INT. J. BIOL. BIOTECH., 3 (2): 375-380, 2006.

JACKKNIFING AND BOOTSTRAPPING SOME INDICES OF PATTERN DETECTION

S. Shahid Shaukat and Sabih-us-Shams

Department of Botany, University of Karachi, Karachi-75270, Pakistan

ABSTARCT

A comparison of the average degree of nonrandomness of patchiness in two or more populations as well as the degree of their statistical accuracy is often desired. Both Jackknife and Bootstrap provide nonparametric estimation of standard error of an estimator and thereby making possible the significance tests and the establishment of confidence intervals and testing the significance. In this paper we examine the statistical accuracy (standard error, bias, mean squared error) and provide methods for estimating confidence intervals for three indices of spatial dispersion pattern.

Key-words: Statistical methods, Jackkniefing, Bootstrapping, pattern detection

INTRODUCTION

Techniques of pattern detection abound in ecological literature (Greig-Smith, 1983; Ludurig and Reynolds, 1988; Upton and Fingleton, 1985). Some methods use statistical distributions of indices of dispersion for detecting and measuring spatial pattern of species populations. An obvious requirement of a good measure of dispersion is the relative insensitiveness to changes in density (Pielou, 1978). Morisitia (1959, 1971) developed an index of dispersion I_δ that is insensitive to changes in density caused by random thinning. This index which is based on a diversity measure proposed by Sinpson (1949) is as follows:

$$I_\delta = \frac{Q \sum_{i=1} x_i (x_i - 1)}{N (N - 1)}$$

where x_i equals the number of individuals in the ith quadrat (i = l, …., Q) $\sum x_i$ – N while N equals the sum of x_i

Lloyd (1967) proposed an ‘index of mean crowding’ and an index of patchiness. Mean crowding is defined as:

$$\overset{*}{m} = \frac{1}{N} \sum_{i=1} x_i (x_i - 1)$$

The degree of crowding as measured by is dependent on the degree of clumping and the population density. However, Lloyds’ index patchiness C which is scaled to cancel out the population density effect out of the measurement is unaffected when some members of the population are removed at randon.

$$C = \overset{*}{m} / \lambda$$

where λ is the mean density per quadrat.

A comparison of the average degree of nonrandomness of patchiness in two or more populations as well as the degree of their statistical accuracy is often desired. Both jackknife and bootstrap provide nonparametric estimation of standard error of an estimator and thereby making possible the significance tests and the establishment of confidence intervals and testing the significance. These computer-intensive technique also peranst bias reduction. The jackknife method (Tukey, 1958) has been applied to ecological diversity indices (Zahl, 1977; Adams and McCune, 1979; Heltshe and Forrestea, 1985), population size estimation (Burnham and Overton, 1979), genetic distance estimates (Mueller, 1979) measures of niche overlap (Mueller and Altenberg, 1985) and similarity index (Smith *et al*., 1986).

The bootstrap method (Efron, 1979a) sidesteps the mathematical difficulties in analyzing many statistics having non-normal distributions. Bootstrap has been used to derive the sampling variance and confidence intervals of measures of niche overlap (Mueller and Altenberg, 1985), similarity index (Smith *et al*., 1986), Gini coefficient of inequality (Dixon *et al*., 1987) and growth rate (Meyer *et al*., 1986).

In this paper we examine the statistical accuracy (standard error, bias, mean squared error) and provide methods for estimating confidence intervals for three indices of spatial dispersion pattern.

MATERIALS AND METHODS

Artificial populations

Three types of spatial patterns of populations were generated i) random, ii) regular, and iii) contagious. The study area was represented by a unit square in each case. The random population was generated by simply drawing random numbers from a uniform distribution for the X and Y coordinates. The population consisted by 500 individuals distributed uniformly at random. The regular population was generated by locating points (individuals) on a square lattice, the exact location being stochastic and depended upon the desired uniformity. The parameter σ^2 created the stochasticity (cf. Hettshe and Bitz, 1979). The value of σ^2 was set at 0.001 and 0.002. The three contagious (clumped) populations were generated using Diggle *et al.*'s (1987) Modified Thaomas Process. Each population consisted of λ_1 = 25 randomly distributed mother plant with λ_2 = 20 poisson distributed offsprings around each parent. The tightness of clumping depended on σ^2 which was set at 0.01, 0.015 and 0.02. Details of the process have appeared in Heltshe and Forrester (1983). Each population was sampled by a grid of 100 quadrats (0.1 x 0.1 unit).

