MEASURING ASSOCIATION BETWEEN TWO TRAITS

Hans-Rolf Gregorius

Institut für Forstgenetik und Forstpflanzenzüchtung, Universität Göttingen, Büsingenweg 2, D-37077 Göttingen

Received 26-II-1997

ABSTRACT

A measure of association is introduced that is based on a conceptual rather than a model approach in order to ensure its broad applicability. The basis of the concept involves two variables or traits α and β of members of a population. The association of the β-state with the α-state is measured by the degree to which members of given α-state share their β-state. This formulation yields an index of association, which is applicable to all categories of traits, including discontinuous and continuous traits as well as combinations of those. Complete association of one trait with the other is equivalent to the existence of a functional relationship of the second to the first trait. Therefore, the degree of association can be understood as the closeness of the relations between two variables to a non-specified functional relationship. This feature in connection with the asymmetry of the index attests its suitability for cause-effect analyses. In fact, the conceptual approach to the measurement of association yields a conclusive method of detection and description of functional relationships between variables together with a method for quantification of the stringency of these relationships. The legitimacy of the correlation coefficient and of disequilibrium indices as measures of association is briefly addressed.

1. INTRODUCTION

In the construction of measures, a common approach consists in modelling specific instances or occurrences of a defined concept, such as association, difference, or diversity. The actual measure then results as a particular transformation of model characteristics representing a specified feature of the concept. The statistical variance (or standard deviation), for example, can be conceived as a measure of diversity which treats the deviation of a metric trait from a reference value as a particular feature of the concept of diversity, and which is modelled on the basis of linear differences from the mean.

Concerning the present topic, Goodman and Kruskal state on p. 5 in the 1979 reprint of their four landmark papers on measures of association: “In order to choose a measure of association which has meaning we propose the construction of probabilistic models of predictive activity”, and the “measure of association will then be a probability, or perhaps some simple function of probabilities, within such a model.” Here the concept is “association”, “predictive activity” is the specific occurrence of the concept, and modelling of predictive activity takes place on a probabilistic basis. Probabilities or functions of these define measures of features of association specified by the respective model of predictive activity.

Other features may be seen in special forms of association. In the (product-moment) correlation coefficient, for example, a linear relationship between two metric variables is considered as complete association while stochastic independence among these variables signifies absence of any association. The correlation coefficient therefore measures a degree of linear association. Allowing both positive and negative values, the coefficient reflects as an additional aspect of the form of association its direction. Closely related to this aspect are measures of concordance or discordance in ordering of two (ordinal) variables (see e.g. p. 49ff in Liebetrau, 1983; instead of attempting a well-balanced citation of the numerous relevant original literature, this booklet will be referred to in the following for presentation and discussion of established methods of measuring association).

Although aspects of form of association play an important role in the design of measures, elementary conceptual characteristics such as definition of completeness of association are frequently ambiguous among measures (Goodman & Kruskal 1979, p. 8). Even stochastic independence, which is generally agreed to describe the complete absence of association, is not unequivocally reflected by a number of established measures (see the pertaining demonstrations in Liebetrau, 1983). A very disturbing consequence of this situation is the loss of comparability among measures of association. Statements such as “a tree’s height shows higher association with its genotype than with its environment” are meaningless when the criterion for complete association of the quantitative trait “height” with the qualitative trait “genotype” differs from that specified for associations with the environmental trait. Measures of association between traits belonging to different categories seem to have received little attention.

In general, problems arise in cases where concepts are vague and definition of features as well as their modelling becomes arbitrary. In fact, most of the criticism concerning appropriateness of measures traces back to insufficient comprehensiveness in specification of the underlying concepts. This gives rise to disagreement on whether an index “really” measures association, diversity, difference, and so on. It also emphasizes the necessity of measures that are identically applicable to all relevant specific instances and occurrences of a comprehensively defined concept. For example, since form of association usually refers to some unique assignment of states of one variable to states of the other, degree of association has to be determined first and independently in order to ensure that the association is strong enough to make considerations of form meaningful. High degrees of association including complete association can therefore be realized for different forms of association. From this point of view, the correlation coefficient appears inadequate, since it confounds form with degree of association, with the result that complete association can be realized for only one form (linearity).

To overcome these problems, an attempt will be made to derive indices of association directly, without recourse to special features or models, from a description of the concept believed to capture its essence. This description relies on the idea that association becomes manifest in the relations (in the mathematical sense) among the states of two variables or traits which determine the extent and possibly the form of “concurrence” between the states of the two traits. The idea of concurrence will be translated into its set theoretic analogue (shared trait characteristics, intersections), and extents of association will be specified in terms of generalized procedures of counting individuals which share trait characteristics. The intrinsically asymmetric nature of the