



A Chi-Square Goodness-of-Fit Analysis of Dependent Resource Selection Data

Nairanjana Dasgupta; J. Richard Alldredge

Biometrics, Vol. 56, No. 2. (Jun., 2000), pp. 402-408.

Stable URL:

<http://links.jstor.org/sici?sici=0006-341X%28200006%2956%3A2%3C402%3AACGAOD%3E2.0.CO%3B2-H>

Biometrics is currently published by International Biometric Society.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ibs.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

A Chi-Square Goodness-of-Fit Analysis of Dependent Resource Selection Data

Nairanjana Dasgupta* and J. Richard Alldredge

Program in Statistics, Washington State University,
Pullman, Washington 99164-3144, U.S.A.

**email*: dasgupta@wsu.edu

SUMMARY. The χ^2 goodness-of-fit test is commonly used for testing if animals use resources in proportion to availability. This method assumes independence of resource selection among animals. In reality, this assumption is violated if animals display antisocial or gregarious behavior. Data from a study of sharp-tailed grouse in eastern Washington suggested some dependency among observations. Realizing that this dependency can have a great influence on inference for resource selection data, we develop a technique to incorporate information on dependent observations through a simple adjustment of the usual goodness-of-fit statistic. We also demonstrate how confidence intervals on proportional use may be modified for dependent observations. Simulation is used to compare our method to other methods.

KEY WORDS: Dependency parameter; Habitat use–availability; Multiple comparisons; Multivariate chi-square; Sharp-tailed grouse; Wald statistics.

1. Introduction

Resource selection studies are commonly employed to compare resources used by animals and the availability of those resources. Results of these studies have implications for endangered species (Layman, Salwasser, and Barrett, 1985), for evaluation of disturbance due to human activity (Bowyer and Bleich, 1984), for habitat management (Mazur, Frith, and James, 1998), and for modeling wildlife populations (Schoen and Kirchoff, 1985).

Alldredge, Thomas, and McDonald (1998) present a review of many methods used in analysis of resource selection data. One of the most commonly used methods to evaluate the null hypothesis of no difference between proportions of use and availability for categorical habitat selection data is the χ^2 goodness-of-fit test (Mazur et al., 1998; McClean et al., 1998; Carriere, Bromley, and Gauthier, 1999). This technique is popular because it is easy to apply and it tests the intuitively appealing hypothesis that resources are used proportionally to their availability. Neu, Byers, and Peek (1974) used the χ^2 goodness-of-fit test in conjunction with the Bonferroni z statistic. Their method evaluates whether use of each specific habitat occurs more or less frequently than expected. Byers, Steinhorn, and Krausman (1984) clarified the use of this method. The Neu et al. (1974) method has been criticized (Aebischer, Robertson, and Kenward, 1993; Cherry, 1996; Wilson, Shackleton, and Campbell, 1998). Nevertheless, the method continues to be commonly used in both appropriate and inappropriate situations. A modification of the method to deal with one important area of misapplication is suggested in this paper.

An assumption made in using the Neu et al. (1974) method to analyze resource selection data is that observations on one animal do not depend on observations of other animals. For example, the animals exhibit neither antisocial nor gregarious behavior. Knowledge of the life history of the species under study should guide the selection of analysis method. If a study is done when individuals would normally be in groups, then groups should be the sampling unit. If individuals are known to behave independently, multiple observations at the same location and time should be treated as independent observations. For other situations, potential dependency among observations should be considered. Some investigators treat multiple observations of animals in the same location at the same time as a single observation (Smith, Hupp, and Ratti, 1982). This tactic discards information and shifts the focus of inference from individual animals to a mixture of individuals and groups of animals.

Dasgupta and Alldredge (1998) devised a test of resource selection based on the maximum of the multivariate χ^2 , which uses all the data and accounts for dependency among observations. The dependency is estimated from data on sightings of pairs or groups of animals with respect to time and location. Their results show that ignoring dependent sightings of animals leads to inflated Type I error rates for the Neu et al. (1974) statistic. However, it must also be noted that the Neu et al. (1974) method detects disproportionate use of habitats for a group of animals by pooling data to calculate the test statistic, whereas the Dasgupta and Alldredge (1998) method focuses on the individual animal. In the latter method, the proportion of each animal's usage of each habitat

is compared to its availability and the Wald χ^2 statistics are calculated. The test statistic is the maximum of the multivariate χ^2 statistics. Hence, disproportionate selection by one animal may result in the test statistic value exceeding the critical value. As noted in Alldredge and Ratti (1992), the choice of a method for analysis of resource selection depends ultimately on which statistical hypothesis is most closely related to the biological question of interest. Therefore, the Dasgupta and Alldredge (1998) method may not be appropriate for hypotheses concerning selection by a group of animals.

