McCullagh, P. (1985). On the asymptotic distribution of Pearson's statistic in linear exponential-family models. *Int. Stat. Rev.* **53** 61–67.
McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models*. Chapman and Hall, London.
Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *J. Roy. Statist. Soc. Ser. A* **135** 370–384.
Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* **61** 439–447.
Williams, D. A. (1982). Extra-binomial variation in logistic linear models. *Appl. Statist.* **31** 144–148.

Department of Biostatistics
University of Washington
Seattle, Washington 98195

James M. Dickey[1]

State University of New York at Albany

1. General. The paper under discussion by Diaconis and Efron (1985) (DE) is impressive and stimulating. I would like to bring forward here a few general questions to which it gives rise and then take a brief look at coherent inference for the models employed by DE.

Toward the end of Section 1, DE state that their "goal is to extend the usefulness of χ^2." I would wish to ask first, how should χ^2 be used? On the one hand, inferences based on tail areas, rather than probability densities or masses, are not coherent. On the other hand, tail areas are naturally interesting facts about the data (and about other nonoccurring data values). I do not know the best answer to this question and I would personally prefer to keep both kinds of tools in our kit.

Recall that the coherent inference in favor of a hypothesis H versus its alternative $\bar{H}$ is given by the Bayes factor $B(H, \bar{H})$ (Jeffreys, 1939; Good, 1950; Edwards, Lindman and Savage, 1963; Dickey and Lientz, 1968). This is the ratio of the coherent posterior odds $P(H \mid \mathbf{x})/[1 - P(H \mid \mathbf{x})]$ to the prior odds $P(H)/[1 - P(H)] > 0$. This ratio depends on the data $\mathbf{x}$, but not on the prior odds, so it serves as a sufficient report of the data for inference regarding H. The Bayes factor also equals the ratio of predictive densities, $B(H, \bar{H}) = p(\mathbf{x} \mid H)/p(\mathbf{x} \mid \bar{H})$, each a function of the respective conditional prior distribution, $p(\mathbf{x} \mid J) = \int p(\mathbf{x} \mid \boldsymbol{\pi})\, \mathrm{dP}(\boldsymbol{\pi} \mid J)$, $J = H, \bar{H}$. The dependence on conditional uncertainty may necessitate a tabular or graphical report of the Bayes factor (Dickey, 1973).

Technical point. In the case of a sharp hypothesis defined by a point value of a constraining parameter, H: $\boldsymbol{\eta} = \mathbf{0}$, where $\boldsymbol{\eta} \equiv \boldsymbol{\eta}(\boldsymbol{\pi})$, it is tempting to use a single joint density $g(\boldsymbol{\pi})$ to specify both of the conditional prior distributions, $p(\boldsymbol{\pi} \mid \bar{H}) = g(\boldsymbol{\pi})$ and $p(\boldsymbol{\pi} \mid H) = g(\boldsymbol{\pi} \mid \boldsymbol{\eta} = \mathbf{0})$, where $g(\boldsymbol{\pi} \mid \boldsymbol{\eta})$ is a lower-dimensional density obtained in the usual way by conditioning in $g(\boldsymbol{\pi})$. For one thing, Savage's

[1] Research supported by U.S. National Science Foundation Grant MCS-8301335.

density-ratio form of the Bayes factor is then available (Dickey and Lientz, 1968). But this must be done with care, because of the Borel-Kolmogorov dependence of a conditional distribution on the choice of conditioning variable. In such an approach, the Bayes factor would depend not only on the sharp-hypothesis event H, but also on the choice of constraint parameter used to define H. For example, the Bayes factor for independence in a 2×2 table would depend on whether "independence," H: $\eta = 0$, referred to

$$\eta(\pi) \equiv \pi_{11}/(\pi_{11} + \pi_{12}) - \pi_{21}/(\pi_{21} + \pi_{22})$$

or to

$$\eta(\pi) \equiv \log\{[\pi_{11}/(\pi_{11} + \pi_{12})]/[\pi_{21}/(\pi_{21} + \pi_{22})]\}.$$

Dickey and Lientz (1968) and Lindley (1971, footnote on page 32) mislead in this regard, a mistake for which I am responsible, and which was eventually corrected in Gûnel and Dickey (1974). (Li (1983) has attempted to avoid dependence on choice of variable by defining the conditional distribution directly as a geometrically induced lower-dimensional Hausdorff measure. However, the induced measure depends on the choice of joint variable π, and the different induced versions are not then merely related by change of variable within H.)

