integrated to find marginal posterior densities for components of interest, using flat priors for the regression coefficients. (It is not clear to me how one can be sure that anomalies of the type that arise in Mitchell's example cited above do not arise in this model.) The conditional approach, carried out approximately, requires an exercise similar to that outlined above to be recomputed for each component parameter of interest. While this is tiresome, it is not particularly complicated.

Casella, DiCiccio and Wells refer in their Section 3.1 to the saddlepoint alternative. I think it is more accurate to refer to it as a marginal alternative, since, as they acknowledge, Field's saddlepoint approximation for M-estimates is an approximation to the marginal sampling distribution of these estimates. It shares the drawback with other marginal solutions that elimination of nuisance parameters is not achieved in models where nuisance parameters are eliminated by conditioning. The saddlepoint method is a technique of approximation, which can be applied to conditional or unconditional models, but is not an inferential methodology.

5. CONCLUSION

The approach of Liang and Zeger is more directly motivated by particular practical applications, and it presents a quite different method for eliminating or minimizing the effect of nuisance parameters. At first glance our two papers seem quite unrelated, but the discussion of Lindsay and Waterman shows that they are more closely related than might appear. In particular, Lindsay and Li give convincing evidence that projection using Bhattacharyya scores can imitate conditioning. It would be interesting to know if this approach reproduces exact results when they are available, and what the connection might be to approximate ancillarity.

Dawid and Goutis refer to some confusion in my use of the terms sufficiency and ancillarity in the presence of nuisance parameters, and they point out that one aspect of this is the use of the phrase "the nuisance parameter," when in fact this parameter is typically not uniquely defined. Severini provides a more careful, and more helpful, definition of S-ancillarity related to this point. Barndorff-Nielsen and Cox (1994, Chapter 8) emphasize the importance of finding procedures which are invariant to interest-respecting reparametrizations. It is possible that, in particular models, a natural form of the nuisance parameter could be constructed using the estimating equations approach; that is, the nuisance parameter in a fully specified parametric model could be chosen to coincide with the nuisance parameter that would arise in a compatible semiparametric estimating equations approach.

In conclusion I find it heartening to see that discussions of theoretical statistics continue to engender lively debate.

ACKNOWLEDGMENTS

I would like to thank the discussants for their contributions and the Editor, Rob Kass, for his encouragement.

Rejoinder

Kung-Yee Liang and Scott L. Zeger

We would like to thank the discussants for their thoughtful comments. We would also like to add our congratulations to Dr. Reid, for her clear exposition on conditional inferences, a tool for reducing the influence of nuisance parameters in a fully parametric setting.

The two papers by Dr. Reid and ourselves address, in part, the common question of how to draw inferences in the presence of nuisance parameters. As pointed out by the discussants, they also both focus on methods most directly applicable to exponential family models. However, the fundamental distinction between these two papers is the degree to which we specify a probability mechanism for the data. Dr. Reid's paper starts with the assumption that a full probability mechanism can be specified. We begin with the assumption that it is not possible nor perhaps desirable to do so.

Peter McCullagh comments on the role of conditional inference when the likelihood is fully specified. He reflects upon the inherent contradiction in the practice of statistics that conditionality and sufficiency are accepted, while the likelihood principle is not. He raises the important point that the likelihood function is most often unknown, forcing applied statisticians to look beyond this measure of evidence. We would add that statisticians address different questions in practice. As summarized by Royall (1992), they ask the following questions:

- What do the data say?
- What is it reasonable to believe given the data?
- What decision should we make in light of the data?

When addressing the first question, we believe most applied statisticians are satisfied with the likelihood principle and will rely on the likelihood function as the best measure of evidence when a sensible probability mechanism for the data is available. Applied statisticians are also content to amend the first question slightly, asking: "What do the part of the data most relevant to a particular subset of parameters say?" Here, they are happy to use a conditional likelihood that is less dependent on nuisance parameters, but again, this conditional measure of evidence is satisfactory.

To address the second and third questions, the Bayesian and frequentist paradigms are used. As we will continue to address all three questions, we should not expect a single paradigm to dominate.

As is evident in the practice of statistics, the majority of questions asked by scientists of their data address only a part of the probability mechanism that generated the data. For example, in regression analysis, we ask whether a mean response is related to a particular set of covariates. In likelihood-based inference, we more often than not complete the specification of the parts of the probability model that are not themselves of scientific interest for mathematical or analytic convenience. Gross departures from these nuisance parts may be detected with large samples, but, as often as not, they receive little or no attention. Hence, we require methods of inferences which, to the extent possible, are insensitive to misspecification of these nuisance parts. Ideally, we would specify those parts of the probability mechanism which are of interest and make inferences in a way that is robust to the other parts. The question naturally arises, exactly what do we mean by "robust" and how can we specify an optimal estimating equation from a robust class?