The method presented here involves a modification of the Neu et al. (1974) χ^2 goodness-of-fit test to incorporate information on the dependency behavior of animals under study. We first describe our motivating data set on Columbian sharp-tailed grouse (*Tympanuchus phasianellus columbianus*) (McDonald, 1998). We use Dasgupta and Alldredge's (1998) method to estimate the dependency parameter. We then establish some notation and provide a theoretical framework for our test statistic. We also modify the Goodman (1965) and Bailey (1980) simultaneous confidence intervals for proportional use when animals exhibit dependent behavior. These confidence intervals may be used without explicit hypothesis testing in those studies where estimation is the focus (Cherry, 1998). Computer simulation results are used to compare Type I error rates for the modified χ^2 goodness-of-fit approach, the multivariate χ^2 , and the usual χ^2 goodness-of-fit approach as well as coverage of the confidence intervals.

2. Data Example

The motivating data is part of a study of Columbian sharp-tailed grouse in eastern Washington State (McDonald, 1998). The objective of the study was to document seasonal habitat use and movements of Columbian sharp-tailed grouse for future conservation of the species. In these studies, the grouse were trapped in April using walk-in traps, then fitted with radio transmitters and released. Trapping periods corresponded

with peak female attendance to maximize the number of females trapped. Most management recommendations focus on female populations due to their reproductive role. Females were located approximately weekly during the spring and summer and more sporadically during fall and winter. Time and location of the radio-tagged individuals were recorded for 1995–1996 in two study areas in Washington State, the Colville Indian Reservation and Swanson Lakes Wildlife Area.

Data from the Colville Indian Reservation for the spring of 1995 are presented in Table 1. In general, we will denote the number of units (animals) by t and the number of habitats by $(h + 1)$. These data consist of $t = 13$ grouse in the $(h + 1 =)$ four habitats, which are grass/forb, grass/shrub, sagebrush, and riparian shrub, respectively. Our table entries are X_{ij} , the number of relocations of animal j in habitat i , $i = 1, \dots, h + 1$ and $j = 1, \dots, t$; N_j , the total number of relocations of animal j ; and $N = \sum_{j=1}^t N_j$, the total number relocations over all t animals. There were a total of $N = 61$ sightings. We can calculate $p_{ij} = X_{ij}/N_j$, the observed proportion of relocations of animal j in habitat i , which is an estimate of θ_{ij} , the probability of animal j using habitat i , assuming relocations are conducted at random times. The availability percentage, denoted by π_i , for these four habitats are known and given in the last row of Table 1. The hypotheses of interest are

$$H_0: \theta_{ij} = \pi_i \quad \forall i, \forall j,$$

versus

$$H_a: \theta_{ij} \neq \pi_i \quad \text{for some } i, j.$$

These hypotheses are often tested using the Neu et al. (1974) univariate χ^2 goodness-of-fit statistic,

$$GF = N \sum_{i=1}^{h+1} \frac{(\bar{p}_i - \pi_i)^2}{\pi_i}, \quad (1)$$

Table 1

Habitat use by individual radio-tagged female sharp-tailed grouse and proportion of area of four habitat types, Colville Indian Reservation, spring 1995 (McDonald 1998)

Bird, j	Habitat, i				Total N_j
	Grass/forb	Grass/shrub	Sagebrush	Riparian/shrub	
1	3	0	0	0	3
2	3	0	0	1	4
3	2	0	0	0	2
4	5	0	0	0	5
5	7	1	0	0	8
6	5	0	0	0	5
7	2	0	3	0	5
8	3	0	0	0	3
9	3	1	2	0	6
10	5	1	0	0	6
11	6	0	0	0	6
12	4	0	0	0	4
13	4	0	0	0	4
Total	52	3	5	1	61
% Usage, \bar{p}_i	85.3	4.9	8.2	1.6	
% Available, π_i	77.9	10.1	9.1	2.9	

where

$$\bar{p}_i = \frac{\sum_{j=1}^t X_{ij}}{\sum_{j=1}^t N_j} = \sum_{j=1}^t p_{ij} a_j,$$

with $a_j = N_j/N$, which, under the null and independence among the units, is assumed to follow a χ^2 distribution with h d.f. For the data in Table 1, $GF = 2.41$ ($p = 0.51$), so we fail to reject the null.

However, the question of independence among the units can be answered based only on auxiliary information. For the data in Table 1, we have the additional information that, out of the 61 total sightings, there were 48 distinct sightings, out of which 36 were individual sightings or singletons, 11 were sightings with two birds together, and one was a sighting of three birds together. The term together was defined as two relocations observed within 1 km and 1 hour. All the together birds were in the grass/forb habitat. This is the same data set described and used by Dasgupta and Alldredge (1998).