The Jackknife method

The jackknife estimations applied to an index of pattern detection is as follows. Let x_1, x_2, …. X_N be the observed distribution of measurements (counts) in N quadrats calculate an index of pattern detection, P_o. Remove ith quadrat from the pool by successively the index P^{-i}. Repeat this by successively removing one quadrat at a time from i = 1, …., N quadrats. This provides N pseudovalues as follows:

$$P_i = NP_o - (N-1)\,P^{-i} \qquad i = 1, 2, \ldots. N$$

Then the jackknife estimate $(\hat{P}_J)$ is

$$\hat{P}_J = \sum P_i / N,$$

and its variance is given by

$$\mathrm{var}\,(\hat{P}_J) = \sum_{i=1}^{N} (P_i - \hat{P}_J) / N\,(N-1)$$

The approximate 100 (1 - α)% confidence interval is computed as follows:

$$\hat{P}_J \pm t\,N-1, \alpha\,[\mathrm{Var}\,(\hat{P}_J)]^{1/2}$$

While some authors use t distribution (Gray and Schucany, 1972; Smith *et al.*, 1986) others proper a standard normal variate Z (Woodwald and Schucany, 1977; Adams and McCune, 1979) for establishing confidence interval.

The Bootstrap method

The bootstrap method (Efron, 1979a,b) uses the sampled observations x_i, the distribution function. This probability distribution assigns mass 1/N to each x_i. From the empirical distribution samples of size N are repeatedly drown sampling with replacement. These samples are known as bootstrap samples. The test statistic, e.g., on index of pattern detection P is computed for each of the B bootstrap samples; call these $P^*_{(i)}$. the number of bootstrap samples B must be large simulation. For artificial populations B was set to 100 so as to make the results of jackknife and bootstrap procedures comparable. The bootstrap estimate of P and its sampling variance are obtained as:

$$\hat{P}_B = \frac{1}{N} \sum_{i=1}^{B} \overset{*}{P}_{(i)}$$

and 5.

$$\mathrm{Var}\,(\hat{P}_B) = (B-1)^{-1} \sum_{i=1}^{B} [\overset{*}{P}_{(i)} - P_B]^2$$

The bias in estimating $\hat{P}$ was found as Bias $(\hat{P}) = P_B - P_o$. the bias adjusted value was found as:

6 -

$$\hat{P}^{*} = \hat{P}_o - \mathrm{Bias}\,(\hat{P}) = 2\hat{P}_o - \hat{P}_B$$

Confidence interval for bootstrap estimated was established in two ways a) symmetric interval and b) asymmetric interval. The symmetric interval is given by:

8 ·

$$\hat{P} \pm \Phi^{-1}(1 - \alpha/2)\ \sqrt{\mathrm{Var}\,(\hat{P}_B)}$$

The asymmetric confidence interval by percentile method requires 1000 or more replications. The bootstrap samples are ordered and the limits (L, U) of the 100 (1 - α)% confidence interval are given by 100 (α/2) and 100 (1 - α/2) percentiles of the ordered values.

RESULTS

Artificial populations

A summary of the results of jackknife estimation of pattern detection indices for the artificial populations are given in Table 1. Both Morisitia's index (I_δ) and index of patchiness were substantially greater than 1 for contagious populations, less than 1 for regular populations and dose to 1 for the random population as expected. Lloyd's index m^* which measures the amount by which variance mean ratio exceeds unity added to mean density was high for the contagious populations lower than the mean density for regular populations and close to mean density for the random population. All three indices yielded higher value for the contagious population with tight clumping (σ^2 = 0.01) and the values declined with the increase in σ^2, i.e., with the decrease in clumping intensity.

In general, the jackknife estimates of all the three indices of pattern detection were close to the actual (population) value and the percentage bias was small. Generally the percentage bias was lowest for Morisitas index, followed by index of patchiness and Lloyd's index of mean crowding in that order (Table 1). The bootstrap estimates of Morisitia's index of patchiness had generally lower bias than that of jackknife estimate while the reverse was true for Lloyd's index of mean crowding. Variance was found closely similar for Morisitia's index and index of patchiness and was much lower than that of index of mean crowding. The values of mean squared error were usually close to those of the corresponding variance due to low bias.