Lindsay and Li offer a sensible answer as well as the basis for connecting the estimating equation approach with the more traditional and desirable likelihood formulation. Their definition (a) of "robust" is our condition (b) that the consistency of the estimation of the nuisance parameters should not affect the consistency of the estimation of the parameters of interest. Among the class that satisfies this property, the optimal estimating equation can be defined as the one nearest to the unconditional score in terms of squared error lost. They point out that, under regularity conditions, inferences based upon the conditional likelihood (Lindsay, 1982) are optimal; that is, the conditional score is the unique estimating equation that satisfies our condition (b) and is as close to the unconditional score as possible.

Lindsay and Li further point out that this strategy does not depend upon the existence of the likelihood or a factorization to admit the conditional argument. One can think of finding the optimal estimating equation by projecting the true score function for the parameters of interest onto Hilbert's base spanned by the first-order statistics Y_i, $1 \le i \le n$, or first- and second-order statistics Y_iY_j, $1 \le y \le j \le n$. Hence, the estimating equations we discussed can be thought of as approximations to score equations, approximations that are formulated in terms of that part of available data whose expectations involve the parameters of interest but not others.

Louise Ryan focuses on the analysis of teratological data in which the likelihood approach based on the beta-binomial distribution and the estimating function approach are nicely compared. It is comforting to learn that the impact of nuisance parameters is considerably less dramatic for the latter approach in the examples Ryan considered.

Ryan identified some interesting statistical issues that deserve further investigation. We agree that Wald-based confidence intervals may be less desirable, especially in the situation where the sample size is not sufficiently large. One alternative approach is to invert the distribution of the standardized estimating function $\hat{\theta}$ as detailed by Godambe (1991a). More recently, some approximate likelihood ratio methods based on the estimating function have been proposed by Hanfelt and Liang (1995a). This method provides a likelihood-based version confidence interval and is shown to behave better than the other two approaches in the context of odds ratio regression with a series of 2×2 tables (Hanfelt and Liang, 1995b).

The issue of dealing with multivariate outcome is a challenging one. The Dirichlet–trinomial distribution may not be desirable, in that intralitter correlations for different outcomes are captured by a single parameter. The estimating function approach is not limited in this way as different correlation parameters may be modeled for different outcomes. While the work on mixture of discrete and continuous outcomes along the line of estimating functions is primitive, the work by Zhao, Prentice and Self

(1992) may serve as a starting point. Recent work by Heagerty and Zeger (1995) illustrates very flexible simultaneous regression modelling of means and associations for multivariate discrete data.

Finally, we thank Professor Godambe for his comments. While our interests in estimating functions with nuisance parameters originated with very practical data analysis problems, Dr. Godambe's foundational research has paved the way and provided valuable insights as to how best to proceed sensibly. He continues to provide leadership in this fruitful area of research.

ADDITIONAL REFERENCES

BARNARD, G. A. (1982). Conditionality versus similarity in the analysis of 2×2 tables. In *Statistics and Probability: Essays in Honor of C. R. Rao* (G. Kallianpur, P. R. Krishnaiah and J. K. Ghosh, eds.) 59–65. North-Holland, Amsterdam.

BASU, D. (1988). *Statistical Information and Likelihood: A Collection of Critical Essays. Lecture Notes in Statist.* **45**. Springer, New York.

BERGER, J. O. and BERNARDO, J. M. (1992). On the development of the reference prior method. In *Bayesian Statistics* **4** (J. O. Berger, J. M. Bernardo, A. P. Dawid and A. F. M. Smith, eds.) 35–49. Oxford Univ. Press.

BERGER, J. O. and YE, K. Y. (1991). Noninformative priors for exponential regression. *Biometrika* **78** 645–656.

BIRNBAUM, A. (1962). On the foundations of statistical inference. *J. Amer. Statist. Assoc.* **57** 269–326.

BOX, G. E. P. (1980). Sampling and Bayes' inference in scientific modeling and robustness. *J. Roy. Statist. Soc. Ser. A* **143** 383–430.

CASELLA, G., WELLS, M. T. and TANNER, M. A. (1994) Using sampling-based calculations for pivotal inference. Technical Report BU-1178-M, Biometrics Unit, Cornell Univ., Ithaca, NY.

CATALANO, P. J. and RYAN, L. M. (1992). Bivariate latent variable models for clustered discrete and continuous outcomes. *J. Amer. Statist. Assoc.* **87** 651–658.

CATALANO, P., SCHARFSTEIN, D. and RYAN, L. (1994). Modeling fetal death and malformation in developmental toxicity. *Risk Analysis* **14** 629–637.