The dependency parameter, ρ , can be estimated in various ways. Based on the method suggested by Dasgupta and Alldredge (1998), we define $\rho = (R - \delta)/(1 - \delta)$, where R is the probability of observing more than one animal within a fixed time and space interval and δ is the same probability when independence is assumed among the animals. R can be estimated by the sample proportion, r , of observing more than one animal within the fixed time and space interval. The threshold parameter, δ , can be estimated by d using the total time and area of sampling and calculating the probability under independence. Replacing R by r and δ by d , we may calculate $\hat{\rho}$. There were 12 observations of two or more animals (including one group of three) considered together in 48 distinct sightings; hence, $r = 12/48$. There were 3 hours of sampling times recorded per day in a 12,000-hectare study area. We estimated the probability of an animal being detected in a circular radius of 1000 m in 1 hour as $(1/3) \times (3.14) \times 1000^2 / 12,000 \times 10,000 = 0.0087$. The approximate probability of observing two or more animals when the 61 observations are considered to be independent Bernoulli trials is given by $d = 0.10$. The estimate of the dependence parameter is $\hat{\rho} = (0.25 - 0.10)/(0.90) = 0.17$, indicating at least a slight departure from the independence assumption.

3. Test Statistics and Calculations

For animal j , X_{ij} ($i = 1, 2, \dots, (h+1)$) can be thought of as observations from a multinomial distribution with N_j trials and probability θ_{ij} . Without loss of generality, we will work with the first h habitats.

Let us define the vectors

$\mathbf{p}_j = (p_{1j}, \dots, p_{hj})$, the proportion of relocations vector for animal $j = 1, \dots, t$,

$\boldsymbol{\theta}_j = (\theta_{1j}, \dots, \theta_{hj})$, the probability of use vector for animal $j = 1, \dots, t$, and

$\boldsymbol{\pi} = (\pi_1, \dots, \pi_h)$, the proportion availability vector.

Let $\mathbf{V}_j (= \boldsymbol{\Sigma}_j/N_j)$ denote the $(h \times h)$ multinomial variance-covariance matrix for \mathbf{p}_j . Under the null hypothesis, $\boldsymbol{\Sigma}_1 =$

$\dots = \boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma} = ((\sigma_{ii'}))$ is the $(h \times h)$ variance-covariance matrix, with

$$\sigma_{ii} = \pi_i(1 - \pi_i) \quad \text{and} \quad \sigma_{ii'} = -\pi_i\pi_{i'}. \quad (2)$$

We do not assume that the vectors \mathbf{p}_j and $\mathbf{p}_{j'}$ are independent. Under the null hypothesis,

$$\text{cov}(\mathbf{p}_j, \mathbf{p}_{j'}) = \rho \boldsymbol{\Sigma} / \sqrt{N_j N_{j'}}. \quad (3)$$

Here $\rho > 0$ indicates gregarious behavior of the species, $\rho < 0$ indicates avoidance, and $\rho = 0$ indicates independence. Let $\bar{\mathbf{p}} = (\bar{p}_1, \dots, \bar{p}_h)$, which represents the mean proportion vector over all t animals. The penultimate row of Table 1 gives the corresponding \bar{p}_i . Applying some straightforward algebra results in

$$GF = (\bar{\mathbf{p}} - \boldsymbol{\pi}) \left(\frac{\boldsymbol{\Sigma}^{-1}}{N} \right) (\bar{\mathbf{p}} - \boldsymbol{\pi})'. \quad (4)$$

Dasgupta and Alldredge (1998) showed that, when the dependence parameter is nonzero, the limiting distribution of the statistic GF no longer follows a χ^2 distribution with h d.f. However, Theorem 1 shows that a simple adjustment to the goodness-of-fit statistic, GF , would follow a χ^2 distribution with h d.f. The proof of the result is provided in the Appendix.

THEOREM 1: Given $N_j, N_{j'} > 0$ for $j \neq j' = 1, \dots, t$, let $b = \sum_{j=1}^t \sum_{j'=1}^t (a_j a_{j'})^{1/2}$, with a_j being defined in (1), $a = \{1 + (b)\hat{\rho}\}$, and the adjusted goodness-of-fit statistic

$$AGF = N \sum_{i=1}^{h+1} \frac{(\bar{p}_i - \pi_i)^2}{\pi_i(1 + b\hat{\rho})} = \frac{GF}{a}. \quad (5)$$

Under the null hypothesis, the AGF statistic is χ^2 distributed with h d.f. It should be noted that, when the number of relocations N_j are equal for all t units, $b = (t-1)$.

For the data in Table 1, we can calculate a , which only depends on N_j , as $a = 2.99 = (1 + 11.72 \times 0.17)$, with $b = \{((3)(4))^{1/2} + \dots + ((4)(4))^{1/2}\}/61 = 11.72$ and $AGF = 0.082$. Comparing AGF to a χ^2 with 3 d.f., we fail to reject the null hypothesis ($p = 0.98$), thereby failing to conclude that the female grouse are selecting habitat disproportionately to availability. However, as noted by Alldredge and Ratti (1986), studies with few observations of few animals should be interpreted cautiously.

