Sankhyā : The Indian Journal of Statistics
2002, Volume 64, Series A, Pt. 1, pp 626-639

DIAGNOSTICS FOR THE HOMOGENEITY OF INCLUSION PROBABILITIES IN A BERNOULLI CENSUS

By CHANG XUAN MAO
University of California, Berkeley, USA
and
BRUCE G. LINDSAY
Pennsylvania State University, USA

SUMMARY. In a Bernoulli census, the inclusion probabilities often vary with individuals. The dispersion score test is applied to test homogeneity of inclusion probabilities. Three graphic diagnostics, the log ratio plot, residual plot and gradient plot, are used to detect the existence of heterogeneity. Confidence bands are available for all three plots. Real data examples are studied as illustration.

1. Introduction

Consider a closed population of finitely many individuals in which one may be interested in estimation of the population size through Bernoulli census. This occurs in fields like health sciences, system reliability and sociological research (Chao, 2001). The problem also arises from capture-recapture experiments if the captured individuals are not removed and time is assumed to have no affect on capture probabilities. Suppose there are N individuals indexed by $i = 1, 2, \ldots N$. We focus on a special case in which each individual is assumed to have the same probability to be included in totally M Bernoulli samples. Let X_i be the number of samples that include the ith individual, which is a binomial random variable with index M and inclusion probability π_i. The X_i's will be called frequencies. If $X_i = 0$, then the ith individual has never been included in any sample; we assume that these values are missing data. Let n_x be the number of individuals that are included

Paper received January 2002.
AMS (1991) *subject classification.* 62G10, 62P10, 62P25.
Keywords and phrases. Binomial mixture, capture-recapture, conditional likelihood, graphic diagnostics, number of species, population size.

exactly in x samples. The n_x's are called counts. Let $n_* = n_1 + n_2 + \cdots + n_M$ be the number of observed individuals. Note that n_0 and $N = n_* + n_0$ are unknown.

Here is an example from prevalence studies. The names of Down's syndrome children in Massachusetts were obtained from five different record sources. The children were born between January 1, 1955 and December 31, 1959, and were still alive on December 31, 1966. The five different sources were: obstetric records, other hospital records, Massachusetts Department of Health, Massachusetts Department of Mental Health and schools. This dataset was analysed by Wittes (1970), Fienberg (1972) and Hook and Regal (1982). In this paper, we assume that the name of each child has the same probability π_i of appearing in each of these five records. The observed frequency counts were $n_1 = 248$, $n_2 = 188$, $n_3 = 81$, $n_4 = 18$ and $n_5 = 2$. Note that $n_* = 537$ and n_0 represents the unknown number of children with Down's syndrome whose names appeared in none of the five sources.

An important issue in model building is whether the inclusion probabilities are heterogeneous. In the homogeneity model, the π_i's are assumed to be identical. In the heterogeneity model, the π_i's are allowed to be different and are often assumed to be a random sample arising from a latent distribution, which leads to a mixture model. When the latent distribution is degenerate, the homogeneity model is obtained, naturally called a one-component model. Testing homogeneity of inclusion probabilities can be put in a hypothesis testing problem: a one-component model versus any mixture alternative. Note that N is a nuisance parameter in the hypothesis testing problem although it is often the primary parameter of interest in estimation.

There have been several testing procedures. The test in Cormack (1966) only works for the case $M = 3$. While no theoretical evidence of optimality properties was found, two moment-based procedures in Lloyd (1992) were shown in simulation to be more efficient and to have higher power than the tests in Otis et al. (1978), Chapman (1952) and Cormack (1966). Since general purpose goodness-of-fit tests usually lack power against important alternatives (Breiman, 2001), likelihood ratio tests were presented in Norris and Pollock (1996) and Pledger (2000). However, inference on the latent distribution was dependent on N both in Norris and Pollock (1996) and in Pledger (2000). Likelihood ratio tests based on mixture alternatives are also complicated in distribution theory and in computing (Lindsay 1995); the results typically depend on the computational procedures. Seidel et al.

(2000) presented other irregularities of likelihood ratio tests. We will show here that the principles of unbiased testing point to methods that eliminate N from the testing problem.