2. Coherent test for independence using Dirichlet prior distributions. Gûnel and Dickey (1974) gave the odds test for independence in two-way frequency tables using Dirichlet conditional prior distributions. This work preceded Good (1976), who considered only symmetric Dirichlet distributions. It would seem even more realistic to use mixtures of Dirichlet distributions or otherwise structured prior distributions to more accurately model real uncertainty under the alternative to independence. This would seem to harmonize with the spirit of statements made in DE.

It would be interesting to see DE's approximation of discrete uniform probabilities by relative volumes generalized to approximation of Dirichlet-multinomial probabilities by Dirichlet probabilities. This can be done very simply through the means and covariances, but DE seem to have more sophisticated tools.

The Bayes factor B is factorized by Gûnel and Dickey (1974) into separate factors based on the marginal count data and based on the conditional distribution of data within the rows and columns. Conditional inference is thus available directly, and conditioning is not required as a device to set up a point null hypothesis. Anticipating Good and Crook (1980), we found that the margins are quite uninformative regarding row-column independence. We also obtained the coherent inference for models conditioning on the margins of only *one* of the two types, a very common situation in practice. See Gûnel (1982) and Atkins and Gûnel (1984) for further work.

3. Normal model. DE treated the normal sampling distribution, $\mathbf{x} \mid \beta \sim N_D(\beta, n^{-1}\mathbf{I})$, with unknown location $\beta \mid \theta \sim N_D(\mathbf{0}, \sigma_\beta^2\mathbf{I})$,

$$\theta \equiv \nu/n \equiv n^{-1}/(n^{-1} + \sigma_\beta^2).$$

This yields the marginal uncertainty conditional on the hyperparameter θ (or ν or σ_β),

$$\mathbf{x} \mid \theta \sim N_D(\mathbf{0}, \nu^{-1}\mathbf{I}). \tag{1}$$

Dickey (1971, 1974) developed coherent inference with Bayes factors for such models, including more extensive families of prior distributions.

Since the statistic $S = \mathbf{x}^T\mathbf{x}$ is sufficient for θ in the model (1), we have, with $S \mid \theta \sim \chi^2_D/\nu$,

$$B(H_\theta, H_1) = p(S \mid \theta)/p(S \mid 1) = (\nu/n)^{(1/2)D}\exp[\tfrac{1}{2}(1 - \nu/n)nS]. \tag{2}$$

This is the ratio of odds in favor of the alternative H_θ versus the usual null hypothesis H_1; whereas in the previous order, $B(H_1, H_\theta) = 1/B(H_\theta, H_1)$. If the alternative value θ is subject to further uncertainty, then the chosen order, H_θ over H_1, allows informal or formal use of a prior-uncertainty mixture conditional on the nonoccurence of H_1. For a mixing distribution $P(\theta \mid \bar{H}_1)$, we have

$$B(\bar{H}_1, H_1) = \int B(H_\theta, H_1)\, dP(\theta \mid \bar{H}_1). \tag{3}$$

In the joint model, with θ also random, the events $\beta = 0$ and $\theta = 1$ are probabilistically equivalent, that is, $p(\mathbf{x}, \beta \mid \beta)_{\beta=\mathbf{0}} = p(\mathbf{x}, \beta \mid \theta)_{\theta=1}$ and $p[\mathbf{x}, \beta \mid (\beta \neq \mathbf{0})] = p(\mathbf{x}, \beta \mid \theta)$ as evaluated at θ where $\theta \neq 1$, and $P(\beta = \mathbf{0}) = P(\theta = 1)$. Hence the factor $B(H_\theta, H_1)$ (or $B(\bar{H}_1, H_1)$) can be used for inference concerning the event $\beta = \mathbf{0}$. The following steps arise naturally in a coherent reference.