CATALANO, P. J., SCHARFSTEIN, D. O., RYAN, L. M., KIMMEL, C. A. and KIMMEL, G. L. (1993). A statistical model for fetal death, fetal weight, and malformation in developmental toxicity studies. *Teratology* **47** 281–290.

CHEN, J. J. and KODELL, R. L. (1989). Quantitative risk assessment for teratologic effects. *J. Amer. Statist. Assoc.* **84** 966–971.

CHEN, J. J., KODELL, R. L., HOWE, R. B. and GAYLOR, D. W. (1991). Analysis of trinomial responses from reproductive and developmental toxicity experiments. *Biometrics* **47** 1049–1058.

CHEN, M.-H. (1994). Importance weighted marginal Bayesian posterior density estimation. *J. Amer. Statist. Assoc.* **89** 818–824.

CLARKE, B. and WASSERMAN, L. (1993). Noninformative priors and nuisance parameters. *J. Amer. Statist. Assoc.* **88** 1427–1432.

COX, D. R. (1971). The choice between alternative ancillary statistics. *J. Roy. Statist. Soc. Ser. B* **33** 251–255.

CRUMP, K. S. (1984). A new method for determining allowable daily intakes. *Fundamental and Applied Toxicology* **4** 854–871.

DAWID, A. P. (1975). On the concepts of sufficiency and ancillarity in the presence of nuisance parameters. *J. Roy. Statist. Soc. Ser. B* **37** 248–258.

DAWID, A. P. (1977). Conformity of inference patterns. In *Recent Developments in Statistics* (J. R. Barra, B. van Cutsen, F. Brodeau and G. Romier, eds.) 245–256. North-Holland, Amsterdam.

DAWID, A. P. (1980). A Bayesian look at nuisance parameters. In *Bayesian Statistics* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 167–184. Valencia Univ. Press.

DAWID, A. P. (1983). Inference, statistical: I. In *Encyclopedia of Statistical Sciences* **4** 89–105. Wiley, New York.

DAWID, A. P. (1986). Discussion of "On principles and arguments to likelihood" by M. J. Evans, D. A. S. Fraser and G. Monette. *Canad. J. Statist.* **14** 196–197.

DICICCIO, T. J. and EFRON, B. (1992). More accurate confidence intervals in exponential families. *Biometrika* **79** 231–245.

DICICCIO, T. J. and EFRON, B. (1995). Better approximate confidence intervals. Technical report, Dept. Statistics, Stanford Univ.

DICICCIO, T. J., KELLER, J. B. and MARTIN, M. A. (1992). Accurate procedures for approximate Bayesian and conditional inference without the need for orthogonal parameters. Technical Report 391, Dept. Statistics, Stanford Univ.

DURBIN, J. (1969). Inferential aspects of the randomness of sample size in survey sampling. In *New Developments in Survey Sampling* (N. L. Johnson and H. Smith, Jr., eds.) 629–651. Wiley, New York.

DURBIN, J. (1970). On Birnbaum's theorem on the relation between sufficiency, conditionality and likelihood. *J. Amer. Statist. Assoc.* **65** 395–398.

EFRON, B. (1987). Better bootstrap confidence intervals (with discussion). *J. Amer. Statist. Assoc.* **82** 171–200.

EFRON, B. (1993). Bayes and likelihood calculations from confidence intervals. *Biometrika* **80** 3–26.

FIELD, C. A. and RONCHETTI, E. (1990). *Small Sample Asymptotics*. IMS, Hayward, CA.

FISHER, R. A. (1930). Inverse probability. *Proc. Cambridge Philos. Soc.* **26** 528–535.

GELFAND, A. E., SMITH, A. F. M. and LEE, T. M. (1992). Bayesian analysis of constrained parameters and truncated data problems. *J. Amer. Statist. Assoc.* **87** 523–532.

GODAMBE, V. P. (1985). Foundations of finite sample estimation in stochastic processes. *Biometrika* **72** 419–428.

GODAMBE, V. P. (1994). Linear Bayes and optimal estimation. Technical Report STAT-94-11, Univ. Waterloo.

GOUTIS, C. and CASELLA, G. (1992). Increasing the confidence in Student's t interval. *Ann. Statist.* **20** 1501–1513.

HAMMERSLEY, J. M. and HANDSCOMB, D. C. (1964). *Monte Carlo Methods*. Chapman and Hall, London.

HANFELT, J. J. and LIANG, K.-Y. (1995a). Approximate likelihood ratio for general estimating functions. *Biometrika*. To appear.

HANFELT, J. J. and LIANG, K.-Y. (1995b). Statistical inference for odds ratio regression models with sparse dependent data. Technical report, Div. Biostatistics, Georgetown Univ.