We present procedures that are simpler to apply but still have good power against mixture alternatives. For testing, we will discuss a conditional dispersion score test, which is a locally most powerful test against mixture alternatives. As a supplement, three graphical diagnostics are presented, each more sensitive to mixture structure than the dispersion score test. The mixture hypothesis deviates from the null hypothesis in very distinctive ways, a fact we will exploit.

Note that in the full binomial data of the X_i's, the unknown N plays the role of sample size. The diagnostics used here have been developed previously only for the case that N is known. In Section 2, it will be shown how key diagnostics can be extended to a relevant conditional distribution that is free of the nuisance parameter N. The dispersion score test is presented in Section 3. In Section 4, some results about the conditional distribution are introduced that allow the development of graphical diagnostics. Besides the Down syndrome incidence dataset, a meadow vole dataset is used as an illustration in Section 5. The test and graphical diagnostics do not require one to model the latent distribution or to estimate the population size N, and so ask for significantly less computation than the methods in Norris and Pollock (1996) and Pledger (2000).

2. The Truncated Binomial Sample

Let $g(x;\lambda)$ be the binomial density with index M and odds $\lambda = \pi/(1-\pi)$:

$$g(x;\lambda) = \frac{M!\lambda^x}{(M-x)!x!(1+\lambda)^M}, \quad x = 0, 1, \ldots M.$$

Let X be a binomial random variable given λ and let λ itself be from a latent distribution $P(\lambda)$. The marginal density of X is a binomial mixture $g(x;P) = \int g(x;\lambda)\mathrm{d}P(\lambda)$. For a full data X_i's, the counts n_0, n_1, ..., n_M are sufficient for the parameters and have a multinomial distribution:

$$\frac{N!}{\prod_{k=0}^{M} n_k!}\prod_{k=0}^{M} p_k^{n_k}, p_k = g(k;P), k = 0, 1, \ldots M.$$

As we do not observe n_0, the observed likelihood is written as:

$$L(N,P) = \frac{N!}{(N-n_*)!p_0^{N-n}\prod_{k=1}^{M} n_k!} p_0^{N-n_*} \prod_{k=1}^{M} p_k^{n_k}.$$

The statistic n_* is sufficient for N given P fixed, leading to a factorization of $L(N,P)$ into a marginal likelihood $L_m(N,P)$ of n_*, which has a binomial distribution with index N and probability $1-p_0$, and a conditional likelihood $L_c(P)$ of n_1, n_2, ..., n_M given n_*, which has a multinomial distribution:

$$L_c(P) = \frac{n_*!}{\prod_{k=1}^{M} n_k!} \prod_{k=1}^{M} q_k^{n_k}, q_k = \frac{g(k;P)}{1-g(0;P)}, k = 1, 2, \ldots, M.$$

It is natural to perform diagnostics for the latent distribution P based on the conditional likelihood $L_c(P)$; the statistic n_* is a complete and sufficient for N, so optimal unbiased procedures can be devised in this way (Lehmann, 1997).

Our next key observation is that the conditional distribution is identical to one that we would obtain from the sufficient statistics of an independent and identical distributed sample of size n_*. The conditional data come from a mixture of truncated binomial densities, although with a different latent distribution Q defined by:

$$\mathrm{d}Q(\lambda) = \frac{[1-(1+\lambda)^M]\mathrm{d}P(\lambda)}{\int [1-(1+\lambda)^M]\mathrm{d}P(\lambda)}.$$

Let $f(x;\lambda)$ be the truncated binomial density:

$$f(x;\lambda) = \frac{M!\lambda^x}{(M-x)!x![(1+\lambda)^M - 1]}, x = 1, 2, \ldots M.$$

Let $f(x;Q) = \int f(x;\lambda)\mathrm{d}Q(\lambda)$ be a mixture of the truncated binomial densities. The simple but key result is:

$$f(x;Q) = q_x = g(x;P)/[1-g(0;P)], x = 1, 2, \ldots M,$$

so that a truncated binomial mixture can be treated as a mixture of truncated binomial densities.

