Journal of Mammalogy, 81(1):245–249, 2000

CLASSIFICATION BIAS IN DISCRIMINANT FUNCTION ANALYSES USED TO EVALUATE PUTATIVELY DIFFERENT TAXA

RICHARD F. LANCE,* MICHAEL L. KENNEDY, AND PAUL L. LEBERG

Department of Biology, University of Louisiana at Lafayette, Lafayette, LA 70504 (RFL, PLL)
Department of Biology, University of Memphis, Memphis, TN 38152 (MLK)
Present address of RFL: Department of Biology, University of Memphis, Memphis, TN 38152

Discriminant function analysis has been used to assess morphological distinctiveness of putatively different taxa. We used randomizations of previously published morphological data for 2 subspecies of the coyote, *Canis latrans frustror* and *C. l. thamnos,* to quantify a recognized but previously unexamined bias in discriminant-function analyses that use resubstitution classification. This bias results in overestimates of intertaxon distinctiveness and is exacerbated when sample sizes are small. An alternative classification technique, jackknife sampling, is relatively unbiased.

Key words: classification, discriminant function analysis, jackknifing, morphological differentiation, resubstitution

Discriminant function analysis (DFA) commonly is used to evaluate morphological distinctiveness of putatively different taxa (Collister and Wicklum 1996; Foottit and MacKauer 1980; Gibson and Kessel 1989; Handford 1985; Humphrey and Setzer 1989; Matsui 1986; Nowak 1979; Tumlison 1993). Classification error, or the proportion of observations that are estimated to belong to a population other than their population of origin, often is used as a measure of the distinctiveness of taxa being compared.

There are a number of ways in which observations can be sampled for DFA classification, and the choice of sampling methods can affect results of classification (Huberty 1994; Johnson and Wichern 1992). A common method of sampling used with DFA classification is resubstitution. In this procedure, a discriminant function is calculated from all observations in a data set for the putatively different populations. Each observation is reassigned to one of the populations according to an estimate of affiliation based on the discriminant function. The proportion of observations that are not reassigned to their original population is known as the apparent error rate (APER—Johnson and Wichern 1992). Because the discriminant function is not independent of the observations that are classified, the resulting APER is smaller than it would be if the function were calculated from an independent set of observations (Huberty 1994; Johnson and Wichern 1992). This bias is expected to decrease as the size of the sample increases (Huberty 1994; Johnson and Wichern 1992).

Cross-validation techniques, such as sample splitting and jackknifing, are alternative sampling methods that are less biased than resubstitution (Huberty 1994; Johnson and Wichern 1992). Sample splitting involves using a portion of the total observations to build the discriminant function, after which the remaining observations are classified (Foottit and MacKauer 1980; Orthmeyer et al. 1995). A drawback to this method is that because some observations are not used in building the discrim-

* *Correspondent: rlance@memphis.edu*

inant function, it is likely to be an inferior discriminator relative to a function derived from the entire data set (Johnson and Wichern 1992). Another drawback to sample splitting is that a relatively large number of samples is required for this technique to be informative.

Jackknife sampling, also known as Lachenbruch's holdout or the leave-one-out method, involves removing 1 observation from the data set and then classifying that observation based on a DFA of the remaining data (Huberty 1994; Johnson and Wichern 1992). That observation is subsequently returned to the data set, and the entire procedure is repeated for each observation. The resulting error rate is referred to as the estimated actual error rate (Johnson and Wichern 1992). Because classification of each observation is based on a function that was calculated from a data set excluding the observation, the estimated actual error rate should be less biased than a corresponding APER. Estimated actual error rates, however, may exhibit a higher variance than APERs (Huberty 1994).

Although the potential for bias in DFA resubstitution has been recognized by many statisticians (Frank et al. 1965; Johnson and Wichern 1992), its effect on taxonomic studies has not been evaluated. Our objectives were to quantify the bias, evaluate the extent to which it might affect taxonomic assessments, and contrast results from DFA resubstitution and DFA jackknife classification procedures. We also examined the effect of sample size on APER and estimated actual error rate.