Table 2 gives the results of bootstrap estimation of pattern detection indices for the artificial populations. Percentage bias was usually lower for the bootstrap compared to that of jackknife estimate (Table 1 and 2). Where ever bias was lower the variance of bootstrap estimate was slightly higher than that of the jackknife. Mean squared errors for the pattern indices of the six artificial populations were either very close for the two estimation methods or slightly lower for the bootstrap method. The bias, the variance and the mean squared error were lower for regular and random populations compared to contagious population in both the methods of estimation.

Bootstrap symmetric confidence intervals were usually slightly narrower than those of jackknife. This is particularly more apparent for index of patchiness (Table 1 and 2). Bootstrap asymmetric confidence intervals were usually found to be close to the symmetric intervals, in some cases through the former were narrow. Simulations were not performed to check the percentage of time the known parameter value was contained within the interval. However, normality of bootstrap samples was tested using Geary's test of normality (D'Agostino, 1970). With the exception of Lloyd's index of mean crowding for regular population with σ^2 = 0.002 (where there was a significant departure from normality) the bootstrap samples for the artificial populations did not depart significantly from normality. Fig. 1 shows the histogram of bootstrap replications for the three indices of pattern detection in relation to the contagious population with σ^2 = 0.01. All the three indices shows normal distribution.

Table 1. Results of jackknife estimation of three pattern detection indices for the artificial populations. Given below are percent bias, variance, mean squared error (MSE) and 95% confidence interval (C.I.).

Populations	Index	Calculated by index	Jackknife estimate	% Bias	Var (P_J)	MSE	95% C.I. L,	U
Contagious δ^2=0.01	I_δ	1.9464	1.9629	0.8486	0.03210	0.03211	1.6074,	2.3183
	m*	9.8099	9.9171	1.09327	1.78521	1.78532	7.2662,	12.5679
	C	1.9425	1.9629	1.04944	0.03196	0.03197	16085,	2.3172
Contagious δ^2=0.015	I_δ	1.7062	1.7091	0.1735	0.01451	0.01454	1.4701,	1.9480
	m*	8.7187	8.7743	0.6373	0.65449	0.65452	7.1692,	10.3793
	C	1.7028	1.7092	0.315	0.01442	0.01442	1.4709,	1.9474
Contagious δ^2=0.02	I_δ	1.4819	1.4847	0.19488	0.00775	0.00775	1.3100,	1.6593
	m*	7.6763	7.7171	0.5322	0.4302	0.43404	6.4100,	9.0241
	C	1.4791	1.4848	0.3895	0.00770	0.00770	1.3107,	1.6588
Regular δ^2=0.001	I_δ	0.8986	0.8977	0.1039	0.000153	0.000153	0.8731,	0.9222
	m*	4.7448	4.7488	0.0861	0.0235	0.0235	4.4446,	5.0529
	C	0.8969	0.8977	0.0855	0.000152	0.000152	0.8732,	0.92216
Regular δ^2=0.002	I_δ	0.9194	0.9188	0.0615	0.000195244	0.000195247	0.8910,	0.9465
	m*	4.8544	4.8601	0.1178	0.0355415	0.0355419	4.4867,	5.2334
	C	0.9176	0.9188	0.1280	0.00019483	0.00019484	0.8911,	0.9464

Table 2. Results of bootstrap estimation (B = 1000) of three pattern detection indices for the artificial populations. Given below are percent bias, variance, mean squared error (MSE) and 95% confidence interval (C.I.).