Note that P is degenerate if and only if Q is degenerate. Hereafter, we replace n_* by n. Our diagnostics and the dispersion score test will thus be

based on treating the observed X_i's, denoted by $Z_1, Z_2, \ldots, Z_n$, as a sample from $f(x;Q)$. The asymptotic theory that we use will be based on n going to infinity. The idea of eliminating nuisance parameters via conditioning has deep roots in statistics (Liang, 1987; Cruddas et al., 1989; Wang and McDermott, 1999). Conditional procedures in this paper have similar applications in the species problem where we deal with Poisson random variables (Mao and Lindsay, 2002). The Poisson mixture model is fully identifiable while Q here is only partially identifiable. However, it is still possible to perform homogeneity tests.

3. The Dispersion Score Test

The $C(\alpha)$ test for homogeneity is a simple but powerful procedure (Neyman and Scott, 1966; Lindsay, 1995). We follow Lindsay (1995) to call it the dispersion score test. Let $\mu_i = \mu_i(\lambda) = \sum_{x=1}^{M} x^i f(x;\lambda)$ be the ith moment of $f(x;\lambda)$, $i = 1, 2, 3$ and 4. Let $\hat{\lambda}$ be the maximum likelihood estimator for λ under homogeneity and let $\hat{\mu}_i = \mu_i(\hat{\lambda})$. The dispersion test statistic is:

$$T = [n\tau(\hat{\lambda})]^{-1/2} \sum_{j=1}^{M} n_j[(j - \hat{\mu}_1)^2 - \hat{\mu}_2 + \hat{\mu}_1^2]$$

with the variance function:

$$\tau(\lambda) = \mu_4 + 8\mu_2\mu_1^2 - 4\mu_1\mu_3 - \mu_2^2 - 4\mu_1^4 - (\mu_2 - \mu_1^2)^{-1}(\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3)^2.$$

The test statistic T is an asymptotically standard normal random variable under the null hypothesis (Neyman and Scott, 1966). Rejecting for a large T indicates a departure from the null hypothesis in the direction of over-dispersion. If T is too small, then there is a lack of fit in the direction of under-dispersion. The dispersion score test is consistent against all mixture alternatives. However, it only examines one predictive feature of the mixture model, namely that the mixture model data are over-dispersed relative to the one-component data. This makes it simple as well as more powerful for small n. Since it is a test statistic, the magnitude of T should not be used as a measure of the degree of over-dispersion; for a large n, it could be quite large despite the presence of a small amount of heterogeneity.

The dispersion score test is related to the non-parametric maximum likelihood estimator for Q in that if the estimator for the latent distribution Q is the one-component model, then T must be negative.

For the Down's syndrome incidence data, we have $T = 1.011$ with P-value= 0.156. So the dispersion score test is not significant even at the level 0.10. This dataset displays only weak heterogeneity relative to the one-component model.

4. **The Graphical Diagnostics**

Graphical diagnostics are often used in mixture modelling (Titterington et al., 1985; Lindsay and Roeder, 1992). They may be used to show the presence of mixture structure, provide raw estimates for underlying parameters, and tell uniqueness of non-parametric maximum likelihood estimator for the latent distribution.

Let $\mu = \mu_1$ be the mean value parameter and $\sigma^2 = \mu_2 - \mu_1^2$ be the variance of a truncated binomial distribution. It is clear that μ is in $[1, M]$ and the mean value parameter space is bounded. The truncated binomial density can also be parameterized by the mean parameter μ. For notational convenience, $f(z;\mu)$ is used to represent a truncated binomial density with mean parameter μ. The latent distribution and the mixture density are also denoted by $Q(\mu)$ and $f(z;Q)$ respectively. Define some key diagnostic functions:

$$H(z) = \log[z!(M-z)!f(z;Q)] - \log M!,$$

$$r(z) = f(z;Q)/f(z;\mu_0) - 1,$$

$$D(\mu) = \sum_{z=1}^{M} r(z)f(z;\mu) \text{ with } \mu_0 = \int \mu \mathrm{d}Q(\mu).$$

Here $H(z)$, $r(z)$ and $D(\mu)$ are called the log ratio function, residual function and gradient function respectively. The function $D(\mu)$ is a smoothed version of $r(z)$.