4. Confidence Intervals Based on the Adjusted Goodness of Fit

It is often of interest to estimate habitat use and to find which specific habitats are being used disproportionately to their availability. Hence, one approach would be to construct confidence intervals on the overall usage parameters, $\bar{\theta}_i (= \sum_{j=1}^t \theta_{ij}/t)$, with the joint coverage of $(1 - \alpha)100\%$. Cherry (1996) compared several simultaneous confidence interval procedures and concluded that the intervals given by Goodman (1965) and Bailey (1980) were superior with respect to Type I and II errors. We present modified versions of both intervals through simple algebraic manipulations of original intervals to incorporate the dependency of observations between animals.

4.1 Adjusted Goodman's Interval (AG)

Goodman's (1965) intervals are based on solutions of quadratic equations and use $\chi^2_{1,\alpha/h}$, the upper α/h point of a χ^2 distribution with 1 d.f. Our adjusted Goodman interval is given by

$$\bar{\theta}_{i\mp} = \frac{a\chi^2_{1,\alpha/h} + 2 \left(\sum_{j=1}^t X_{ij} \mp 0.5 \right)}{2(a\chi^2_{1,\alpha/h} + N)} \mp \frac{\sqrt{a\chi^2_{1,\alpha/h} \left\{ a\chi^2_{1,\alpha/h} + 4 \sum_{j=1}^t (X_{ij} \mp 0.5)B/N \right\}}}{2(a\chi^2_{1,\alpha/h} + N)}, \quad (6)$$

where

$$B = \left(N - \sum_{j=1}^t X_{ij} \pm 0.5 \right)$$

and $\bar{\theta}_{i-} = 0$ if $\sum_{j=1}^t X_{ij} = 0$, $\bar{\theta}_{i+} = 1$ if $\sum_{j=1}^t X_{ij} = N$.

For our data example, the adjusted Goodman 90% confidence interval for the mean usage parameters is given by (0.604, 0.959), (0.005, 0.277), (0.015, 0.319), and (0.001, 0.232), respectively. It is to be noted that each of these intervals contains the corresponding availability proportions agreeing with our results from the hypothesis test.

4.2 Adjusted Bailey's Interval (AB)

Bailey (1980) provided a modification on Goodman's intervals by using a square-root transform of the data. He has shown that his intervals are in general shorter than Goodman's intervals. His method assumes that the square root of the data is normally distributed. However, he has also shown that, under certain data configurations, his method does not provide very good coverage. For the adjusted Bailey's interval,

$$A^- = \sum_{j=1}^t (X_{ij} - 1/8)/(N + 1/8),$$

$$A^+ = \sum_{j=1}^t (X_{ij} + 7/8)/(N + 1/8),$$

and

$$C = (a\chi^2_{1,\alpha/h}/4N).$$

Table 2
Number of rejections out of 10,000 replications at $\alpha = 0.10$

Sample size	ρ	Test											
		Number of habitats = 4, number of units = 10						Number of habitats = 7, number of units = 20					
		Parameter (0.25 (4))			Parameter (0.1 (3), 0.7)			Parameter (0.15 (6), 0.1)			Parameter (0.1 (6), 0.4)		
		MCS	GF	AGF	MCS	GF	AGF	MCS	GF	AGF	MCS	GF	AGF
10	0.000	1124	949	949	1429	975	975	795	986	986	2137	980	980
	0.100	850	3563	1015	1782	3373	927	1028	7294	970	1687	7259	971
	0.200	877	5310	950	1848	5290	950	1120	9130	906	1607	9118	939
	0.300	978	6464	1001	1610	6513	899	1181	9619	930	1625	9633	975
	0.400	1035	7270	907	1621	7372	926	1255	9800	902	1668	9843	912
	0.500	1345	7910	954	1424	8053	852	1328	9863	872	1721	9905	999
25	0.000	914	1010	1010	1267	1061	1061	1183	1001	1001	1354	997	997
	0.100	1003	3529	944	1345	3434	987	1023	7351	977	1332	7217	945
	0.200	1000	5305	974	1320	5253	982	1121	8972	944	1390	9035	1012
	0.300	1055	6507	998	1327	6422	980	1186	9545	936	1403	9518	978
	0.400	1138	7204	960	1433	7237	928	1220	9735	946	1438	9792	911
	0.500	1223	7672	966	1548	7770	986	1430	9867	914	1563	9870	938
50	0.000	1042	970	970	1338	1067	1067	1071	1007	1007	1166	1031	1031
	0.100	966	3447	1022	1147	3532	1003	1094	7162	1012	1203	7265	962
	0.200	1081	5335	1004	1119	5196	947	1081	9049	939	1282	8985	1026
	0.300	1130	6442	1023	1231	6399	983	1173	9542	987	1279	9547	981
	0.400	1164	7203	987	1208	7190	949	1240	9761	1000	1297	9779	950
	0.500	1331	7674	979	1379	7728	964	1430	9840	994	1453	9868	986
100	0.000	1012	1007	1007	1131	991	991	1022	1008	1008	1131	1018	1018
	0.100	1052	3513	986	1164	3491	987	1032	7207	966	1205	7195	980
	0.200	1026	5180	1012	1164	5226	999	1129	9001	994	1140	9032	935
	0.300	1122	6431	970	1198	6402	1037	1151	9543	941	1137	9543	939
	0.400	1238	7094	1010	1253	7215	976	1275	9772	990	1368	9744	992
	0.500	1355	7761	997	1407	7671	990	1432	9848	936	1477	9861	986