Materials and Methods

Data for our study were taken from previously published work by Lydeard and Kennedy (1988), in which DFA resubstitution was used to evaluate taxonomic designation of 2 subspecies of the coyote, *Canis latrans thamnos* and *C. l. frustror*. Those data consisted of cranial measurements taken from 364 museum specimens. We were able to obtain data for 315 of those specimens, of which 158 were *C. l. frustror* (79 males, 79 females) and 157 were *C. l. thamnos* (82 males, 75 females). Use of those observations allowed us to demonstrate with biologically relevant data how bias in resubstitution classification can affect evaluation of morphological differentiation between putative taxa. Because coyotes are sexually dimorphic (Lydeard and Kennedy 1988), males and females were analyzed separately.

To contrast different results obtained from resubstitution and jackknife classification, we performed 5,000 iterations of DFA classification of randomized data for coyotes using both sampling techniques. Data sets for each iteration were created by randomly reassigning individuals to *C. l. frustror* and *C. l. thamnos*. The same proportions of subspecies that occurred in the original data sets were maintained for each iteration. To be consistent with Lydeard and Kennedy (1988), analyses of males and females were based on 7 and 8 characters, respectively. We used the PROC DISCRIM procedure in SAS (SAS Institute Inc. 1989) to perform DFA and conducted resubstitution and jackknife sampling using the LISTERR and CROSSLISTERR options. Because populations comprised of randomly assigned individuals should be undifferentiated, it was expected that classification error rates would average 50%. Any deviation from a mean error rate of 50% would indicate a bias in the DFA classification procedure.

To examine the effect of sample size on APER and estimated actual error rate, we conducted 5,000 DFAs on randomized subsets of data totaling 20, 30, 40, 50, and 60 male coyotes. Those smaller data sets consisted of equal numbers of individuals from each putative subspecies randomly reassigned to *C. l. thamnos* and *C. l. frustror*.

Results and Discussion

Mean APERs were lower than the expected 50% (Figs. 1a and 1b), indicating overestimation of intertaxon morphological differentiation. That bias increased as number of samples became smaller (Fig. 2). Although bias in APER decreased as size of samples increased (Huberty 1994; Johnson and Wichern 1992), DFA resubstitution still resulted in a sizable bias when used with the largest sets of samples that we evaluated ($n = 161$ and $n = 154$; Figs. 1 and 2).

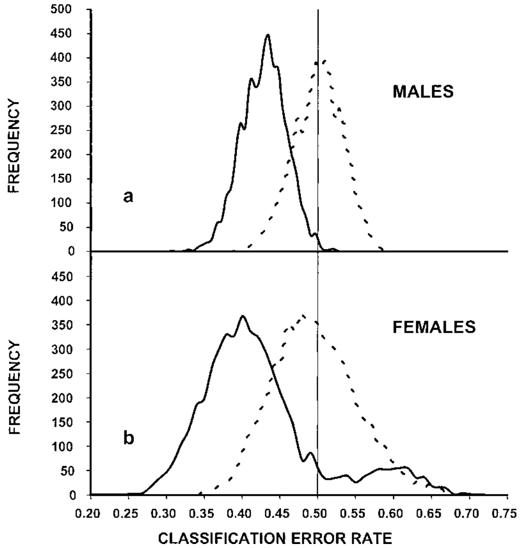

FIG. 1.—Distribution of resubstitution (APER; solid line) and jackknife (dashed line) classification error rates resulting from 5,000 discriminant function analyses of randomized morphological data from 2 subspecies of *Canis latrans.* Data sets for a) males and b) females consist of 161 and 154 observations, respectively. The expected mean classification error (50%) is indicated by the vertical line.

Mean estimated actual error rates did not differ notably from the expected mean error rate of 50% (Fig. 1) regardless of size of sample (Fig. 2). Because our results show that estimated actual error rates are less biased than APERs, jackknife sampling should be preferred over resubstitution sampling in DFA classifications of putative taxa. However, several statistical packages either do not offer jackknife classification as an option in DFA or do not offer it as the default classification method.

Bias resulting from resubstitution sampling can influence perceptions of morphological distinctiveness between putatively different taxa. For example, when we conducted DFAs on our data for *C. l. frustror* and *C. l. thamnos*, estimated actual error rates were 32% and 35% for males and females, respectively. However, APERs for those same observations were 28% and 23%, respectively. In the worst case, DFA resubstitution overestimated distinctiveness of female coyotes from those 2 populations by ≥12%. That overestimate would have been greater if classification had been based on a smaller number of observations.