THEOREM 1 *The functions $H(z)$ and $r(z)$ are convex in z and $D(\mu)$ is convex in μ. They are linear if and only if Q is degenerate. If Q is not degenerate, then $r(z)$ has the sign sequence $+,-,+$ as z traverses the real line.*

The convexity of $r(z)$ and $H(z)$ are relatively easy to obtain (Lindsay 1995). The sign sequence was first discussed in Shaked (1980). The convexity of $D(\mu)$ was proved in Lindsay and Roeder (1992).

Let $\bar{Z}$ be the sample mean of $Z_1, Z_2, \ldots, Z_n$. The maximum likelihood estimator $\hat{\mu}$ for μ equals $\bar{Z}$ under the one-component model. Define the empirical log ratio function $\widehat{H}(z)$, the empirical residual function $\hat{r}(z)$ and the empirical gradient function $\widehat{D}(\mu)$ as:

$$\widehat{H}(z) = \log[z!(M-z)!n_z] - \log(nM!),$$

$$\hat{r}(z) = n_z/[nf(z;\hat{\mu})] - 1,$$

$$\widehat{D}(\mu) = \sum_{z=1}^{M} \hat{r}(z) f(z;\mu).$$

Note that $\widehat{D}(\mu)$ is a smoothed version of $\hat{r}(z)$. When n is sufficient large, $\widehat{H}(z)$, $\hat{r}(z)$ and $\widehat{D}(\mu)$ can be good estimators for their theoretic counterparts $H(z)$, $r(z)$ and $D(\mu)$ respectively (Lindsay and Roeder, 1992). We will call plots $(z, \widehat{H}(z))$, $(z, \hat{r}(z))$ and $(\mu, \widehat{D}(\mu))$ the log ratio plot, residual plot and gradient plot respectively. Under the heterogeneity model, all these plots will show convexity and $\hat{r}(z)$ will have a sign sequence $+, -, +$. The plots are approximately linear under the homogeneity model. If neither the linearity nor convexity holds, then the binomial model should be in doubt. Pointwise confidence bands for these plots can also be constructed to account stochastic variation.

THEOREM 2 *Each $\widehat{H}(z)$ is approximately normally distributed. An approximate pointwise standard error of $\widehat{H}(z)$ valid under the one-component model or any mixture model is given by $e_H(z) = (1/n_z - 1/n)^{1/2}$.*

PROPOSITION 1 *Each $\hat{r}(z)$ is approximately normally distributed. The approximate pointwise standard error $e_r(z)$ of the residual plot under the one-component model is given by*

$$e_r^2(z) = n^{-1}[1/f(z;\hat{\mu}) - 1 - (z-\hat{\mu})^2/\sigma^2(\hat{\mu})].$$

PROPOSITION 2 *Each $\widehat{D}(\mu)$ is approximately normally distributed. The approximate pointwise standard error $e_D(\mu)$ of the gradient plot under the one-component model is given by*

$$e_D^2(\mu) = n^{-1}\{E_z[f(z;\mu)/f(z;\hat{\mu})]^2 - 1 - (\mu-\hat{\mu})^2/\sigma^2(\hat{\mu})\}.$$

The results about the residual plot and gradient plot are direct applications of the theory in Lindsay and Roeder (1992). The proof for the log