1. Evaluate $B(H_\theta, H_1)$, (i) for one value or (ii) for multiple values of θ. The latter allows evaluation of $B(\bar{H}_1, H_1)$ by (3). The corresponding decision-theoretic criterion is to choose $d = d_{H_1}$ if $B(\bar{H}_1, H_1)$ exceeds the threshold,

$$\{P(H_1)/[1 - P(H_1)]\}\{E_{\theta \mid H_{1,x}} W_1(\theta)/E_{\theta \mid \bar{H}_{1,x}} W_2(\theta)\}, \tag{4}$$

   where the utility differences, $W_1(\theta) = U(d_{H_1}, \theta) - U(d_{\bar{H}_1}, \theta)$ and $W_2(\theta) = -W_1(\theta)$, satisfy $W_1(\theta) > 0$ for all $\theta \in H_1$ and $W_2(\theta) > 0$ for all $\theta \in \bar{H}_1$. The threshold will not depend on the data $\mathbf{x}$ if $W_1(\theta)$ is constant within H_1 and $W_2(\theta)$ is constant within $\bar{H}_1$. (See Kadane and Dickey, 1980, for discussion.)
2. If H_1 is rejected (i) in an analysis using fixed θ within $\bar{H}_1$, one can then use the usual posterior density $p(\beta \mid \mathbf{x}, \theta)$ obtained from the prior $p(\beta \mid \theta)$ to estimate β. If H_1 is rejected (ii) by using a mixture $\bar{H}_1$, one can first obtain the inference $p(\beta \mid \mathbf{x}, \theta)$ and then use

$$p(\beta \mid \mathbf{x}, \bar{H}_1) = \int p(\beta \mid \mathbf{x}, \theta)\, dP(\theta \mid \mathbf{x}, \bar{H}_1). \tag{5}$$

What posterior distribution to use for θ in (5)? DE consider confidence intervals based on the pivot, $\rho \equiv \nu S$, $\rho \mid \nu \sim \mathrm{X}^2_D$. (What is the logic of the "confidence" property for a "random effects" model?) For these intervals to be posterior credible intervals, that is, to have $\rho \mid S \sim \mathrm{X}^2_D$, would require a prior density given by $p(\nu \mid \bar{H}_1) \propto 1/\nu$, or

$$p(\sigma_\beta^2 \mid \bar{H}_1) \propto 1/(\sigma_\beta^2 + n^{-1}). \tag{6}$$

This density is, of course, nonintegrable and has the further objection of depending on the experiment through n. I do not know whether real uncertainty densities tend to give approximately the same inference as (6).

The expectand $B(H_\theta, H_1)$ (2) in (3) offers the further advantage of being proportional to the marginal (weighted) likelihood function of σ_β; and hence it points out the values of σ_β supported by the data. As $\sigma_\beta \to 0+$, $B \to 1$ ($H_\theta \to H_1$); and as σ_β increases, B has available two modes of behavior. If $S \le D/n$, B starts at its maximum, 1 at $\sigma_\beta = 0$, and decreases to zero as $\sigma_\beta \to \infty$. In this case, no conceivable prior value σ_β could give favor to H_θ over H_1. If $S > D/n$, B increases to its maximum, $[D/(nS)]^{(1/2)D}\exp[\frac{1}{2}(nS - D)]$ at $\sigma_\beta = (S/D - n^{-1})^{1/2}(\nu = D/S)$, and then B decreases to zero from there. B is strictly increasing in the statistic S.

One can obtain approximate highest-posterior-density intervals for any transformation τ of σ_β, based on an approximate constant prior density for τ. The interval end points are obtained by specifying likelihood-ratio values, $k = B/\max B$, and then solving for τ in the equation $\ln(\nu S/D) = \nu S/D - [1 + (2/D)\ln(1/k)]$.