Although classification based on jackknifing is not completely unbiased, this bias should only affect results when number of samples is smaller than number of independent variables used for discrimination (McLachlan 1974). In our analyses, any bias associated with estimated actual error rate was minimal, even for the smallest number of samples. In every case, bias in estimated actual error rate was smaller than bias in APER.

There is some concern that estimated actual error rate is an overly variable estimate of error rate relative to APER (Huberty 1994). Although our comparison of APER and estimated actual error rate exhibited that disparity (Fig. 2), the difference in standard deviation between the 2 estimates never exceeded 2.5% (Fig. 2), except when the number of samples equaled 10. However, that disparity was not substantive because the standard deviation of APER was hedged by the zero boundary (Fig. 2). The disparity in variance between APER and estimated actual error rate was much smaller for large samples (Fig. 2). Our results indicate that variance is more affected by sample size than by classification technique. This issue becomes important when considering the use of sample splitting rather than jackknifing to reduce bias in estimating morphological differentiation, because the former will require more samples.

The problem of biased estimates of interpopulation differentiation has implications for conservation policy when DFA is used to assess taxonomic status of endangered or threatened populations, such as the Louisiana black bear, *Ursus americanus luteolus* (M. L. Kennedy, in litt.), and the red wolf, *Canis rufus* (Nowak 1979). When taxonomic assignments have legal and economic implications, it is important that they be based on sound statistical procedures.

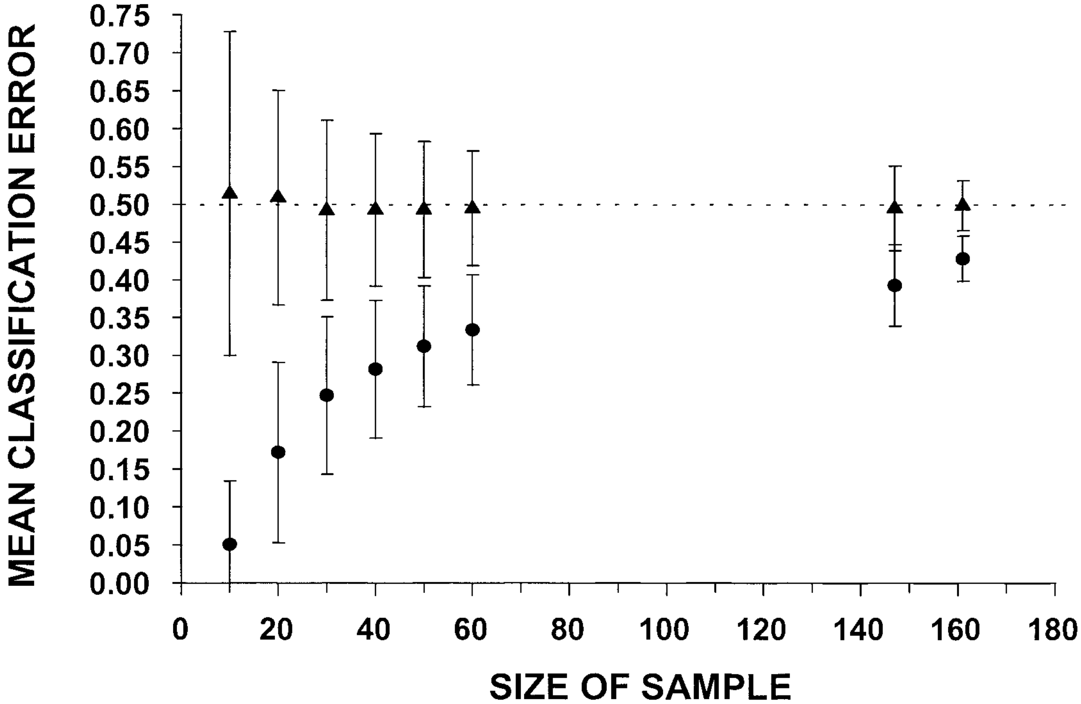

FIG. 2.—Sample size versus mean classification error rates (±1 *SD*) from resubstitution-based (circles) and jackknife-based (triangles) discriminant function analyses of randomized morphological data from 2 subspecies of *Canis latrans.* Sample sets of ≤60 observations were created by drawing random observations from the total data set for male coyotes; one-half of the observations were designated as *C. l. frustror* and the other one-half were designated as *C. l. thamnos.* For each sample size, 5,000 iterations of randomization and DFA classification were performed. The expected mean classification error (50%) is indicated by the dotted line.