ratio plot is given in the Appendix. The gradient plot is directly related to the theory of non-parametric maximum likelihood. In particular, the one-component model is the non-parametric maximum likelihood estimator for Q if and only if $\widehat{D}(\mu)$ is no greater than 0 for any μ. The sign of $\widehat{D}''(\hat{\mu})$ is the same as that of the dispersion score test statistic T. If the one-component model is true, the residual plot and the gradient plot should be horizontal lines through the origin while the log ratio plot should be a non-horizontal straight-line. Since it is easier to recognize a straight-line if its slope is near to zero than a straight-line with nonzero slope, we suggest adjusting the slope and intercept on the log ratio plot. Suppose $H(z) = a + bz + E(z)$. One kind of simple estimators for a and b, denoted by $\hat{a}$ and $\hat{b}$ respectively, can be obtained using weighted least square estimation by regressing $\widehat{H}(z)$ on z with weights n_z for each z. Note that the statistic $\hat{b}$ is the first order efficient estimator for the parameter $\log \lambda$ in the homogeneity model (Lindsay 1995). Let $\widehat{E}(z) = \widehat{H}(z) - \hat{a} - \hat{b}z$. The only difference between $\widehat{H}(z)$ and $\widehat{E}(z)$ is the intercept and slope. Since we can treat $\hat{a}$ and $\hat{b}$ as constants although they are obtained from the data, the functions $\widehat{H}(z)$ and $\widehat{E}(z)$ share the same confidence bands. The strict convexity does not change if it exists. Let $Z_{\alpha/2}$ be the upper $\alpha/2$ quantile of the standard normal distribution, we have α-level pointwise confidence bands

$$(z, \widehat{E}(z) \pm Z_{\alpha/2} e_H(z)), (z, \hat{r}(z) \pm Z_{\alpha/2} e_r(z)), \text{ and } (\mu, \widehat{D}(\mu) \pm Z_{\alpha/2} e_D(\mu)).$$

If a horizontal line is contained in the confidence bands the homogeneity model may be the first choice due to the principle of parsimony. As will be seen, the log ratio plot and residual plot tend to be very similar in appearance. An advantage to the log ratio plot is that the assessment of convexity is more stringent; that is, if the log ratio plot is convex, then the residual plot must also be, but not vice versa.

For the Down's syndrome incidence data, the empirical residuals are $\hat{r}(1) = 0.025$, $\hat{r}(2) = -0.046$, $\hat{r}(3) = 0.009$, $\hat{r}(4) = 0.101$ and $\hat{r}(5) = 0.502$. There is a sign sequence $+, -, +$, which indicates that there is some mixing effect. In the log ratio plot and residual plot, the curves themselves show slight convexity. There is a horizontal line through zero lying in the confidence bands of the residual plot and that of the gradient plot. The results from the residual plot, gradient plot and the dispersion score test are consistent. Although the log ratio plot confidence bands does not contain a straight line, its wide fan-like right-tail of confidence bands does not suggest a mixture model.