Populations	Index	Calculated by index	Bootstrap estimate	% Bias	Var (P_J)	MSE	95% C.I. Symmetric L,	U	Asymmetric L,	U
Contagious δ^2=0.01	I_δ	1.9464	1.9339	0.6415	0.03403	0.03403	1.5848	2.3079	1.5766	2.2943
	m*	9.8099	9.6963	1.1577	1.60306	1.61596	7.3283	12.2914	7.2240	12.2306
	C	1.9425	1.9300	0.6447	0.03387	0.03403	1.5693	2.2907	1.5785	2.2641
Contagious δ^2=0.015	I_δ	1.7062	1.7037	0.1452	0.01437	0.01438	1.4712	1.9411	1.4750	1.9658
	m*	8.7185	8.6373	0.9338	0.72950	0.73613	7.0444	10.3925	7.0839	10.2095
	C	1.7028	1.7003	0.1486	0.01428	0.01429	1.4685	1.9370	1.4749	1.9558
Contagious δ^2=0.02	I_δ	1.4819	1.4748	0.4776	0.00705	0.00710	1.3173	1.6464	1.3152	1.6413
	m*	7.6763	7.5284	1.9264	0.35948	0.38135	6.5011	8.8514	6.2394	8.7957
	C	1.4790	1.4719	0.4814	0.00701	0.00707	1.3148	1.6431	1.3127	1.6383
Regular δ^2=0.001	I_δ	0.8986	0.8985	0.0087	0.00014	0.00014	0.8754	0.9218	0.8754	0.9257
	m*	4.7448	4.7558	0.2328	0.02467	0.02479	4.4369	5.0526	4.4484	5.0274
	C	0.8969	0.8968	0.0084	0.00014	0.00014	0.8737	0.9201	0.8737	0.9239
Regular δ^2=0.002	I_δ	0.9194	0.9192	0.0141	0.00022	0.00022	0.8903	0.9484	0.8921	0.9451
	m*	4.8544	4.8352	0.3959	0.03577	0.03614	4.4837	5.2251	4.4939	5.1855
	C	0.9176	0.9175	0.0150	0.00022	0.00022	0.8885	0.9467	0.8904	0.9433
Random (uniform)	I_δ	1.0509	1.0508	0.0090	0.00104	0.00104	0.9877	1.1141	0.9865	1.1144
	m*	5.2440	5.3182	1.4163	0.08056	0.08608	4.6877	5.8003	4.6351	5.8566
	C	1.0488	1.0487	0.0066	0.00104	0.00104	0.9856	1.1120	0.9845	1.1120

DISCUSSION

For many indices used in ecology such as diversity, similarity, niche overlap, remains unknown. Computer simulation in such situations is one alternative (e.g., Ricklefs and Lau, 1980; Smith and Zaret, 1982). This is in essence bootstrap estimation of the parameters. Both bootstrap and jackknife provide the alternative means for determining the statistical accuracy.

The results showed that both the jackknife and bootstrap are effective at reducing bias and at lower sample size bootstrap is more effective in this respect. It is evident, however, that there is a cost to reducing bias in the form of increased variance of the estimator. The bias, variance and mean squared errors are all reduced when the populations are either regular or random. Thus, the statistical accuracy of indices of pattern detection are greater when the populations are distributed are greater when the populations are distribution regularly or randomly and lower when the populations follow aggregated distributions.

Even though the symmetric jackknife and bootstrap confidence interval estimates were calculated differently the results were usually closely similar. For constructing the jackknife interval the conservative to t-variate.

Schenker (1985) has pointed out that bootstrap confidence intervals should be used with coution in complex problems was used while for bootstrap interval Z-variate was employed. The accuracy of confidence intervals was not checked by simulation. However, these interval estimates are expected to be accurate when the resampled values (jackknife pseudovalces and bootstrap samples) are normally distributed. Geary's test of normality showed that with a few exceptions they in fact followed normal distributions for the populations considered. The asymmetric confidence interval based on percentile method also yielded closely similar results to those of jackknife and bootstrap. In the context of a similarity index Smith *et al*., (1986) demonstrated that the bootstrap percentile method works well regardless of the skewness. As the number of observations increases, the frequency distribution of bootstrap values approaches the frequency distribution of original data (Nash, 1981). Therefore, bootstrap interval estimates are expected to be more accurate for the larger data sets. Besides the bootstrap confidence interval methods used in this study other methods have also been proposed (cf. Efron and Tibshirani, 1986; Hall, 1988; DiCiecio and Ramano, 1990). On the basis of our results based on natural and artificial populations it appears that both jackknife and bootstrap perform more or less equally well when the sample size is small while bootstrap out performs jackknife when the sample size is large, particularly for Moristia's index and index of patchiness.

In ecology as well as in certain other disciplines it is often required to compare the dispersion pattern of two or more populations. Once the variances of an index of pattern detection of two populations are known by either jackknife or bootstrap method the significance of difference in their dispersion pattern can easily be tested using the familiar Z-statistic.

ACKNOWLEDGEMENTS

We owe our thanks to Mr. Mudassir Uddin, Department of Statistics, University of Karachi for providing some useful references.

REFERENCES

Adams, J.E. and E.D. McCune (1979). Application of the generalized jackknife to Shannon's measure of information used as an index of diversity. In: *Ecological diversity in theory and practice*. (Eds.): J.F. Grassle, G.P. Patil, W. Smith, C. Taille. pp. 117-131. International Co-operative Publ. House, Fairland, Maryland.

Burnham, K.P. and W.S. Overton (1979). Robust estimation of population size when capture probabilities vary among animals. *Ecology*, 60: 927-936.