Table 3
Number of trials out of 10,000 simulations when the parameter value was not included in the 90% confidence interval

Sample size (<i>n</i>)	ρ	Number of habitats = 4, number of units = 10				Number of habitats = 7, number of units = 20			
		Parameter (0.25 (4))		Parameter (0.1 (3), 0.7)		Parameter (0.15 (6), 0.1)		Parameter (0.1 (6), 0.4)	
		AG	AB	AG	AB	AG	AB	AG	AB
10	0.00	561	750	494	500	898	937	820	813
	0.10	722	764	735	596	859	701	926	696
	0.20	621	810	677	598	997	619	925	504
	0.30	756	935	815	955	948	565	1031	491
	0.40	722	1018	877	365	985	503	1004	503
	0.50	800	1009	708	343	945	654	947	267
25	0.00	717	881	586	708	748	848	925	930
	0.10	776	808	813	802	800	780	856	829
	0.20	784	968	830	845	952	885	952	833
	0.30	832	890	735	799	888	837	913	787
	0.40	785	859	770	972	889	976	913	878
	0.50	785	908	767	948	878	1115	861	1056
50	0.00	721	826	773	787	799	804	874	880
	0.10	772	794	743	788	883	863	927	868
	0.20	880	836	768	782	852	891	856	848
	0.30	844	893	787	831	816	848	908	934
	0.40	862	884	821	879	900	931	920	914
	0.50	815	868	826	893	897	975	893	960
100	0.00	774	840	862	850	854	866	831	844
	0.10	852	889	765	803	877	892	924	930
	0.20	889	869	819	839	917	928	901	898
	0.30	875	898	830	859	935	894	867	851
	0.40	856	860	790	862	929	946	891	926
	0.50	859	872	825	861	949	985	855	990

Then

$$\bar{\theta}_{i-} = \left[\sqrt{A^-} - \sqrt{\{C(C+1-A^-)\}} \right]^2 / (C+1)^2,$$

$$\bar{\theta}_{i+} = \left[\sqrt{A^+} + \sqrt{\{C(C+1-A^+)\}} \right]^2 / (C+1)^2, \quad (7)$$

and $\bar{\theta}_{i-} = 0$ if $\sum_{j=1}^t X_{ij} \leq (N+1/8)C$, $\bar{\theta}_{i+} = 1$ if $\sum_{j=1}^t X_{ij} = N$. For our data example, the adjusted Bailey's 90% confidence interval for the mean usage parameters is given by (0.581, 0.959), (0.000, 0.220), (0.001, 0.270), and (0.000, 0.164), respectively. These contain the corresponding availability parameters, too.

5. Simulation Study

Monte Carlo simulation was used to compare Type I error rates for our adjusted goodness-of-fit statistic, *AGF*, to the usual goodness-of-fit statistic, *GF*, and the maximum of the multivariate χ^2 statistic (*MCS*) proposed by Dasgupta and Alldredge (1998). Data were generated for combinations of 4 habitats with 10 units (animals) and 7 habitats with 20 units (animals), dependency parameters varying from 0 (no dependency) to 0.50, and two different patterns of use. The two patterns of simulated use for four habitats were (0.25 (4)) and (0.10 (3), 0.70), and for seven habitats, the patterns were (0.15 (6), 0.1) and (0.10 (6), 0.40). For convenience, we refer to these two patterns as uniform and skewed, respectively.

Simulations were run for sample sizes of 10, 25, 50, and 100 observations per unit (animal). The number of rejections of the true null hypothesis out of 10,000 simulations for the nominal value of $\alpha = 0.10$ is reported in Table 2.

The *AGF* and *GF* have the same error rate, as they should, when the dependency parameter is zero. As the dependency parameter increases, the number of Type I errors tends to increase for the *MCS* statistic, increases dramatically for the *GF* statistic, and remains fairly close to the nominal level for the *AGF* statistic for all combinations of number of habitats and number of units considered. Increasing the sample size appears to have little effect on Type I error rates except decreasing the *MCS* error rate for the skewed pattern.