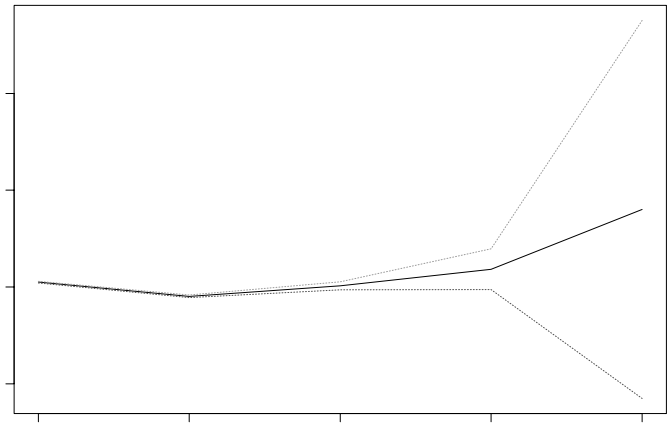

Figure 1: Log ratio plot: Down's syndrome incidence data

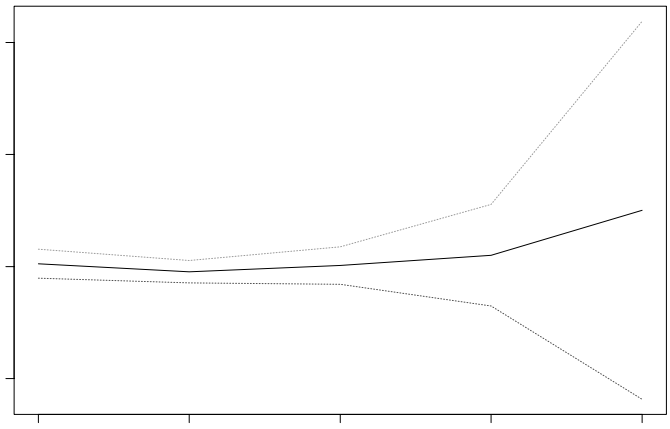

Figure 2: Residual plot: Down's syndrome incidence data

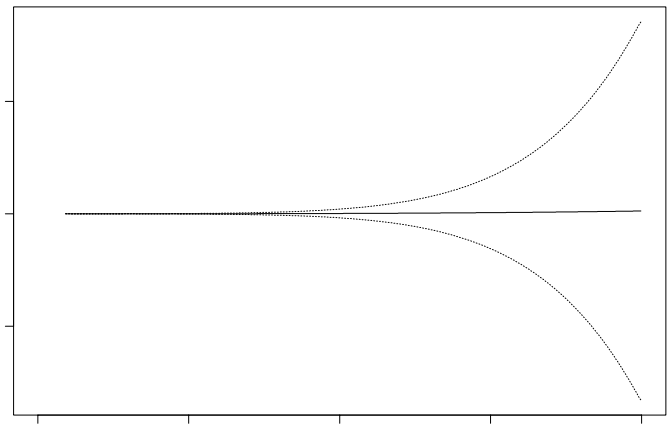

Figure 3: Gradient plot: Down's syndrome incidence data

5. **Meadow Vole Data**

A capture-recapture dataset of meadow vole was analysed by Pollock et al. (1990) and Lee and Chao (1994). There were five consecutive trapping days and totally $n = 102$ distinct voles were captured. The frequency counts were $n_1 = 29$, $n_2 = 15$, $n_3 = 15$, $n_4 = 16$ and $n_5 = 27$.

The residuals are $\hat{r}(1) = 2.290$, $\hat{r}(2) = -0.401$, $\hat{r}(3) = -0.579$, $\hat{r}(4) = -0.368$ and $\hat{r}(5) = 2.753$. There is a sign sequence $+, -, +$ to the empirical residuals. All graphical diagnostics show strong convexity. For the dispersion score test we have $T = 10.554$ with P-value 0. So the dispersion score test is highly significant. Both the graphical diagnostics and the dispersion score test confirm the conclusion that there existed heterogeneity of inclusion probabilities that was made by Pollock et al. (1990) and Lee and Chao (1994).

6. **Discussion**

The heterogeneity model introduces a larger variance than does the homogeneity model, a phenomenon called over-dispersion. The log ratio func-

