Indian Statistical Institute, Calcutta

# A Note on Parent-Offspring Correlation and Inbreeding

## Ranajit Chakraborty

SUMMARY

For a diallelic autosomal character, the correlation between parental total score and offspring total score (known as the parent-offspring correlation) given by $\sqrt{\{s(1 + 3F)/[(s + 1) + F(3s - 1)]\}}$, is used here to estimate $F$, the constant which is commonly interpreted as coefficient of inbreeding. When the family records include varying number of children, a reasonably good estimate of $F$ is exhibited which agrees fairly with other available estimates.

Parent-offspring correlation may now be considered as one of the most important statistical tools for studying the inheritance of metric characters. Li (1955) was the first to take into account the measurement on both parents and all their children, and obtain the canonical correlation between these two sets of measurements as the correlation between the parental total and the offspring total. A method of obtaining the parent-offspring correlation for any mating scheme keeping the population in equilibrium, satisfying Wright's equilibrium law, was given by the author (Chakraborty, 1970). In this note, the use of parent-offspring correlation is demonstrated to provide an alternative estimate of $F$, the constant which can be interpreted as a coefficient of departure from random mating (commonly known as inbreeding coefficient). As an illustration, family data (Taylor and Prior, 1938; Race et al, 1942) on MN blood groups are taken, the metrization being M-gene content for an individual. This result shows that the estimate of $F$, obtained by this technique, is fairly close to the estimate obtained by Li's procedure (Li and Horvitz, 1953).

## Parent-Offspring Correlation and $F$

For a character expressed by two codominant genes, $A$ and $a$, the phenotypically distinct genotypes are $AA$, $Aa$ and $aa$. Assigning metric values 2, 1, and 0 to these three genotypes, Chakraborty (1970) derived that the canonical correlation between the parental scores and the offspring scores is the correlation between the parental total score and the total score of their children, $s$ in number. The derivation is as follows.

Denoting the measurements on parents by $x_1$ and $x_2$, and that for the $j$th child (in order of birth) by $y_j$, we have the expressions for variances and covariances as:

$$\sigma^2_{x_1} = \sigma^2_{x_2} = \sigma^2_{y_1} = ... = \sigma^2_{y_s} = 2pq(1 + F) \qquad \text{(Kempthorne, 1957)};$$

$$\sigma_{x_1 x_2} = 4Fpq \qquad \text{(Li, 1955); and}$$

$$\sigma_{x_i y_j} = pq(1 + 3F) \text{ and } \sigma_{y_j y_{j'}} = pq(1 + 3F) \qquad \text{(Chakraborty, 1970).}$$

Now the parental total, $X = x_1 + x_2$ and offspring total score, $Y = y_1 + ... + y_s$, will have correlation

$$\varrho_s = \sum_i \sum_j \sigma_{x_i y_j} / \sqrt{\{[\sum_i \sigma^2_{x_i} + \sum_i \sum_{i'} \sigma_{x_i x_{i'}}] [\sum_j \sigma^2_{y_j} + \sum_j \sum_{j'} \sigma_{y_j y_{j'}}]\}}$$
$$= \sqrt{\{s(1 + 3F)/[(s + 1) + F(3s - 1)]\}} \qquad [1]$$

using the above expressions for variances and covariances.

From [1] one can easily obtain:

$$(1 + 3F)/(1 - F) = 1/s \times \varrho_s^2/(1 - \varrho_s^2). \qquad [2]$$

Now consider a random sample of $N$ families, out of which $N_s$ families are with $s$ children each $(s = 1, 2, ..., r)$, such that

$$\sum_{s=1}^{r} N_s = N.$$

Let $r_s$ denote the sample correlation coefficient between parental total score and offspring total score. Then, it is easy to see that

$$E[r_s^2/(1 - r_s^2)] = \varrho_s^2/(1 - \varrho_s^2) + o(1/N_s) \qquad [3]$$

and

$$\mathrm{Var}[r_s^2/(1 - r_s^2)] = 4\varrho_s^2/[N_s(1 - \varrho_s^2)^2] + o(1/N_s^2). \qquad [4]$$

(For an analogous treatment one can refer to Hotelling, 1953, p. 214.)

It is evident now that for large $N_s$, $r_s^2/(1 - r_s^2)$ can be taken as an unbiased estimate of $\varrho_s^2/(1 - \varrho_s^2)$ and consequently one gets an unbiased estimate of

$$(1 + 3F)/(1 - F) \text{ as}$$

$$t_s = (1/s)[r_s^2/(1 - r_s^2)] \qquad [5]$$

for a fixed $s$.