Another simulation study was undertaken to compare the coverage of the two confidence intervals suggested in conjunction with the *AGF* statistic. Data were generated under the same conditions as for the hypothesis test. The number of times the 90% confidence intervals failed to include the parameter value for at least one habitat is reported. Results for 10,000 simulations are summarized in Table 3. The adjusted Bailey intervals (*AB*) have coverage rates generally close to the nominal level except when the sample size is small and the pattern of use is skewed. This conservative behavior of the intervals for skewed data for small samples has been noted by Bailey (1980). The adjusted Goodman method performed quite well throughout.

6. Discussion

A small amount of dependency among observations can dramatically increase Type I error rates for a *GF* analysis of resource selection data. The high Type I error rates are a result of the inflation of the expected value of *GF* even for small values of the dependency parameter, ρ . For our sharp-tailed grouse example (see Section 3), the expected value of *GF* is inflated by a factor of 2.99. As a consequence, the *p* value increased from 0.51 to 0.98 when dependent behavior among grouse was incorporated into the analysis. This change in *p* value is consistent with the increase in Type I errors observed in our simulation of field studies. Type I errors have implications for endangered species conservation efforts and habitat management. Some habitats, incorrectly identified as preferred, may be needlessly enhanced, wasting limited resources. Other habitats identified as avoided may be eliminated or reduced in size, impacting animals using the habitat.

Our *AGF* method, based on a minor modification of the χ^2 goodness-of-fit statistic, tests the intuitively appealing hypothesis that habitats are used proportionally to their availability. The multiplicative adjustment factor is a function of the number of observations per individual and a measure of the dependency among sightings of individuals. Information about dependent behavior of animals may also be used to modify joint confidence intervals on proportional use to identify which habitats are used significantly more or less than expected.

Simulated field studies indicate that the *AGF* method controls the probability of Type I error much better than *GF*. The Type I error rates for the *AGF* rarely deviate from the nominal level by more than 1%, while the *GF* error rates increase to more than 90% as dependency increases. The error rates for the *AGF* statistic are clearly closer to the nominal level than the MCS for the skewed patterns of use. The *AGF* also has a computational advantage over the MCS statistic, which is based on the maximum of the multivariate χ^2 distribution and would require special tables. Another difference between these two methods is that the *AGF* evaluates selection based on data pooled over all animals, whereas the MCS evaluates the resource selection of each animal.

In this paper, we have assumed all animals in the study have the same availability, repeated observations on the same animal are not serially correlated, and all animals exhibit the same preferences for habitats. If these assumptions are satisfied, it is desirable to investigate selection at the population level, and if animals exhibit dependent behavior, the *AGF* method presented here effectively reduces Type I error rates compared to the usual χ^2 goodness of fit. If these assumptions are violated or selection by individuals is to be examined, other analyses should be used. The MCS statistic provides one alternative when animals exhibit dependent behavior. Future work is planned to modify other methods of analysis to incorporate information on dependent observations.

RÉSUMÉ

Le test d'adéquation du chi-deux est fréquemment utilisé pour tester si les animaux utilisent les ressources proportionnellement à leur disponibilité. Cette méthode suppose l'indépendance de la sélection des ressources parmi les animaux. En réalité, cette hypothèse n'est pas réaliste si les animaux ont

des comportements asociaux ou grégaires. Les données d'une étude portant sur des téttras à queue fine dans la partie est de l'état de Washington suggèrent une certaine dépendance entre les observations. Ayant réalisé que cette dépendance peut avoir une influence majeure sur l'inférence faite à partir de données de sélection des ressources, nous avons développé une technique qui permet de tenir compte de la dépendance entre observations par un simple ajustement de la statistique d'adéquation habituelle. Nous montrons aussi comment les intervalles de confiance usuels sur l'utilisation proportionnelle peuvent être modifiés quand les observations sont dépendantes. Nous comparons notre méthode à d'autres méthodes par des simulations.

ACKNOWLEDGEMENTS

The authors wish to thank the editor, an associate editor, and two referees, who provided suggestions that led to substantial improvements in the paper. We would also like to thank M. W. McDonald and K. P. Reese for sharing their data and knowledge about sharp-tailed grouse.