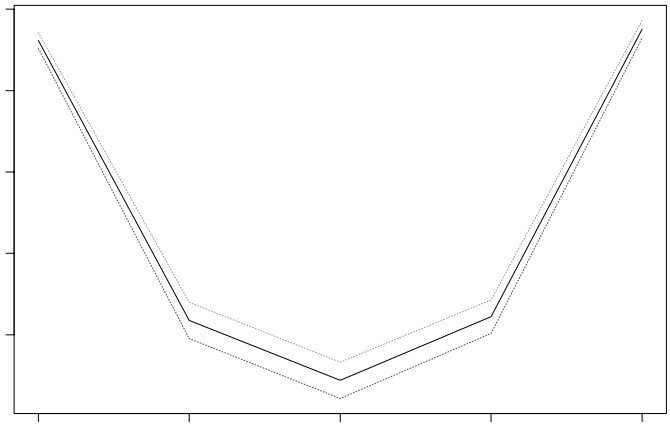

Figure 4: Log ratio plot: Meadow vole data

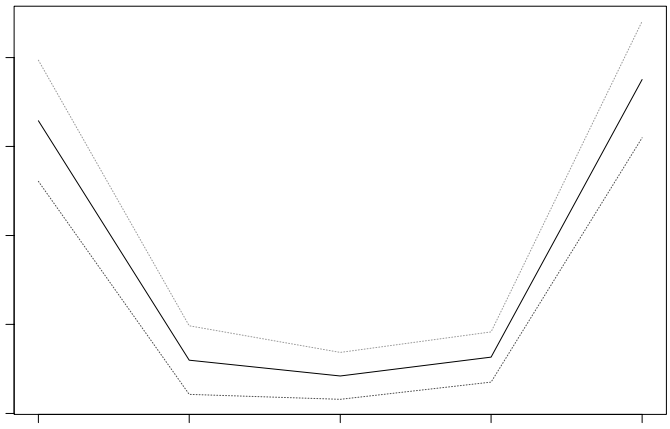

Figure 5: Residual plot: Meadow vole data

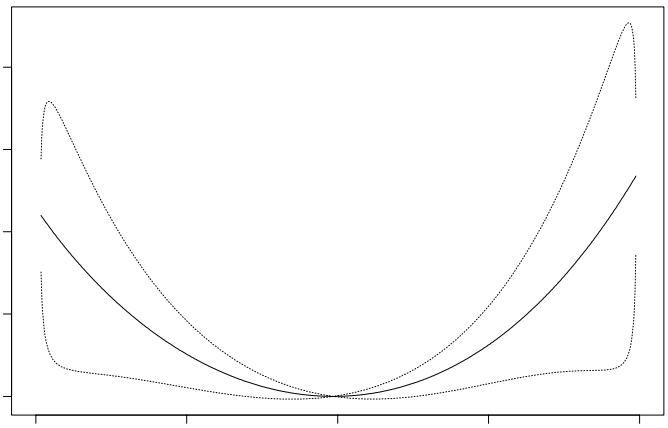

Figure 6: Gradient plot: Meadow vole data

tion, residual function and gradient function provide a refined assessment of over-dispersion. The conceptually simple graphical diagnostic procedures can display heterogeneity if it exists. We recommend the log ratio plot as being the best single plot due to its simplicity. Over-dispersion is also the basis for the dispersion score test. While the graphical procedures may provide satisfactory informal assessments, the dispersion score test provides a locally most powerful test for the existence of heterogeneity. The dispersion score test is not specific for the mixture model, as it has power against any over-dispersion alternative; the plots can give guidance as to whether the over-dispersion takes a mixture form.

Note that the log ratio plot and residual plot are only informative for frequencies with non-zero counts. At a frequency with zero-count, the empirical residual function is -1 and the empirical log ratio function is undefined. A simple but useful strategy is to treat the frequencies with zero count as missing points in the two plots. A more systematic approach would be to calculate the upper confidence limit for the probability p in a binomial distribution with index n and observed value zero. This means to solve $(1-p)^n = \alpha/2$. Plugging this into the log ratio function or residual function then gives an upper confidence limit to the curves and indication to whether the zero count is consistent with convexity.

Appendix: Confidence Bands of the Log Ratio Plot

Let Z_i given ϕ_i follow a discrete distribution $f(z;\phi_i)$, $i = 1, 2, \ldots n$, where $f(z;\phi) = c(\phi)h(z)e^{z\phi}$ is the density with respect to counting measure, such as binomial distributions, Poisson distributions and their truncated versions. If the ϕ_i's are a random sample from $Q(\phi)$, then the Z_i are a random sample from $f(z;Q) = \int f(z;\phi)dQ(\phi)$. Note that n_z stands for the number of Z_i's with value z and, as n goes to infinity,

$$n_z/n \xrightarrow{P} f(z;Q),$$

$$n^{1/2}(n_z/n - f(z;Q)) \xrightarrow{d} N(0, f(z;Q)[1 - f(z;Q)]),$$

$$n^{1/2}(\log n_z/n - \log f(z;Q)) \xrightarrow{d} N(0, [f(z;Q)]^{-1}[1 - f(z;Q)]),$$

The asymptotic result is obtained by replacing $f(z;Q)$ with n_z/n.

Acknowledgments. Professor Lindsay's research was supported by U. S. National Science Foundation grant DMS-0104443.

References

Breiman, L. (2001). Statistical modeling: the two cultures, *Statist. Sci.*, **16**, 199-231.

Chao, A. (2001). An overview of closed capture-recapture models, *J. Agric. Biol. Environ. Stat.*, **6**, 138-155.

Chapman, D.G. (1952). Inverse multiple and sequential sample census, *Biometrics*, **8**, 286-306.