HEAGERTY, P. and ZEGER, S. L. (1995). Marginal regression models for clustered ordinal measurements. *J. Amer. Statist. Assoc.* To appear.

KALTER, H. (1974). Choice of the number of sampling units in teratology. *Teratology* **9** 257–258.

KASS, R. E. and WASSERMAN, L. (1994). Formal rules for selecting prior distributions: a review and annotated bibliography. Technical Report 583, Dept. Statistics, Carnegie Mellon Univ.

KREWSKI, D. and ZHU, Y. (1994). Application of multinomial dose response modelings in developmental toxicity risk assessment. *Risk Analysis* **14** 613–627.

LINDSAY, B. G. and LI, B. (1995). On second order optimality of the observed Fisher information. Technical report, Dept. Statistics, Pennsylvania State Univ.

LISEO, B. (1993). Elimination of nuisance parameters with reference priors. *Biometrika* **80** 295–304.

MANTEL, H. J. and GODAMBE, V. P. (1993). Estimating functions for conditional inference: many nuisance parameter case. *Ann. Inst. Statist. Math.* **45** 55–67.

MCCULLAGH, P. (1992). Conditional inference and Cauchy models. *Biometrika* **79** 247–259.

MITCHELL, A. F. S. (1967). Discussion of "A Bayesian significance test for multinomial distributions," by I. J. Good. *J. Roy. Statist. Soc. Ser. B* **29** 423–424.

NEYMAN, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philos. Trans. Roy. Soc. London Ser. A* **236** 333–380.

PEERS, H. W. (1965). On confidence points and Bayesian probability points in the case of several parameters. *J. Roy. Statist. Soc. Ser. B* **27** 9–16.

REID, N. (1995). Likelihood and Bayesian asymptotics. In *Bayesian Statistics 5* (J. O. Berger, J. M. Bernardo, D. V. Lindley and A. F. M. Smith, eds.). Oxford Univ. Press. To appear.

ROBERT, C. P. (1994). *The Bayesian Choice*. Springer, New York.

ROYALL, R. M. (1992). The illusive concept of statistical evidence. In *Bayesian Statistics 4* (J. O. Berger, J. M. Bernardo, A. P. Dawid, D. V. Lindley and A. F. M. Smith, eds.) 405–418. Oxford Univ. Press.

RYAN, L. M. (1992a). The use of generalized estimating equations for risk assessment in developmental toxicity. *Risk Analysis* **12** 439–447.

RYAN, L. M. (1992b). Quantitative risk assessment for developmental toxicity. *Biometrics* **48** 163–174.

SMALL, C. G. and MCLEISH, D. L. (1988). Generalizations of ancillarity, completeness and sufficiency in an inference function space. *Ann. Statist.* **16** 534–551.

SMALL, C. G. and MCLEISH, D. L. (1989). Projection as a method for increasing sensitivity and eliminating nuisance parameters. *Biometrika* **76** 693–703.

STAPLES, R. E. and HASEMAN, J. K. (1974). Selection of appropriate experimental units in teratology. *Teratology* **9** 259–260.

STEIN, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.* **16** 155–160.

STRAWDERMAN, R. L., CASELLA, G. and WELLS, M. T. (1995). Practical small sample asymptotics for regression problems. Technical Report BU-1269-M, Biometrics Unit, Cornell Univ., Ithaca, NY. (To appear in *J. Amer. Statist. Assoc.*)

TANNER, M. A. (1993). *Tools for Statistical Inference*, 2nd ed. Springer, New York.

WATERMAN, R. P. and LINDSAY, B. G. (1995). Projected score methods for nuisance parameters: asymptotics and Neyman–Scott problems. Technical report, Pennsylvania State Univ.

WILLIAMS, D. A. (1975). The analysis of binary responses from toxicology experiments involving reproduction and teratogenicity. *Biometrics* **31** 949–952.

WILLIAMS, D. A. (1982). Extra-binomial variation in logistic linear models. *J. Roy. Statist. Soc. Ser. C* **31** 144–148.

WILLIAMS, D. A. (1987). Dose–response models for teratological experiments. *Biometrics* **43** 1013–1016.

WILLIAMS, D. A. (1988). Estimation bias using the beta-binomial distribution in teratology. *Biometrics* **44** 305–309.

ZHAO, L. P., PRENTICE, R. L. and SELF, S. G. (1992). Multivariate mean parameter estimation by using a partly exponential model. *J. Roy. Statist. Soc. Ser. B* **54** 805–812.

ZHU, Y., KREWSKI, D. and ROSS, W. H. (1994). Dose response models for correlated data from developmental toxicity studies. *J. Roy. Statist. Soc. Ser. C* **43** 583–589.