and Mange (1965) and Crow (1980) consider the problem of estimating $\phi_\tau^{(t)}$ from surname frequencies obtained from a single generation. As is conventional, I assume that surnames are always inherited unchanged from the biologic father and that the population is closed to immigration.

In generation t let $M^{(t)}$ denote the number of males, and let $P_{im}^{(t)}$ and $P_{if}^{(t)}$ be the frequencies of the ith surname among males and females, respectively. Throughout, uppercase letters denote quantities referring to surname frequencies and lowercase letters refer to allele frequencies. Further, let

$$I^{(t)} = \sum_i P_{im}^{(t)} P_{if}^{(t)}, \tag{1}$$

$$Q^{(t)} = \frac{\sum_i (P_{im}^{(t)})^2 - 1/M^{(t)}}{1 - 1/M^{(t)}}. \tag{2}$$

$I^{(t)}$ is the conditional probability, given current surname frequencies, that a random male and female from generation t have the same surname. $Q^{(t)}$ is the equivalent probability for two males chosen at random without replacement. Crow and Mange (1965) suggested that $\phi_\tau^{(t)}$ could be estimated by either $Q^{(t)}/4$ or $I^{(t)}/4$. In the appendix I show that these formulas hold only if each male in generation τ had a unique surname, that is, only under the strong monophyletic assumption.

It is difficult to justify the strong monophyletic assumption, even when it is known that a population is descended from a small number of founders. For example, the Hutterite population discussed by Crow (1983) is thought to be largely descended from 17 couples, who immigrated together from Russia in 1762. It is more reasonable to postulate unique surnames for a population of 17 males than for a population of 17,000, but still the issue seems doubtful. Human migration is highly kin structured, and new populations are often founded by groups of relatives (Neel and Salzano 1967; Fix 1978; Smouse et al. 1981). Thus duplicate surnames are likely to exist even in small founding populations. The
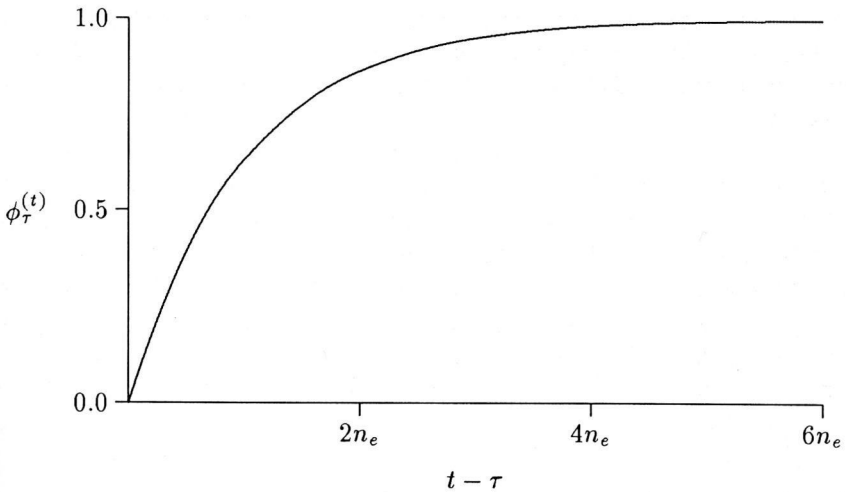
Figure 1. The effect of τ on $\phi_\tau^{(t)}$. The mean inbreeding coefficient, relative to a founding stock τ generations before the present, is approximately $1 - e^{-\tau/2n_e}$, where n_e is the effective population size.

strong monophyletic assumption severely restricts the applicability of the isonymy method.

It is tempting to define this problem away by interpreting $\phi_\tau^{(t)}$ as a measure of inbreeding relative to a hypothetical and unspecified founding stock. Although this does (by definition) remove the problem of violated assumptions, it raises another that is just as serious: When the founding stock is unspecified, estimates of inbreeding provide no information whatsoever. To see why this is so, consider Figure 1, where $\phi_\tau^{(t)}$ is graphed as a function of $t - \tau$. Note that $\phi_\tau^{(t)}$ can be made to vary from zero to unity—the entire range of permissible values—simply by adjusting the value of τ. This conclusion does not depend on the shape of the particular curve drawn in the figure but follows from the considerations that (1) $\phi_{-\infty}^{(t)} = 1$, because if life had a single origin, then *all* genes are ultimately identical by descent; and (2) $\phi_t^{(t)} = 0$, because distinct genes drawn from generation t cannot be copies of the same gene in generation t. As τ varies from 0 to $-\infty$, $\phi_\tau^{(t)}$ varies from 0 to 1. Consequently, no information is provided by an estimator that leaves τ unspecified, as the isonymy method does. Even Crow's (1983) cautious statement that "surnames provide a quick, easy, cheap, and crude way to study human inbreeding and migration" seems overly optimistic.

In view of all this, the generally poor correspondence between estimates of inbreeding obtained from isonymy and pedigrees is not

surprising. Estimates from pedigree data refer to a well-defined founding stock: the oldest generation in the pedigree (Jacquard, 1975). Estimates from surname data, on the other hand, refer to an unknown founding stock. In Jorde's (1989) study of Utah Mormons, the estimates from isonymy were several times larger than those obtained from pedigrees. Similar results have been reported in other studies [e.g., Lasker (1985), Rogers (1987), and references cited by Jorde (1989)]. Apparently, the isonymy method estimates inbreeding relative to a more ancient founding stock than is usually available in pedigrees.