REFERENCES

- Aebischer, N. J., Robertson, P. A., and Kenward, R. E. (1993). Compositional analysis of habitat use from animal radio-tracking data. *Ecology* **74**, 1313–1325.
- Allredge, J. R. and Ratti, J. T. (1986). Comparison of some statistical techniques for analysis of resource selection. *Journal of Wildlife Management* **50**, 157–165.
- Allredge, J. R. and Ratti, J. T. (1992). Further comparison of some statistical techniques for analysis of resource selection. *Journal of Wildlife Management* **56**, 1–9.
- Allredge, J. R., Thomas, D. L., and McDonald, L. L. (1998). Survey and comparison of methods for study of resource selection. *Journal of Agricultural, Biological, and Environmental Statistics* **3**, 237–253.
- Arnold, S. F. (1990). *Mathematical Statistics*. Englewood Cliffs, New Jersey: Prentice Hall.
- Bailey, B. J. R. (1980). Large sample simultaneous confidence intervals for the multinomial probabilities based on transformation of the cell frequencies. *Technometrics* **22**, 583–589.
- Bowyer, R. T. and Bleich, V. C. (1984). Effects of cattle grazing on selected habitats of southern mule deer. *California Fish and Game* **70**, 240–247.
- Byers, C. R., Steinhorst, R. K., and Krausman, P. R. (1984). Clarification of a technique for analysis of utilization-availability data. *Journal of Wildlife Management* **48**, 1050–1053.
- Carriere, S., Bromley, R. G., and Gauthier, G. (1999). Comparative spring habitat and food use by two Arctic nesting geese. *Wilson Bulletin* **111**, 166–180.
- Cherry, S. (1996). A comparison of confidence interval methods for habitat use-availability studies. *Journal of Wildlife Management* **60**, 653–658.
- Cherry, S. (1998). Statistical tests in publications of The Wildlife Society. *Wildlife Society Bulletin* **26**, 947–953.
- Dasgupta, N. and Allredge, J. R. (1998). A multivariate chi-square analysis of resource selection data. *Journal of Agricultural, Biological and Environmental Statistics* **3**, 323–334.

- Goodman, L. A. (1965). On simultaneous confidence intervals for multinomial proportions. *Technometrics* **7**, 247–254.
- Layman, S. A., Salwasser, H., and Barrett, R. H. (1985). *Habitat suitability index models: Spotted owl*. Biological Report 82(10.113), U.S. Fish and Wildlife Service, Biological Services Program, Washington, D.C.
- Mazur, K. M., Frith, S. D., and James, P. C. (1998). Barred owl home range and habitat selection in the boreal forest of central Saskatchewan. *The Auk* **115**, 746–754.
- McClellan, S. A., Rumble, M. A., King, R. M., and Baker, W. L. (1998). Evaluation of resource selection methods with different definitions of availability. *The Journal of Wildlife Biology* **62**, 793–801.
- McDonald, M. W. (1998). Ecology of Columbian sharp-tailed grouse in eastern Washington. M.S. thesis, University of Idaho, Moscow.
- Neu, C. W., Byers, C. R., and Peek, J. M. (1974). A technique for analysis of utilization–availability data. *Journal of Wildlife Management* **38**, 541–545.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd edition. New York: Wiley.
- Schoen, J. W. and Kirchoff, M. D. (1985). Seasonal distribution and home range patterns of Sitka black-tailed deer on Admiralty Island, southeast Alaska. *Journal of Wildlife Management* **49**, 96–103.
- Smith, L. M., Hupp, J. W., and Ratti, J. T. (1982). Habitat use and home range of grey partridge in eastern South Dakota. *Journal of Wildlife Management* **46**, 580–587.
- Wilson, S. F., Shackleton, D. M., and Campbell, K. L. (1998). Making habitat-availability estimates spatially explicit. *Wilson Society Bulletin* **26**, 626–631.

Received January 1999. Revised September 1999.

Accepted November 1999.

APPENDIX

Proof of Theorem 1

Given $N_j, N_{j'} > 0$ for $j \neq j' = 1, \dots, t$, $a_j < 1$ for all j from the asymptotic properties of proportions, we have, under the null hypothesis (cf., Arnold, 1990, p. 500),

$$\mathbf{W}_j = \sqrt{N_j}[\mathbf{p}_j - \boldsymbol{\pi}] \xrightarrow{d} N_h[\mathbf{0}, \boldsymbol{\Sigma}], \quad j = 1, \dots, t, \quad (\text{A.1})$$

where N_h represents an h -variate normal distribution and \xrightarrow{d} indicates converges in distribution to N . Also from (3) and (A.1), we have $\text{cov}(\mathbf{W}_j, \mathbf{W}_{j'}) = \rho \boldsymbol{\Sigma}$ and

$$\mathbf{V} = \begin{pmatrix} \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_t \end{pmatrix} \xrightarrow{d} N \left[\begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma} & \rho \boldsymbol{\Sigma} & \cdots & \rho \boldsymbol{\Sigma} \\ \vdots & \vdots & \ddots & \vdots \\ \rho \boldsymbol{\Sigma} & \rho \boldsymbol{\Sigma} & \cdots & \boldsymbol{\Sigma} \end{pmatrix} \right]. \quad (\text{A.2})$$

Let \mathbf{I} denote the $(h \times h)$ identity matrix and let $u_j = (a_j)^{1/2}$, with $\mathbf{A} = (u_1 \mathbf{I} \ u_2 \mathbf{I} \ \cdots \ u_t \mathbf{I})$. Now we have $N^{1/2}(\bar{\mathbf{p}} - \boldsymbol{\pi}) = \mathbf{A}\mathbf{V}$. Hence, considering the vector $\mathbf{A}\mathbf{V}$, we have

$$\mathbf{E}(\mathbf{A}\mathbf{V}) = \mathbf{0}$$

$$\begin{aligned} \text{var}(\mathbf{A}\mathbf{V}) &= \boldsymbol{\Sigma}^* = \left(1 + \rho \sum_{j=1}^t \sum_{\substack{j'=1 \\ j \neq j'}}^t \sqrt{a_j a_{j'}} \right) \boldsymbol{\Sigma} \\ &= (1 + \rho b) \boldsymbol{\Sigma}. \end{aligned} \quad (\text{A.3})$$