Jackknifing reduces bias in classification error rates; however, the level of classification error that is sufficiently low enough to consider putative taxa as distinct is arbitrary. Alternative approaches, such as identifying phylogenetic subdivisions based on concordant patterns in multiple characters (Ball and Avise 1992), may provide a less arbitrary basis for assigning subspecific taxonomy.

Although we focused on how resubstitution-based DFA may overestimate levels of differentiation between putative taxa, this problem is not restricted to taxonomic analyses. If we had used ecological or simulated data for this analysis, we would have found the same biases resulting from resubstitution that we found with morphological data. Therefore, when reporting classification results from DFA, the method of classification should be fully described. In most cases, jackknife classification should be preferred over resubstitution.

ACKNOWLEDGMENTS

We thank C. Lydeard for providing the data used in this analysis. J. Abdalla, J. Akin, J. Ferrence, C. Lance, D. Rogowski, C. Spencer, and J. Waits provided many helpful suggestions during preparation of the manuscript. R. Lance was funded by Board of Regents Louisiana Enhancement of Quality Science Fund Fellowship (1994–1999)-GF-29, awarded through R. G. Jaeger. Analysis and manuscript preparation were supported by National Science Foundation grants DEB-AC09-76SR00-819 and DEB-9123943.

Literature Cited

Ball, R. M., Jr., and J. C. Avise. 1992. Mitochondrial DNA phylogeographic differentiation among avian populations and the evolutionary significance of subspecies. The Auk 109:626–636.

Collister, D. M., and D. Wicklum. 1996. Intraspecific variation in loggerhead shrikes: sexual dimorphism and implication for subspecies classification. The Auk 113:221–223.

Foottit, R. G., and M. MacKauer. 1980. Morphometric variation between populations of the balsam woolly aphid, *Adelges picae* (Ratzeburg) (Homoptera: Adelgidae), in North America. Canadian Journal of Zoology 58:1494–1503.

Frank, R. E., W. F. Massy, and D. G. Morrison. 1965. Bias in multiple discriminant analysis. Journal of Marketing Research 7:250–258.

Gibson, D. D., and B. Kessel. 1989. Geographic variation in the marbled godwit and description of an Alaskan subspecies. The Condor 91:436–443.

Handford, P. 1985. Morphological relationships among subspecies of the rufous-collared sparrow, *Zonotrichia capensis.* Canadian Journal of Zoology 63:2383–2388.

Huberty, C. J. 1994. Applied discriminant analysis. John Wiley & Sons, Inc., New York.

Humphrey, S. R., and H. W. Setzer. 1989. Geographic variation and taxonomic revision of rice rats (*Oryzomys palustris* and *O. argentatus*) of the United States. Journal of Mammalogy 70:557–570.

Johnson, R. A., and D. W. Wichern. 1992. Applied multivariate statistical analysis. 3rd ed. Prentice-Hall, Englewood Cliffs, New Jersey.

Lydeard, C., and M. L. Kennedy. 1988. Morphologic assessment of recently founded populations of coyote, *Canis latrans,* in Tennessee. Journal of Mammalogy 69:773–781.

Matsui, M. 1986. Geographic variation in toads of the *Bufo bufo* complex from the Far East, with a description of a new subspecies. Copeia 1986:561–579.

McLachlan, G. J. 1974. An asymptotic unbiased technique for estimating the error rates in discriminant analysis. Biometrics 30:239–249.

Nowak, R. M. 1979. North American Quartenary *Canis.* Monographs of the Museum of Natural History, The University of Kansas 6:1–154.

Orthmeyer, D. L., J. Y. Takekawa, C. R. Ely, M. L. Wege, and W. E. Newton. 1995. Morphological differences in Pacific Coast populations of greater white-fronted geese. The Condor 97:123–132.

SAS Institute Inc. 1989. SAS/STAT user's guide. Version 6, 4th ed. SAS Institute Inc., Cary, North Carolina.

Tumlison, R. 1993. Geographic variation in the lappet-eared bat, *Idionycteris phyllotis,* with description of subspecies. Journal of Mammalogy 74:412–421.

Submitted 8 May 1998. Accepted 5 March 1999.

Associate Editor was Troy L. Best.