An estimate of variance of $t_s$ is obviously given by

$$\sigma_s^2 = 4r_s^2/[N_s(1 - r_s^2)s^2]. \qquad [6]$$

A pooled estimate of $(1 + 3F)/(1 - F)$, obtained from the whole sample, is given by

$$T = \left[ \sum_{s=1}^{r} (t_s/\sigma_s^2) \right] \Big/ \left[ \sum_{s=1}^{r} (1/\sigma_s^2) \right] \qquad [7]$$

and

$$\mathrm{Var}(T) = 1 \Big/ \left[ \sum_{s=1}^{r} (1/\sigma_s^2) \right]. \tag{8}$$

Note that $T$ is also an unbiased estimate of $(1 + 3F)/(1 - F)$ and hence the estimate of $F$ can be written as

$$\hat{F} = (T - 1)/(T + 3) \tag{9}$$

an approximate variance of which is obtained as

$$\mathrm{Var}(\hat{F}) = 4 \Big/ \left[ (T + 3)^2 \sum_{s=1}^{r} (1/\sigma_s^2) \right] \tag{10}$$

by using [8] and [9].

## Numerical Illustration

Taylor and Prior (1938) and Race et al (1942) analysed two series of family data on MN blood groups from England. This illustration is also based on these two series of family data.

For MN blood group system there are three phenotypically distinct genotypes: MM, MN, and NN. Assigning values 2, 1, and 0 (or, equivalently, M-gene content) to each individual, frequencies of parental total and offspring total are presented in the Table for family sizes 1, 2, 3, and 4. Other families are left out from this analysis, because of small numbers when classified according to the family size. The canonical correlations are also presented in the Table.

Using [5], one now has:

$$\begin{array}{llll} t_1 = 1.0036 & & t_3 = 1.9917 \\ t_2 = 1.1367 & \text{and} & t_4 = 1.2163 \end{array}$$

From [6]:

$$\begin{array}{llll} 1/\sigma_1^2 = 8.4542 & & 1/\sigma_3^2 = 2.3765 \\ 1/\sigma_2^2 = 10.0783 & \text{and} & 1/\sigma_4^2 = 2.5232 \end{array}$$

Thus, $T = 27.7417/23.4322 = 1.1839$ and $\mathrm{Var}(T) = 1/23.4322 = 0.0427$.

Hence, $\hat{F} = 0.1839/4.1839 = 0.0440$ by [9], and $\mathrm{SE}_{\hat{F}} = 0.0236$ by [10].

From the combined sample one also gets 69 M, 112 MN, and 54 N individuals among the 235 fathers (the family with serial number 200 is excluded, due to the reasons mentioned by the authors). From this, an estimate of $F$ (using any of the five methods suggested by Li and Horvitz, 1953) is obtained as

$$F^* = 0.0429,$$

which is fairly close to the estimate obtained from parent-offspring correlations.

## Table. Frequency distribution of the families according to the parental and the offspring total

$(s = 1)$

| Parental total | Offspring total | | | Total frequency |
|---|---|---|---|---|
| | 2 | 1 | 0 | |
| 4 | 5 | | | 5 |
| 3 | 6 | 15 | | 21 |
| 2 | 3 | 15 | 2 | 20 |
| 1 | | 8 | 13 | 21 |
| 0 | | | 1 | 1 |
| Total frequency | 14 | 38 | 16 | 68 |

Correlation $(r_1) = 0.7077$

$(s = 2)$

| Parental total | Offspring total | | | | | Total frequency |
|---|---|---|---|---|---|---|
| | 4 | 3 | 2 | 1 | 0 | |
| 4 | 7 | | | | | 7 |
| 3 | 5 | 11 | 9 | | | 25 |
| 2 | | 4 | 12 | 5 | 1 | 22 |
| 1 | | | 4 | 11 | 2 | 17 |
| 0 | | | | | 4 | 4 |
| Total frequency | 12 | 15 | 25 | 16 | 7 | 75 |

Correlation $(r_2) = 0.8334$

$(s = 3)$

| Parental total | Offspring total | | | | | | | Total frequency |
|---|---|---|---|---|---|---|---|---|
| | 6 | 5 | 4 | 3 | 2 | 1 | 0 | |
| 4 | 3 | | | | | | | 3 |
| 3 | 2 | 3 | 2 | 1 | | | | 8 |
| 2 | | 1 | | 16 | 2 | | | 19 |
| 1 | | | | | 6 | 3 | 1 | 10 |
| 0 | | | | | | | 4 | 4 |
| Total frequency | 5 | 4 | 2 | 17 | 8 | 3 | 5 | 44 |

Correlation $(r_3) = 0.9255$

$(s = 4)$

| Parental total | Offspring total | | | | | | | | | Total frequency |
|---|---|---|---|---|---|---|---|---|---|---|
| | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | |
| 4 | 1 | | | | | | | | | 1 |
| 3 | | 1 | | 2 | | | | | | 3 |
| 2 | | | | | 5 | 2 | | | | 7 |
| 1 | | | | | | 3 | 1 | 1 | 1 | 6 |
| 0 | | | | | | | | | 1 | 1 |
| Total frequency | 1 | 1 | | | 2 | 5 | 1 | 1 | 2 | 18 |

Correlation $(r_4) = 0.9100$

# Discussion

Though in the title $F$ is referred as the coefficient of inbreeding, it is better to avoid this interpretation of $F$, since it can only measure the departure from random mating, which is the effect of several causes, out of which inbreeding turns to be one.