From the Mann–Wald theorem (Rao, 1973, p. 124) and from (A.1) and (A.3),

$$\sqrt{N}[\bar{\mathbf{p}} - \boldsymbol{\pi}] \xrightarrow{d} N_h[\mathbf{0}, \boldsymbol{\Sigma}^*]$$

and

$$GF^* = (\bar{\mathbf{p}} - \boldsymbol{\pi})' \left(\frac{\boldsymbol{\Sigma}^{*-1}}{N} \right) (\bar{\mathbf{p}} - \boldsymbol{\pi}) \xrightarrow{d} \chi_h^2. \quad (\text{A.4})$$

But

$$GF^* \frac{1}{1 + b\rho} (\bar{\mathbf{p}} - \boldsymbol{\pi})' \left(\frac{\boldsymbol{\Sigma}^{-1}}{N} \right) (\bar{\mathbf{p}} - \boldsymbol{\pi}) = \frac{GF}{1 + b\rho}. \quad (\text{A.5})$$

In general, ρ is not known and is estimated by $\hat{\rho}$ and r and d are both consistent for the parameters R and δ . The consistency of ratio estimators (Arnold, 1990, p. 240, Theorem 6-8) implies $1 + \hat{\rho}b \xrightarrow{P} 1 + \rho b$, where \xrightarrow{P} denotes convergence in probability. This implies

$$AGF = \frac{GF}{(1 + \hat{\rho}b)} \xrightarrow{d} \chi_h^2.$$

LINKED CITATIONS

- Page 1 of 2 -



You have printed the following article:

A Chi-Square Goodness-of-Fit Analysis of Dependent Resource Selection Data

Nairanjana Dasgupta; J. Richard Alldredge

Biometrics, Vol. 56, No. 2. (Jun., 2000), pp. 402-408.

Stable URL:

<http://links.jstor.org/sici?sici=0006-341X%28200006%2956%3A2%3C402%3AACGAOD%3E2.0.CO%3B2-H>

This article references the following linked citations. If you are trying to access articles from an off-campus location, you may be required to first logon via your library web site to access JSTOR. Please visit your library's website or contact a librarian to learn about options for remote access to JSTOR.

References

Compositional Analysis of Habitat Use From Animal Radio-Tracking Data

Nicholas J. Aebischer; Peter A. Robertson; Robert E. Kenward

Ecology, Vol. 74, No. 5. (Jul., 1993), pp. 1313-1325.

Stable URL:

<http://links.jstor.org/sici?sici=0012-9658%28199307%2974%3A5%3C1313%3ACAOHUF%3E2.0.CO%3B2-C>

Survey and Comparison of Methods for Study of Resource Selection

J. Richard Alldredge; Dana L. Thomas; Lyman L. McDonald

Journal of Agricultural, Biological, and Environmental Statistics, Vol. 3, No. 3, Resource Selection Using Data from Geographical Information Systems (GIS). (Sep., 1998), pp. 237-253.

Stable URL:

<http://links.jstor.org/sici?sici=1085-7117%28199809%293%3A3%3C237%3ASACOMF%3E2.0.CO%3B2-B>

Large Sample Simultaneous Confidence Intervals for the Multinomial Probabilities Based on Transformations of the Cell Frequencies

B. J. R. Bailey

Technometrics, Vol. 22, No. 4. (Nov., 1980), pp. 583-589.

Stable URL:

<http://links.jstor.org/sici?sici=0040-1706%28198011%2922%3A4%3C583%3ALSSCIF%3E2.0.CO%3B2-Q>

LINKED CITATIONS

- Page 2 of 2 -



A Multivariate # 2 Analysis of Resource Selection Data

Nairanjana Dasgupta; J. Richard Alldredge

Journal of Agricultural, Biological, and Environmental Statistics, Vol. 3, No. 3, Resource Selection Using Data from Geographical Information Systems (GIS). (Sep., 1998), pp. 323-334.

Stable URL:

<http://links.jstor.org/sici?sici=1085-7117%28199809%293%3A3%3C323%3AAMAORS%3E2.0.CO%3B2-V>

On Simultaneous Confidence Intervals for Multinomial Proportions

Leo A. Goodman

Technometrics, Vol. 7, No. 2. (May, 1965), pp. 247-254.

Stable URL:

<http://links.jstor.org/sici?sici=0040-1706%28196505%297%3A2%3C247%3AOSCIFM%3E2.0.CO%3B2-O>