On the other hand, surname data do contain information. When surname frequencies are available from two different generations, $Q^{(t)}$ can be calculated for each, and Eq. (A.2) (see appendix) can then be evaluated without ambiguity. As Crow and Mange (1965) observed, this leads to a better estimate of inbreeding—one that relates to a well-defined founding stock.

Appendix

Let $\Phi_\tau^{(t)}$ denote the probability that the surnames of two males, drawn without replacement from generation t, are identical by descent from generation τ. The conditional expectations of $I^{(t)}$ and $Q^{(t)}$, given male surname frequencies of some ancestral generation τ, are

$$E_\tau[I^{(t)}] = E_\tau[Q^{(t)}] = \Phi_\tau^{(t)} + (1 - \Phi_\tau^{(t)})Q^{(\tau)}, \tag{A.1}$$

because the probability that two distinct males have the same name is the sum of the probabilities that (1) their names are identical by descent from generation τ and (2) they are descended from distinct males in generation τ with the same surname.

Equation (A.1) can be rewritten as

$$\Phi_\tau^{(t)} = \frac{E_\tau[Q^{(t)}] - Q^{(\tau)}}{1 - Q^{(\tau)}}. \tag{A.2}$$

Here we encounter a problem. Without knowledge of $Q^{(\tau)}$, we cannot relate $Q^{(t)}$ to the probability of identity by descent. In the absence of data from generation τ, we can proceed only by making some assumption about the value of $Q^{(\tau)}$. The greatest simplicity is achieved if $Q^{(\tau)}$ is assumed to equal 0. Note that, because $Q^{(\tau)}$ is the probability that two distinct random males in generation τ have the same surname, it equals 0 if and only if *each* male in generation τ has a unique surname. Under this rather strong assumption,

$$E_\tau[I^{(t)}] = E_\tau[Q^{(t)}] = \Phi_\tau^{(t)}. \tag{A.3}$$

This result (for surnames) can be related to $\phi_\tau^{(t)}$, the inbreeding coefficient (for autosomal genes), through two well-known formulas (Hartl 1980). If $t - \tau \ll M_e$, then

$$\Phi_\tau^{(t)} \approx \frac{t - \tau}{M_e} \quad \text{and} \quad \phi_\tau^{(t)} \approx \frac{t - \tau}{2N_e}, \tag{A.4}$$

where M_e, the effective number of males, is the reciprocal of the probability that two males have the same father and N_e is the effective population size. These approximations are discussed in detail by Crow (1980). If $t - \tau \ll M_e$ and $M_e = N_e/2$, then

$$E_\tau[I^{(t)}] = E_\tau[Q^{(t)}] \approx 4\phi_\tau^{(t)}. \tag{A.5}$$

Therefore, to the extent that these assumptions are justified, $Q/4$ (or $I/4$) can be interpreted as an estimate of the inbreeding coefficient relative to generation τ. This is the formula proposed by Crow and Mange (1965).

This shows that the assumptions are sufficient to justify Crow and Mange's result, but are they also necessary? Note that, if some males in generation τ had the same surname, then $Q^{(\tau)} > 0$, Eq. (A.2) does not reduce to Eq. (A.3), and the formula of Crow and Mange does not hold. This proves the necessity of the strong monophyletic assumption. The formula of Crow and Mange is valid only if in some ancestral founding stock all males had different surnames.

Literature Cited

Crow, J.F. 1980. The estimation of inbreeding from isonymy. *Hum. Biol.* 52:1–14.

Crow, J.F. 1983. Discussion of "Surnames as markers of inbreeding and migration." *Hum. Biol.* 55:383–397.

Crow, J.F., and A.P. Mange. 1965. Measurement of inbreeding from the frequency of marriages between persons of the same surname. *Eugen. Q.* 12:199–203.

Fix, A. 1978. The role of kin-structured migration in genetic microdifferentiation. *Ann. Hum. Genet.* 41:329–339.

Hartl, D. 1980. *Principles of Population Genetics.* Sunderland, Mass.: Sinauer Associates.

Holgate, P. 1971. Drift in the random component of isonymy. *Biometrics* 21:448–451.

Jacquard, A. 1975. Inbreeding: One word, several meanings. *Theor. Popul. Biol.* 7:338–363.

Jorde, L.B. 1989. Inbreeding in the Utah Mormons: An evaluation of estimates based on pedigrees, isonymy, and migration matrices. *Ann. Hum. Genet.* 53:339–355.

Lasker, G.W. 1985. *Surnames and Genetic Structure*. New York: Cambridge University Press.

Neel, J.V., and F. Salzano. 1967. Further studies on the Xavante Indians. X. Some hypotheses-generalizations resulting from these studies. *Am. J. Hum. Genet.* 19:554–574.

Rogers, L.A. 1987. Concordance in isonymy and pedigree measures of inbreeding: The effects of sample composition. *Hum. Biol.* 59(5):753–767.

Smouse, P.E., V.J. Vitzthum, and J.V. Neel. 1981. The impact of random and lineal fission on the genetic divergence of small human groups: A case study among the Yanomama. *Genetics* 98:179–197.