The parent-offspring correlations, as obtained for four different family sizes, show that they are increasing functions of $s$, the family size (Chakraborty, 1970), though family sizes 3 and 4 give a dismal result. This may as well be a case of sampling fluctuation. Of course, one may note that there are only 18 families with family size 4.

At this point it can also be mentioned that for an inbred population the parent-offspring correlation will always be greater than $\sqrt{[s/(s+1)]}$, where $s$ is the size of the family. The significance of this result is also worth noting. In a population where certain extent of inbreeding is prevailing, the mating partners are also correlated. (One may recall that $F$ can also be interpreted as the correlation between uniting gametes.) Presence of this correlation among the mating partners is reflected in a high magnitude of parent-offspring correlation, $\sqrt{[s/(s+1)]}$ being the parent-offspring correlation for a random mating population.

After a complete discussion on this alternative method of estimating $F$, the coefficient of departure from random mating, it is natural to ask why one should prefer this method in spite of the existence of simpler ways of estimating it from a random sample of individuals (as in Li and Horvitz, 1953). The serious disadvantage with their method is that one cannot have any idea about the standard error of the estimate of $F$. The use of parent-offspring correlation enables us to know the standard error of $F$ also.

In case of codominance, Wright (1921) had also a method of estimating $F$, since in such a case correlation between mates is given by $2F/(1+F)$. Use of parent-offspring correlation pays devident in the sense that the standard error of $F$, thus obtained, is much less than that obtained by Wright's method, since the former not only uses the information on mates but also the information from their offspring. Because of these theoretical accounts, the use of parent-offspring correlation is advocated here to estimate $F$, although, from the practical viewpoint of human genetics, it is becoming more and more difficult to collect family data than to collect data on unrelated individuals.

# References

CHAKRABORTY R. (1970). Parent-offspring correlation in an equilibrium population Amer. J. Hum. Genet., 22: 605-610.

HOTELLING H. (1953). New light on the correlation coefficient and its transformations. J.R.S.S., B., 15: 193-232.

KEMPTHORNE O. (1957). An Introduction to Genetic Statistics. John Wiley and Sons, Inc., New York.

LI C. C., HORVITZ D. G. (1953). Some methods of estimating the inbreeding coefficient. Amer. J. Hum. Genet., 5: 107-17.

LI C. C. (1955). Population Genetics. University of Chicago Press, Chicago.

RACE R. R., IKIN E. W., TAYLOR G. L., PRIOR A. M. (1942). A second series of families examined in England for the $A_1A_2BO$ and MN blood group factors. Ann. Eugen., 11: 385-94.

TAYLOR G. L., PRIOR A. M. (1938). Blood groups in England. I. Examination of family and unrelated material. Ann. Eugen., 8: 343-55.

WRIGHT S. (1921). System of mating. Genetics, 6: 111-178.

RIASSUNTO

Per un carattere diallelico autosomico, la correlazione tra genitori e figli, data da $\sqrt{\{s(1 + 3F)/[(s + 1) + F(3s - 1)]\}}$, viene qui adoperata per la stima di $F$, la costante che è comunemente interpretata come coefficiente di consanguineità.

Quando le famiglie includono un numero variante di figli, si ha una stima ragionevolmente buona di $F$ che concorda con le altre stime disponibili.


RÉSUMÉ

Pour un caractère diallélique autosomique, la corrélation entre parents et enfants, donnée par $\sqrt{\{s(1 + 3F)/[(s + 1) + F(3s - 1)]\}}$, est ici introduite pour l'estime de $F$, la constante qui est généralement interprétée comme coefficient de consanguinité.

Pour les cas où les familles présentent un nombre variable de fils, une estime raisonnablement fidèle de $F$ est présentée, qui concorde avec les autres estimes disponibles.


ZUSAMMENFASSUNG

Zur Schätzung von $F$, der allgemein als Blutsverwandtschaftskoeffizient angesehenen Konstante wird hier die durch die Formel $\sqrt{\{s(1+3F)/[(s+1)+F(3s-1)]\}}$ ausgedrückte Korrelation zwischen Eltern und Kindern für ein autosomes dialleles Merkmal angewandt.

Wenn in den Familien eine unterschiedliche Kinderzahl inbegriffen ist, so erhält man eine ziemlich gute Schätzung für $F$, die mit den anderen zur Verfügung stehenden Schätzungen übereinstimmt.

RANAJIT CHAKRABORTY, Indian Statistical Institute, 203 B.T. Road, Calcutta 35, India.