4. Two-way frequency tables. Following DE, we now apply our normal-theory analysis to the limiting form of the conditional multinomial model. Note that the prior-uncertainty variance matrix, $\mathrm{var}(\boldsymbol{\pi} \mid H_\theta, (\mathbf{r}, \mathbf{c})) = \sigma_\beta^2 \hat{\boldsymbol{\Sigma}}_{\cdot(\mathbf{r},\mathbf{c})}$ (where $\hat{\boldsymbol{\Sigma}} = \mathrm{diag}(\hat{\boldsymbol{\pi}})$), is required in the analysis to be approximately proportional to the sampling conditional variance matrix, $\mathrm{var}(\mathbf{p} \mid \boldsymbol{\pi}, (\mathbf{r}, \mathbf{c})) \doteq n^{-1} \hat{\boldsymbol{\Sigma}}_{\cdot(\mathbf{r},\mathbf{c})}$. Since, when both π_{i+} and π_{+j} are small, $\mathrm{var}(p_{ij} \mid \boldsymbol{\pi}, (\mathbf{r}, \mathbf{c}))_{\pi=\hat{\pi}} \doteq n^{-1}\hat{\pi}_{ij}$, we have the prior variance of a coordinate π_{ij} approximately given by the product $\sigma_\beta^2 \hat{\pi}_{ij}$.

Consider the arithmetic average of the prior variances over the categories. This will involve the constant average of the probabilities: $\bar{\sigma}_\pi^2 = \sigma_\beta^2 av(\hat{\pi}_{ij}) = \sigma_\beta^2 (IJ)^{-1}$. Denoting the known average probability by $\bar{\pi} = av(\pi_{ij}) = (IJ)^{-1}$, we obtain the intuitively meaningful prior parameter,

$$\bar{\sigma}_\pi/\bar{\pi} = \sigma_\beta (IJ)^{1/2}. \tag{7}$$

Note that $0 \le \bar{\sigma}_\pi/\bar{\pi} \le IJ$.

We use this hyperparameter $\bar{\sigma}_\pi/\bar{\pi}$ to tabulate, in Tables 1′ and 2′, the Bayes factor (2) for the data of DE's Tables 1 and 2, respectively. It is apparent that for moderate and large values of one's conditional prior standard deviation under $\bar{H}_1$, one would strongly reject H_1 with the present joint uncertainty model. No values of prior standard deviation would yield support for H_1, except values beyond the range of the model. Just as with the standard tail-area test on these data, one would apparently reject H_1 more strongly for the Table 2 data than for the Table 1 data.

I agree with DE that it seems more promising to carry out analyses of structured alternatives to independence than the present limiting-normal analyses or the similar Dirichlet-prior analyses.

TABLE 1′

Bayes factor against independence for eye color and hair color. Data from Table 1 of Diaconis and Efron (1985) ($I = 4$, $J = 4$, $D = 9$, $n = 529$, $nS = 138.29$). Approximate normal sampling model and centered normal uncertainty model under the alternative hypothesis. B by equation (2) with $\nu/n = (1 + n\sigma_\beta^2)^{-1}$ and $\sigma_\beta = (\bar{\sigma}_\pi/\bar{\pi})(IJ)^{-1/2}$.

$\bar{\sigma}_\pi/\bar{\pi}$	ν/n	B
0.001	0.99997	1.002
0.01	0.997	1.24
0.04	0.950	25.6
0.05	0.924	137
0.10	0.752	8.00×10^6
0.659	6.51×10^{-2}	5.44×10^{22} (max)
16.0 (max)	1.18×10^{-4}	2.25×10^{12}
(800)	(4.73×10^{-8})	(1.16×10^{-3})

TABLE 2′

Bayes factor against independence for yearly income and number of children. Data from Table 2 of Diaconis and Efron (1985) ($I = 5$, $J = 4$, $D = 12$, $n = 25{,}263$, $nS = 568.57$). Approximate normal sampling model and centered normal uncertainty model under the alternative hypothesis. B by equation (2) with $\nu/n = (1 + n\sigma_\beta^2)^{-1}$ and $\sigma_\beta = (\bar{\sigma}_\pi/\bar{\pi})(IJ)^{-1/2}$.

$\bar{\sigma}_\pi/\bar{\pi}$	ν/n	B
0.0001	0.999990	1.004
0.001	0.9990	1.42
0.003	0.989	22.8
0.005	0.969	4.99×10^3
0.01	0.888	3.44×10^{13}
0.192	2.10×10^{-2}	$\geq 10^{100}$ (max)
20.0 (max)	1.98×10^{-6}	$\geq 10^{100}$
(6×10^6)	(2.20×10^{-17})	($\leq 10^{-100}$)

REFERENCES

ATKINS, J. and GÛNEL, E. (1984). Baycat: Bayes factors for categorical data analysis. *Amer. Statist.* **38** 156.