Cormack, R.M.(1996). A test for equal catchability, *Biometrics*, **22**, 330-342.

Cruddas, A.M., Reid, N. and Cox, D.R. (1989). A time-series illustration of approximate conditional likelihood, *Biometrika*, **76**, 231-237.

Fienberg, S. (1972). The multiple recapture census for closed populations and incomplete 2^k contingency tables, *Biometrika*, **59**, 591-603.

Hook, E. and Regal, R. (1982). Validity of Bernoulli census, log-linear, and truncated binomial models for correcting under-estimates in prevalence studies, *Amer. J. Epidemiology*, **116**, 168-176.

Jiang, J.M. (1999). Conditional inference about generalized linear mixed models, *Ann. Statist.*, **27**, 1974-2007.

Lee, S.M. and Chao, A. (1994). Estimating Population size via sample coverage for closed capture-recapture models, *Biometrics*, **50**, 88-97.

Lehmann, E.L. (1997). *Testing Statistical Hypotheses*, Second edition, Springer-Verlag, New York.

Liang, K.Y. (1987). Estimating functions and approximate conditional likelihood, *Biometrika*, **74**, 695-702.

LINDSAY, B.G. (1995). *Mixture Models: Theory, Geometry and Applications*, Institute of Mathematical Statistics, Beachwood, OH.

LINDSAY, B.G. and ROEDER, K. (1992). Residual Diagnostics for Mixture Models, *J. Amer. Statist. Assoc.*, **87**, 785-794.

CHRIS J.L. (1992). Testing capture homogeneity in a recapture model, *Biometrika*, **79**, 555-561.

MAO, C.X. and LINDSAY, B.G. (2002). Tests and diagnostics for the heterogeneity in the species problem, *Comput. Statist. Data Anal.*, To appear.

NEYMAN, J. and SCOTT, E.L. (1966). On the use of $C(\alpha)$ optimal tests of composite hypothesis, *Bull. Inst. Internat. Statist.*, **41 I**, 477-497.

NORRIS, J.L. and POLLOCK, K.H. (1996). Nonparametric MLE under two closed capture recapture models with heterogeneity, *Biometrics*, **52**, 639-649.

OTIS, D.L., BURNHAM, K.P., WHITE, G.C. and ANDERSON, D.R. (1978). Statistical inference from capture data on closed animal populations, *Wildlife Monographs*, **62**, 1-135.

PLEDGER, S. (2000). Unified maximum likelihood estimates for closed capture-recapture models using mixtures, *Biometrics*, **56**, 434-442.

POLLOCK, K.H., NICHOLS, J.D., BRONWIE, C. and HINES, J.E. (1990). Statistical inference for capture-recapture experiments, *Wildlife Monographs*, **107**, 1-97.

SEIDEL, W., MOSLER, K. and ALKER, M. (2000). A Cautionary Note on Likelihood Ratio Tests in Mixture Models, *Ann. Inst. Statist. Math.*, **52**, 481-487.

SHAKED, M. (1980). On mixtures from exponential families, *J.R. Statist. Soc. Ser. B*, **42**, 192-198.

TITTERINGTON, D.M., SMITH, A.F. and MAKOV, U.E. (1985). *Statistical Analysis of Finite Mixture Distributions*, Wiley, New York.

WANG, Y.N. and MCDERMOTT, M.P. (1999). Conditional likelihood ratio test for a nonnegative normal mean vector, *J. Amer. Statist. Assoc.*, **94**, 660-660.

WITTES, J.T. (1970). Estimation of population size: the Bernoulli census, *Technical Report*, Harvard University, Department of Statistics.

CHANG XUAN MAO
SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF CALIFORNIA, BERKELEY
367 EVANS HALL
BERKELEY, CA 94720-3860
USA
E-mail: cmao@stat.berkeley.edu
cmao@statserv.ucr.edu

BRUCE G. LINDSAY
DEPARTMENT OF STATISTICS
PENNSYLVANIA STATE UNIVERSITY
326 THOMAS BUILDING
UNIVERSITY PARK, PA 16802-2111
USA
E-mail: bgl@psu.edu