D'Agostino, R.B. (1970). Simple compact portable test of normality: Geary's test revisited. *Psychol. Bull.*, 74: 138-140.

DiCiccio, T.J. and J.P. Romano (1990). Nonparametric confidence limits by resampling methods and least favorable families. *Int. Statist. Rev*., 58: 59-76.

Diggle, P.J., J.E. Besag and J.T. Gleaves (1976). Statistical analysis of spatial point patterns by means of distance methods. *Biometrics*, 32: 659-667.

Dixon, P.M., J. Weiner, T. Mitchell-Olds and R. Woodley (1987). Bootstrapping the Gini coefficient of inequality. *Ecology*, 68: 1548-1551.

Efron, B. (1979a). Bootstrap methods: another look at the jackknife. *Ann. Statist.*, 7: 1-26.

Efron, B. (1979b). Computers and the theory of statistics: thinking the unthinkable. *SIAM Review*, 21: 460-480.

Efron, B. and R. Tibshirani (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Sci*., 1: 54-77.

Gray, H.L. and W.R. Schucany (1972). The generalized jackknife statistic. Marcel Dekker, New York, NY.

Greig-Smith, P. (1983). Quantitative plant ecology, 3rd ed. Blackwell, Oxford.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals (with discussion). *Ann. Statist*., 16: 927-1985.

Heltshe, J.F. and D.W. Bitz (1979). Comparing diversity measures in sampled communities. In: *Ecological diversity in theory and practive*. (Eds.): J.F. Grassle, G.P. Patil, W. Smith, C. Taillie. pp. 133-144. International Co-operative Publ. House, Fairland, Maryland.

Heltshe, J.F. and N.E. Forrester (1983). The jackknife estimate of species richness. *Biometrics*, 39: 1-11.

Heltshe, J.F. and N.E. Forrester (1985). Statistical evaluation of the jackknife estimate of diversity when using quadrat samples. *Ecology*, 66: 107-111.

Lloyd, M. (1967). Mean crowding. *J. Anin. Ecol*., 36: 1-30.

Ludwig, J.A. and J.F. Reynolds (1988). Statistical ecology: A primer on methods and computing. Wiley, New York.

Meyers, J.S., C.G. Ingersof, L.L. Mc Donald and M.S. Boyce (1986). Estimating uncertainty in population growth rates: jackknife vs. bootstrap techniques. *Ecology*, 67: 1156-1166.

Morisita, M. (1959). Measuring of the dispersion of individuals and analysis of the distributional pattern. *Mem. Fac. Sci. Kyushu Univ. Ser. E*., 2: 215-235.

Morisitia, M (1971). Composition of the I_δ index. *Res. Pobul. Ecol*., 13: 1-27.

Mueller, L.D. (1979). A comparison of two methods for making statistical inferences on Nei's measure of genetic distance. *Biometrics*, 35: 757-763.

Mueller, L.D. and L. Altenberg (1985). Statistical inference on measures of niche overlap. *Ecology*, 66: 1204-1210.

Nash, S.W. (1981). Discussion of Efron paper. *Can. J. Statist*., 9: 163-164.

Pielou, E.C. (1977). Mathematical ecology. Wiley, New York.

Ricklefs, R.E. and M. Lau (1980). Bias and dispersion of overlap indices: results of some Monte Carlo simulations. *Ecology*, 61: 1019-1024.

Schenker, N. (1985). Qualms about bootstrap confidence intervals. *J. Amer. Statist. Assoc*., 80: 360-361.

Simpson, E.H. (1949). Measurement of diversity. *Nature Lond*., 163: 688.

Smith, E.P., R.B. Genter and J. Cairns, Jr (1986). Confidence interval for the similarity between algal communities. *Hydrobiologia*, 139: 237-245.

Smith, E.P. and T.M. Zaret (1982). Bias in estimating niche overlap. *Ecology*, 63: 1248-1253.

Tukey, J.W. (1958). Bias and confidence in not quite large samples. *Abst. Ann. Math. Statist*., 29: 614.

Upton, G.J.G. and B. Fingleton (1985). Spatial data analysis of example. Vol I. Wiley, New York.

Woodward, W.A. and W.R. Schucany (1977). Combination of preference pattern with the triangle taste test. *Biometrics*, 33: 31-39.

Zahl, S. (1977). Jackknifing an index of diversity. *Ecology*, 58: 907-913.

(*Accepted for publication January 2006*)