DIACONIS, P. and EFRON, B. (1985). Testing for independence in a two-way table: New interpretations of the chi-square statistic. *Ann. Statist.* **13** 845–874.

DICKEY, J. M. (1971). The weighted likelihood ratio, linear hypotheses on normal location parameters. *Ann. Math. Statist.* **42** 204–223.

DICKEY, J. M. (1973). Scientific reporting and personal probabilities: Students hypothesis. *J. Roy. Statist. Soc. Ser. B* **35** 285–305.

DICKEY, J. M. (1974). Bayesian alternatives to the *F*-test and least squares estimate in the normal linear model. In *Studies in Bayesian Econometrics and Statistics.* (S. E. Fienberg and A. Zellner, eds.) pp. 515–554. North-Holland, Amsterdam.

DICKEY, J. M. and LIENTZ, B. P. (1968). The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain. *Ann. Math. Statist.* **41** 214–226.

EDWARDS, W., LINDMAN, H. and SAVAGE, L. J. (1963). Bayesian statistical inference for psychological research. *Psychol. Rev.* **70** 193–242.

GOOD, I. J. (1950). *Probability and the Weighing of Evidence.* Hafner, New York.
GOOD, I. J. (1976). On the application of symmetric Dirichlet distributions and their mixtures to contingency tables. *Ann. Statist.* **4** 1159–1189.
GOOD, I. J. and CROOK, J. F. (1980). The information in the margins of a 2 × 2 contingency table. Unpublished report.
GÛNEL, E. (1982). Bayes factors in three-way contingency tables. *Comm. Statist. A—Theory Methods.* **11** 911–931.
GÛNEL, E. and DICKEY, J. (1974). Bayes factors for independence in contingency tables. *Biometrika* **61** 545–557.
JEFFREYS, H. (1939). *Theory of Probability* (3rd ed. 1961) Clarendon Press, Oxford.
KADANE, J. B. and DICKEY, J. M. (1980). Bayesian decision theory and the simplification of models. In *Evaluation of Econometric Models.* (J. Kmenta and J. Ramsey, eds.) pp. 245–268. Academic, New York.
LI, L.-A. (1983). Decomposition theorems, conditional probability, and finite mixture distributions. Ph.D. dissertation, Dept. of Mathematics and Statistics, SUNY at Albany.
LINDLEY, D. V. (1971). *Bayesian Statistics, A Review.* SIAM, Philadelphia.

DEPARTMENT OF MATHEMATICS
AND STATISTICS
STATE UNIVERSITY OF NEW YORK AT ALBANY
ALBANY, NEW YORK 12222

STEPHEN E. FIENBERG[1]

Carnegie-Mellon University

Diaconis and Efron's (henceforth DE) goal in this paper is a laudable one—to help interpret the classical chi-square statistic used to test for independence in two-way contingency tables in cases where independence clearly does not hold. Their mathematical statistics results are impressive, their theorems are seemingly impeccable, and their writing style is lucid. Yet, even after several readings, I came away from the paper with a feeling of disquiet, and a belief that they had failed to achieve their goal for most practical purposes. This comment provides some explanations for my disquiet and raises questions about the immediate utility of DE's results. The claim here is not so much that DE's results will not be of use to someone in the future (for their elegant results and geometrical interpretations will surely be put to good use), but rather that they will not be useful for the purpose originally proposed.

The statistical model for the counts in a two-way contingency table has two components: (1) a sampling model for the generation of the counts given a set of cell probabilities or expected values; (2) a structural model (corresponding to a curved manifold in the simplex) for the cell probabilities that is typically tied to the relationship between the categorical variables underlying the rows and

[1] Research partially supported by Office of Naval Research Contract N00014-84-K-0588, by a fellowship from the John Simon Guggenheim Foundation, and by NSF Grant BNS-8011494 while the author was a fellow at the Center for Advanced Study in the Behavioral Sciences.